Analysis and Modeling of Manufacturing Systems

edited by

Stanley B. Gershwin
Massachusetts Institute of Technology
Cambridge, MA, U.S.A.

Yves Dallery
École Centrale Paris
Chatenay-Malabry, France

Chrissoleon T. Papadopoulos
University of the Aegean
Chios Island, Greece

J. MacGregor Smith
University of Massachusetts at Amherst
Amherst, MA, U.S.A.

KLUWER ACADEMIC PUBLISHERS
Boston / Dordrecht / London

Chapter 10

DUE-TIME PERFORMANCE OF PRODUCTION SYSTEMS WITH MARKOVIAN MACHINES*

Jingshan Li
General Motors Research & Development Center
Warren, MI 48090-9055
jingshan.li@gm.com

Semyon M. Meerkov
Department of Electrical Engineering and Computer Science
University of Michigan, Ann Arbor, MI 48109-2122
smm@eecs.umich.edu

Abstract Customer demand satisfaction in serial production lines with Markovian machines and Finished Goods Buffers (FGB) is analyzed. The level of demand satisfaction is measured by the probability to ship to the customer a required number of parts during a fixed interval of time. This performance measure, referred to as Due-Time Performance (DTP), is analyzed using a simplification procedure termed stationarization. As a result, an iteration-based method for DTP calculation is obtained and utilized for analysis of various properties of DTP. In particular, it is shown that when the demand is relatively low (~ 95% of the production capacity), FGB of 2-4 shipments is sufficient to ensure high DTP (~ 0.98). When the demand is high (~ 99%), FGB of 6-9 shipments is required. Further increase of FGB leads to highly diminishing return. Also, it is shown that a ramp, rather than an inverted bowl, of machine efficiency allocation leads to optimization of DTP.

Keywords: Production systems, unreliable machines, due-time performance, finished goods buffer.

*This work was supported by NSF grant No. DMI-9820580.

1. Motivation

The number of parts produced by a manufacturing system with unreliable machines during a fixed interval of time is a random variable. Its distribution defines both production volume and production variability. In practice, the production volume is characterized by the production rate (PR), i.e., the average number of parts produced per unit of time. The production variability is often characterized by the due-time performance (DTP), which is the probability to ship to the customer a required number of parts during a fixed interval of time. This is the case, for instance, in the automotive industry where car and truck assembly plants are expected to be supplied by engine and seat plants on a four-hour delivery schedule. Similarly, engine plants receive ignition and injection units from the second tier suppliers on a fixed shipping schedule, etc.

PR was analyzed in many publications during the last 50 years (see reviews by Dallery and Gershwin 1992, Papadopoulos and Harvey 1996, Govil and Fu 1999, and monographs by Viswanadham and Narahari 1992, Buzacott and Shanthikumar 1993, Papadopoulos et al 1993, Gershwin 1994, Altiok 1997). In contrast, DTP has been addressed just recently. Specifically, Jacobs and Meerkov (1995a) analyzed DTP for a one-machine system without a finished goods buffer (FGB). The case of multiple machines, but also without a FGB, was addressed by Tan (1998, 1999). In practice, however, most production systems do have FGBs, intended to both improve DTP and limit finished goods inventory. Production systems with multiple machines and FGBs were investigated by Li and Meerkov (1998, 2000a, 2000b, 2001). The development was carried out under the assumption that the machines obey the so-called Bernoulli reliability model. According to this model, a machine is up or down during each cycle time (i.e., the time necessary to produce a part) with a probability independent of the status of the machine (up or down) in the previous cycle time. Although this model does take place in some assembly operations, most machining lines are better described by Markovian reliability model whereby the state of the machine in a cycle time is conditioned on the state of the machine in the previous cycle time. DTP in production systems with such machines and FGB has not been analyzed in the literature. This work is intended to contribute to this end.

The outline of the paper is as follows: Section 2 provides the problem formulation. A formal definition of DTP is also introduced. It turns out that to evaluate DTP, the calculation of PR is necessary. Since a method for PR calculation for production system with machines considered in this work is not available in the literature, Section 3 presents an aggregation technique for PR evaluation. In Section 4, a method for DTP calculation in serial lines with Markovian machines and FGB is developed. Section 5 addresses structural properties of DTP. The conclusions are formulated in Section 6. All proofs are given in the Appendix.

2. Problem Formulation

The structure of the production system analyzed in this work is shown in Figure 1, where the circles represent the machines and rectangles are the buffers. Assumptions concerning the machines, buffers, interactions between the machines, and the demand are defined as follows:

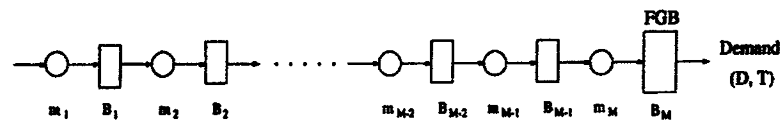

Figure 1. Serial production line with finished goods buffer

<u>Machines</u>:

(i). Each machine, m_i, $i = 1, \cdots, M$, requires a fixed unit of time to process a part. This unit is referred to as the *cycle time.* All machines have identical cycle time. The time axis is slotted with the slot duration equaled to the cycle time.

Remark 1 The assumption that the cycle time is deterministic and identical for all machines in the system is sufficiently realistic for production systems with automated material handling. It may not hold for other systems, e.g., job shop-type environment. Thus, the results of this research contribute to the area of automated production lines, such as typically encountered in the automotive industry. □

(ii). During a cycle time, each machine can be either "up" or "down". When up, the machine could process a part. When down, no processing can take place.

(iii). The state of the machines is defined by a discrete time Markov process: If machine m_i, $i = 1, \cdots, M$, is up, it will be down during the next cycle with probability P_i and up with probability $1 - P_i$; if it is down, it will be up during the next cycle with probability R_i and down with probability $1 - R_i$. In other words, the up- and downtime are distributed geometrically with the parameters P_i and R_i, respectively.

<u>Buffers</u>:

(iv). Each buffer B_i, $i = 1, \cdots, M$, has capacity $1 \leq N_i < \infty$, $i = 1, \cdots, M$. Buffers $B_1, \cdots, B_{M-1}$ are called in-process buffers. Buffer B_M is the finished goods buffer.

<u>Starvation rule</u>:

(v). If B_i, $i = 1, \cdots, M-1$, is empty at the beginning of the time slot, then m_{i+1}, $i = 2, \cdots, M-1$, is starved during this time slot. The first machine is never starved.

Blockage rule:

(vi). If B_i, $i = 1, \cdots, M-1$, is full at the beginning of a time slot and m_{i+1}, $i = 1, \cdots, M-1$, does not take a part from B_i at the beginning of this slot, then m_i, $i = 1, \cdots, M-1$, is blocked during this time slot. The last machine, m_M, can be blocked during a time slot if the FGB is full at the beginning of this time slot.

Remark 2 Assumptions (iii), (v) and (vi) are formulated in terms of the so-called time-dependent failures, i.e., machines can go down even when blocked or starved (Buzacott and Shanthikumar 1993). Another possible model is that of operation-dependent failures, where no breakdowns of starved or blocked machines is possible (Buzacott and Shanthikumar 1993, Gershwin 1994). Both models are practical, depending on the production system at hand: For automated palletized material handling systems, the time-dependent model is more applicable. In case of manual material handling, operation-dependent failures often take place. Both failure modes, however, result in similar behavior. Studies show that throughputs of a line with time-dependent or operation-dependent failures differ at most by 3% - 4% (Buzacott and Shanthikumar 1993), which is well within the accuracy of the data describing production lines. □

Demand:

(vii). From the point of view of the demand, the time axis is divided into "epochs", each containing T time slots.

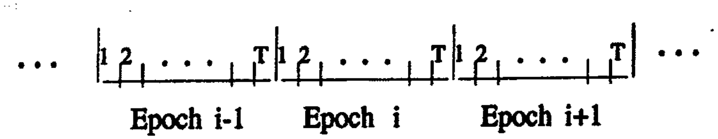

Figure 2. Epochs

(viii). At the end of each epoch, a shipment of D parts has to be available for the customer. If p_a is the production rate of the system (without FGB), then

$$D \leq T p_a. \tag{1}$$

Demand satisfaction policy:

(ix). At the beginning of epoch i, parts are removed from the FGB in the amount of $\min(H(i-1), D)$, where $H(i-1)$ is the number of parts in the FGB at the end of $(i-1)$-th epoch. If $H(i-1) \geq D$, the shipment is complete; if $H(i-1) < D$, the balance of the shipment, i.e., $D - H(i-1)$ parts, is to be produced by m_M during the shipping period T. Parts produced are immediately removed from the FGB and prepared for shipment, until the shipment is complete, i.e., D parts are available. If the shipment is complete before the end of the epoch, the system continues operating, but with the parts being accumulated in the FGB, either until the end of the epoch or until the last machine, m_M, is blocked, whatever occurs first. If the shipment is not complete by the end of the epoch, an incomplete shipment is sent to the customer. No backlog is allowed.

Remark 3 Assumption of no backlog is introduced to simplify the analysis. However, since backlog is typically satisfied by overtime, this assumption does not reduce the generality. □

Assumptions (i)-(ix) define the serial production line under consideration. In the time scale of the epoch and in an appropriately defined state space, the system (i)-(ix) is a stationary ergodic Markov chain. We refer to its steady state as the "normal system operation".

Let $\bar{t}_i$ be the number of parts produced by machine m_M in epoch i during the normal system operation. Then DTP can be expressed as

$$DTP = Pr\left(\bar{t}_i + H(i-1) \geq D\right). \tag{2}$$

In the framework of model (i)-(ix), DTP is a function of all system parameters. In other words,

$$DTP = DTP(P, R, N, N_M, D, T), \tag{3}$$

where N_M, T, and D are defined in (iv), (vii), and (viii), respectively, P, R and N are vectors of the machines and in-process buffers parameters,

$$P = [P_1, \cdots, P_M], \quad R = [R_1, \cdots, R_M], \quad N = [N_1, \cdots, N_{M-1}].$$

The goal of this paper is to develop a method evaluating function (3) and to investigate its structural (e.g., monotonicity) properties.

3. Production Rate Evaluation

This Section describes a method for PR calculation in serial lines defined by assumptions (i)-(vi), i.e., having no FGB. In Section 4, this method is used for DTP calculation in corresponding systems with FGBs.

3.1 Two-machine Line

3.1.1 Calculations.

Theorem 1 *The production rate in a serial production line (i)-(vi) with $M = 2$ is given by*

$$PR = e_2\left[1 - Q(P_1, R_1, P_2, R_2, N)\right], \quad (4)$$

where

$$e_i = \frac{R_i}{P_i + R_i}, \quad i = 1, 2,$$

$$Q(P_1, R_1, P_2, R_2, N) = \begin{cases} \frac{P_1\beta_2}{(R_1+R_2-R_1R_2)(R_1+P_1)}, & \text{if } N = 1, \\ \frac{P_1\alpha_1\alpha_2\beta_2^2(R_2+P_2)}{A+B+C+D}, & \text{if } N > 1, \end{cases} \quad (5)$$

and

$$\begin{aligned} \alpha_1 &= P_1 + P_2 - P_1P_2 - R_1P_2, \\ \alpha_2 &= P_1 + P_2 - P_1P_2 - R_2P_1, \\ \beta_1 &= R_1 + R_2 - R_1R_2 - P_1R_2, \\ \beta_2 &= R_1 + R_2 - R_1R_2 - P_2R_1, \\ \sigma &= \frac{\alpha_2\beta_1}{\alpha_1\beta_2}, \\ A &= P_1R_2\alpha_1\alpha_2\beta_2(P_2 + \beta_2), \\ B &= P_1R_1R_2\alpha_2\left[\beta_2^2 + P_2(\alpha_1 + \beta_1)(\alpha_2 + 2\beta_2)\right], \\ C &= \sum_{k=2}^{N-1} P_1P_2R_1R_2(\alpha_2 + \beta_2)^3\sigma^{k-1}, \\ D &= P_2R_1\alpha_1\beta_2\left[R_2(\alpha_1 + \beta_1) + \alpha_2(P_1 + R_1)\right]\sigma^{N-1}. \end{aligned}$$

Moreover, the average probability of buffer occupancy, $E(h)$, and the probabilities of manufacturing blockage of m_1, mb_1, and starvation of m_2, ms_2, are given by

$$E(h) = \begin{cases} \frac{P_1[(R_1+R_2-R_1R_2)(P_2+R_2)+P_1P_2]}{(R_1+R_2-R_1R_2)(R_1+P_1)(R_2+P_2)} & N = 1, \\ \frac{B+\sum_{k=2}^{N-2} P_1P_2R_1R_2k(\alpha_1+\beta_1)^3\sigma^{k-1}+ND}{A+B+C+D} & N > 1, \end{cases}$$

$$ms_2 = \begin{cases} \frac{P_1R_2\beta_2}{(R_1+R_2-R_1R_2)(R_1+P_1)(R_2+P_2)} & N = 1, \\ \frac{P_1R_2\alpha_2\alpha_2\beta_2^2}{A+B+C+D} & N > 1, \end{cases} \quad (6)$$

$$mb_1 = \begin{cases} \frac{P_2R_1\beta_1}{(R_1+R_2-R_1R_2)(R_1+P_1)(R_2+P_2)} & N = 1, \\ \frac{P_2R_1\alpha_1\alpha_2\beta_1\beta_2\sigma^{N-1}}{A+B+C+D} & N > 1. \end{cases} \quad (7)$$

Proof: See Appendix. ■

Remark 4 The PR calculation for two-machine line with time- dependent failure was studied by Buzacott and Shanthikumar (1993), Sheskin (1976), however the assumptions defined were different from (i)-(vi). More specifically, it is either assumed that machine m_2 could produce a part if both machines are up and buffer is empty (Buzacott and Shanthikumar 1993), or an additional constraint that $P_i + R_i = 1$ is needed (Sheskin 1976). □

3.1.2 Monotonicity.

Proposition 1 *For $N = 1$, function $Q(P_1, R_1, P_2, R_2, N)$ is monotonically increasing with respect to P_1, R_2, and decreasing with respect to P_2, R_1.*

Proof: See Appendix. ■

Unfortunately, the proof of this property for $N > 1$ seems to be all but impossible. Therefore, based on numerical simulation, we formulate

Hypothesis 1 *Function $Q(P_1, R_1, P_2, R_2, N)$, $N > 1$, is monotonically increasing with respect to P_1, R_2, and decreasing with respect to P_2, R_1.*

An illustration is shown in Figure 3.

3.2 *M*-machine Line

3.2.1 Aggregation. No closed form expression for *PR* in *M*-machine line is available. Therefore, we develop an aggregation procedure, based on

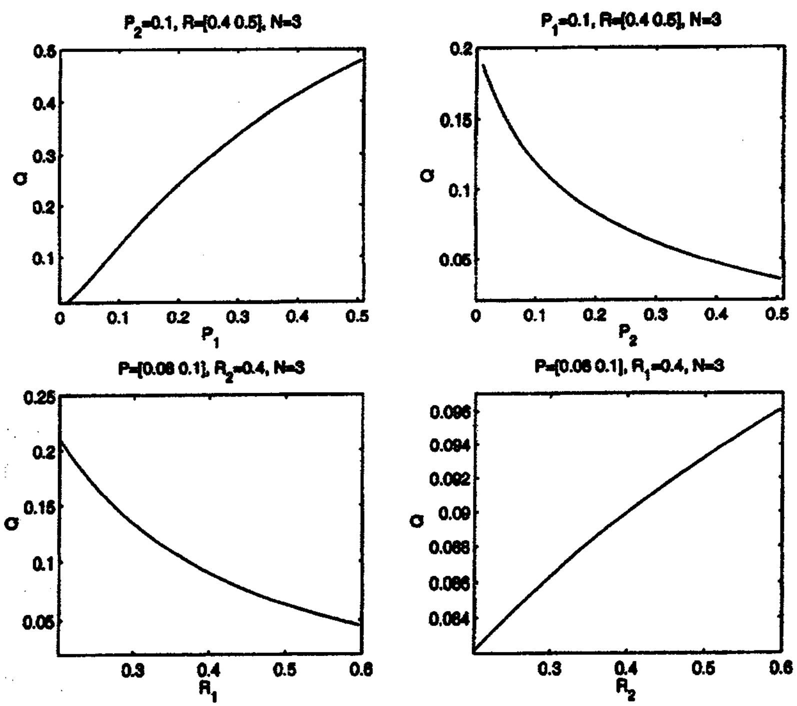

Figure 3. Monotonicity of $Q(\cdot)$

the result of the previous Subsections. Specifically, we aggregate the first two machines into a single machine, m_2^f, with parameter R_2^f defined as

$$R_2^f = R_2[1 - Q(P_1, R_1, P_2, R_2, N_1)],$$

and P_2^f selected so that

$$\frac{R_2^f}{P_2^f + R_2^f} = \frac{R_2}{P_2 + R_2}[1 - Q(P_1, R_1, P_2, R_2, N_1)],$$

i.e.,

$$P_2^f = P_2 + R_2 Q(P_1, R_1, P_2, R_2, N_1),$$

where $Q(\cdot)$ is defined in (5). Next m_2^f is a aggregated with m_3 to result in m_3^f, with the parameters defined as shown above, and so on until all M machines are aggregated in a single one, m_M^f. This constitutes the forward aggregation (superscript f is used to denote this fact). Then, in the backward aggregation, the last machine, m_M, is aggregated with m_{M-1}^f to result in m_{M-1}^b and so on until all machines are again aggregated in a single machine, m_1^b. Then the procedure is repeated again. Formally, this process can be represented as follows:

Procedure 1

$$\begin{aligned}
R_i^b(s+1) &= R_i - R_i Q\big(P_{i+1}^b(s+1), R_{i+1}^b(s+1), P_i^f(s), R_i^f(s), N_i\big), \\
&\qquad\qquad i = 1, \cdots, M-1, \\
P_i^b(s+1) &= P_i + R_i Q\big(P_{i+1}^b(s+1), R_{i+1}^b(s+1), P_i^f(s), R_i^f(s), N_i\big), \\
&\qquad\qquad i = 1, \cdots, M-1, \\
R_i^f(s+1) &= R_i - R_i Q\big(P_{i-1}^f(s+1), R_{i-1}^f(s+1), P_i^b(s+1), R_i^f(s+1), N_{i-1}\big), \\
&\qquad\qquad i = 2, \cdots, M, \qquad (8) \\
P_i^f(s+1) &= P_i + R_i Q\big(P_{i-1}^f(s+1), R_{i-1}^f(s+1), P_i^b(s+1), R_i^f(s+1), N_{i-1}\big), \\
&\qquad\qquad i = 2, \cdots, M,
\end{aligned}$$

with boundary conditions

$$\begin{aligned}
P_1^f(s) = P_1, &\qquad R_1^f(s) = R_1, \\
P_M^b(s) = P_M, &\qquad R_M^b(s) = R_M, \\
s &= 0, 1, 2, \cdots,
\end{aligned}$$

and initial conditions

$$P_i^f(0) = P_i, \quad R_i^f(0) = R_i, \qquad i = 2, \cdots, M-1$$

and function $Q(\cdot)$ defined in (5). □

The question of convergence of the resulting sequences $P_i^b(s)$, $R_i^b(s)$, $P_i^f(s)$, $R_i^f(s)$, $i = 1, \cdots, M$, $s = 0, 1, \cdots$, is answered in the following:

Theorem 2 *Under Hypothesis 1, recursive procedure 1 is convergent and, therefore, the following limits exist:*

$$\begin{aligned} \lim_{s\to\infty} P_i^f(s) &:= P_i^f, \qquad \lim_{s\to\infty} P_i^b(s) := P_i^b, \\ \lim_{s\to\infty} R_i^f(s) &:= R_i^f, \qquad \lim_{s\to\infty} R_i^b(s) := R_i^b, \\ i &= 1, 2, \cdots, M. \end{aligned} \tag{9}$$

Moreover, the following relationship holds:

$$\frac{R_M^f}{P_M^f} = \frac{R_1^b}{P_1^b}. \tag{10}$$

Proof: See Appendix. ∎

The limits in (9) can be used to define estimates of performance measures for line (i)-(vi) without FGB. Indeed, since the last machine is not blocked and the first is not starved, production rate can be estimated as

$$\widehat{PR}(P_1, R_1, \cdots, P_M, R_M, N_1, \cdots, N_{M-1}) = \frac{R_M^f}{P_M^f + R_M^f} = \frac{R_1^b}{P_1^b + R_1^b}. \tag{11}$$

The accuracy of the estimate (11) is discussed next.

3.2.2 Accuracy. The accuracy of the estimate (11) has been evaluated numerically. We simulate dozens of systems defined by assumptions (i)-(vi) without FGB with various machine and buffer parameters assumed. Twenty of them with 3 - 8 machines, are shown in Table 1. In each simulation run, zero initial occupancy of all buffers has been assumed, and 50,000 time slots of warm up period has been carried out. The next 500,000 slots of stationary regime have been used to statistically evaluate the PR. Confidence intervals have been evaluated with 20 runs. The 95% confidence intervals were consistently around ±0.0015. Same procedures are followed for all numerical simulations throughout this paper. In Table 1, PR denotes the actual production rate obtained by simulation, whereas $\widehat{PR}$ denotes the estimate of production rate calculated according to (11). As it can be seen from Table 1, the estimate results in relatively high precision, comparable with that of Gershwin (1994), Dallery et al (1978), Jacobs (1993), Jacobs and Meerkov (1995b), Chiang (1999).

Table 1. Numerical justification of production rate estimation for serial lines (err$=\frac{|PR-\widehat{PR}|}{PR}\cdot$ 100%)

P_i	R_i	N_i	PR	$\widehat{PR}$	err
0.06 0.07 0.08	0.27 0.28 0.29	1 2	0.561	0.562	0.21
0.12 0.15 0.10	0.43 0.46 0.50	2 2	0.606	0.605	0.18
0.05 0.10 0.15	0.50 0.45 0.40	2 2	0.627	0.629	0.32
0.10 0.10 0.10	0.42 0.42 0.42	3 3	0.668	0.668	0.04
0.10 0.05 0.20	0.90 0.85 0.75	3 2	0.776	0.786	1.24
0.10 0.02 0.06	0.60 0.04 0.09	1 1	0.360	0.349	3.11
0.11 0.08 0.08 0.11	0.40 0.41 0.41 0.40	2 3 2	0.608	0.606	0.39
0.11 0.12 0.13 0.10	0.39 0.38 0.36 0.43	4 5 3	0.606	0.605	0.20
0.08 0.09 0.07 0.06	0.37 0.43 0.41 0.39	3 2 2	0.644	0.645	0.22
0.15 0.04 0.30 0.02	0.50 0.80 0.40 0.70	3 2 3	0.550	0.557	1.26
0.06 0.08 0.05 0.10	0.36 0.39 0.42 0.37	3 3 4	0.682	0.689	0.94
0.04 0.07 0.10 0.13	0.40 0.43 0.37 0.46	2 3 3	0.646	0.645	0.13
0.10 0.07 0.09 0.12 0.11	0.40 0.35 0.33 0.42 0.39	4 4 4 3	0.607	0.607	0.15
0.10 0.12 0.13 0.11 0.12	0.45 0.42 0.43 0.46 0.44	4 3 4 3	0.613	0.612	0.03
0.12 0.09 0.12 0.09	0.41 0.36 0.41 0.36	3 4 4 3	0.627	0.631	0.54
0.05 0.09 0.13 0.17 0.21	0.42 0.45 0.48 0.51 0.54	2 2 2 2	0.542	0.540	0.31
0.80 0.80 0.80 0.80 0.80 0.80	0.42 0.42 0.42 0.42 0.42 0.42	3 3 3 3 3	0.638	0.644	0.97
0.06 0.08 0.07 0.10 0.12 0.09	0.43 0.46 0.45 0.48 0.47 0.44	2 2 3 2 3	0.616	0.617	0.13
0.09 0.08 0.07 0.10 0.12 0.10 0.07	0.35 0.37 0.32 0.38 0.39 0.41 0.36	3 2 3 4 3 2	0.544	0.547	0.58
0.06 0.07 0.09 0.10 0.12 0.08 0.11 0.09	0.43 0.42 0.41 0.41 0.43 0.45 0.44 0.40	3 3 2 3 4 3 2	0.575	0.582	1.34

4. DTP Calculation

The Due Time Performance measure, introduced in Section 2, is defined as

$$DTP = Pr\left(\bar{t}_i + H(i-1) \geq D\right),$$

where $\bar{t}_i$ be the number of parts produced during i-th epoch, $H(i-1)$ is the number of parts that remained in the FGB at the end of $(i-1)$-th epoch. Unlike the case of production rate, the calculation of DTP even for a one-machine system is a nontrivial problem. Therefore, before addressing the general case, we study a single machine system.

4.1 One-machine Case

Consider a one-machine production system with the FGB of capacity N (Figure 4). Introduce notations:

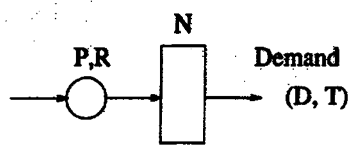

Figure 4. One-machine system with FGB

t_i = number of parts produced during epoch i if no blockage occurs.

$\pi(j,s,n)$ = Probability that in the time slot j of an epoch ($j = 1,\cdots,T$) the machine is in state s ($s = 0$ when the machine is down and $s = 1$ when machine is up and n parts ($n = 0,1,\cdots,j$) have been produced during $[0,j]$.

$z_k = Pr\big(H(i-1) = k\big),\ k = 0,1,\cdots,N.$

$r_{k,j} = Pr(t_i = D+k-j),\ k = 1,\cdots,N-1,\ j = 0,1,\cdots,N.$

$\bar{r}_{N,j} = Pr(t_i \geq D+N-j),\ \ j = 0,1,\cdots,N.$

Probabilities $\pi(j,s,n)$ can be calculated recursively as follows (see Gershwin 1993):

$$\begin{aligned}\pi(j,0,n) &= (1-R)\pi(j-1,0,n) + P\pi(j-1,1,n), \\ &\qquad 0 < n < T,\quad 1 \leq j \leq T, \\ \pi(j,1,n) &= R\pi(j-1,0,n-1) + (1-P)\pi(j-1,1,n-1), \\ &\qquad 0 < n < T,\quad 1 \leq j \leq T,\end{aligned} \tag{12}$$

with initial conditions

$$\begin{aligned}&\pi(j,1,0) = 0, \qquad \pi(j,0,j) = 0, \\ &\pi(j,0,0) = \tfrac{P}{P+R}(1-R)^{j-1},\quad \pi(j,1,j) = \tfrac{R}{P+R}(1-P)^{j-1}, \\ &\qquad j = 1,\cdots,T.\end{aligned} \tag{13}$$

Then, $r_{k,j}$ and $\bar{r}_{N,j}$ can be calculated as follows:

$$\begin{aligned} r_{k,j} &= \sum_{s=0}^{1} \pi(T,s,D+k-j), \\ \bar{r}_{k,j} &= \sum_{n=D+k-j}^{T} \sum_{s=0}^{1} \pi(T,s,n), \end{aligned} \tag{14}$$

Using the above notations, the DTP of a one-machine system with the FGB can be calculated as follows:

Theorem 3 *Let* $Z = [z_1,\cdots,z_N]^T$ *be a vector defined by*

$$Z = -\mathbf{R}^{-1}Z_0, \tag{15}$$

with matrix $\mathbf{R}$ *and vector* Z_0 *as follows:*

$$\mathbf{R} = \begin{pmatrix} r_{1,1}-r_{1,0}-1 & r_{1,2}-r_{1,0} & \cdots & r_{1,N}-r_{1,0} \\ r_{2,1}-r_{2,0} & r_{2,2}-r_{2,0}-1 & \cdots & r_{2,N}-r_{2,0} \\ \cdots & \cdots & \cdots & \cdots \\ \bar{r}_{N,1}-\bar{r}_{N,0} & \bar{r}_{N,2}-\bar{r}_{N,0} & \cdots & \bar{r}_{N,N}-\bar{r}_{N,0}-1 \end{pmatrix}, \tag{16}$$

$$Z_0 = \begin{pmatrix} r_{1,0} \\ r_{2,0} \\ \cdots \\ r_{N-1,0} \\ \bar{r}_{N,0} \end{pmatrix}. \tag{17}$$

Then, under assumptions (i)-(ix) with $M = 1$, *the DTP is given by:*

$$DTP = \sum_{k=0}^{N} z_k \sum_{n=D-k}^{T} \sum_{s=0}^{1} \pi(T,s,n), \tag{18}$$

where $\pi(T,s,n)$ *is calculated by (12).*

Proof: Similar to the proof for the Bernoulli case of Li and Meerkov (2000a).
■

4.2 Approach to *M*-machine Line

4.2.1 Idea of the approach. The main difficulty in developing a method for calculating DTP is that the performance of a line within an epoch (i.e., in the time scale of the machine cycle time) is not described by a stationary random process. For instance, the probability that the last machine, m_M is blocked during a time slot j, $j = 1,\cdots,T$, clearly depends on j and is a monotonically increasing function of j. This makes exact calculations of DTP all but impossible. Therefore, a simplification is necessary. The simplification used in this work is based on two ideas: stationarization and subsequent recursive iterations. These are described below.

4.2.2 Stationarization. The stationarization procedure is intended to approximate the nonstationary behavior of the production line within an epoch (i.e., in the time scale of the cycle time) by a stationary process, so that the behavior of the system in the time scale of the epoch is not changed too dramatically. To accomplish this, consider a one-machine line, shown in Figure 4. Assume that the probability that the machine, m, is up and FGB is full, during a time slot j, $j = 1, \cdots, T$, is known. Denoted this probability as $P_b(j)$. Then the probability, P_{k_b}, that m is up and FGB is full during an epoch exactly k_b times, $k_b = 1, \cdots, T$, can be expressed as

$$P_{k_b} = \begin{cases} P_b(T-k_b+1) - P_b(T-k_b), & 1 \le k_b \le T-1, \\ 0, & k_b = T. \end{cases} \qquad (19)$$

Thus, the expected number in an epoch, $\bar{k}_b$, is

$$\bar{k}_b = \sum_{k_b=0}^{T} k_b P_{k_b} = \sum_{k_b=1}^{T-1} k_b P_{k_b}. \qquad (20)$$

Using this expression, the approximate probability that m is up and FGB is full during *any* slot j can be introduced as follows:

$$\widehat{mb} = \frac{\bar{k}_b}{T}. \qquad (21)$$

Then the approximate probability that FGB is full during any time slot j is:

$$\widehat{P}_f = \frac{\widehat{mb}}{\frac{R}{P+R}} = \frac{\bar{k}_b}{T\frac{R}{P+R}}. \qquad (22)$$

We use these expressions as the stationarized probabilities of the appropriate events.

To evaluate $\widehat{mb}$ and $\widehat{P}_f$, the distribution $P_b(j)$ is necessary. It can be calculated as follows:

Introduce the probabilities:

$$h(i,j,s) = \text{Prob}\{i \text{ parts in FGB at the beginning of slot } j, \text{ and the machine is in state } s\},$$
$$i = 1, \cdots, N, j = 1, \cdots, T, s = 0, 1,$$
$$h_0(0,j,s,k) = \text{Prob}\{\text{FGB is empty at the beginning of slot } j, \text{ the machine is in state } s, \text{ and } k \text{ parts have been sent to customer}\},$$
$$i = 1, \cdots, N, j = 1, \cdots, T, k = 0, \cdots, D, s = 0, 1.$$

Under assumptions (i)-(ix) and $M = 1$, function $h(i,j,s)$ is defined recursively by

$$h(i,j,0) = \begin{cases} h(1,j-1,0)(1-R) + h_0(0,j-1,1,D)P, & \text{for } i = 1, \\ h(i,j-1,0)(1-R) + h(i-1,j-1,1)P, & \text{for } 1 < i < N, \\ h(N,j-1,0)(1-R) + h(N,j-1,1)P \\ \quad +h(N-1,j-1,1)P, & \text{for } i = N, \end{cases} \qquad (23)$$

$$h(i,j,1) = \begin{cases} h(1,j-1,0)R + h_0(0,j-1,1,D)(1-P), & \text{for } i = 1, \\ h(i,j-1,0)R + h(i-1,j-1,1)(1-P), & \text{for } 1 < i < N, \\ h(N,j-1,0)R + h(N,j-1,1)(1-P) \\ \quad +h(N-1,j-1,1)(1-P), & \text{for } i = N, \end{cases} \qquad (24)$$

where

$$h_0(0,j,\lambda,k) = \begin{cases} h_0(0,j-1,0,k)(1-R) + h_0(0,j-1,1,k-1)P, & \text{for } \lambda = 0, k > 0, \\ h_0(0,j-1,1,k-1)(1-P) + h_0(0,j-1,0,k)R, & \text{for } \lambda = 1, k > 0, \end{cases} \qquad (25)$$

with initial conditions

$$h_0(0,j,\lambda,0) = \begin{cases} z_0\frac{P}{P+R}(1-R)^{j-1}, & \text{for } \lambda = 0, \\ z_0\frac{R}{P+R}(1-R)^{j-1}, & \text{for } \lambda = 1, \end{cases} \qquad (26)$$

$$h_0(0,1,\lambda,k) = \begin{cases} z_k\frac{P}{P+R}, & \text{for } \lambda = 0, 0 \le k \le D, N > D, \\ z_k\frac{R}{P+R}, & \text{for } \lambda = 1, 0 \le k \le D, N > D, \\ 0, & \text{for } N+1 \le k \le D, N < D, \end{cases} \qquad (27)$$

$$h(i,1,0) = \begin{cases} \frac{P}{P+R}, & \text{for } i = 0, D \ge N, \\ \sum_{i=1}^{D} z_i\frac{P}{P+R}, & \text{for } i = 0, D < N, \\ z_{D+i}\frac{P}{P+R}, & \text{for } 0 < i \le N-D, \\ 0, & \text{for } i > N-D > 0, \text{ or } i > 0, D \ge N, \end{cases} \qquad (28)$$

$$h(i,1,1) = \begin{cases} \frac{R}{P+R}, & \text{for } i = 0, D \ge N, \\ \sum_{i=1}^{D} z_i\frac{R}{P+R}, & \text{for } i = 0, D < N, \\ z_{D+i}\frac{R}{P+R}, & \text{for } 0 < i \le N-D, \\ 0, & \text{for } i > N-D > 0, \text{ or } i > 0, D \ge N, \end{cases} \qquad (29)$$

and z_i, $i = 0, 1, \cdots, N$, can be calculated from (15). Therefore, the probability that m is up and FGB is full during slot j is:

$$P_b(j) = \sum_{s=0}^{1} h(N, j, s)\frac{R}{P+R}. \qquad (30)$$

Thus, $\widehat{mb}$ can be calculated as follows:

$$\begin{aligned}
\widehat{mb} &= \frac{1}{T}\sum_{k_b=1}^{T-1} k\big[P_b(T-k_b+1) - P_b(T-k_b)\big] \\
&= \frac{R}{T(P+R)}\sum_{k_b=1}^{T-1}\Big[\sum_{s=0}^{1} h(N, T-k_b+1, s) - \sum_{\lambda=0}^{1} h(N, T-k_b, s)\Big]k_b \\
&= \frac{R}{T(P+R)}\sum_{j=2}^{T}\Big[\sum_{s=0}^{1} h(N, j, s) - \sum_{s=0}^{1} h(N, j-1, \lambda)\Big](T-j+1).
\end{aligned} \qquad (31)$$

Analogously,

$$\widehat{P}_f = \frac{1}{T}\sum_{j=2}^{T}\Big[\sum_{s=0}^{1} h(N, j, s) - \sum_{s=0}^{1} h(N, j-1, s)\Big](T-j+1). \qquad (32)$$

To summarize, it is convenient to introduce an operator of stationarization, Φ_1, defined as follows:

$$\begin{aligned}
\widehat{P}_f &= \Phi_1(P, R, N, T, D) \\
&:= \frac{1}{T}\sum_{j=2}^{T}\Big[\sum_{s=0}^{1} h(N, j, s) - \sum_{s=0}^{1} h(N, j-1, s)\Big](T-j+1).
\end{aligned} \qquad (33)$$

where $h(N, j, s)$ is calculated according to (23)-(29). Expression (33) is used in subsequent subsections for DTP calculation.

4.2.3 Iterations. Consider the serial or assembly line shown in Figures 1. To adopt the above stationarization procedure, assume that the probability that the last machine m_M is starved in the communication sense is known. Denote the estimate of this probability as $\widehat{p}_s$. Modify R_M by multiplying it by $1-\widehat{p}_s$ and P_M by adding it by $R_M\widehat{p}_s$, denote the modified machine as m'_M. Using this machine and procedure (23)-(29) and (33), calculate the stationarized probability, $\widehat{P}_f$. Introduce now another fictitious machine, m''_M, defined by $R_M(1-\widehat{P}_f)$ and $P_M + R_M\widehat{P}_f$. Using this machine and deleting FGB, calculate a new $\widehat{p}_s$ by employing the recursive procedure for performance analysis of serial lines developed in Section 3. Having this probability, repeat the process described above anew.

In other words, in the case of serial lines, the iterations have the form:

$$\begin{aligned}
R'_M(n+1) &= R_M[1-\widehat{p}_s(n)], \\
P'_M(n+1) &= P_M + R_M\widehat{p}_s(n), \\
\widehat{P}_f(n+1) &= \Phi_1\big(P'_M(n+1), R'_M(n+1), N_M, T, D\big), \\
R''_M(n+1) &= R_M[1-\widehat{P}_f(n+1)], \\
P''_M(n+1) &= P_M + R_M\widehat{P}_f(n+1), \\
\widehat{p}_s(n+1) &= \Phi_2\big(P_1, R_1, \cdots, P''_M(n+1), R''_M(n+1), N_1, \cdots, N_{M-1}\big),
\end{aligned} \qquad (34)$$

where Φ_1 is the stationarization operator defined in (33) and Φ_2 is the operator that represents the mapping from P''_M, R''_M to $\widehat{p}_s$. This operator is described Section 4.3.

If procedure (34) is convergent and the following limits exist,

$$\lim_{n\to\infty} \widehat{p}_s(n) := \widehat{p}_s,$$

the DTP of a M-machine serial line can be evaluated, using (18), as the DTP of a one-machine system defined by $R_M(1-\widehat{p}_s)$, $P_M + R_M\widehat{p}_s$ and FGB of capacity N_M, i.e.,

$$\widehat{DTP}_M = DTP_1(P_M + R_M\widehat{p}_s, R_M(1-\widehat{p}_s), N_M, D, T). \qquad (35)$$

In Subsections 4.3-4.5, operators Φ_2 is described, convergence of the iterations is considered, and the accuracy of DTP calculations is investigated.

4.3 Operator Φ_2

Operator Φ_2 can be defined through the aggregation procedure for performance analysis of serial lines developed in Section 3. Indeed, assume that $P''_M(n+1)$ and $R''_M(n+1)$, calculated according to (34), are known. Consider the serial line consisting of M machines with parameters $P_1, R_1, \cdots, P_{M-1}, R_{M-1}$, $P''_M(n+1)$, $R''_M(n+1)$ and $M-1$ in-process buffers $N_1, \cdots, N_{M-1}$. According to Section 3, its performance can be analyzed using the convergent recursive procedure 1. In terms of the steady state of this procedure, the probability that buffer B_{M-1} is empty can be estimated as

$$\widehat{p}_s(n+1) = 1 - \frac{\frac{R_M^f(n+1)}{P_M^f(n+1)+R_M^f(n+1)}}{\frac{R_M}{P_M+R_M}}. \quad (36)$$

4.4 Convergence

To prove the convergence of recursive procedure (34), it is important to know whether function $\widehat{P}_f$, defined by (23)–(27) and (33), is monotonic with respect to P and R. Intuitively, this property clearly takes place. However, a rigorous proof of this fact seems to be all but impossible. Therefore, although in every example numerically analyzed this function was found to be monotonically increasing, we introduce this property as a hypothesis:

Hypothesis 2 *For system (i)-(ix) with $M = 1$, the stationarized probability that FGB is full, $\widehat{P}_f$, defined by (23)-(27) and (33), is a monotonically increasing function of R and monotonically decreasing function of P.*

An illustration of this property is given in Figure 5.

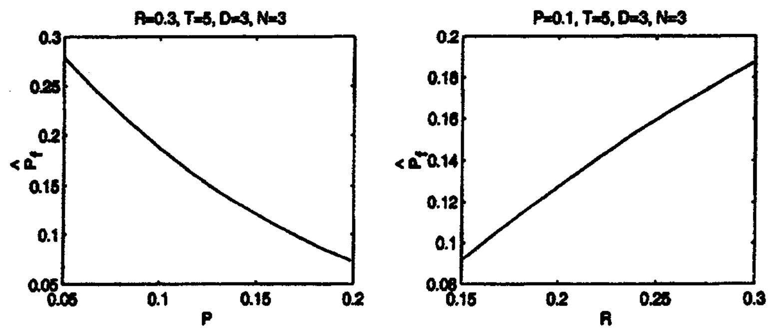

Figure 5. $\widehat{P}_f$: Monotonicity with respect to P and R

Theorem 4 *Under Hypothesis 2, iteration procedure (34), (36) is convergent, i.e., the following limits exist:*

$$\lim_{n\to\infty} \widehat{P}_f(n) := \widehat{P}_f,$$
$$\lim_{n\to\infty} \widehat{p}_s(n) := \widehat{p}_s. \quad (37)$$

Proof: See Appendix. ■

Thus, the DTP_M of a serial production defined by (i)-(ix) can be calculated using (35) with $\widehat{p}_s$ defined by (37).

4.5 Accuracy

The accuracy of $\widehat{DTP}_M$ has been evaluated using discrete event simulation. We simulated dozens of systems defined by (i)-(ix) with various machine, buffer, demand, and epoch parameters. Table 2 presents 20 of them with 2 - 6 machines. In each run of the corresponding discrete event model, zero initial conditions for all buffers have been assumed and $10,000T$ time slots of warm up period has been carried out (T, as before, is the length of the epoch). The next $100,000T$ slots of stationary regime have been used to statistically evaluate the DTP. In Table 2, DTP denotes the actual DTP obtained by simulation, whereas $\widehat{DTP}$ denotes the estimate of DTP calculated according to (8), (34)-(36). As it can be seen from Table 2, the estimates result in a relatively high precision, with errors ranging from 0.26% to 2.82%.

Remark 5 Calculation of DTP using expressions (8), (34)-(36) is orders of magnitude faster than using discrete event simulations. For example, discrete event simulation of every system included in Table 2 takes, on the average, 40 minutes using HP C360 workstation, whereas calculation according to (8), (34)-(36) takes about 1 second. □

5. Structural Properties

5.1 Load Factor

This Section illustrates the utility of DTP calculations by analyzing structural properties of production lines with FGBs. Two questions are addressed: Which capacity of FGB is necessary to ensure sufficiently high DTP? What are monotonicity properties of DTP, in particular with respect to the shipping period T? It turns out that answers to both questions depend on how large is the demand, D, vis-a-vis the production capacity of the system, Tp_a, where, as before, p_a is the production rate of the corresponding system without the FGB. We formalize this relationship by the load factor, L, defined as follows:

$$L = \frac{D}{Tp_a}. \quad (38)$$

Due to (1), $0 \leq L \leq 1$. Obviously, large L implies that the demand is heavy, relative to the average production capacity of the system; small L means that the production capacity is under-utilized.

The load factor (38) was introduced and analyzed by Jacobs and Meerkov (1995a) for a single machine production system without FGB. Here we use this notion in the general case.

Table 2. Numerical justification of DTP estimation for serial line (err$=\frac{|DTP-\widehat{DTP}|}{DTP}\cdot 100\%$)

P_i	R_i	N_i	T	D	DTP	$\widehat{DTP}$	err
0.12 0.14	0.43 0.38	1 1	5	3	0.713	0.705	1.05
0.05 0.10	0.37 0.42	2 1	4	3	0.796	0.798	0.33
0.12 0.12	0.76 0.76	2 2	5	4	0.907	0.904	0.26
0.10 0.10	0.40 0.40	1 1	3	2	0.774	0.771	0.34
0.05 0.05	0.30 0.30	1 2	4	3	0.805	0.817	1.45
0.12 0.08	0.35 0.39	2 3	3	2	0.875	0.892	2.00
0.10 0.12	0.33 0.40	2 2	5	3	0.848	0.850	0.26
0.10 0.05 0.20	0.90 0.85 0.75	3 2 3	4	3	0.962	0.968	0.60
0.10 0.20 0.06	0.41 0.58 0.32	2 2 3	5	3	0.879	0.855	2.82
0.10 0.10 0.10	0.42 0.42 0.42	3 3 3	3	2	0.907	0.915	0.80
0.12 0.15 0.10	0.43 0.46 0.50	2 2 2	5	3	0.857	0.849	0.93
0.08 0.12 0.10	0.45 0.44 0.42	1 2 2	5	3	0.835	0.820	1.77
0.10 0.12 0.10 0.09	0.46 0.49 0.47 0.48	3 3 3 2	3	2	0.912	0.911	0.30
0.11 0.08 0.08 0.11	0.40 0.41 0.41 0.40	2 3 2 2	5	3	0.850	0.831	2.22
0.06 0.08 0.05 0.10	0.36 0.39 0.42 0.37	3 3 4 3	3	2	0.918	0.935	1.88
0.15 0.04 0.30 0.02	0.50 0.80 0.40 0.70	3 2 3 2	6	3	0.948	0.965	1.83
0.06 0.07 0.09 0.08	0.42 0.41 0.44 0.43	2 3 2 3	3	2	0.908	0.915	0.77
0.10 0.07 0.09 0.12 0.11	0.40 0.35 0.33 0.42 0.39	4 4 4 3 2	5	3	0.867	0.843	2.43
0.10 0.10 0.10 0.10 0.10	0.41 0.41 0.41 0.41 0.41	3 3 3 3 3	5	3	0.901	0.892	1.07
0.08 0.10 0.12 0.13 0.09 0.11	0.45 0.46 0.44 0.47 0.43 0.42	3 2 4 3 4 3	5	3	0.924	0.918	0.62

5.2 FGB Capacity for High DTP

Consider a serial line with five machines and four in-process buffers, each with capacity 5. Assume that parameters P_i of the machines are identical ($P_i = P = 0.05, i = 1, \cdots, 5$), while R_i are constrained by $\sum_{i=1}^{5} R_i = R^*$ and consider four types of R_i allocations:

uniform: $R_i = R, \quad \forall i, i = 1, \cdots, 5;$
inverted bowl: $R_3 > R_2 = R_4 > R_1 = R_5;$
ramp: $R_1 < R_2 < R_3 < R_4 < R_5;$
inverted ramp: $R_1 > R_2 > R_3 > R_4 > R_5.$

Specific values of R_i for each of these allocations are shown in Figures 6 and 7 with $R^* = 2$ and 1.75, respectively, along with the corresponding production rate, *PR* (i.e., the average number of parts produced by the last machine per cycle time when no FGB is present), and the load factor of each line. Note that in Figure 6 the load is relatively low whereas in Figure 7 it is high.

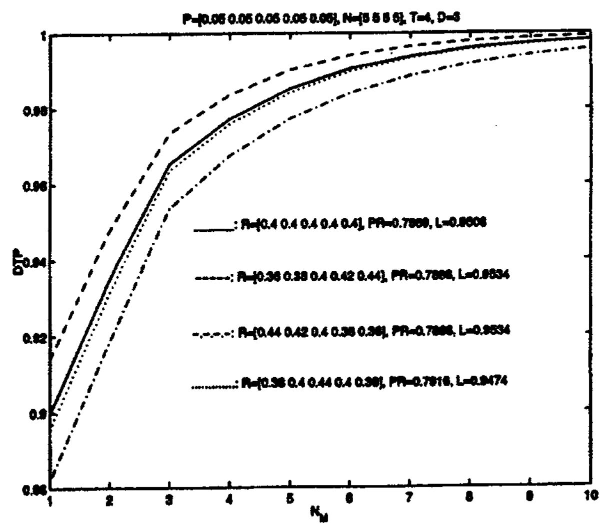

Figure 6. Lower load case (P_i=const, R_i=var)

Next, we assume that parameters R_i of the machines are constant, while P_i are constrained by $\sum_{i=1}^{5} P_i = P^*$ and consider four types of P_i allocations. (To be consistent with the case of constraint on R_i, we categorized these types from the point of view of machine efficiency: $e_i = \frac{R_i}{P_i+R_i}$.)

uniform: $P_i = P, \quad \forall i, i = 1, \cdots, 5;$
inverted bowl: $P_3 < P_2 = P_4 < P_1 = P_5;$
ramp: $P_1 > P_2 > P_3 > P_4 > P_5;$
inverted ramp: $P_1 < P_2 < P_3 < P_4 < P_5.$

Specific values of P_i for each of these allocations are shown in Figures 8 and 9 with $P^* = 0.25$, but $R_i = 0.4, i = 1, \cdots, 5$, and $R_i = 0.35, i = 1, \cdots, 5$, respectively, along with the corresponding production rate, *PR*, of each line. Figures 8 and 9 correspond to relatively low and relatively high loads, respectively.

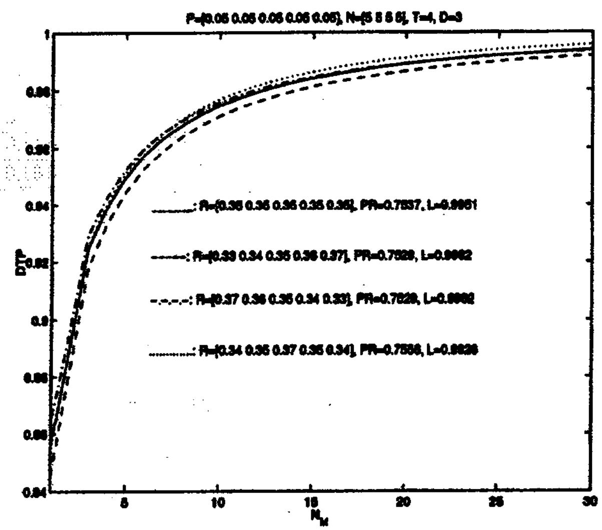

Figure 7. Higher load case (P_i=const, R_i=var)

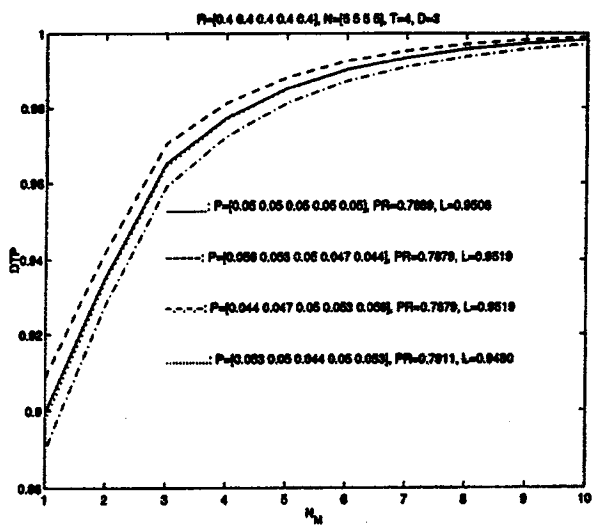

Figure 8. Lower load case (R_i=const, P_i=var)

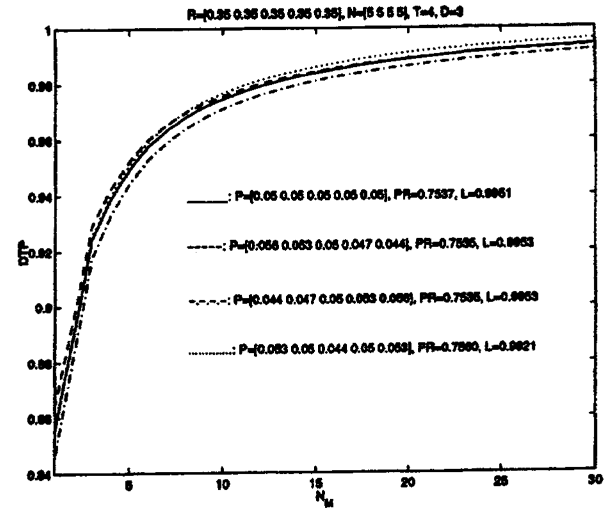

Figure 9. Higher load case (R_i=const, P_i=var)

Within this scenario, using (8), (34)-(36), we calculate the $\widehat{DTP}$ for all lines involved. The results are shown in Figures 6 - 9. Examining these data, we arrive at the following observations:

1. For lighter load and for all R_i and P_i allocations, FGB capacity of about 2 - 4 shipments is sufficient to provide high DTP.

2. For the heavy load, FGB of capacity 6 - 9 shipments is necessary to provide high DTP for all R_i and P_i allocations.

3. Although, as it is well known (Hiller and Boling 1966), *PR* is maximized by the inverted bowl e_i allocation (where $e_i = \frac{R_i}{P_i+R_i}$), the *DTP* is maximized by the ramp allocation, with the exception of relatively large N_M in the case of heavy load. Thus, a ramp, rather than an inverted bowl, is the appropriate e_i allocation if DTP is to be optimized. The advantage of ramp allocation is particularly clear for relatively small N_M, i.e., in the case of the so-called lean operation.

Remark 6 Although the above observations are based on the data obtained via approximate calculations (8), (34)-(36), we believe that relative properties of various allocations remain valid for the real systems as well. Discrete event simulation support these conclusions. □

5.3 Monotonicity

Intuitively, it is clear that DTP is monotonically decreasing as a function of the load factor, L, and P_i, and increasing as a function of R_i and N_i, i =

$1, \cdots, M$. Although analytical proof of these facts seems to be impossible, all systems analyzed in this research, using (8), (34)-(36), exhibited these properties. Examples are shown in Figures 10 and 11. (In Figure 10, values of $L > 1$ are included for the sake of mathematical completeness.)

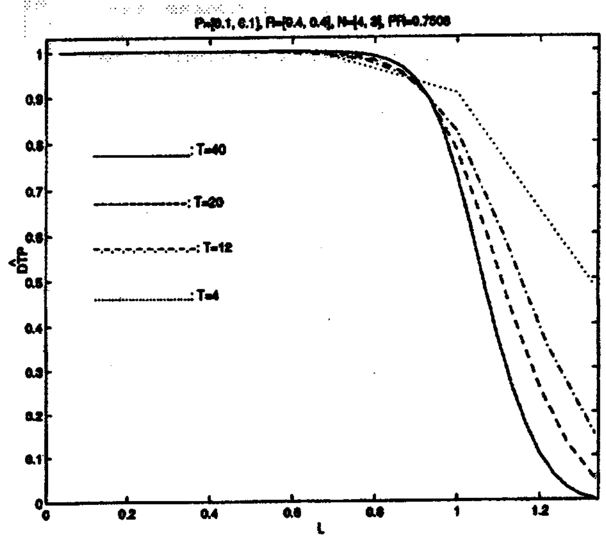

Figure 10. Monotonicity with respect to load factor

On the other hand, the intuitively expected property of DTP monotonicity with respect to the shipping period, T, does not take place. Several examples are shown in Figures 12 and 13. (In Figure 12, the value of D/T is kept constant, while T is changing; in Figure 13, the value of PR is fixed.) The lack of monotonicity is especially pronounced for high load factors and decreased for relatively small L. Discrete event simulations also support this conclusion.

The counterintuitive behavior of Figures 12 and 13 may be explained as follows: For smaller T (which implies smaller D), the ratio N_M/D is larger, and the probability to satisfy the demand by the contents of FGB is larger than that for relatively larger T. This explains the decrease of DTP. Only when T becomes so large that the averaging effect of the shipping period becomes dominant, the system ships reliably all demands, D, at or below its capacity.

Thus, smaller but more frequent shipments are easier to meet than larger but less frequent ones. This conclusion is important for lean operations, since the regime when DTP is increasing as a function of T occurs for relatively large T, typically associated with mass production operations.

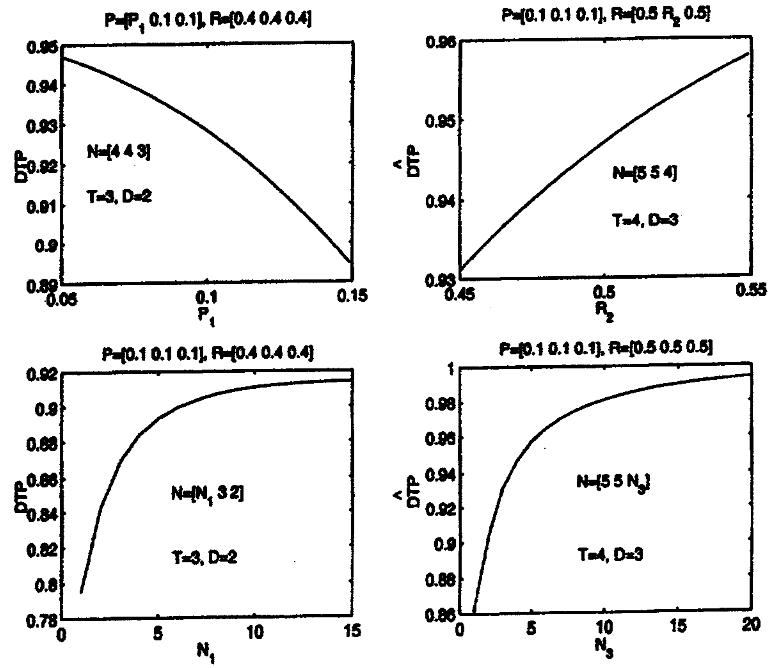

Figure 11. Monotonicity with respect to P_i, R_i and N_i

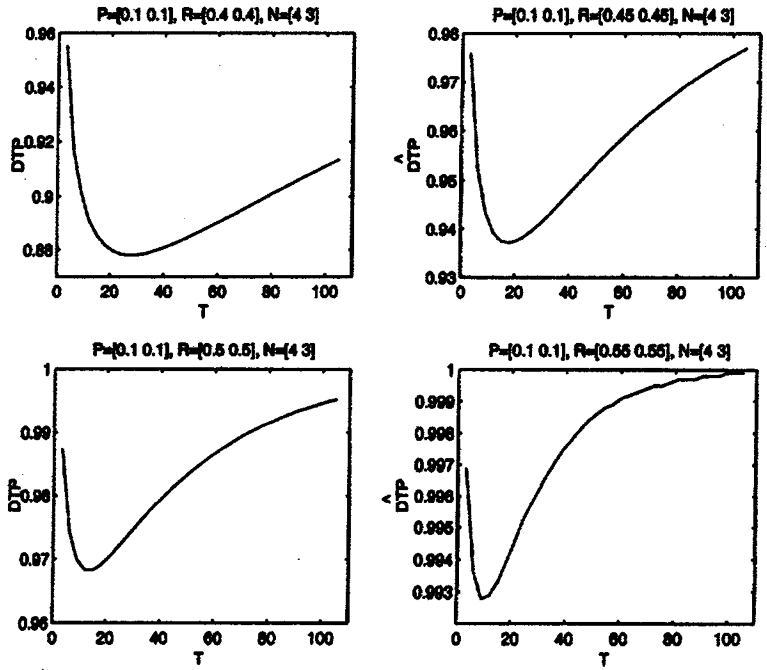

Figure 12. Lack of monotonicity with respect to shipping period (the case of D/T=constant)

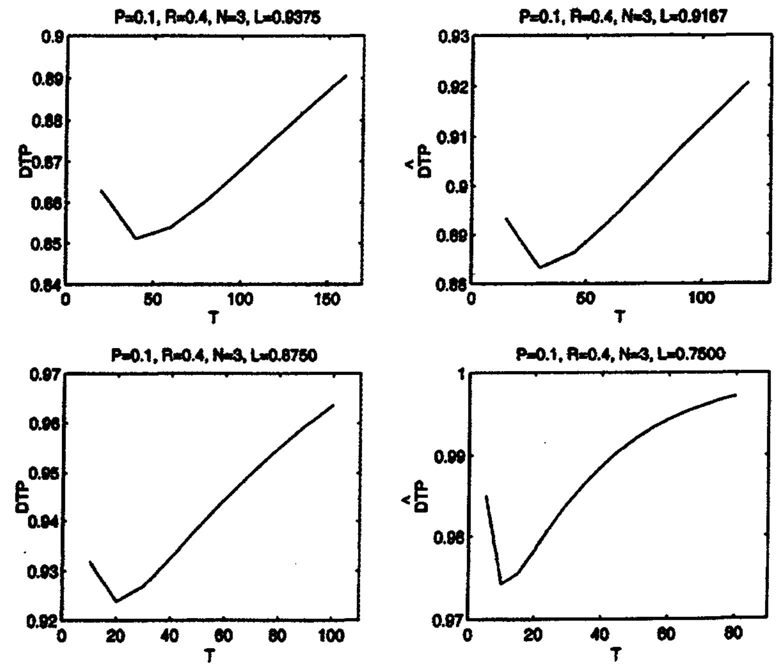

Figure 13. Lack of monotonicity with respect to shipping period (the case of PR=constant)

6. Conclusions

This paper provides a method for calculating Due-Time Performance measure in serial production lines with Markovian machines, finite in-process buffering, and finite Finished Goods Buffer. The method is computationally efficient and results in sufficiently high accuracy. Based on this method, system properties of DTP are analyzed, and FGB capacity necessary for high DTP is investigated.

The results obtained may be useful for analysis and design of production systems from the point of view of customer demand satisfaction and for design of efficient supply chains.

Appendix: Proofs

Due to page limitation, we provide here only sketch of the proofs. Complete proofs can be found in Li and Meerkov (2000c).

Proof of Theorem 1

The proof of Theorem 1 consists of the following three steps:

Step 1: Derivation of the steady state balance equations.

First, introduce the following steady state probabilities

$$Y_{k,s_1s_2} = \text{Prob}\{k \text{ parts in the buffer}, m_1 \text{ and } m_2 \text{ are in states } s_1 \text{ and } s_2 \text{ respectively at beginning of the slot}\}, \quad k = 0, 1, \cdots, N,$$

where

$$s_i = \begin{cases} 1, & m_i \text{ is up}, \\ 0, & m_i \text{ is down}, \end{cases} \quad i = 1, 2.$$

Next, write the balance equations for empty buffer, buffer occupancy equaled to 1 ($N = 1$, $N > 1$, respectively), buffer occupancy equaled to k, $1 < k < N$, and the full buffer, respectively.

Step 2: Analysis of case $N = 1$.

- Write $Y_{0,00}$, $Y_{0,10}$, $Y_{0,01}$, $Y_{1,11}$, $Y_{1,10}$, $Y_{1,01}$, $Y_{1,00}$ in terms of $Y_{0,11}$.
- From the fact that the total probability is equaled to 1, calculate $Y_{0,11}$ and $Q(P_1, R_1, P_2, R_2, N)$.

$$Y_{0,11} = \frac{R_1R_2P_1\beta_2}{(R_1 + R_2 - R_1R_2)(R_1 + P_1)(R_2 + P_2)}.$$

It follows then that

$$Q(P_1, R_1, P_2, R_2, N) = \frac{Y_{0,11}}{e_2R_1} = \frac{P_1\beta_2}{(R_1 + R_2 - R_1R_2)(R_1 + P_1)}.$$

Step 3: Analysis of the case $N > 1$.

- Write $Y_{k,11}$, $Y_{k,10}$, $Y_{k,01}$, $Y_{k,00}$, $k = 1, \cdots, N$, in terms of $Y_{0,11}$.

- From the fact that the total probability is equaled to 1, calculate $Y_{0,11}$ and $Q(P_1, R_1, P_2, R_2, N)$.

$$\begin{aligned} Y_{0,11} &= P_1R_1R_2\alpha_1\alpha_2\beta_2^2\Big[P_1R_2\alpha_1\alpha_2\beta_2(P_2+\beta_2) \\ &\quad + P_1R_1R_2\alpha_2\big[\beta_2^2 + P_2(\alpha_1+\beta_1)(\alpha_2+2\beta_2)\big] \\ &\quad + \sum_{k=2}^{N-1} P_1P_2R_1R_2(\alpha_2+\beta_2)^3\sigma^{k-1} \\ &\quad + P_2R_1\alpha_1\beta_2\big(R_2(\alpha_1+\beta_1) + \alpha_2(P_1+R_1)\big)\sigma^{N-1}\Big]^{-1} \\ &= \frac{P_1R_1R_2\alpha_1\alpha_2\beta_2^2}{A+B+C+D}, \end{aligned}$$

where,

$$\begin{aligned} A &= P_1R_2\alpha_1\alpha_2\beta_2(P_2+\beta_2), \\ B &= P_1R_1R_2\alpha_2\big[\beta_2^2 + P_2(\alpha_1+\beta_1)(\alpha_2+2\beta_2)\big], \\ C &= \sum_{k=2}^{N-1} P_1P_2R_1R_2(\alpha_2+\beta_2)^3\sigma^{k-1}, \\ D &= P_2R_1\alpha_1\beta_2\big[R_2(\alpha_1+\beta_1) + \alpha_2(P_1+R_1)\big]\sigma^{N-1}. \end{aligned}$$

It follows then that,

$$Q(P_1, R_1, P_2, R_2, N) = \frac{P_1\alpha_1\alpha_2\beta_2^2(R_2+P_2)}{A+B+C+D}.$$

- Calculate $E(h)$, ms_2, mb_1, where

$$\begin{aligned} E(h) &= \sum_{k=1}^{N} k(Y_{k,11} + Y_{k,10} + Y_{k,01} + Y_{k,00}), \\ ms_2 &= Y_{0,11} + Y_{0,01}, \\ mb_1 &= Y_{N,10}. \end{aligned}$$

Theorem 1 is proved. ■

Proof of Proposition 1

Rewrite (5) as

$$Q(P_1, R_1, P_2, R_2, N) = \frac{R_1 + R_2 - R_1R_2 - P_2R_1}{(R_1 + R_2 - R_1R_2)(1 + \frac{R_1}{P_1})}.$$

It follows then that $Q(P_1, R_1, P_2, R_2, N)$ is monotonically increasing with respect to P_1 and monotonically decreasing with respect to P_2.

Rewrite (5) as

$$Q(P_1, R_1, P_2, R_2, N) = \frac{P_1}{R_1 + P_1}\left(1 - \frac{P_2}{1 + \frac{R_2}{R_1} - R_2}\right).$$

It follows then that $Q(P_1, R_1, P_2, R_2, N)$ is monotonically decreasing with respect to R_1, and monotonically increasing with respect to R_2. Proposition 1 is proved. ■

The proof of Theorem 2 is based on the following 5 Lemmas, the proofs of which can be found in Li and Meerkov (2000c).

Lemma 1 *Function $Q(P_1, R_1, P_2, R_2, N)$, defined in (5) is monotonically decreasing with respect to N.*

Lemma 2 *Function $Q(P_1, R_1, P_2, R_2, N)$, defined in (5) takes values on* $(0, 1)$.

Lemmas 3 and 4 are used to analyze the monotonicity properties of sequences $P_j^b(s)$, $R_j^b(s)$ and $P_j^f(s)$, $R_j^f(s)$, as formulated next.

Lemma 3 *Consider $P_i^f(s)$, $R_i^f(s)$, $P_i^b(s)$ and $R_i^b(s)$, $i = 1, \cdots, M$, defined by recursive procedure 1. If for all $j = 2, \cdots, M$, $R_j^f(s) < R_j^f(s-1)$ and $P_j^f(s) > P_j^f(s-1)$, then for all $j = 1, \cdots, M-1$, $R_j^b(s) > R_j^b(s-1)$ and $P_j^b(s) < P_j^b(s-1)$.*

Lemma 4 *If for all $j = 1, \cdots, M-1$, $R_j^b(s+1) > R_j^b(s)$ and $P_j^b(s+1) < P_j^b(s)$, then for all $j = 2, \cdots, M$, $R_j^f(s+1) < R_j^f(s)$ and $P_j^f(s+1) > P_j^f(s)$.*

Lemma 5 *Sequences $P_j^f(s)$ and $P_j^b(s)$ are monotonically increasing and sequences $R_j^f(s)$ and $R_j^b(s)$ are monotonically decreasing.*

Proof of Theorem 2: Under the assumptions of the Theorem, since the sequences $P_i^f(s)$, $R_i^f(s)$, $P_i^b(s)$ and $R_i^b(s)$, $i \leq j \leq M$, are monotonic (Lemma 5) and bounded from above and below (Lemma 2), they are convergent. This proves (9).

To prove (10), consider the steady state equations of the recursive procedure (1) and define

$$e_i^f = \frac{R_i^f}{R_i^f + P_i^f}, \quad i = 1, \cdots, M,$$
$$e_i^b = \frac{R_i^b}{R_i^b + P_i^b}, \quad i = 1, \cdots, M.$$

The following property holds (see Li and Meerkov 2000c):

$$\frac{e_i^f e_i^b}{e_i} = \frac{e_j^f e_j^b}{e_j}, \quad i, j = 1, \cdots, M, \forall i \neq j.$$

Therefore, $\frac{R_M^f}{P_M^f} = \frac{R_1^b}{P_1^b}$. Theorem 2 is proved. ∎

Proof of Theorem 4

By induction: For $n = 0$, $\widehat{p}_s(0) = 0$, then

$$P_M'(1) = P_M, \quad R_M'(1) = R_M,$$
$$\widehat{P}_f(1) = \Phi_1(P_M, R_M, N_M, T, D),$$
$$R_M''(1) = R_M[1 - \widehat{P}_f(1)],$$
$$P_M''(1) = P_M + R_M\widehat{P}_f(1).$$

Thus,

$$\widehat{p}_s(1) = \Phi_2(P_1, R_1, \cdots, P_M + R_M\widehat{P}_f(1), R_M[1 - \widehat{P}_f(1)], N_1, \cdots, N_{M-1}) > 0,$$
$$R_M'(2) = R_M[1 - \widehat{p}_s(1)] < R_M'(1),$$
$$P_M'(2) = P_M + R_M\widehat{p}_s(1) > P_M'(1).$$

From Hypothesis 2, we have

$$\widehat{P}_f(2) < \widehat{P}_f(1).$$

Assume that for $n > 0$,

$$\widehat{p}_s(n) > \widehat{p}_s(n-1).$$

Then

$$\widehat{P}_f(n+1) < \widehat{P}_f(n),$$
$$R_M''(n+1) > R_M''(n),$$
$$P_M''(n+1) < P_M''(n).$$

Due to the monotonicity of operator Φ_2, we have

$$\widehat{p}_s(n+1) > \widehat{p}_s(n),$$

which implies that

$$\widehat{P}_f(n+2) < \widehat{P}_f(n+1).$$

Therefore, the sequences $\widehat{p}_s(n)$ and $\widehat{P}_f(n)$ are monotonically increasing and decreasing, respectively. From (5) and (31), $\widehat{p}_s(n)$ and $\widehat{P}_f(n)$ are bounded from above and below. Therefore, they are convergent. Theorem 4 is proved. ∎

References

[1] Altiok, T. (1997). *Performance Analysis of Manufacturing Systems*, Springer.

[2] Buzacott, J. A., and J. G. Shanthikumar. (1993). *Stochastic Models of Manufacturing Systems*, Prentice Hall.

[3] Chiang, S. -Y. (1999). "Bottlenecks in Production Systems with Markovian Machines: Theory and Applications", *Ph.D Thesis*, Dept. of EECS, Univ. of Michigan, Ann Arbor, MI.

[4] Dallery, Y., R. David and X. L. Xie. (1988). "An Efficient Algorithm for Analysis of Transfer Lines with Unreliable Machines and Finite Buffers", *IIE Transactions*, Vol. 20, pp. 280-283.

[5] Dallery, Y., and S. B. Gershwin. (1992). "Manufacturing Flow Line Systems: A Review of Models and Analytical Results", *Queuing Systems*, Vol. 12, pp. 3-94.

[6] Gershwin, S. B. (1993). "Variance of Output of a Tandem Production System", in *Queuing Networks with Finite Capacity*, pp. 291-304, ed. R. O. Onvaral and I. F. Akyildi, Elsevier Science Publishers.

[7] Gershwin, S. B. (1994). *Manufacturing Systems Engineering*, Prentice Hall.

[8] Govil, M. C., and M. C. Fu. (1999). "Queueing Theory in Manufacturing: A Survey", *Journal of Manufacturing Systems*, Vol. 18, pp. 214-240.

[9] Hillier, F. S., and R. W. Boling. (1966). "The Effect of Some Design Factors on the Efficiency of Production Lines with Variable Operation Times", *Journal of Industrial Engineering*, Vol. 17, pp. 651-658.

[10] Jacobs, D. A. (1993). "Improvability in Production Systems: Theory & Case Studies", *Ph.D. Thesis*, Dept. of EECS, Univ. of Michigan, Ann Arbor, MI.

[11] Jacobs, D. A., and S. M. Meerkov. (1995a). "System-Theoretic Analysis of Due-Time Performance in Production Systems", *Mathematical Problems in Engineering*, Vol. 1, pp. 225-243.

[12] Jacobs, D. A., and S. M. Meerkov. (1995b). "A System-theoretic Property of Serial Production Lines: Improvability", *International Journal of System Sciences*, Vol. 26, pp. 755-785.

[13] Li, J., and S. M. Meerkov. (1998). "Production Variability in Manufacturing Systems: Problem Formulation and Performance Bounds", *Proc. of 37th IEEE CDC*, pp. 2730-2735, Tampa, FL.

[14] Li, J., and S. M. Meerkov. (2000a). "Production Variability in Manufacturing Systems: Bernoulli Reliability Model", *Annals of Operations Research*, Vol. 93, pp. 299-324.

[15] Li, J., and S. M. Meerkov. (2000b). "Bottleneck with respect to Due-Time Performance in Pull Serial Production Lines", *Mathematical Problems in Engineering*, Vol. 6, pp. 479-498.

[16] Li, J., and S. M. Meerkov. (2000c). "Due-Time Performance in Markovian Production Systems with Finished Goods Buffers", *Control Group Report No. CGR 00-05*, EECS Dept., Univ. of Michigan.

[17] Li, J., and S. M. Meerkov. (2001). "Customer Demand Satisfaction in Production Systems: A Due-Time Performance Approach", *IEEE Transactions on Robotics and Automation*, Vol. 17, pp. 472-482.

[18] Papadopoulos, H. T., and C. Harvey. (1996). "Queueing Theory in Manufacturing Systems Analysis and Design: A Classification of Models for Production and Transfer Lines", *European Journal of Operational Research*, Vol. 92, pp. 1-27.

[19] Papadopoulos, H. T., C. Harvey and J. Browne. (1993). *Queueing Theory in Manufacturing Systems Analysis and Design*, Chapman & Hall.

[20] Sheskin, T. J. (1976). "Allocation of Interstage Along an Automatic Production Line", *AIIE Transactions*, Vol. 8, pp. 146-152.

[21] Tan, B. (1998). "Effects of Variability on the Due-Time Performance of Production Lines", *International Journal of Production Economics*, Vol. 54, pp. 87-100.

[22] Tan, B. (1999). "Variance of the Output as a Function of Time: Production Line Dynamics", *European Journal of Operational Research*, Vol. 117, pp. 470-484.

[23] Viswanadham, N., and Y. Narahari. (1992). *Performance Modeling of Automated Manufacturing System*, Practice Hall.