A POTPOURRI ON THE THEME OF LEAN BUFFERING

Emre Enginarlar[a], Jingshan Li[b] and Semyon M. Meerkov[a]

[a] *Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109-2122, USA, 734-763-6349, eenginar@engin.umich.edu, smm@eecs.umich.edu*

[b] *Enterprise Systems Laboratory, GM Research and Development Center, Warren, MI 48090-9055, USA 810-986-7071, jingshan.li@gm.com*

Abstract: *This paper provides an overview of recently developed methods for design of lean in-process and finished goods buffering in production lines.*

Keywords: *Serial production lines, lean buffering, empirical law.*

1 INTRODUCTION

Production lines with unreliable machines usually contain in-process and finished goods buffering. The former is used to decouple the machines and thereby reduce mutual interference due to machine breakdowns. The latter is used to filter out random fluctuations of the production and, thus, ensure timely delivery of shipments to the customer. In modern manufacturing environment both in-process and finished goods buffering should be as small as possible, i.e. lean. But **how lean must be lean?** The purpose of this paper is to provide an answer to this question. To this end, Section 2 below addresses the issue of lean in-process buffering and in Section 3 finished goods buffering is considered. More details can be found in [1]-[5].

2 LEAN IN-PROCESS BUFFERING

2.1 Problem Formulation

We consider here a serial production line with M unreliable machines, $M-1$ finite in-process buffers and no finished goods buffer. Due to space limitation, the reader is referred to [1]-[3] for a detailed model description. For simplicity, we assume that all machines have identical cycle time (i.e., the time necessary to process a part), obey identical reliability model (i.e., the probability density functions of uptime and downtime of all machines are identical), and, therefore, have identical efficiency (i.e., $e = T_{up}/(T_{up} + T_{down})$, where T_{up} and T_{down} are the average up- and downtimes, respectively). The capacity of all buffers are also assumed to be identical. (These assumptions are generalized to non-identical machines and non-identical buffers in [2] and [3]; the conclusions, however, are quite similar to those described below.)

In order to obtain the results reported in this paper, buffer capacity, N, has to be normalized appropriately. In the case of in-process buffers, this normalization is:

$$k = \frac{N}{T_{down}}, \tag{2.1}$$

where T_{down} is measured in units of the cycle time. The number k, i.e. the buffer capacity in units of the average downtime, is referred to as the *level of buffering.*

Another measure, necessary for our analysis is the *line efficiency* E defined as

$$E = \frac{PR_k}{PR_\infty}, \tag{2.2}$$

where PR_k and PR_∞ are the production rates of the line with the level of buffering $k < \infty$ and infinity, respectively. Obviously, $E < 1$.

The problem addressed in this Section is: What is the smallest k that guarantees the required line efficiency? Below, we provide an answer to this question for the case of both exponential and non-exponential machines.

2.2 Exponential Machines

2.2.1 Two-machine lines: Assume that each machine in a two-machine line has up- and downtime distributed exponentially with the average values T_{up} and T_{down}, respectively. As before, e denotes the machine efficiency. Then it is possible to show, [2], that

$$k_E^{ex} = \begin{cases} \frac{2e(E-e)}{1-E}, & \text{if } E > e, \\ 0, & \text{otherwise,} \end{cases} \tag{2.3}$$

where k_E^{ex} is the lean level of buffering necessary to attain line efficiency E. Thus, if $E \leq e$, just-in-time operation represents the lean system. Otherwise, a lean system **must** contain in-process buffering. For instance, if $e = 0.83$ and $E = 0.95$, the lean level of buffering is 4 (i.e., buffer of four downtimes is required); if $e = 0.9$ and $E = 0.95$, the lean level of buffering is roughly 1.5.

2.2.2 $M > 2$-machine lines: A general analytical expression for k_E^{ex} in this case has not been derived. However, using the aggregation procedure of [1] and the subsequent polynomial approximations, the following expressions can be obtained for ten-machine lines:

$$k_E^{ex}(e) = \begin{cases} -2.74e^3 - 64.8e^2 + 68.2e + 0.63, & \text{for } E = 0.95, \\ -3.14e^3 - 26.6e^2 + 30.3e + 0.58, & \text{for } E = 0.9, \\ -7.02e^3 - 6.5e^2 + 12.7e + 1.65, & \text{for } E = 0.85. \end{cases} \tag{2.4}$$

As it follows from these equations, the lean buffering may be non-zero even if $E < e$. For instance, if $e = 0.9$ and $E = 0.85$, the lean level of buffering is roughly 2.5; if $e = 0.95$ and $E = 0.85$, it is roughly 1.5.

The behavior of k_E^{ex} as a function of the number of machines in the system, M, also can be analyzed. Using the aggregation procedure of [1] and exponential approximation, the following can be obtained for $e = 0.9$:

$$\begin{aligned} k_{0.95}(M) &= 1.8 + 6.255\left(1 - \exp\left(-\frac{(M-2)}{3}\right)\right), \\ k_{0.90}(M) &= 0.045 + 4.365\left(1 - \exp\left(-\frac{(M-2)}{3.5}\right)\right), \\ k_{0.85}(M) &= 0.045 + 3.061\left(1 - \exp\left(-\frac{(M-2)}{3.75}\right)\right), \\ & M \in \{2, 3, \ldots\}. \end{aligned} \tag{2.5}$$

These expressions indicate that k_E^{ex} is a monotonically increasing function of M, however, with an exponentially decreasing rate, saturating at around $M = 10$. In other words, the level of buffering, which is good for 10-machine lines, is also good for $M > 10$-machine lines. Perhaps, this happens due to the fact that the machines, separated by 9 buffers of appropriate (i.e., consistent with (2.4)) capacity, become practically decoupled, and the lean level of buffering necessary for ten-machine lines is also satisfactory for longer lines.

2.3 Non-Exponential Machines

2.3.1 Scenario: There are several difficulties in analyzing lean buffering in production lines with non-exponential machines. First, there are no analytical methods for production rate evaluation in such lines; therefore, only discrete event simulations are available as a tool for analysis. Second, the distributions of up- and downtime are typically not available on the factory floor. Third, and most importantly, even if these distributions were available and k_E^A for machines with reliability model denoted as A could be analyzed, the question of lean level of buffering for machines with reliability models other than A would still remain open. Obviously, analysis of k_E for **every** reliability model is impossible. Luckily, it is unnecessary as well: We show below that a simple *empirical law* can be introduced that provides a sufficiently precise estimate of k_E for every reliability model as a function of k_E^{ex}.

This empirical law has the form:

$$k_E = CV_{down} k_E^{ex}, \tag{2.6}$$

where CV_{down} is the coefficient of variation of the downtime, i.e.

$$CV_{down} = \frac{\sigma_{down}}{T_{down}}, \tag{2.7}$$

and σ_{down} is the standard deviation of the downtime. Empirical law (2.6) implies that the lean level of buffering for **any** model of machine reliability may be calculated as the product of the lean level of buffering for exponential machines and the downtime coefficient of variation of the machine at hand. Since the latter is typically available (or may be available) on the factory floor and the former was characterized above, (2.6) offers a simple tool for selecting lean buffering in serial production lines with any type of machine reliability. Below, we show that (2.6) indeed holds.

2.3.2 Method of analysis: The validity of empirical law (2.6) has been investigated using the gamma and Weibull distributions of machine up- and downtime. The gamma distribution is:

$$f_{up}^{G_P}(t) = pe^{-pt}\frac{(pt)^{P-1}}{\Gamma(P)}, \quad f_{down}^{G_R}(t) = re^{-rt}\frac{(rt)^{R-1}}{\Gamma(R)}, \tag{2.8}$$

where $\Gamma(x)$ is the gamma function, $\Gamma(x) = \int_0^\infty s^{x-1}e^{-s}ds$, and p, r, P and R are positive real numbers. This distribution is denoted as $G_P(p)$ or $G_R(r)$.

The Weibull distribution is:

$$f_{up}^{W_P}(t) = p^P e^{-(pt)^P} Pt^{P-1}, \quad f_{down}^{W_R}(t) = r^R e^{-(rt)^R} Rt^{R-1}, \tag{2.9}$$

where p, r, P and R are positive real numbers. This distribution is denoted as $W_P(p)$ or $W_R(r)$.

The expected values, variances, and coefficients of variation, of each of these distributions are given in Table 2.1.

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

Table 2.1: Expected values, variances, and coefficients of variation of up- and downtime distributions considered.

Distributions (2.8) and (2.9) have been selected for the analysis reported here since, being defined by two parameters, they allow us to place independently and at will the expected value and the variance (or the coefficient of variation). Thus, they allow for the possibility of having different distributions but with identical coefficients of variation. An example of such distributions, all with $CV_{down} = 0.5$, is shown in Figure 2.1.

To investigate the accuracy of empirical law (2.6), a set of 30 downtime distributions shown in Table 2.2 has been considered. The uptime, corresponding to each of these distributions,

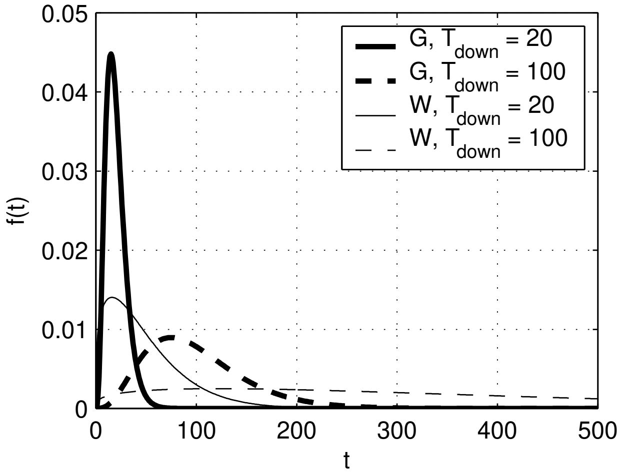


Figure 2.1: Different distributions with identical coefficients of variation ($CV_{down} = 0.5$).

CV_{down}	$T_{down} = 20$	$T_{down} = 100$
0.25	$G_{16}(0.8),\ W_{1.6}(0.007)$	$G_{16}(0.16),\ W_{1.6}(0.0005)$
0.50	$G_{4}(0.2),\ W_{1.3}(0.02)$	$G_{4}(0.04),\ W_{1.3}(0.002)$
0.75	$G_{1.8}(0.09),\ W_{1.1}(0.04)$	$G_{1.8}(0.018),\ W_{1.1}(0.006)$
1.00	$ex(0.05)$	$ex(0.01)$
1.25	$G_{0.6}(0.03),\ W_{0.9}(0.06)$	$G_{0.6}(0.006),\ W_{0.9}(0.017)$
1.50	$G_{0.4}(0.02),\ W_{0.87}(0.07)$	$G_{0.4}(0.004),\ W_{0.87}(0.02)$
1.75	$G_{0.3}(0.015),\ W_{0.82}(0.09)$	$G_{0.3}(0.003),\ W_{0.82}(0.025)$
2.00	$G_{0.25}(0.012),\ W_{0.79}(0.11)$	$G_{0.25}(0.0025),\ W_{0.79}(0.029)$

Table 2.2: Downtime distributions considered.

has been selected as follows: Assume that machine efficiency is e. Then,

$$T_{up} = \frac{e}{1-e} T_{down}. \tag{2.10}$$

The coefficient of variation of the uptime was selected either as

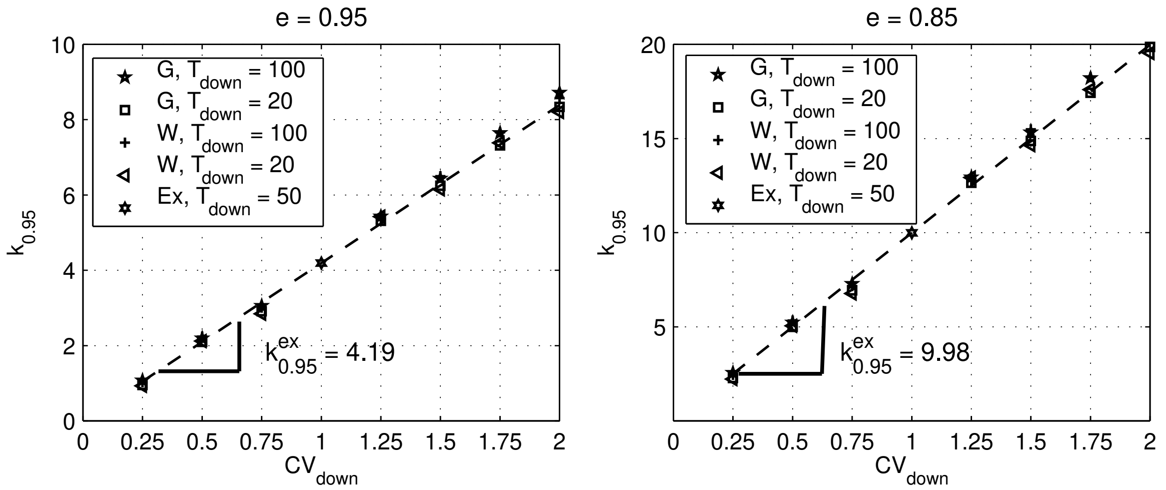
$$CV_{up} = CV_{down}, \tag{2.11}$$

or as

$$CV_{up} = \text{const.} \tag{2.12}$$

Given these T_{up} and CV_{up}, the nature of uptime distribution was selected to be the same as that for the downtime. (This selection was generalized in [3], where the up- and the downtime may obey different distributions.)

For each of the distributions of Table 2.2, using discrete event simulations (see [2], [3] for details), we calculate $k_E^{\mathcal{A}_s}$ for $CV_{down} = 0.25s$, $s = 1, \ldots, 8$, where $\mathcal{A}_s$ is the set of four



(a) (b)

Figure 2.2: Equal lean level of buffering and empirical law (2.6) ($M = 10$).

different distributions corresponding to $CV_{down} = 0.25s$, $s = 1, \ldots, 8$ (see Table 2.2). Then, we evaluate the accuracy of the empirical law (2.6) by

$$\mu(0.25s) = \frac{\max_{\mathcal{A}_s} \left|0.25sk_E^{ex} - k_E^A\right|}{\min_{\mathcal{A}_s} k_E^B}, \quad A, B \in \mathcal{A}_s. \tag{2.13}$$

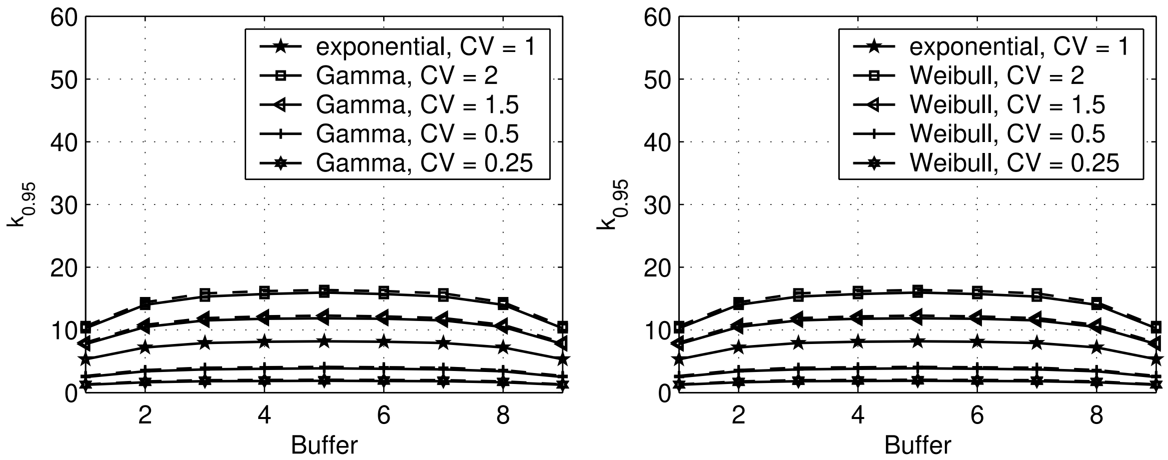
2.3.3 Results: Buffers of equal and non-equal ("inverted bowl") capacity have been investigated. The results are shown in Figures 2.2 for equal capacity buffering and 2.3 for inverted bowl buffering (with CV_{up} selected according to (2.11)). In the case of non-equal capacity buffering, the normalization and the empirical law have been introduced as a generalization of (2.1) and (2.6):

$$k_{E,i} = \frac{N_{E,i}}{T_{down}}, \tag{2.14}$$

$$k_{E,i} = CV_{down} k_{E,i}^{ex}, \quad i = 1, \ldots, M-1, \tag{2.15}$$

where M is the number of machines in the system. In Figure 2.2, empirical law (2.6) (respectively, in Figure 2.3, empirical law (2.15)) are plotted as the broken lines. Obviously, the accuracy is quite high. This is quantified by function $\mu(0.25s)$ of (2.13), which takes values less than 0.1 for both equal and non-equal buffer capacity. Thus, we conclude that empirical law (2.6) (or (2.15)) provides a simple tool for calculating lean buffering in serial production lines with unreliable machines for any model of machine reliability.

Remark 2.1: Reference [3] offers a generalization of (2.6) and (2.15) to systems with non-identical machines. Due to space limitation, this generalization is not described here,



(a) Exponential and gamma machine reliability

(b) Exponential and Weibull machine reliability

Figure 2.3: Non-equal capacity lean level of buffering and empirical law (2.15) ($M = 10$).

however, for the convenience of the reader we provide below the form of buffer capacity normalization and the form of the empirical law, used in the case of non-identical machines:

$$k^{E,i} = \frac{N_{E,i}}{\max(T_{down_i}, T_{down_{i+1}})}, \quad i = 1, \ldots, M-1,$$
$$k_{E,i} = \max(CV_{down_i}, CV_{down_{i+1}}) k^{ex}_{E,i}, \quad i = 1, \ldots, M-1.$$

3 LEAN FINISHED GOODS BUFFERING

3.1 Problem Formulation

We consider here a single machine production line with a finished goods buffer (FGB). As before, the machine is unreliable, and T_{up} and T_{down} are its average up- and downtime. The capacity of FGB is assumed to be N. The demand is specified by a pair (D, T) where D is the shipment size and T is the shipping period. The performance is characterized by the so-called due-time performance measure (DTP) defined as

$$DTP = \text{Prob}\{H(i-1) + \overline{t_i} \geq D\}, \tag{3.1}$$

where $H(i-1)$ is the occupancy of the FGB at the end of the $(i-1)$st shipping period and $\overline{t_i}$ is the number of parts produced during the ith shipping period (see [4], [5] for details). Analysis of lean N and lean T that lead to the desired DTP is the goal of this study.

To determine the lean N and T, the following normalizations are introduced:

$$\nu = \frac{N}{D} \tag{3.2}$$

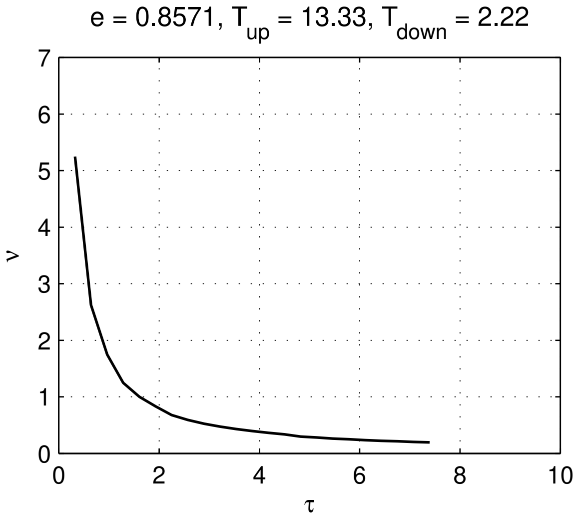


Figure 3.1: ν vs τ for $DTP = 0.99$ and $L = 0.9333$.

and

$$\tau = \frac{T}{T_{up} + T_{down}}. \tag{3.3}$$

In addition, the so-called load factor,

$$L = \frac{D}{Te} \tag{3.4}$$

is necessary to characterize lean N and T. It is assumed that D and T are selected so that $L < 1$.

In terms of the variables introduced,

$$DTP = F(\nu, \tau, L, e, T_{up}, T_{down}). \tag{3.5}$$

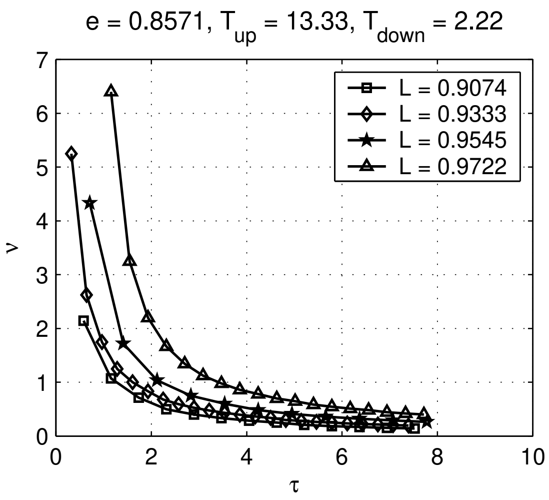
The problem addressed below is: How ν and τ should be selected so that $DTP \geq 0.99$? Below an answer to this question is provided for the case of exponential machines and deterministic or random demand D.

3.2 Method of Analysis

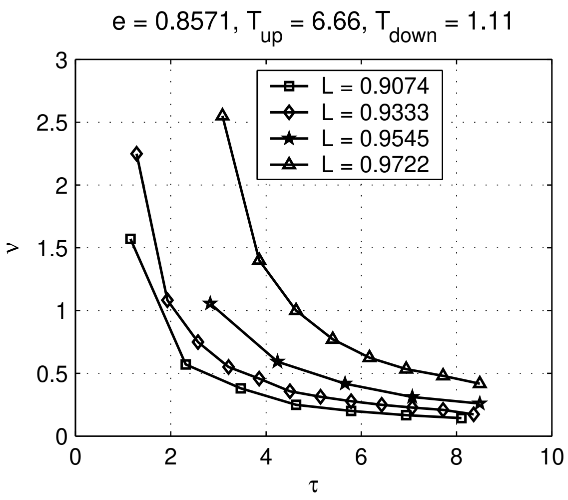
The results described below have been obtained analytically, based on expressions for DTP as a function of system parameters. For the deterministic demand, these expressions have been obtained in [4]. A generalization for random D has been derived in [5].

3.3 Results

3.3.1 Deterministic demand: Figure 3.1 illustrates a typical behavior of the (τ, ν)-curve for $L =$ const. It exhibits the trade-off between filtering random production by the finished goods buffer with capacity N and by the shipping period T. For example,



(a)



(b)

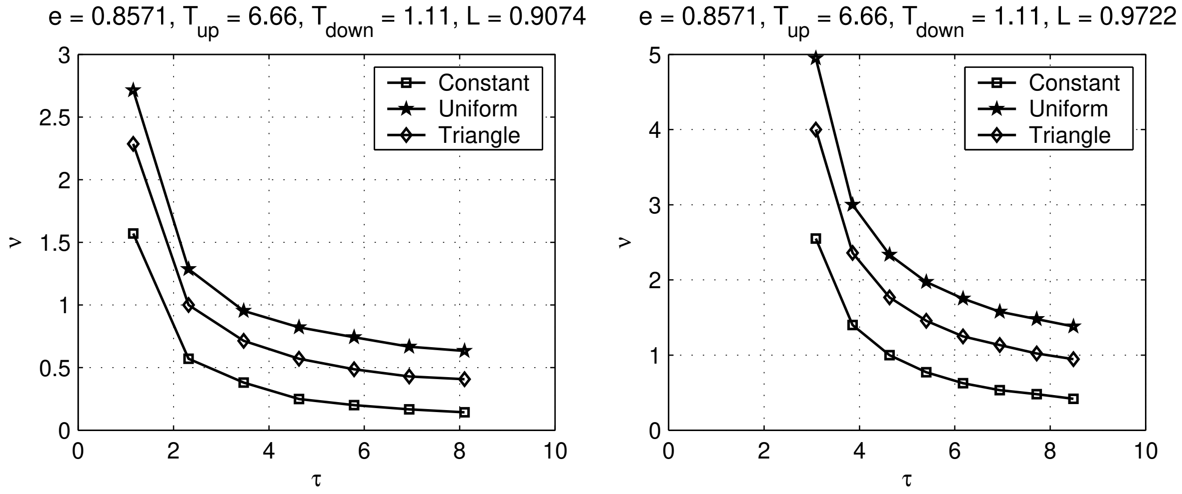
Figure 3.2: ν vs τ for $DTP = 0.99$ (constant demand).

if $\tau = 4$, i.e., T is $4(T_{up} + T_{down})$, FGB capacity of less than 1/2 of the shipment size is sufficient to provide 0.99 probability of customer demand satisfaction. Such a small FGB is sufficient because "filtering" by T is strong enough to reject production fluctuations. However, if $\tau = 0.75$, FGB should be about 5 shipment sizes in order to provide the same quality of service; in this case, the filtering takes place mostly in space, rather than in time. If one wants to share the filtering roughly equally between the time and the space, τ and ν should be selected as 1.2 and 1, respectively.

Figure 3.2 illustrates the dependence of lean ν and τ on L and T_{up} and T_{down} (for the same e). Surprisingly, shorter T_{up} and T_{down} require larger ν. Smaller L requires less filtering, as quantified in Figure 3.2.

3.3.2 Random demand: Figure 3.3 illustrates lean (τ, ν) for the same system as that of Figure 3.1, but for random demand (with the average value equal to the demand of Figure 3.1), distributed according to a uniform and triangular distribution. As one can see, randomness of the demand leads to an increase in either τ or ν or both. For instance, for $\tau = 2$, $L = 0.9$ and uniformly distributed demand, ν should be increased by a factor of two in comparison with the corresponding constant demand (see Figure 3.3(a)). For a heavier loaded system, similar results are shown in Figure 3.3(b).

Extension of these results to systems with arbitrary model of machine reliability is the focus of current work.



(a) (b)

Figure 3.3: ν vs τ for $DTP = 0.99$ (random demand).

REFERENCES

[1] S.-Y. Chiang, C.-T. Kuo and S.M. Meerkov, "DT-bottlenecks in serial production lines", *IEEE Trans. on Robotics and Automation*, Vol. 16, pp. 567-580, 2000.

[2] E. Enginarlar, J. Li., S.M. Meerkov and R.Q. Zhang, "Buffer capacity for accommodating machine downtime in serial production lines", *Control Group Report No. CGR-00-07,* Department of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, MI, 2000; also submitted to *Int. J. Prod. Res.*, 2000.

[3] E. Enginarlar, J. Li. and S.M. Meerkov, "A method for designing lean buffering in serial production lines", *Control Group Report No. CGR-01-03,* Department of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, MI, 2001; also submitted to *Production Planning and Control*, 2001.

[4] J. Li, *Production variability in manufacturing systems: A systems approach,* Ph.D. Thesis, Dept. of EECS, Univ. of Michigan, Ann Arbor, MI, 2000.

[5] E. Enginarlar, J. Li. and S.M. Meerkov, "Lean finished goods buffering in serial production lines," *Control Group Report No. CGR-01-04,* Department of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, MI, 2001.