OR Spectrum (2005) 27: 195–219
DOI: 10.1007/s00291-004-0187-1

OR Spectrum

Lean buffering in serial production lines with non-exponential machines

Emre Enginarlar[1], **Jingshan Li**[2], **and Semyon M. Meerkov**[3]

[1] Decision Applications Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA
[2] Manufacturing Systems Research Laboratory, GM Research and Development Center, Warren, MI 48090-9055, USA
[3] Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109-2122, USA (e-mail: smm@eecs.umich.edu)

Abstract. In this paper, lean buffering (i.e., the smallest level of buffering necessary and sufficient to ensure the desired production rate of a manufacturing system) is analyzed for the case of serial lines with machines having Weibull, gamma, and log-normal distributions of up- and downtime. The results obtained show that: (1) the lean level of buffering is not very sensitive to the type of up- and downtime distributions and depends mainly on their coefficients of variation, CV_{up} and CV_{down}; (2) the lean level of buffering is more sensitive to CV_{down} than to CV_{up} but the difference in sensitivities is not too large (typically, within 20%). Based on these observations, an empirical law for calculating the lean level of buffering as a function of machine efficiency, line efficiency, the number of machines in the system, and CV_{up} and CV_{down} is introduced. It leads to a reduction of lean buffering by a factor of up to 4, as compared with that calculated using the exponential assumption. It is conjectured that this empirical law holds for any unimodal distribution of up- and downtime, provided that CV_{up} and CV_{down} are less than 1.

Keywords: Lean production systems – Serial lines – Non-exponential machine reliability model – Coefficients of variation – Empirical law

1 Introduction

1.1 Goal of the study

The smallest buffer capacity, which is necessary and sufficient to achieve the desired throughput of a production system, is referred to as lean buffering. In (Enginarlar et al., 2002, 2003a), the problem of lean buffering was analyzed for the case of

Correspondence to: S.M. Meerkov

serial production lines with exponential machines, i.e., the machines having up- and downtime distributed exponentially. The development was carried out in terms of normalized buffer capacity and production system efficiency. The normalized buffer capacity was introduced as

$$k = \frac{N}{T_{down}}, \tag{1}$$

where N denoted the capacity of each buffer and T_{down} the average downtime of each machine in units of cycle time (i.e., the time necessary to process one part by a machine). Parameter k was referred to as the *Level of Buffering* (LB). The production line efficiency was quantified as

$$E = \frac{PR_k}{PR_\infty}, \tag{2}$$

where PR_k and PR_∞ represented the production rate of the line (i.e., the average number of parts produced by the last machine per cycle time) with LB equal to k and infinity, respectively. The smallest k, which ensured the desired E, was denoted as k_E and referred to as the *Lean Level of Buffering* (LLB).

Using parameterizations (1) and (2), Enginarlar et al., (2002, 2003a) derived *closed* formulas for k_E as a function of system characteristics. For instance, in the case of two-machines lines, it was shown that (Enginarlar et al., 2002)

$$k_E^{exp} = \begin{cases} \dfrac{2e(E-e)}{1-E}, & \text{if } e < E, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

Here the superscript exp indicates that the machines have exponentially distributed up- and downtime, and e denotes machine efficiency in isolation, i.e.,

$$e = \frac{T_{up}}{T_{up} + T_{down}}, \tag{4}$$

where T_{up} is the average uptime in units of cycle time. For the case of $M > 2$-machine serial lines, the following formula had been derived (Enginarlar et al., 2003a):

$$k_E^{exp}(M \geq 3) = \begin{cases} \dfrac{e(1-Q)(eQ+1-e)(eQ+2-2e)(2-Q)}{Q(2e-2eQ+eQ^2+Q-2)} \times \\ \ln\left(\dfrac{E-eE+eEQ-1+e-2eQ+eQ^2+Q}{(1-e-Q+eQ)(E-1)}\right), & \text{if } e < E^{\frac{1}{M-1}}, \\ 0, & \text{otherwise,} \end{cases} \tag{5}$$

where

$$Q = 1 - E^{\frac{1}{2}\left[1+\left(\frac{M-3}{M-1}\right)^{M/4}\right]} + \left(E^{\frac{1}{2}\left[1+\left(\frac{M-3}{M-1}\right)^{M/4}\right]} - E^{\frac{M-2}{M-1}}\right) \times \exp\left\{-\left(\frac{E^{\frac{1}{M-1}} - e}{1-E}\right)\right\}. \tag{6}$$

This formula is exact for $M = 3$ and approximate for $M > 3$.

Initial results on lean buffering for non-exponential machines have been reported in (Enginarlar et al., 2002). Two distributions of up- and downtime have been considered (Rayleigh and Erlang). It has been shown that LLB for these cases is smaller than that for the exponential case. However, (Enginarlar et al., 2002) did not provide a sufficiently complete characterization of lean buffering in non-exponential production systems. In particular, it did not quantify how different types of up- and downtime distributions affect LLB and did not investigate relative effects of uptime vs. downtime on LLB.

The goal of this paper is to provide a method for selecting LLB in serial lines with non-exponential machines. We consider Weibull, gamma, and log-normal reliability models under various assumptions on their parameters. This allows us to place their coefficients of variations at will and study LLB as a function of up- and downtime variability. Moreover, since each of these distributions is defined by two parameters, selecting them appropriately allows us to analyze the lean buffering for 26 various shapes of density functions, ranging from almost delta-function to almost uniform. This analysis leads to the quantification of both influences of distribution shapes on LLB and effects of up- and downtime on LLB. Based of these results, we develop a method for selecting LLB in serial lines with Weibull, gamma, and log-normal reliability characteristics and conjecture that the same method can be used for selecting LLB in serial lines with arbitrary unimodal distributions of up- and downtime.

1.2 Motivation for considering non-exponential machines

The case of non-exponential machines is important for at least two reasons:

First, in practice the machines often have up- and downtime distributed non-exponentially. As the empirical evidence (Inman, 1999) indicates, the coefficients of variation, CV_{up} and CV_{down} of these random variables are often less than 1; thus, the distributions cannot be exponential. Therefore, an analytical characterization of k_E for non-exponential machines is of theoretical importance.

Second, such a characterization is of practical importance as well. Indeed, it can be expected that k_E^{exp} is the upper bound of k_E for $CV < 1$ and, moreover, k_E might be substantially smaller than k_E^{exp}. This implies that a smaller buffer capacity is necessary to achieve the desired line efficiency E when the machines are non-exponential. Thus, selecting LLB based on realistic, non-exponential reliability characteristics would lead to increased leanness of production systems.

1.3 Difficulties in studying the non-exponential case

Analysis of lean buffering in serial production lines with non-exponential machines is complicated, as compared with the exponential case, by the reasons outlined in Table 1 . Especially damaging is the first one, which practically precludes analytical investigation. The other reasons lead to a combinatorially increasing number of cases to be investigated. In this work, we partially overcome these difficulties by

Table 1. Difficulties of the non-exponential case as compared with the exponential one

Exponential case	Non-exponential case
Analytical methods for evaluating PR are available	No analytical methods for evaluating PR are available
Machine up- and downtimes are distributed identically (i.e., exponentially).	Machine up- and downtimes may have different distributions.
Coefficients of variation of machine up- and downtimes are identical and equal to 1.	Coefficients of variation of machine up- and downtimes may take arbitrary positive values and may be non-identical.
All machines in the system have the same type of up- and downtime distributions (i.e., exponential).	Each machine in the system may have different types of up- and downtime distributions.

using numerical simulations and by restricting the number of distributions and coefficients of variation analyzed.

1.4 Related literature

The majority of quantitative results on buffer capacity allocation in serial production lines address the case of exponential or geometric machines (Buzacott, 1967; Caramanis, 1987; Conway et al., 1988; Smith and Daskalaki, 1988; Jafari and Shanthikumar, 1989; Park, 1993; Seong et al., 1995; Gershwin and Schor, 2000). Just a few numerical/empirical studies are devoted to the non-exponential case. Specifically, two-stage coaxian type completion time distributions are considered by Altiok and Stidham (1983), Chow (1987), Hillier and So (1991a,b), and the effects of log-normal processing times are analyzed by Powell (1994), Powell and Pyke (1998), Harris and Powell (1999). These papers consider lines with reliable machines having random processing time. Another approach is to develop methods to extend the results obtained for such cases to unreliable machines with deterministic processing time (Tempelmeier, 2003). Phase-type distributions to model random processing time and reliability characteristics are analyzed by Altiok (1985, 1989), Altiok and Ranjan (1989), Yamashita and Altiok (1998), but the resulting methods are computationally intensive and can be used only for short lines with small buffers (e.g., two-machine lines with buffers of capacity less than six). Finally, as it was mentioned in the Introduction, initial results on lean level of buffering in serial lines with Rayleigh and Erlang machines have been reported in (Enginarlar et al., 2002).

1.5 Contributions of this paper

The main results derived in this paper are as follows:

- LLB is not very sensitive to the type of up- and downtime distributions and depends mostly on their coefficients of variation (CV_{up} and CV_{down}).
- LLB is more sensitive to CV_{down} than to CV_{up}, but this difference in sensitivities is not too large (typically, within 20%).
- In serial lines with M machines having Weibull, gamma, and log-normal distributions of up- and downtime with CV_{up} and CV_{down} less than 1, LLB can be selected using the following upper bound:

$$k_E(M, E, e, CV_{up}, CV_{down}) \leq \frac{\max\{0.25, CV_{up}\} + \max\{0.25, CV_{down}\}}{2} k_E^{exp}(M, E, e), \quad (7)$$

  where k_E^{exp} is given by (5), (6). This bound is referred to as the *empirical law*. It is conjectured that this bound holds for all unimodal up- and downtime distributions with $CV_{up} < 1$ and $CV_{down} < 1$.
- Although for some values of CV_{up} and CV_{down}, bound (7) may not be too tight, it still leads to a reduction of lean buffering by a factor of up to 4, as compared to LLB based on the exponential assumption.

1.6 Paper organization

In Section 2, the model of the production system under consideration is introduced and the problems addressed are formulated. Section 3 describes the approach of this study. Sections 4 and 5 present the main results pertaining, respectively, to systems with machines having identical and non-identical coefficients of variation of up- and downtime. In Section 6, serial lines with machines having arbitrary, i.e., general, reliability models are discussed. Finally, in Section 7, the conclusions are formulated.

2 Model and problem formulation

2.1 Model

The block diagram of the production system considered in this work is shown in Figure 1, where the circles represent the machines and the rectangles are the buffers. Assumptions on the machines and buffers, described below, are similar to those of (Enginarlar et al., 2003a) with the only difference that up- and downtime distributions are not exponential. Specifically, these assumptions are:

(i) Each machine m_i, $i = 1, \ldots, M$, has two states: up and down. When up, the machine is capable of processing one part per cycle time; when down, no production takes place. The cycle times of all machines are the same.

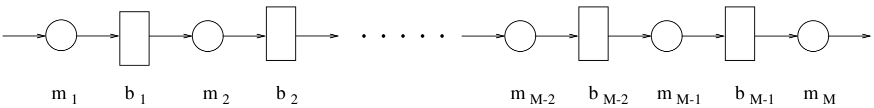

Fig. 1. Serial production line

(ii) The up- and downtime of each machine are random variables measured in units of the cycle time. In other words, uptime (respectively, downtime) of length $t \geq 0$ implies that the machine is up (respectively, down) during t cycle times. The up- and downtime are distributed according to one of the following probability density functions, referred to as *reliability models:*

(a) Weibull, i.e.,

$$f_{up}^{W}(t) = p^P e^{-(pt)^P} P t^{P-1},$$
$$f_{down}^{W}(t) = r^R e^{-(rt)^R} R t^{R-1}, \tag{8}$$

where $f_{up}^{W}(t)$ and $f_{down}^{W}(t)$ are the probability density functions of up- and downtime, respectively and (p, P) and (r, R) are their parameters. (Here, and in the subsequent distributions, the parameters are positive real numbers). These distributions are denoted as $W(p, P)$ and $W(r, R)$, respectively.

(b) Gamma, i.e.,

$$f_{up}^{g}(t) = p e^{-pt} \frac{(pt)^{P-1}}{\Gamma(P)},$$
$$f_{down}^{g}(t) = r e^{-rt} \frac{(rt)^{R-1}}{\Gamma(R)}, \tag{9}$$

where $\Gamma(x)$ is the gamma function, $\Gamma(x) = \int_0^\infty s^{x-1} e^{-s} ds$. These distributions are denoted as $g(p, P)$ and $g(r, R)$, respectively.

(c) Log-normal, i.e.,

$$f_{up}^{LN}(t) = \frac{1}{\sqrt{2\pi} P t} e^{-\frac{(ln(t)-p)^2}{2P^2}},$$
$$f_{down}^{LN}(t) = \frac{1}{\sqrt{2\pi} R t} e^{-\frac{(ln(t)-r)^2}{2R^2}}. \tag{10}$$

We denote these distributions as $LN(p, P)$ and $LN(r, R)$, respectively.

The expected values, variances, and coefficients of variation of distributions (8)–(10) are given in Table 2.

(iii) The parameters of distributions (8)–(10) are selected so that the machine efficiencies, i.e.,

$$e = \frac{T_{up}}{T_{up} + T_{down}}, \tag{11}$$

and, moreover, T_{up}, T_{down}, CV_{up}, and CV_{down} of all machines are identical for all reliability models, i.e.,

$$T_{up} = p^{-1} \Gamma\left(1 + \frac{1}{P}\right) \qquad \text{(Weibull)}$$

Table 2. Expected value, variance, and coefficient of variation of up- and downtime distributions considered

	Gamma	Weibull	Log-normal
T_{up}	P/p	$p^{-1}\Gamma(1+1/P)$	$e^{p+P^2/2}$
T_{down}	R/r	$r^{-1}\Gamma(1+1/R)$	$e^{r+R^2/2}$
σ^2_{up}	P/p^2	$p^{-2}[\Gamma(1+2/P)-\Gamma^2(1+1/P)]$	$e^{2p+P^2}(e^{P^2}-1)$
σ^2_{down}	R/r^2	$r^{-2}[\Gamma(1+2/R)-\Gamma^2(1+1/R)]$	$e^{2r+R^2}(e^{R^2}-1)$
CV_{up}	$1/\sqrt{P}$	$\sqrt{\Gamma(1+2/P)-\Gamma^2(1+1/P)}\big/\Gamma(1+1/P)$	$\sqrt{e^{P^2}-1}$
CV_{down}	$1/\sqrt{R}$	$\sqrt{\Gamma(1+2/R)-\Gamma^2(1+1/R)}\big/\Gamma(1+1/R)$	$\sqrt{e^{R^2}-1}$

$$
\begin{aligned}
&= \frac{P}{p} \quad \text{(gamma)}\\
&= e^{p+P^2/2} \quad \text{(log-normal);}\\
T_{down} &= r^{-1}\Gamma\left(1+1/R\right) \quad \text{(Weibull)}\\
&= \frac{R}{r} \quad \text{(gamma)}\\
&= e^{r+R^2/2} \quad \text{(log-normal);}
\end{aligned}
$$

$$
\begin{aligned}
CV_{up} &= \frac{\sqrt{\Gamma(1+2/P)-\Gamma^2(1+1/P)}}{\Gamma(1+1/P)} \quad \text{(Weibull)}\\
&= \frac{1}{\sqrt{P}} \quad \text{(gamma)}\\
&= \sqrt{e^{P^2}-1} \quad \text{(log-normal);}\\
CV_{down} &= \frac{\sqrt{\Gamma(1+2/R)-\Gamma^2(1+1/R)}}{\Gamma(1+1/R)} \quad \text{(Weibull)}\\
&= \frac{1}{\sqrt{R}} \quad \text{(gamma)}\\
&= \sqrt{e^{R^2}-1} \quad \text{(log-normal).}
\end{aligned}
$$

(iv) Buffer b_i, $i = 1, \ldots, M-1$ is of capacity $0 \le N \le \infty$.

(v) Machine m_i, $i = 2, \ldots, M$, is starved at time t if it is up at time t, buffer b_{i-1} is empty at time t and m_{i-1} does not place any work in this buffer at time t. Machine m_1 cannot be starved.

(vi) Machine m_i, $i = 1, \ldots, M-1$, is blocked at time t if it is up at time t, buffer b_i is full at time t and m_{i+1} fails to take any work from this buffer at time t. Machine m_M cannot be blocked.

Remark 1.

- Assumptions (i)–(iii) imply that all machines are identical from all points of view except, perhaps, for the nature of up- and downtime distributions. The buffers are also assumed to be of equal capacity (see (iv)). We make these assumptions in order to provide a compact characterization of lean buffering.
- Assumption (ii) implies, in particular, that time-dependent, rather than operation-dependent failures, are considered. This failure mode simplifies the analysis and results in just a small difference in comparison with operation-dependent failures.

2.2 Notations

Each machine considered in this paper is denoted by a pair

$$[D_{up}(p, P), D_{down}(r, R)]_i, \qquad i = 1, \dots, M, \tag{12}$$

where $D_{up}(p, P)$ and $D_{down}(r, R)$ represent, respectively, the distributions of up- and downtime of the i-th machine in the system, D_{up} and $D_{down} \in \{W, g, LN\}$. The serial line with M machines is denoted as

$$\{[D_{up}, D_{down}]_1, \dots, [D_{up}, D_{down}]_M\}. \tag{13}$$

If all machines have identical distribution of uptimes and downtimes, the line is denoted as

$$\{[D_{up}(p, P), D_{down}(r, R)]_i, i = 1, \dots, M\}. \tag{14}$$

If, in addition, the types of up- and downtime distributions are the same, the notation for the line is

$$\{[D(p, P), D(r, R)]_i, i = 1, \dots, M\}. \tag{15}$$

Finally, if up- and downtime distributions of the machines are not necessarily W, g, or LN but are general in nature, however, unimodal, the line is denoted as

$$\{[G_{up}, G_{down}]_1, \dots, [G_{up}, G_{down}]_M\}. \tag{16}$$

2.3 Problems addressed

Using the parameterizations (1), (2), the model (i)–(vi), and the notations (12)–(16), this paper is intended to

- develop a method for calculating *Lean Level of Buffering* in production lines (13)–(15) under the assumption that the coefficients of variation of up- and downtime, CV_{up} and CV_{down}, are identical, i.e., $CV_{up} = CV_{down} = CV$;
- develop a method of calculating LLB in production lines (13)–(15) for the case of $CV_{up} \neq CV_{down}$;
- extend the results obtained to production lines (16).

Solutions of these problems are presented in Sections 4–6 while Section 3 describes the approach used in this work.

3 Approach

3.1 General considerations

Since LLB depends on line efficiency E, the calculation of k_E requires the knowledge of the production rate, PR, of the system. Unfortunately, as it was mentioned earlier, no analytical methods exist for evaluating PR in serial lines with either Weibull, or gamma, or log-normal reliability characteristics. Approximation methods are also hardly applicable since, in our experiences, even 1%-2% errors in the production rate evaluation (due to the approximate nature of the techniques) often lead to much larger errors (up to 20%) in lean buffering characterization. Therefore, the only method available is the Monte Carlo approach based on numerical simulations. To implement this approach, a MATLAB code was constructed, which simulated the operation of the production line defined by assumptions (i)–(vi) of Section 2. Then, a set of representative distributions of up- and downtime was selected and, finally, for each member of this set, PR and LLB were evaluated with guaranteed statistical characteristics. Each of these steps is described below in more detail.

3.2 Up- and downtime distributions analyzed

The set of 26 downtime distributions analyzed in this work is shown in Table 3, where the notations introduced in Section 2.1 are used. These distributions are classified according to their coefficients of variation, CV_{down}, which take values from the set $\{0.1, 0.25, 0.5, 0.75, 1.0\}$. The analysis of LLB for this set is intended to reveal the behavior of k_E as a function of CV_{down}.

To investigate the effect of the average downtime, the distributions of Table 3 have been classified according to T_{down}, which takes values 20 and 100.

An illustration of a few of the downtime distributions included in Table 3 is given in Figure 2 for $CV_{down} = 0.5$. As one can see, the shapes of the distributions included in Table 3 range from "almost" uniform to "almost" δ-function.

Table 3. Downtime distributions considered

CV_{down}	$T_{down} = 20$	$T_{down} = 100$
0.1	$g(5, 100)$, $W(0.048,\ 12.15)$, $LN(2.99, 0.1)$	$g(1, 100)$, $W(0.01,\ 12.15)$, $LN(4.602, 0.1)$
0.25	$g(0.8, 16)$, $W(0.046,\ 4.54)$, $LN(2.97, 0.25)$	$g(0.16, 16)$, $W(0.009,\ 4.54)$, $LN(4.57, 0.25)$
0.5	$g(0.2, 4)$, $W(0.044,\ 2.1)$, $LN(2.88, 0.49)$	$g(0.04, 4)$, $W(0.009,\ 2.1)$, $LN(4.49, 0.49)$
0.75	$g(0.09, 1.8)$, $W(0.046,\ 1.35)$, $LN(2.77, 0.66)$	$g(0.018, 1.8)$, $W(0.009,\ 1.35)$, $LN(4.38, 0.66)$
1.00	$LN(2.65, 0.83)$	$LN(4.26, 0.83)$

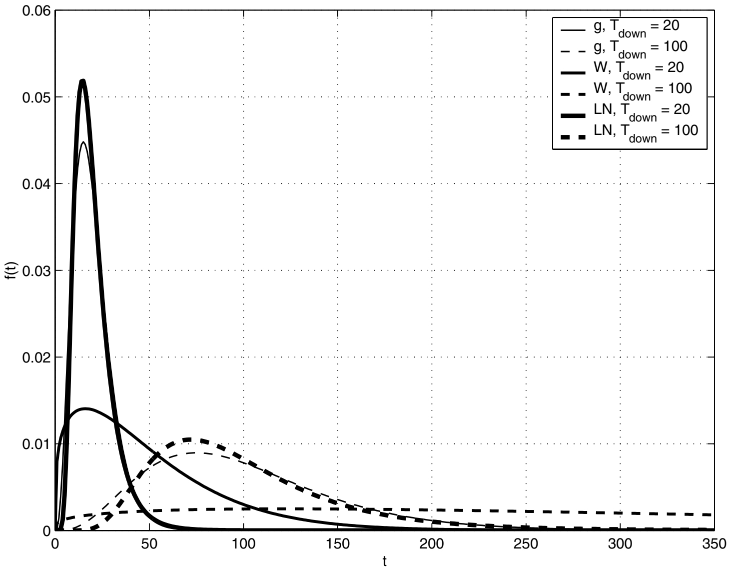

Fig. 2. Different distributions with identical coefficients of variation ($CV_{down} = 0.5$)

The uptime distributions, corresponding to the downtime distributions of Table 3, have been selected as follows: For a given machine efficiency, e, the average uptime was chosen as

$$T_{up} = \frac{e}{1-e} T_{down}.$$

Next, CV_{up} was selected as $CV_{up} = CV_{down}$, when the case of identical coefficients of variation of up- and downtime was considered; otherwise CV_{up} was selected as a constant independent of CV_{down}. Finally, using these T_{up} and CV_{up}, the distribution of uptime was selected to be the same as that of the downtime, if the case of identical distributions was analyzed; otherwise it was selected as any other distribution from the set $\{W, g, LN\}$. For instance, if the downtime was distributed according to $D_{down}(r, R) = g(0.018, 1.8)$ and e was 0.9, the uptime distribution was selected as

$$D_{up}(p, P) = \begin{cases} g(0.002, 1.8) & \text{for} \quad CV_{up} = CV_{down}, \\ g(0.0044, 4) & \text{for} \quad CV_{up} = 0.5, \end{cases}$$

or

$$D_{up}(p, P) = \begin{cases} LN(6.69, 0.47) & \text{for} \quad CV_{up} = CV_{down}, \\ LN(2.88, 0.49) & \text{for} \quad CV_{up} = 0.5. \end{cases}$$

Remark 2. Both CV_{up} and CV_{down} considered are less than 1 because, according to the empirical evidence of (Inman, 1999), the equipment on the factory floor often satisfies this condition. In addition, it has been shown by Li and Meerkov (2005) that CV_{up} and CV_{down} are less than 1 if the breakdown and repair rates of the machines are increasing functions of time, which often takes place in reality.

3.3 Parameters selected

In all systems analyzed, particular values of M, E, and e have been selected as follows:

(a) The number of machines in the system, M: Since, as it was shown in (Enginarlar et al., 2002), k_E^{exp} is not very sensitive to M if $M \geq 10$, the number of machines in the system was selected to be 10. For verification purposes, we analyzed also serial lines with $M = 5$.

(b) Line efficiency, E: In practice, production lines are often operated close to their maximum capacity. Therefore, for the purposes of simulation, E was selected to belong to the set $\{0.85, 0.9, 0.95\}$. For the purposes of verification, additional values of E analyzed were $\{0.7, 0.8\}$.

(c) Machine efficiency, e: Although in practice e may have widely different values (e.g., smaller in machining operations and much larger in assembly), to obtain a manageable set of systems for simulation, e was selected from the set $\{0.85, 0.9, 0.95\}$. For verification purposes, we considered $e \in \{0.6, 0.7, 0.8\}$.

3.4 Systems analyzed

Specific systems of the form (15) considered in this work are:

$$\begin{aligned} \{[W(p,P), W(r,R)]_i, i = 1, \ldots, 10\}, \\ \{[g(p,P), g(r,R)]_i, i = 1, \ldots, 10\}, \\ \{[LN(p,P), LN(r,R)]_i, i = 1, \ldots, 10\}. \end{aligned} \tag{17}$$

Systems of the form (13) have been formed as follows: For each machine $m_i,\ i = 1, \ldots, 10$, the up- and downtime distributions were chosen from the set $\{W, g, LN\}$ equiprobably and independently of each other and all other machines in the system. As a result, the following two lines were selected:

$$\begin{aligned} &\text{Line 1:} && \{(g,W), (LN,LN), (W,g), (g,LN), (g,W), \\ &&& (LN,g), (W,W), (g,g), (LN,W), (g,LN)\}, \\ &\text{Line 2:} && \{(W,LN), (g,W), (LN,W), (W,g), (g,LN), \\ &&& (g,W), (W,W), (LN,g), (g,W), (LN,LN)\}. \end{aligned} \tag{18}$$

We will use notations $A \in \{(17)\}$, $A \in \{(18)\}$ or $A \in \{(17), (18)\}$ to indicate that line A is one of (17), or one of (18), and one of (17) and (18), respectively.

Lines (17) and (18) are analyzed in Sections 4 and 5 for the cases of $CV_{up} = CV_{down}$ and $CV_{up} \neq CV_{down}$, respectively.

3.5 Evaluation of the production rate

To evaluate the production rate in systems (17) and (18), using the MATLAB code and the up- and downtime distributions discussed in Sections 3.1–3.3, zero initial conditions of all buffers have been assumed and the states of all machines at the initial time moment have been selected "up". The first 100,000 cycle times were considered as warm-up period. The subsequent 1,000,000 cycle times were used for statistical evaluation of PR. Each simulation was repeated 10 times, which resulted in 95% confidence intervals of less than 0.0005.

3.6 Evaluation of LLB

The lean buffering, k_E, necessary and sufficient to ensure line efficiency E, was evaluated using the following procedure:

For each model of serial line (13)–(15), the production rate was evaluated first for $N = 0$, then for $N = 1$, and so on, until the production rate $PR = E \cdot PR_\infty$ was achieved. Then k_E was determined by dividing the resulting N_E by the machine average downtime (in units of the cycle time).

Remark 3. Although, as it is well known (Hillier and So, 1991b), the optimal allocation of a fixed total buffer capacity is non-uniform, to simplify the analysis we consider only uniform allocations. Since the optimal (i.e., inverted bowl) allocation typically results in just $1 - 2\%$ throughput improvement in comparison with the uniform allocation, for the sake of simplicity we consider only the latter case.

4 LLB in serial lines with $CV_{up} = CV_{down} = CV$

4.1 System $\{[D(p, P), D(r, R)]_i, i = 1, \ldots, 10\}$

Figures 3 and 5 present the simulation results for production lines (17) for all distributions of Table 3. These figures are arranged as matrices where the rows and columns correspond to $e \in \{0.85, 0.9, 0.95\}$ and $E \in \{0.85, 0.9, 0.95\}$, respectively. Since, due to space considerations, the graphs in Figures 3 and 5 are congested and may be difficult to read, one of them is shown in Figure 4 in a larger scale. (The dashed lines in Figs. 3–5 will be discussed in Sect. 4.3.) Examining these data, the following may be concluded:

- As expected, k_E for non-exponential machines is smaller than k_E^{exp}. Moreover, k_E is a monotonically increasing function of CV. In addition, $k_E(CV)$ is convex, which implies that reducing larger CV's leads to larger reduction of k_E than reducing smaller CV's.
- Function $k_E(CV)$ seems to be polynomial in nature. In fact, each curve of Figures 3 and 5 can be approximated by a polynomial of an appropriate order. However, since these approximations are "parameter-dependent" (i.e., different polynomials must be used for different e and E), they are of small practical importance, and, therefore, are not reported here.
- Since for every pair (E, e), corresponding curves of Figures 3 and 5 are identical, it is concluded that k_E is not dependent of T_{up} and T_{down} explicitly but only through the ratio e. In other words, the situation here is the same as in lines with exponential machines (see (5), (6)).

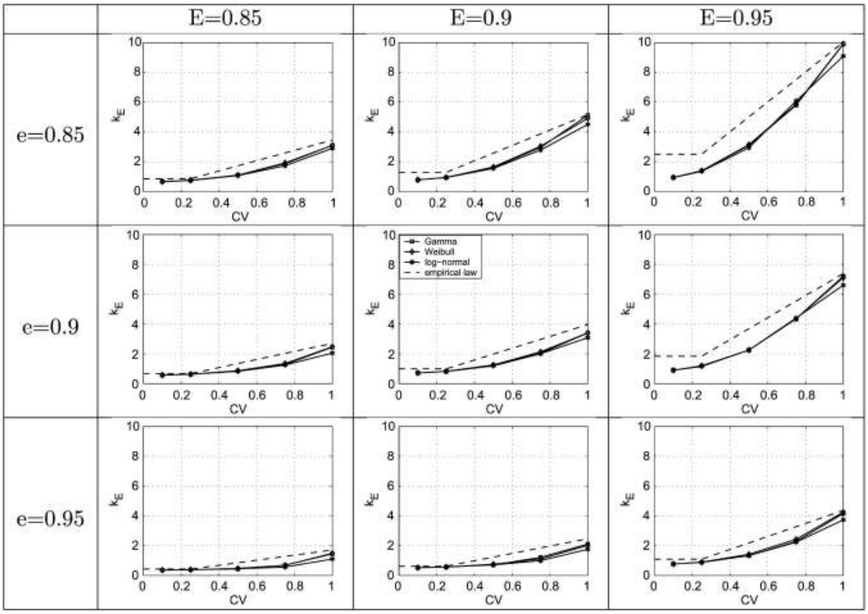

Fig. 3. LLB versus CV for systems (17) with $T_{down} = 20$

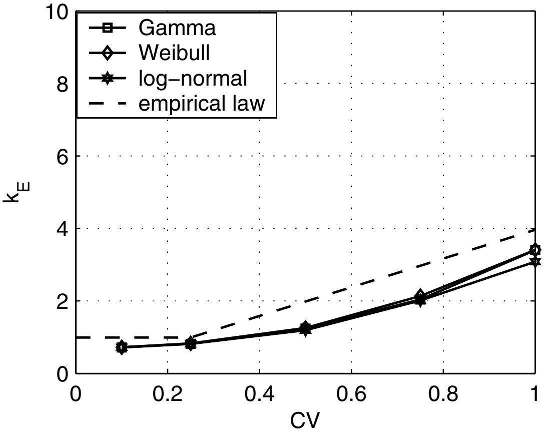

Fig. 4. LLB versus CV for system $\{(D(p, P), D(r, R))_i, i = 1, \dots, 10\}$ with $T_{down} =$ 20, $e = 0.9$, $E = 0.9$

– Finally, and perhaps most importantly, the behavior of k_E as a function of CV is almost independent of the type of up- and downtime distributions considered. Indeed, let $k_E^A(CV)$ denote LLB for line $A \in \{(17)\}$ with $CV \in \{0.1, 0.25, 0.5, 0.75, 1.0\}$. Then the sensitivity of k_E to up- and downtime distributions may be characterized by

$$\epsilon_1(CV) = \max_{A,B \in \{(17)\}} \left| \frac{k_E^A(CV) - k_E^B(CV)}{k_E^A(CV)} \right| \cdot 100\%. \tag{19}$$

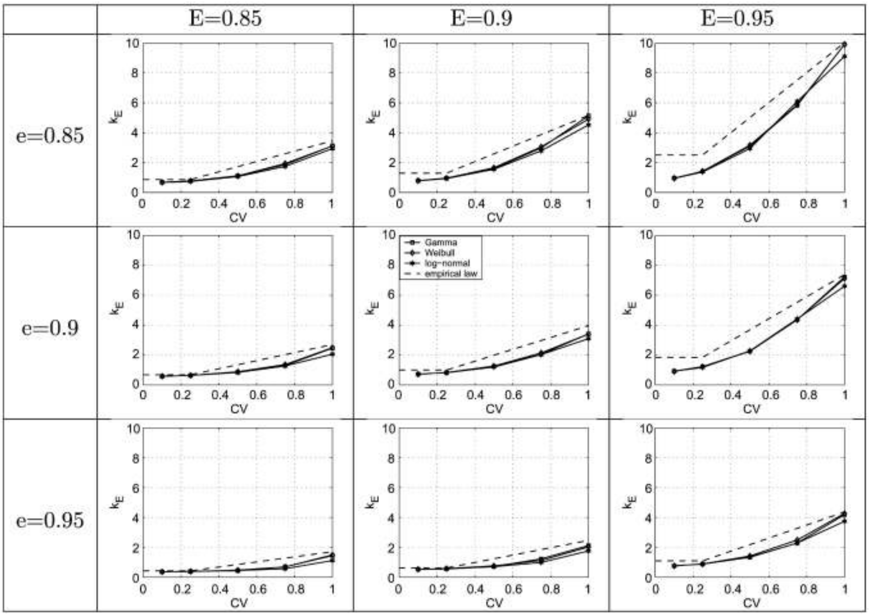

Fig. 5. LLB versus CV for systems (17) with $T_{down} = 100$

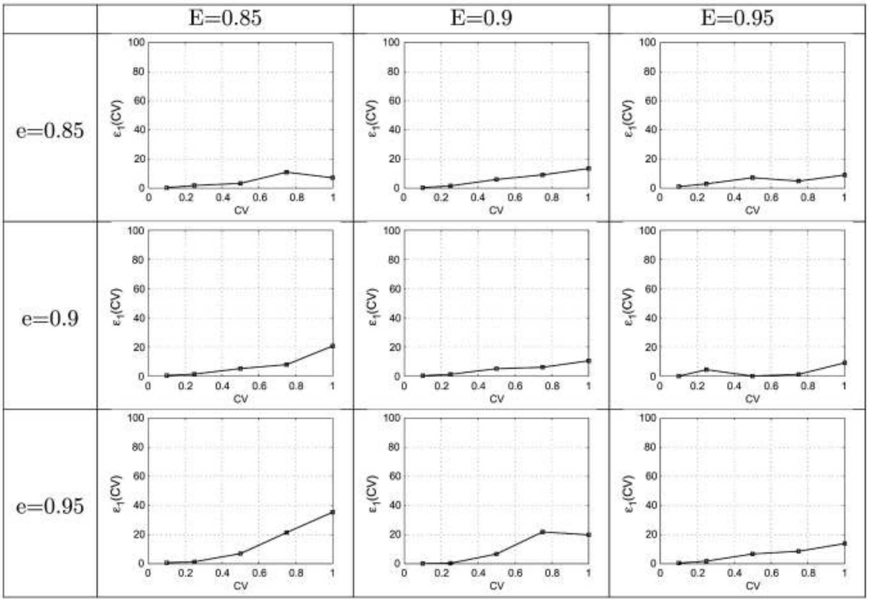

Fig. 6. Sensitivity of LLB to the nature of up- and downtime distributions for systems (17)

Function $\epsilon_1(CV)$ is illustrated in Figure 6. As one can see, in most cases it takes values within 10%. Thus, it is possible to conclude that for all practical purposes k_E depends on the coefficients of variation of up- and downtime, rather than on actual distribution of these random variables.

4.2 System $\{[D(p,P),D(r,R)]_1,\ldots,[D(p,P),D(r,R)]_{10}\}$

Figures 7 and 8 present the simulation results for lines (18), while Figure 9 characterizes the sensitivity of k_E to up- and downtime distributions. This sensitivity is calculated according to (19) with the only difference that the max is taken over $A,B \in \{(18)\}$. Based on these data, we affirm that the conclusions formulated in Section 4.1 hold for production lines of the type (13) as well.

4.3 Empirical law

4.3.1 Analytical expression

Simulation results reported above provide a characterization of k_E for $M = 10$ and E and $e \in \{0.85, 0.9, 0.95\}$. How can k_E be determined for other values of M, E, and e? Obviously, simulations for all values of these variables are impossible. Even for particular values of M, E, and e, simulations take a very long time: Figures 3 and 5 required approximately one week of calculations using 25 Sun workstations working in parallel. Therefore, an analytical method for evaluating k_E for all values of M, E, e, and CV is desirable. Although an exact characterization of the function $k_E = k_E(M,E,e,CV)$ is all but impossible, results of Sections 4.1 and 4.2 provide an opportunity for introducing an upper bound of k_E as a function of all four variables. This upper bound is based on the expression of $k_E^{exp} = k_E^{exp}(M,E,e)$, given by (5), (6), and the fact that all curves of Figures 3, 5 and 7, 8 are below the linear function of CV with the slope k_E^{exp}, if $0.25 < CV \leq 1$. For $0 < CV \leq 0.25$, all curves are below the constant $0.25k_E^{exp}$. Thus, the following piece-wise linear upper bound for k_E may be introduced:

$$k_E(M,E,e,CV) \leq \max\{0.25, CV\} k_E^{exp}(M,E,e), \qquad CV \leq 1. \tag{20}$$

This expression, referred to as *the empirical law*, is illustrated in Figures 3-5 and 7, 8 by the broken lines.

The tightness of this bound can be characterized by the function

$$\epsilon_2(CV) = \max_{A \in \{(17),(18)\}} \frac{k_E^{\text{upper bound}} - k_E^A}{k_E^A} \cdot 100\%, \qquad CV \leq 1, \tag{21}$$

where $k_E^{\text{upper bound}}$ is the right-hand-side of (20). Function $\epsilon_2(CV)$ is illustrated in Figure 10. Although, as one can see, the empirical law is quite conservative, its usage still leads to up to 400% reduction of buffering, as compared with that based on the exponential assumption (see Figs. 3, 5 and 7, 8).

Remark 4. As it was pointed out above, the curves of Figures 3, 5 and 7, 8 are polynomial in nature. This, along with the quadratic dependence of performance

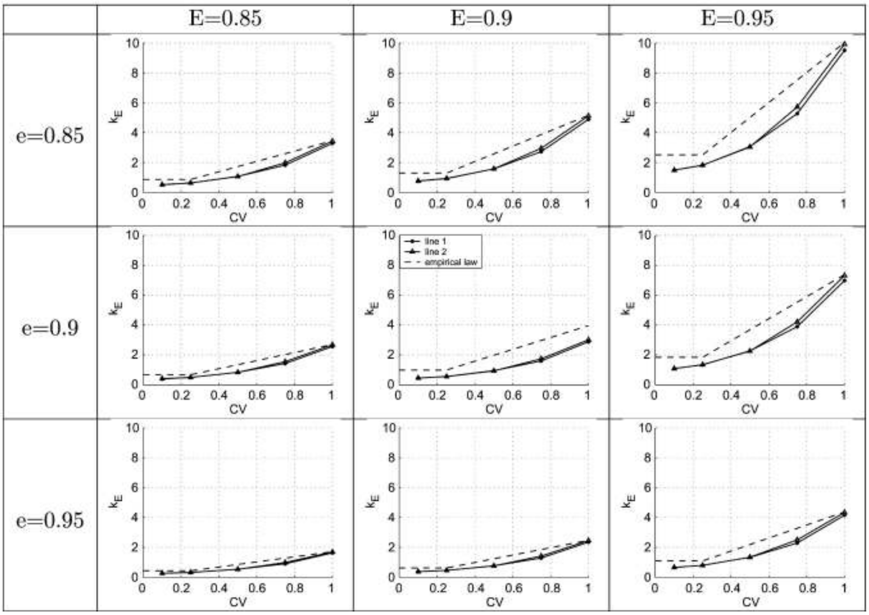

Fig. 7. LLB versus CV for systems (18) with $T_{down} = 20$

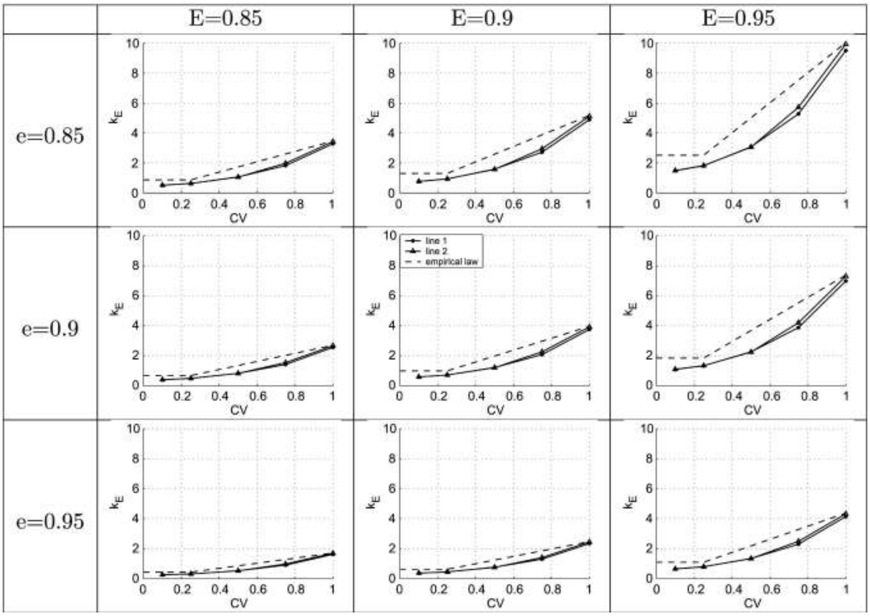

Fig. 8. LLB versus CV for systems (18) with $T_{down} = 100$

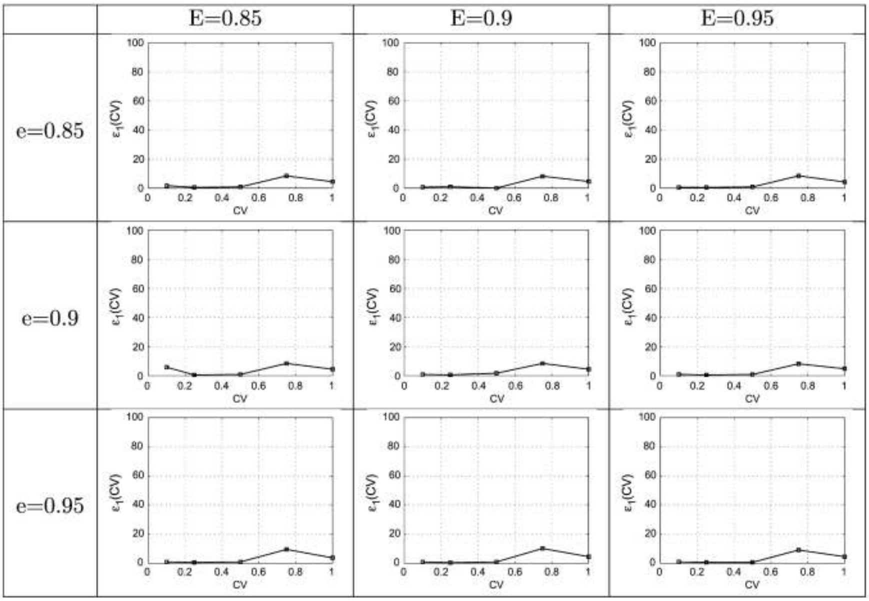

Fig. 9. Sensitivity of LLB to the nature of up- and downtime distributions for systems (18)

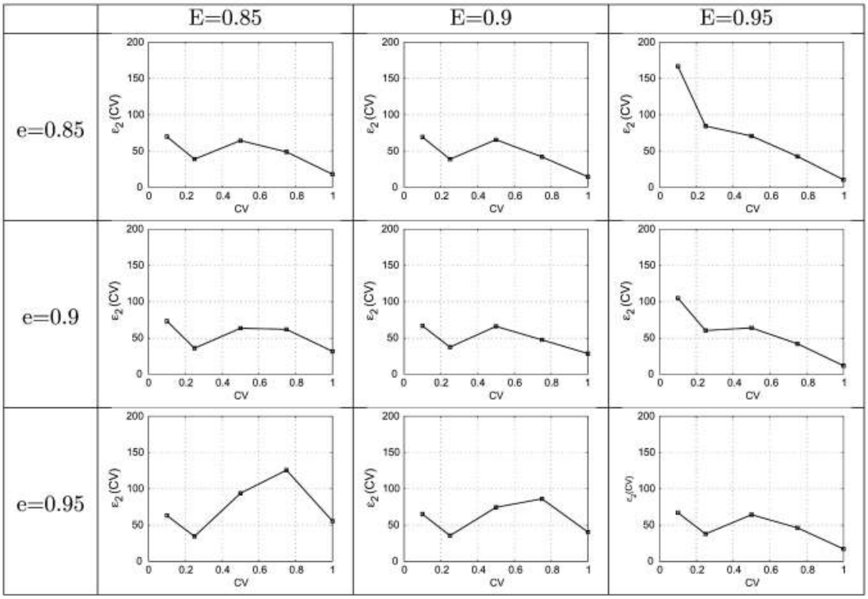

Fig. 10. The tightness of the empirical law (20)

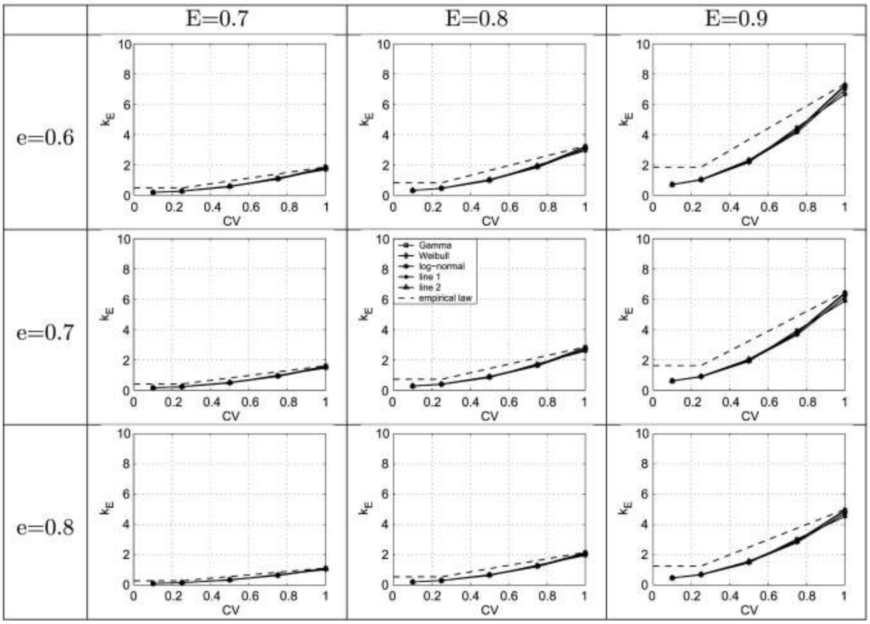

Fig. 11. Verification: LLB versus CV for system $\{(D(p,P), D(r,R))_i, i = 1,\dots,5\}$ with $T_{down} = 10$

measures on CV in $G/G/1$ queues, might lead to a temptation to approximate these curves by polynomials. This, however, proved to be practically impossible, since for various values of M, E, and e, the order and the coefficients of the polynomials would have to be selected differently. This, together with the fact that only one point is known analytically (i.e., k_E^{exp}), leads to the selection of the piece-wise linear approximation (20).

4.3.2 Verification

To verify the empirical law (20), production lines (17) and (18) were simulated with parameters M, E, and e other than those considered in Sections 4.1 and 4.2. Specifically, the following parameters have been selected: $M = 5$, $E \in \{0.7, 0.8, 0.9\}$, $e \in \{0.6, 0.7, 0.8\}$, $T_{down} = 10$. (In lines (18), the first 5 machines were selected.) The results are shown in Figure 11. As one can see, the upper bound given by (20) still holds.

5 LLB in serial lines with $CV_{up} \neq CV_{down}$

5.1 Effect of CV_{up} and CV_{down}

The case of $CV_{up} \neq CV_{down}$ is complicated by the fact that CV_{up} and CV_{down} may have different effects on k_E. If this difference is significant, it would be difficult to expect that the empirical law (20) could be extended to the case of unequal coefficients of variation. On the other hand, if CV_{up} and CV_{down} affect k_E in a somewhat similar manner, it would seem likely that (20) might be extended to the case under consideration. Therefore, analysis of effects of CV_{up} and CV_{down} on k_E is of importance. This section is devoted to such an analysis.

To investigate this issue, introduce two functions:

$$k_E(CV_{up}|CV_{down} = \alpha) \tag{22}$$

and

$$k_E(CV_{down}|CV_{up} = \alpha), \tag{23}$$

where

$$\alpha \in \{0.1, 0.25, 0.5, 0.75, 1.0\}. \tag{24}$$

Function (22) describes k_E as a function of CV_{up} given that $CV_{down} = \alpha$, while (23) describes k_E as a function of CV_{down} given that $CV_{up} = \alpha$. If for all α and $\beta \in \{0.1, 0.25, 0.5, 0.75, 1.0\}$,

$$k_E(CV_{down} = \beta|CV_{up} = \alpha) < k_E(CV_{up} = \beta|CV_{down} = \alpha) \tag{25}$$

when $\alpha > \beta$, it must be concluded that CV_{down} has a larger effect on k_E than CV_{up}. If the inequality is reversed, CV_{up} has a stronger effect. Finally, if (25) holds for some α and β from (24) and does not hold for others, the conclusion would be that, in general, neither has a dominant effect.

To investigate which of these situations takes place, we evaluated functions (22) and (23) using the approach described in Section 3. Some of the results for Weibull distribution are shown in Figure 12 (where the broken lines and CV_{eff} will be defined in Sect. 5.2). Similar results were obtained for gamma and log-normal distributions as well (see Enginarlar et al., 2003b for details). From these results, the following can be concluded:

- For all α and β, such that $\alpha > \beta$, inequality (25) takes place. Thus, CV_{down} has a larger effect on k_E than CV_{up}.
- However, since each pair of curves (22), (23) corresponding to the same α are close to each other, the difference in the effects of CV_{up} and CV_{down} is not too dramatic. To analyze this difference, introduce the function

$$\begin{aligned}&\epsilon_3^A(CV|CV_{up} = CV_{down} = \alpha)\\&= \frac{k_E^A(CV_{up}{=}CV|CV_{down} = \alpha) - k_E^A(CV_{down}{=}CV|CV_{up}{=}\alpha)}{k_E^A(CV_{up}{=}CV|CV_{down}{=}\alpha)} \cdot 100\,,\end{aligned} \tag{26}$$

  where $A \in \{W, g, LN\}$. The behavior of this function for Weibull distribution is shown in Figure 13 (see Enginarlar et al., 2003b for gamma and log-normal distributions). Thus, the effects of CV_{up} and CV_{down} on k_E are not dramatically different (typically within 20% and no more than 40%).

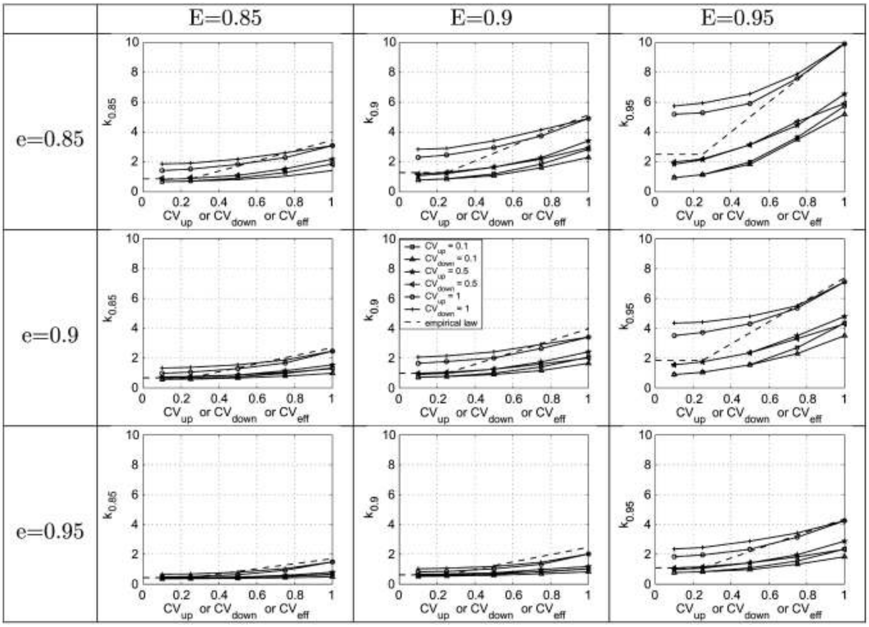

Fig. 12. LLB versus CV for $M = 10$ Weibull machines

5.2 Empirical law

5.2.1 Analytical expression

Since the upper bound (20) is not too tight (and, hence, may accommodate additional uncertainties) and the effects of CV_{up} and CV_{down} on k_E are not dramatically different, the following extension of the empirical law is suggested:

$$k_E(M, E, e, CV_{up}, CV_{down}) \leq \frac{\max\{0.25, CV_{up}\} + \max\{0.25, CV_{down}\}}{2} k_E^{exp}(M, E, e), \quad CV_{up} \leq 1, \quad CV_{down} \leq 1, \tag{27}$$

where, as before, k_E^{exp}, is defined by (5), (6). If $CV_{up} = CV_{down}$, (27) reduces to (20); otherwise, it takes into account different values of CV_{up} and CV_{down}.

The first factor in the right-hand-side of (27) is denoted as CV_{eff}:

$$CV_{eff} = \frac{\max\{0.25, CV_{up}\} + \max\{0.25, CV_{down}\}}{2}. \tag{28}$$

Thus, (27) can be rewritten as

$$k_E \leq CV_{eff} k_E^{exp}(M, E, e). \tag{29}$$

The right-hand-side of (29) is shown in Figure 12 by the broken lines.

The utilization of this law can be illustrated as follows: Suppose $CV_{up} = 0.1$ and $CV_{down} = 1$. Then $CV_{eff} = 0.625$ and, according to (27),

$$k_E \leq 0.625 k_E^{exp}(M, E, e).$$

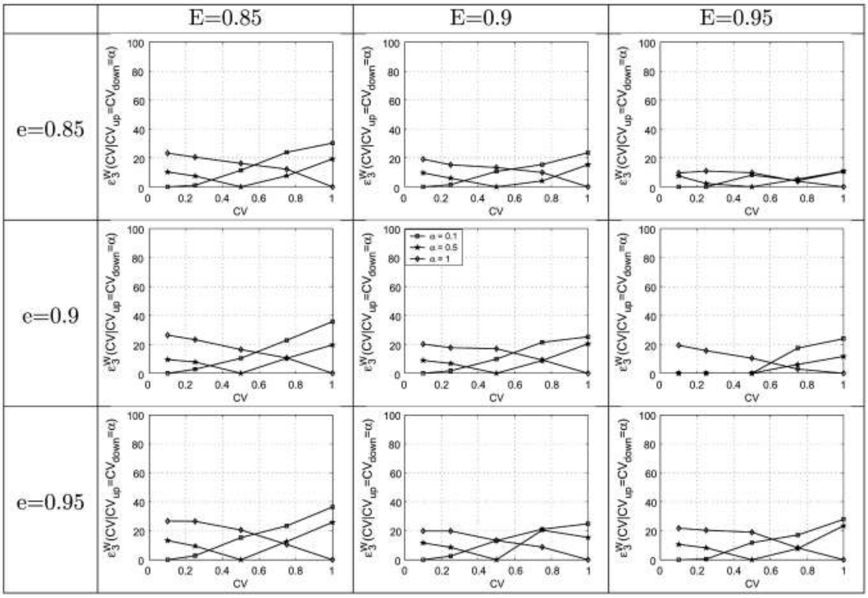

Fig. 13. Function $\epsilon_3^W(CV|CV_{up} = CV_{down} = \alpha)$

Table 4. $\Delta(10, E, e)$ for all $CV_{up} \neq CV_{down}$ cases considered

	E=0.85	E=0.9	E=0.95
$e = 0.85$	0.1016	0.0386	0.0687
$e = 0.9$	0.0425	0.1647	0.1625
$e = 0.95$	0.0402	0.0488	0.1200

To investigate the validity of the empirical law (27), consider the following function:

$$\Delta(M, E, e) = \min_{A \in \{(17)\}} \min_{CV_{up}, CV_{down} \in \{(24)\}} \Big[k_E^{\text{upper bound}}(M, E, e, CV_{eff}) - k_E^A(M, E, e, CV_{up}, CV_{down})\Big], \quad (30)$$

where $k_E^{\text{upper bound}}$ is the right-hand-side of (29), i.e.,

$$k_E^{\text{upper bound}}(M, E, e, CV_{eff}) = CV_{eff} k_E^{exp}(M, E, e).$$

If for all values of its arguments, function $\Delta(M, E, e)$ is positive, the right-hand-side of inequality (27) is an upper bound. The values of $\Delta(10, E, e)$ for $E \in \{0.85, 0.9, 0.95\}$ and $e \in \{0.85, 0.9, 0.95\}$ are shown in Table 4. As one can see, function $\Delta(10, E, e)$ indeed takes positive values. Thus, the empirical law (27) takes place for all distributions and parameters analyzed.

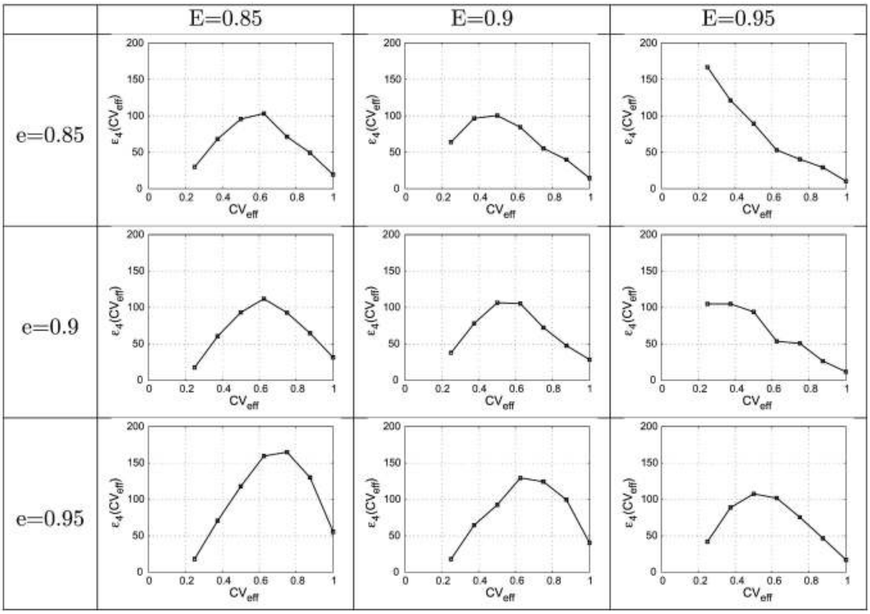

Fig. 14. The tightness of the empirical law (27)

To investigate the tightness of the bound (27), consider the function

$$\epsilon_4(CV_{eff}) = \max_{A \in \{(17)\}} \max_{CV_{up}, CV_{down} \in \{(24)\}} \frac{k_E^{\text{upperbound}}(M, E, e, CV_{eff}) - k_E^A(M, E, e, CV_{up}, CV_{down})}{k_E^A(M, E, e, CV_{up}, CV_{down})} \cdot 100\,. \tag{31}$$

Figure 14 illustrates the behavior of this function. Comparing this with Figure 10, we conclude that the tightness of bound (27) appears to be similar to that of (20).

5.2.2 Verification

To evaluate the validity of the upper bound (27), serial production lines with $M = 5$, $E \in \{0.7, 0.8, 0.9\}$, $e \in \{0.6, 0.7, 0.8\}$, and $T_{up} = 10$ were simulated. For each of these parameters, systems (17) and (18) have been considered. (For system (18), the first 5 machines were selected.) Typical results are shown in Figure 15 (see Enginarlar et al., 2003b for more details). The validity of empirical law (27) for these cases is analyzed using function $\Delta(M, E, e)$, defined in (30) with the only difference that the first min is taken over $A \in \{(17), (18)\}$. Since the values of this function, shown in Table 5, are positive, we conclude that empirical law (27) is indeed verified for all values of M, E, e, and all distributions of up- and downtime considered.

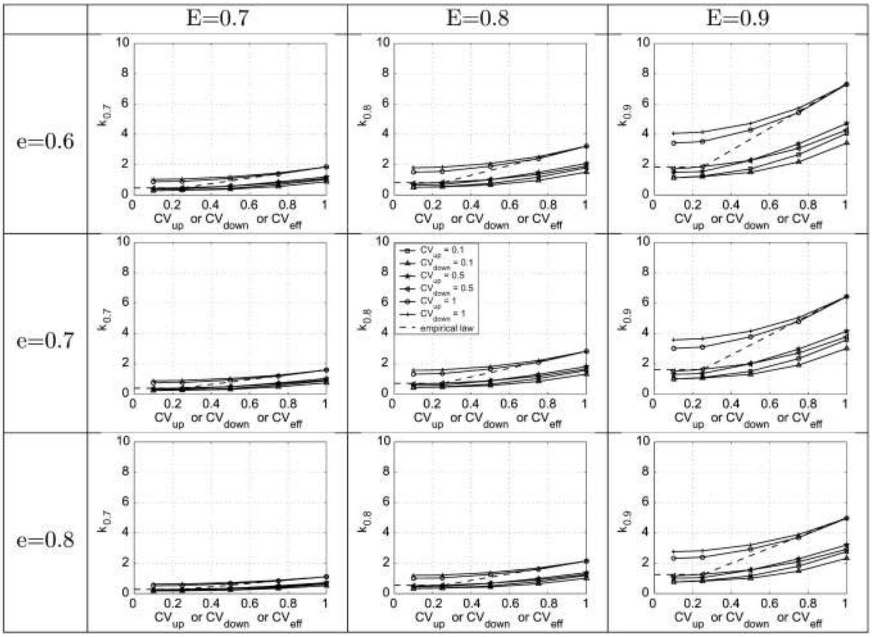

Fig. 15. Verification: LLB versus CV for $M = 5$ Weibull machines

Table 5. Verification: $\Delta(5, E, e)$ for all $CV_{up} \neq CV_{down}$ cases considered

	E=0.7	E=0.8	E=0.9
$e = 0.6$	0.0039	0.0242	0.0547
$e = 0.7$	0.0102	0.0213	0.0481
$e = 0.8$	0.0084	0.0162	0.0355

6 SYSTEM $\{[G_{up}, G_{down}]_1, \ldots, [G_{up}, G_{down}]_M\}$

So far, serial production lines with Weibull, gamma, and log-normal reliability models have been analyzed. It is of interests to extend this analysis to general probability density functions. Based on the results obtained above, the following conjecture is formulated:

The empirical laws (20) and (27) hold for serial production lines satisfying assumptions (i), (iii)–(vi) with up- and downtime having arbitrary unimodal probability density functions.

The verification of this conjecture is a topic for future research.

7 Conclusions

Results described in this paper suggest the following procedure for designing lean buffering in serial production lines defined by assumptions (i)–(vi):

1. Identify the average value and the variance of the up- and downtime, T_{up}, T_{down}, σ^2_{up}, and σ^2_{down}, for all machines in the system (in units of machine cycle time). This may be accomplished by measuring the duration of the up- and downtimes of each machine during a shift or a week of operation (depending on the frequency of occurrence). If the production line is at the design stage, this information may be obtained from the equipment manufacturer (however, typically with a lower level of certainty).
2. Using (5), (6), and T_{up}, T_{down}, determine the level of buffering, necessary and sufficient to obtain the desired efficiency, E, of the production line, if the downtime of all machines were distributed exponentially, i.e., k_E^{exp}.
3. Finally, if $CV_{up} = \frac{\sigma_{up}}{T_{up}} \leq 1$ and $CV_{down} = \frac{\sigma_{down}}{T_{down}} \leq 1$, evaluate the level of buffering for the line with machines under consideration using the empirical law

$$k_E \leq \frac{\max\{0.25, CV_{up}\} + \max\{0.25, CV_{down}\}}{2} \cdot k_E^{exp}.$$

As it is shown in this paper, this procedure leads to a reduction of lean buffering by a factor of up to 4, as compared with that based on the exponential assumption.

References

Altiok T (1985) Production lines with phase-type operation and repair times and finite buffers. International Journal of Production Research 23: 489–498

Altiok T (1989) Approximate analysis of queues. In: Series with phase-type service times and blocking. Operations Research 37: 601–610

Altiok T, Stidham SS (1983) The allocation of interstage buffer capacities in production lines. IIE Transactions 15: 292–299

Altiok T, Ranjan R (1989) Analysis of production lines with general service times and finite buffers: a two-node decomposition approach. Engineering Costs and Production Economics 17: 155–165

Buzacott JA (1967) Automatic transfer lines with buffer stocks. International Journal of Production Research 5: 183–200

Caramanis M (1987) Production line design: a discrete event dynamic system and generalized benders decomposition approach. International Journal of Production Research 25: 1223–1234

Chow W-M (1987) Buffer capacity analysis for sequential production lines with variable processing times. International Journal of Production Research 25: 1183–1196

Conway R, Maxwell W, McClain JO, Thomas LJ (1988) The role of work-in-process inventory in serial production lines. Operations Research 36: 229–241

Enginarlar E, Li J, Meerkov SM, Zhang RQ (2002) Buffer capacity to accommodating machine downtime in serial production lines. International Journal of Production Research 40: 601–624

Enginarlar E, Li J, Meerkov SM (2003a) How lean can lean buffers be? Control Group Report CGR 03-10, Deptartment of EECS, University of Michigan, Ann Arbor, MI; accepted for publication in IIE Transactions on Design and Manufacturing (2005)

Enginarlar E, Li J, Meerkov SM (2003b) Lean buffering in serial production lines with non-exponential machines. Control Group Report CGR 03-13, Deptartment of EECS, University of Michigan, Ann Arbor, MI

Gershwin SB, Schor JE (2000) Efficient algorithms for buffer space allocation. Annals of Operations Research 93: 117–144

Harris JH, Powell SG (1999) An algorithm for optimal buffer placement in reliable serial lines. IIE Transactions 31: 287–302

Hillier FS, So KC (1991a) The effect of the coefficient of variation of operation times on the allocation of storage space in production line systems. IIE Transactions 23: 198–206

Hillier FS, So KC (1991b) The effect of machine breakdowns and internal storage on the performance of production line systems. International Journal of Production Research 29: 2043–2055

Inman RR (1999) Empirical evaluation of exponential and independence assumptions in queueing models of manufacturing systems. Production and Operation Management 8: 409–432

Jafari MA, Shanthikumar JG (1989) Determination of optimal buffer storage capacity and optimal allocation in multistage automatic transfer lines. IIE Transactions 21: 130–134

Li J, Meerkov SM (2005) On the coefficients of variation of up- and downtime in manufacturing equipment. Mathematical Problems in Engineering 2005: 1–6

Park T (1993) A two-phase heuristic algorithm for determining buffer sizes in production lines. International Journal of Production Research 31: 613–631

Powell SG (1994) Buffer allocation in unbalanced three-station lines. International Journal of Production Research 32: 2201–2217

Powell SG, Pyke DF (1998) Buffering unbalanced assembly systems. IIE Transactions 30: 55–65

Seong D, Change SY, Hong Y (1995) Heuristic algorithm for buffer allocation in a production line with unreliable machines. International Journal of Production Research 33: 1989–2005

Smith JM, Daskalaki S (1988) Buffer space allocation in automated assembly lines. Operations Research 36: 343–357

Tempelmeier H (2003) Practical considerations in the optimization of flow production systems. International Journal of Production Research 41: 149–170

Yamashita H, Altiok T (1988) Buffer capacity allocation for a desired throughput of production lines. IIE Transactions 30: 883–891