Customer Demand Satisfaction in Production Systems: A Due-Time Performance Approach

Jingshan Li, Member, IEEE, and Semyon M. Meerkov, Fellow, IEEE

Abstract—The problem of customer demand satisfaction in production systems with unreliable machines and finite finished goods buffers (FGB) isaddressed. The measure of customer demand satisfaction is characterized by the probability to ship to the customer a required number of parts during a fixed time interval. This measure, referred to as the due-time performance (DTP), is often used to characterize the quality of a supplier in the automotive industry supply chain. In this paper, a method for evaluating DTP in serial and assembly lines isdeveloped and the problem of selecting capacity of the FGB is discussed. The results obtained are illustrated by a case study at an automotive component plant.

Index Terms—Due-time performance, finished goods buffer, production systems, supply chain, unreliable machines.

I. INTRODUCTION

A. Problem Addressed

This paper addresses the problem of customer demand satisfaction in supply chain management. Each supplier is assumed to be a production system with unreliable machines. Each customer requires the supplier to ship him a required number of parts during a fixed interval of time. This arrangement is typical in the automotive industry. For instance, many engine and seat plants supply car and truck assembly plants on a four-hour delivery schedule. Similarly, engine plants are expected to receive ignition and injection parts from the second tier suppliers on a fixed-interval delivery schedule, etc.

If the machines were absolutely reliable, perfect customer demand satisfaction would be possible. In reality, however, the machines are never completely reliable and experience random breakdowns. In this situation, the number of parts (units or subsystems) produced by a production system during a fixed time interval is a random variable. Its distribution characterizes both production volume and production variability. The production volume is typically defined as the expected value of this random variable. The production variability can be defined as the probability to ship to the customer a required number of parts during a fixed time interval. This probability is referred to as the due-time performance (DTP). Section I-B gives a formal definition of this performance measure.

Production volume has been studied in numerous publications over the last 50 years (see, for instance, review [1] and monographs [2]–[6]). In contrast, DTP was discussed in just a few recent articles [7]–[11], although other variability measures, such as production variance, have been studied more intensively [12]–[20]. The notion of DTP was introduced in [7] in the context of a single machine production system without a finished goods buffer (FGB). References [8] and [9] also addressed DTP in production lines without FGBs. DTP in production lines with FGBs was analyzed in [10] and [11]. In particular, a formula for DTP calculation in a single machine system with FGB was derived and it was shown that even a relatively small FGB dramatically improves the DTP. However, no methods for calculating DTP in production systems with more than one machine have been developed.

The purpose and the contribution of this paper are in providing a method for calculating DTP in multimachine serial and assembly lines with FGBs.

To this end, Sections I-B and I-C give the problem formulation and review results for the single machine case, respectively. Section II describes the idea of the approach developed in this paper. Section III is devoted to DTP calculations in serial and assembly lines with FGBs. Several structural properties of DTP are considered in Section IV. The case study is described in Section V. Finally, the conclusions are formulated in Section VI. The proofs are given in the Appendix.

B. Problem Formulation

The structure of production systems analyzed in this work is shown in Figs. 1 (serial line) and 2 (assembly line), where the circles represent the machines and rectangles are the buffers. Obviously, Fig. 2 reduces to Fig. 1 when the number of parallel lines in Fig. 2 is one. Assumptions formulated below define the operation of both serial and assembly lines. These assumptions are classified into four groups: those pertaining to the machines, the buffers, the interaction between the machines and buffers, and the demand.

Machines:

i) Each machine, \( m_i \), \( i = 1, \ldots, M \), in serial line (respectively, \( m_{ij} \), \( i = 0, \ldots, S, j = 1, \ldots, M_j \), in assembly...
line) 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 equal to the cycle time.

ii) During a cycle time, each machine can be in either “up” or “down.” When up, the machine could process a part. When down, no processing takes place.

iii) The status of the machines, i.e., up or down, is determined by the process of Bernoulli trials. In other words, it is assumed that during each slot, machine \( m_i \), \( i = 1, \ldots, M \) (respectively, \( m_{ij} \), \( i = 0, \ldots, S, j = 1, \ldots, M_i \)), is up with probability \( p_i \) (respectively, \( p_{ij} \)) and down with probability \( 1 - p_i \) (respectively, \( 1 - p_{ij} \)); the status of the machine is determined at the beginning of each cycle, independent of the status of this machine in the previous cycle.

**Remark 1:** Assumption i) on the fixed processing time is appropriate for large volume, automated manufacturing environment. On the other hand, the assumption on identical cycle time may or may not hold, depending on the nature of the production system. Typically, in systems with automated material handling, this assumption is satisfied.

**Remark 2:** Assumption iii) defines the Bernoulli statistics of machine breakdowns. The Bernoulli model is appropriate for modular production lines where operators use the push-buttons and stop a module of the operational conveyor in order to accomplish the operation with the highest possible quality. The duration of this “breakdown” is short and of the order of the cycle time and, therefore, the probability to produce a part during a cycle time arises naturally. Another frequent perturbation is pallet jam on the operational conveyor; to correct for this problem also a short period of time is required. In many car and engine assembly lines these are the predominant perturbations.

**Remark 3:** Assumptions iii), v), and vi) are formulated in terms of time-dependent failures, i.e., machines can go down even when blocked or starved [3], [4]. Another possible model is that of operation-dependent failures, where no breakdowns of starved or blocked machines is possible [3], [4]. 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.

**Buffers:**

iv) Each buffer \( B_i, i = 1, \ldots, M \) in serial line (respectively, \( B_{ij}, i = 0, \ldots, S, j = 1, \ldots, M_i \) in assembly line) has capacity \( 0 \leq N_i < \infty \) (respectively, \( 0 \leq N_{ij} < \infty \)). Buffer \( B_{M_i} \) (respectively, \( B_{0M_0} \)) is referred to as the finished goods buffer. All other buffers are called the in-process buffers.

**Starvation Rule:**

v) If buffer \( B_i, i = 1, \ldots, M - 1 \) (respectively, \( B_{ij}, i = 0, \ldots, S, j = 1, \ldots, M_i - 1 \)) is empty at the beginning of a time slot and machine \( m_{i+1} \), \( i = 1, \ldots, M - 1 \) (respectively, \( m_{ij+1}, i = 0, \ldots, S, j = 1, \ldots, M_i - 1 \)) is starved during this time slot. The first machine \( m_1 \) (respectively, \( m_{i+1}, i = 1, \ldots, S \)) is never starved. In the assembly line, machine \( m_{M_i} \) is starved during the time slot, if at least one of the buffers \( B_{3M_i} \), \( i = 1, \ldots, S \), is empty at the beginning of the time slot.

**Blockage Rule:**

vi) If buffer \( B_i, i = 1, \ldots, M - 1 \) (respectively, \( B_{ij}, i = 0, \ldots, S, j = 1, \ldots, M_i - 1 \)) is full at the beginning of a time slot and machine \( m_{i+1} \), \( i = 1, \ldots, M - 1 \) (respectively, \( m_{ij+1}, i = 0, \ldots, S, j = 1, \ldots, M_i - 1 \)) does not take a part from \( B_i \) at the beginning of this slot, then \( m_i \), \( i = 1, \ldots, M - 1 \) (respectively, \( m_{ij}, i = 0, \ldots, S, j = 1, \ldots, M_i - 1 \)) is blocked during this time slot. Machine \( m_{M_i} \) (respectively, \( m_{M_i} \)) is blocked if buffer \( B_{M_i} \) (respectively, \( B_{0M_0} \)) is full. In the assembly line, machine \( m_{i+1} \), \( i = 1, \ldots, S \), is full and machine \( m_{0M} \) does not take a part from \( B_{3M_i} \) at the beginning of this slot.

**Demand:**

vii) From the point of view of the demand, the time axis is divided into “epochs,” each consisting of \( T \) time slots.

viii) At the end of each epoch, a shipment of \( D \) parts has to be available for the customer. If \( PR \) is the production rate (PR) of the system, then

\[
D \leq T \cdot PR.
\]
Remark 4: A method for calculating the production rate in the system defined by i)–vi) but without finished goods buffer has been developed in [22]–[24]. Thus, the upper bound of $D$ is readily available.

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 $m_M$ (respectively, $m_M0$) 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, $m_M$ (respectively, $m_M0$), is blocked, whichever 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 5: Assumption of no backlog is introduced to simplify the analysis. However, since backlog is typically satisfied by overtime, this assumption does not substantially reduce the generality.

Assumptions i)–ix) define the production lines under consideration. In the time scale of the epoch and in an appropriately defined state space, 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 the last machine $m_0$ in serial line or $m_M0$ in assembly line in epoch $i$ during the normal system operation. Then DTP can be expressed as

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

where, as before, $H(i-1)$ is the number of parts in the FGB at the end of $(i-1)$th epoch and $D$ is the demand.

The main problem addressed in this paper is as follows: Given production system i)–ix), develop a method for evaluating DTP as a function of system’s parameters.

A solution to this problem is given in Section III.

C. Available Results

As it was pointed out above, a method for DTP calculation in a single-machine system with FGB was developed in [10] and [11]. Since this result is used for DTP calculations in multimachine systems, we briefly outlined it.

Consider a one-machine production system with the FGB of capacity $N$ (Fig. 3). Following [11], introduce the notations:

- $\bar{t}_i$ = number of parts produced during epoch $i$ if no blockage occurs,
- $z_k = \Pr(H(i-1) = k)$, $k = 0, 1, \ldots, N$,
- $\tau_{k,j} = \Pr(\bar{t}_i = D+k-j)$, $k = 1, 2, \ldots, N-1$,
- $j = 0, 1, \ldots, N$,
- $\tau_{N,j} = \Pr(\bar{t}_i \geq D+N-j)$, $j = 0, 1, \ldots, N$.

Probabilities $\tau_{k,j}$ and $\tau_{N,j}$ can be easily calculated since they refer to the system without the FGB. Specifically

$$
\tau_{k,j} = \left(\frac{T}{D+k-j}\right)^{p^{D+k-j}}(1-p)^{T-(D+k-j)}
$$

$$
\tau_{N,j} = \sum_{k=D+N-j}^{T} \binom{T}{k} p^k(1-p)^{T-k}.
$$

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

**Theorem 1[11]:** Let $Z = [z_1, \ldots, z_N]^T$ be a vector defined by

$$
Z = -R^{-1}Z_0
$$

where matrix $R$ and vector $Z_0$ are shown as (7) and (8) at the bottom of the next page. Then, under assumptions i)–ix), the DTP of the system with $M = 1$, is given by the formula

$$
DTP = \sum_{k=0}^{N} \sum_{j=D-k}^{T} z_k(T-j) p^j(1-p)^{T-j}.
$$

II. APPROACH TO DTP CALCULATIONS

The main difficulty in developing a method for calculating DTP is that the behavior of a production 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, $m_M$ (in serial line) or $m_M0$ (in assembly line), is blocked during a time slot $j$, $j = 1, \ldots, T$, clearly depends on $j$ and is a monotonically increasing function of $j$. This makes exact calculations of DTP intractable. Therefore, a simplification is necessary. The simplification used in this work is based on two steps: stationarization and subsequent recursive iterations. These steps are next described.

A. Stationarization

Consider a one-machine line, shown in Fig. 3. Assume that the probability that the machine $m$ is up and FGB is full during a time slot $j$, $j = 1, \ldots, T$, is known. Denoted this probability as $P_k(j)$. Then the probability $P_k$ that $m$ is up and FGB is full during an epoch exactly $k$ time slots $k = 1, \ldots, T$ can be expressed as

$$
P_k = \begin{cases} P_k(T-k+1) - P_k(T-k) & 1 \leq k \leq T-1 \\ 0 & k = T \end{cases}
$$

Thus, the expected value of $k$ is

$$
\bar{k}_b = \sum_{k_b=0}^{T} k_b P_{k_b} = \sum_{k_b=1}^{T-1} k_b P_{k_b},
$$
Using this expression, the approximate probability that \( m \) is up and FGB is full during any slot \( j \) can be introduced as follows:

\[
P_b = \frac{\bar{T}}{T}. \tag{12}
\]

Then the approximate probability that FGB is full during any time slot \( j \), is

\[
P_f = \frac{\bar{P}}{P} = \frac{\bar{T}}{T}P. \tag{13}
\]

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

To evaluate \( \hat{P}_b \) and \( \hat{P}_f \), the distribution \( P_b(j) \) is necessary. It can be calculated as follows. Introduce the probabilities:

\[
\begin{align*}
&h(i, j) = \text{Prob}\{i \text{ parts in FGB at beginning of slot } j \text{ in an epoch}\} \\
&h_0(0, j, k) = \text{Prob}\{0 \text{ parts in FGB at beginning of slot } j \text{ and } k \text{ parts have been shipped in an epoch}\} \\
&q = 1 - p,
\end{align*}
\]

Under assumptions i)–ix) and \( M = 1 \), function \( h(i, j) \) is defined recursively by

\[
h(i, j) = \begin{cases} 
 h(1, j-1)q + h_0(0, j-1, k)p, & \text{for } i = 1 \\
 h(i, j-1)q + h(i-1, j-1)p, & \text{for } 1 < i < N \\
 h(N, j-1), & \text{for } i = N \\
 h(N-1, j-1)p, & \text{for } i = N
\end{cases} \tag{14}
\]

where

\[
h_0(0, j, k) = \begin{cases} 
 z_0, & \text{for } k = 0 \\
 h_0(0, j-1, k)q + h_0(0, j-1, k-1)p, & \text{for } k > 0
\end{cases} \tag{15}
\]

with initial conditions

\[
h(i, 1) = \begin{cases} 
 z_{D+i}, & \text{for } i > 0, \quad N > D \\
 0, & \text{for } i > 0, \quad N \leq D \\
 1, & \text{for } i = 0, \quad N \leq D
\end{cases} \tag{16}
\]

and

\[
h_0(0, 1, k) = \begin{cases} 
 z_k, & \text{for } 0 \leq k \leq D, \quad N > D \\
 z_k, & \text{for } 0 \leq k \leq N, \quad N \leq D \\
 0, & \text{for } N+1 \leq k \leq D, N < D
\end{cases} \tag{17}
\]

and \( z_k, k = 0, 1, \ldots, N, \) is calculated according to (6). Therefore, the probability that \( m \) is up and FGB is full during slot \( j \) is

\[
P_b(j) = h(N, j)p. \tag{18}
\]

Thus, \( \hat{P}_b \) can be calculated as follows:

\[
\hat{P}_b = \frac{1}{T} \sum_{k=1}^{T-1} k[P_b(T-k+1) - P_b(T-k)]
\]

\[
= \frac{p}{T} \sum_{i=1}^{T-1} [h(N, T-k+1) - h(N, T-k)]k
\]

\[
= \frac{p}{T} \sum_{j=2}^{T} [h(N, j) - h(N, j-1)](T-j+1). \tag{19}
\]

Analogously,

\[
\hat{P}_f = \frac{1}{T} \sum_{j=2}^{T} [h(N, j) - h(N, j-1)](T-j+1). \tag{20}
\]
To summarize, it is convenient to introduce an operator of stationarization $\Phi_4$ defined as

$$
\hat{P}_f = \Phi_4(p, N, T, D)
$$

where

$$
\hat{P}_f := \frac{1}{T} \sum_{j=2}^{T} [h(N, j) - h(N, j - 1)](T - j + 1)
$$

Expression (21) is used in Section III for DTP calculation.

B. Iterations

Consider the serial and assembly lines shown in Figs. 1 and 2, respectively. To adopt the above stationarization procedure, assume that the probability that the last machine $m_M$ (for the serial line) or $m_{OM_0}$ (for the assembly line) is not starved is known. Denote an estimate of this probability as $\hat{P}_{ns}$. Modify $p_M$ (respectively, $p_{OM_0}$) by multiplying it by $\hat{P}_{ns}$ and denote the modified machine as $m'_M$ (respectively, $m'_{OM_0}$). Using this machine and procedures (14)–(17), calculate the stationarized probability, $\hat{P}_f$. Introduce now another fictitious machine, $m''_M$ (respectively, $m''_{OM_0}$), defined by $p_M(1 - \hat{P}_f)$ (respectively, $p_{OM_0}(1 - \hat{P}_f)$). Using this machine and deleting FGB, calculate the new $\hat{P}_{ns}$ by employing the recursive procedure for performance analysis of serial lines developed in [22] (respectively, in [23] and [24] for assembly lines). Having this probability, repeat the process described earlier anew.

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

$$
\hat{P}_f(n + 1) = \Phi_4(\hat{P}_{ns}(n), N_M, T, D)
$$

Expression (21) is used in Section III for DTP calculation.

In Section III, operators $\Phi_2$ and $\Phi_3$ are described, convergence of the iterations is considered, and the accuracy of DTP calculations is investigated.

III. DTP CALCULATIONS

A. Operator $\Phi_2$

Operator $\Phi_2$ can be defined through the aggregation procedure for performance analysis of serial lines developed in [22]. Indeed, assume that $\hat{P}_{ns}(n + 1)$, calculated according to (24), is known. Consider the serial line consisting of $M$ machines with parameters $p_1, \ldots, p_{M-1}, p_M(n + 1)$ and $M - 1$ in-process buffers $N_0, \ldots, N_{M-1}$. According to [22], its performance can be analyzed using the following convergent recursive procedure (see [22] for details):

$$
\hat{p}_M(s + 1) = p_M\left[1 - Q\left(p_M^{h+1}(s + 1), p_M^{h}(s), N_i\right)\right],
$$

$$
1 \leq i \leq M - 1
$$

$$
\hat{p}_M(s + 1) = p_M\left[1 - Q\left(p_M^{h-1}(s + 1), p_M^{h}(s + 1), N_{i-1}\right)\right],
$$

$$
2 \leq i \leq M
$$

with boundary conditions

$$
\hat{p}_1(s) = p_1, \quad \hat{p}_M(s) = p_M(n + 1), \quad s = 1, 2, 3, \ldots,
$$

and initial conditions

$$
\hat{p}_i(0) = p_i, \quad i = 1, \ldots, M
$$

In terms of the steady states of this procedure, i.e.,

$$
\lim_{n \to \infty} \hat{p}_{ns}(n) := \hat{P}_{ns}
$$

If procedure (22)–(25) [respectively, (26)–(29)] is convergent and the following limit exists

$$
\lim_{s \to \infty} \hat{p}_i(s) := \hat{p}_i
$$

the DTP of a $M$-machine serial line [respectively, $(M_1 + \cdots + M_S + M_0)$-machine assembly line] can be evaluated, using (6)–(9), as the DTP of a one-machine system defined by $p_M\hat{P}_{ns}$ (respectively, $p_{OM_0}\hat{P}_{ns}$) and FGB of capacity $N_M$ (respectively, $N_{OM_0}$), i.e.,

$$
\hat{DTP}_{serial} = \hat{DTP}_1(p_M\hat{P}_{ns}, N_M, D, T)
$$

$$
\left(\text{respectively, } \hat{DTP}_{assembly} = \hat{DTP}_1(p_{OM_0}\hat{P}_{ns}, N_{OM_0}, D, T)\right).
$$

In Section III, operators $\Phi_2$ and $\Phi_3$ are described, convergence of the iterations is considered, and the accuracy of DTP calculations is investigated.
the probability that buffer $B_{N-1}$ is not empty can be evaluated as
\[ \hat{p}_{ns}(n+1) = \frac{p_M(n+1)}{P_M}. \] (36)

Thus, (32)–(36) represent the mapping from $p_M(n+1)$ to $\hat{p}_{ns}(n+1)$, denoted in (25) as $\Phi_2$.

**B. Operator $\Phi_3$**

In order to be consistent with the application described in Section V and to simplify notations, we consider below assembly line with a single machine in all component lines and a single machine in the final assembly line (Fig. 4). Note that in this case the second subscript becomes superfluous and, therefore, is omitted from Fig. 4 and the following discussion.

For the system shown in Fig. 4, operator $\Phi_3$ of (29) can be defined through the recursive procedure introduced in [23], [24]. Specifically, assume $p_M(n+1)$ is known and consider the assembly line shown in Fig. 4 with no FGB and machine $m_0$ defined by the parameter $\phi_1$. Then, as it follows from [23], [24], estimate of the probability that buffers $B_i$, $i = 1, \ldots, S$, is empty can be calculated using the following recursive procedure [23]:
\[ \hat{X}_i(0, s + 1) = Q \left( p_i, p_M(n+1) \prod_{j=1}^{i-1} [1 - \hat{X}_j(0, s+1)] \right. \]
\[ \left. \cdot \prod_{j=i+1}^S [1 - \hat{X}_j(0, s)], N_i \right), \]
\[ i = 1, \ldots, S, \quad s = 0, 1, 2, \ldots \] (37)

with initial conditions
\[ \hat{X}_i(0, 0) = 0, \quad i = 1, \ldots, S \]
and $Q(x, y, N)$ defined in (33).

In terms of the steady state of this procedure, i.e., in terms of
\[ \lim_{s \to \infty} \hat{X}_i(0, s) = \hat{X}_i(0), \quad i = 1, \ldots, S \] (38)
the probability that the assembly machine is not starved is evaluated as
\[ \hat{p}_{ns}(n+1) = \prod_{i=1}^S [1 - \hat{X}_i(0)]. \] (39)

Thus, (37)–(39) represent the mapping from $p_M(n+1)$ to $\hat{p}_{ns}(n+1)$, denoted as $\Phi_3$ in (29).

**C. Convergence**

To prove the convergence of recursive procedures (22)–(25), (32)–(36) and (26)–(29), (37)–(39), it is important to know whether the probability $\hat{P}_f$ defined by (14)–(21), is monotonic with respect to $p$. Intuitively, this property clearly takes place. However, a rigorous proof of this fact seems to be intractable. Therefore, although in every example analyzed numerically this function was found to be monotonically increasing, we introduce this property as a hypothesis.

**Hypothesis 1:** For system i)–ix) with $M = 1$, the stationarized probability that the FGB is full, $\hat{P}_f$, is a monotonically increasing function of $p$.

An illustration of this property for two typical systems is given in Fig. 5. Based on this hypothesis, we formulate the following.

**Proposition 1:** Under Hypothesis 1, iteration procedures (22)–(25), (32)–(36) and (26)–(29), (37)–(39) are convergent, i.e., the following limits exist
\[ \lim_{n \to \infty} \hat{P}_f(n) := \hat{P}_f \]
\[ \lim_{n \to \infty} \hat{p}_{ns}(n) := \hat{p}_{ns}. \] (40)

**Proof:** See the Appendix.

Thus, the $\hat{DTP}_{\text{serial}}$ and $\hat{DTP}_{\text{assembly}}$ can be calculated using (30) and (31) with $\hat{p}_{ns}$ defined in (40).

**D. Accuracy**

The accuracy of $\hat{DTP}_{\text{serial}}$ and $\hat{DTP}_{\text{assembly}}$ have been evaluated using discrete event simulation. We simulated dozens of systems defined by i)–ix) with various machine, buffer, demand, and epoch parameters. Tables I and II present 20 serial and 20 assembly lines, respectively. In each run of the corresponding discrete event model, zero initial conditions for all buffers have been assumed and 10,000 time slots of warm up period has been carried out ($T$, as before, is the length of the epoch). The next 100,000 time slots of stationary regime have been used to statistically evaluate the DTP. In Tables I and II, DTP denotes the DTP obtained by simulation, whereas $\hat{DTP}$ denotes the estimates calculated according to (22)–(25), (30), (32)–(36), and (26)–(29), (31), (37)–(39), respectively. As it can be seen from Tables I and II, the estimates provide relatively high precision, with errors ranging from 0.23% to 3.38% for serial lines and 0.49% to 4.83% for assembly lines.

**Remark 6:** Calculation of DTP using expressions (22)–(25), (30), (32)–(36), or (26)–(29), (31), (37)–(39), is many orders of magnitude faster than discrete event simulations. For example,
TABLE I
ACCURACY OF DTP ESTIMATION IN SERIAL LINES
(%err = |DTP - \bar{DTP}|/DTP 100%)

<table>
<thead>
<tr>
<th>(p_i)</th>
<th>(N_i)</th>
<th>(\bar{DTP})</th>
<th>(\bar{DTP})</th>
<th>(%err)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.80 0.80</td>
<td>1 1</td>
<td>2</td>
<td>0.8799</td>
<td>0.8717</td>
</tr>
<tr>
<td>0.70 0.85</td>
<td>1 3</td>
<td>2</td>
<td>0.9109</td>
<td>0.9190</td>
</tr>
<tr>
<td>0.85 0.80</td>
<td>2 1</td>
<td>4</td>
<td>3</td>
<td>0.8894</td>
</tr>
<tr>
<td>0.75 0.85</td>
<td>2 3</td>
<td>10</td>
<td>7</td>
<td>0.9330</td>
</tr>
<tr>
<td>0.90 0.65</td>
<td>2 3</td>
<td>10</td>
<td>6</td>
<td>0.9435</td>
</tr>
<tr>
<td>0.55 0.65</td>
<td>3 3</td>
<td>10</td>
<td>5</td>
<td>0.9338</td>
</tr>
<tr>
<td>0.86 0.86 0.86</td>
<td>1 1 1</td>
<td>3</td>
<td>2</td>
<td>0.9142</td>
</tr>
<tr>
<td>0.85 0.85 0.85</td>
<td>2 2 2</td>
<td>4</td>
<td>3</td>
<td>0.9402</td>
</tr>
<tr>
<td>0.76 0.76 0.76</td>
<td>3 3 3</td>
<td>5</td>
<td>3</td>
<td>0.9305</td>
</tr>
<tr>
<td>0.68 0.82 0.76</td>
<td>1 2 1</td>
<td>3</td>
<td>2</td>
<td>0.9072</td>
</tr>
<tr>
<td>0.75 0.90 0.83</td>
<td>1 2 1</td>
<td>4</td>
<td>3</td>
<td>0.7198</td>
</tr>
<tr>
<td>0.81 0.86 0.89</td>
<td>2 1 1</td>
<td>4</td>
<td>3</td>
<td>0.9234</td>
</tr>
<tr>
<td>0.87 0.83 0.87</td>
<td>1 1 2</td>
<td>3</td>
<td>2</td>
<td>0.9061</td>
</tr>
<tr>
<td>0.80 0.85 0.90 0.80</td>
<td>3 2 3 1</td>
<td>4</td>
<td>3</td>
<td>0.9006</td>
</tr>
<tr>
<td>0.85 0.70 0.75 0.90</td>
<td>3 2 1 2</td>
<td>5</td>
<td>3</td>
<td>0.9252</td>
</tr>
<tr>
<td>0.84 0.85 0.85 0.84</td>
<td>1 1 1 2</td>
<td>5</td>
<td>3</td>
<td>0.9224</td>
</tr>
<tr>
<td>0.86 0.87 0.87 0.86</td>
<td>2 2 2 2 1</td>
<td>4</td>
<td>3</td>
<td>0.9232</td>
</tr>
<tr>
<td>0.80 0.85 0.90 0.75 0.85</td>
<td>2 1 3 2 1</td>
<td>3</td>
<td>2</td>
<td>0.9409</td>
</tr>
<tr>
<td>0.75 0.90 0.80 0.90</td>
<td>2 2 2 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.90 0.70 0.75 0.80</td>
<td>2 2 2 2</td>
<td>5</td>
<td>3</td>
<td>0.9363</td>
</tr>
<tr>
<td>0.65 0.90 0.80 0.90</td>
<td>2 2 2 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.90 0.70 0.75 0.80</td>
<td>2 2 2 2</td>
<td>5</td>
<td>3</td>
<td>0.9190</td>
</tr>
</tbody>
</table>

discrete event simulation of every system included in Tables I and II takes on the average about 40 minutes using HP C360 workstation, whereas calculation according to (22)–(25), (30), (32)–(36), or (26)–(29), (31), (37)–(39), takes less than a second.

**Remark 7:** Note that the accuracy of DTP evaluation, illustrated in Tables I and II, is comparable to that of production rate evaluation in production lines without FGBs (see [1]–[6], [22]–[24]).

Fig. 6. Moderate load.

IV. STRUCTURAL PROPERTIES

A. Load Factor

Using the method developed above, this section illustrates several structural properties of DTP in 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, \(D\), vs. the production capacity of the system, \(T \cdot PR\), where, as before, \(PR\) is the production rate of the concerning system without the FGB. We formalize this relationship by the load factor, \(L\), defined as

\[
L = \frac{D}{T \cdot PR} \tag{41}
\]

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 system is underutilized.

B. FGB Capacity for High DTP

Consider a serial line with five machines and four in-process buffers, each with capacity 3. Assume that parameters \(p_i\) of the machines are constrained by \(\sum_{i=1}^{5} p_i = p^H\) and consider four types of \(p_i\) allocations

- uniform: \(p_i = p_i, \forall i, i = 1, \ldots, 5\);
- inverted bowl: \(p_4 > p_2 = p_1 > p_3 = p_5\);
- ramp: \(p_1 < p_2 < p_3 < p_4 < p_5\);
- inverted ramp: \(p_4 > p_2 > p_3 > p_4 > p_5\).

Specific values of \(p_i\) for each of these allocations are shown in Figs. 6 and 7 with \(p^H = 4.5\) and 4.2, respectively, along with the corresponding production rate and load factor of each line. Note that in Fig. 6 the load is moderate whereas in Fig. 7 it is heavy.
Within this scenario, using (22)–(25), (30), (32)–(36), we calculate \( D_{\text{TP}} \) for all lines involved. The results are shown in Figs. 6 and 7. Examining these data, we arrive at the following observations.

1) For moderate load and for all \( p_i \) allocations, the FGB capacity of about 1–1.5 shipments is sufficient to provide high DTP.

2) For heavy load, the FGB of capacity 4–5 shipments is necessary to provide high DTP for all \( p_i \) allocations.

3) For both heavy and moderate loads, the best \( p_i \)'s allocation is the ramp; the worst is the inverted ramp. This implies that DTP, unlike PR, does not obey the reversibility property [25] and is not maximized by the inverted bowl allocation [26].

Remark 8: Although the above observations are based on the data obtained via approximate calculations (22)–(25), (30), (32)–(36), discrete event simulation support these conclusions.

C. Monotonicity Properties

Intuitively, it is clear that DTP is monotonically decreasing as a function of the load factor, \( L \), and monotonically increasing as a function of \( p_i \) and \( N_i \), \( i = 1, \ldots, M \). Although no analytical proofs of these facts are available, all systems analyzed using (22)–(25), (30), (32)–(36), exhibited these properties. Examples are shown in Figs. 8 and 9. (In Fig. 8, values of \( L > 1 \) are included for the sake of mathematical completeness.)

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 Figs. 10 and 11. (In Fig. 10, the value of \( D/T \) is kept constant, while \( T \) is changing; in Fig. 11, the value of \( PR \) is fixed.) The lack of monotonicity is especially pronounced for high load factors and practically disappears for relatively small \( L \). Discrete event simulations also support this conclusion.

The counterintuitive behavior of Figs. 10 and 11 may be explained as follows: For smaller \( T \) (which implies smaller \( D \) and larger \( N_M/D \)) the probability to satisfy the demand by the contents of FGB is larger than that for larger \( T \). Only when \( T \) be-
comes 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.

Remark 9: Note that the lack of monotonicity of the variance of the number of parts produced during a fixed interval of time was also found in [9], [16], [19], and [27].

V. CASE STUDY: DTP ANALYSIS OF INJECTION MOLDING—ASSEMBLY SYSTEM IN AUTOMOTIVE COMPONENT PLANT

A. System Description and Physical Model

The layout of the injection molding–assembly production system under consideration is shown in Fig. 12. It consists of 13 presses, which mold seven part-types, and 3 identical assembly lines. The storage between the injection molding and the assembly serves as the in-process buffer; the storage after the assembly serves as a finished goods buffer.

Each mold can be placed on any of the 13 presses, depending on the availability and schedule. The assembly uses each of the seven part-types to assemble the final product. Thus, the physical model of the system can be represented as shown in Fig. 13.

The goal of this project was to select the capacity of the FGB so that DTP \( \geq 0.99 \), based on the shipping schedule defined by \( T = 12, D = 9 \). Before this study, no strict limits on FGB was formally imposed, and the system was managed on the basis of the intuitive principle “the more finished goods the better.”

B. Virtual Model

Direct analysis of the model in Fig. 13 is quite complicated. Therefore, a simplification is necessary. The simplification used in this work can be described as follows:

From the point of view of the in-process storage, injection molding part of the system consists of seven virtual machines, each producing a single part-type. The parts are placed in the storage and then drawn by the assembly, one part of each part-type to assemble the product. This simplified model, referred to as the virtual model, is shown in Fig. 14.

Based on a two-month performance data, the parameters of the virtual model have been identified as shown in Table III. The subsequent analysis is carried out in terms of this model.

C. Design of FGB Capacity

Using the recursive procedure (26)–(29), (31), (37)–(39) and the parameters of Table III, we calculate DTP of the virtual model for various values of the capacity of FGB, \( N_0 \). The results are shown in Table IV. As it can be seen, for \( N_0 = 9 \), DTP > 0.9997 and subsequent increase of \( N_0 \) results in a very small improvement of DTP. Therefore, the capacity of \( N_0 \) has been recommended as one shipment size (i.e., \( N_0 = 9 \)). This recommendation was accepted by the plant management.

Remark 10: The load factor, \( L \), of the assembly system, defined in Fig. 14 and Table III, is 0.95. Therefore, a relatively small FGB capacity of one shipment size, obtained above, is in agreement with the results for the serial line, discussed in Section IV.

VI. CONCLUSION

This paper provides a recursive method for calculating DTP in production systems with FGBs. Based on a hypothesis, convergence of the recursive procedure is justified analytically. The accuracy of the method is evaluated numerically. It is shown that the accuracy is comparable to that of production rate calculation in production systems without FGBs.

The method developed can be used for design of production systems from the point of view of the customer demand satisfaction in supply chain management. Such a utilization is illustrated by a case study of an injection molding–assembly line at an automotive component plant.
**APPENDIX**

**Proof of Proposition 1:** By induction: In the case of serial line, since for $n = 0$, $\hat{p}_{ns}(0) = 1$, then

$$\hat{p}_M(1) = p_M$$
$$\hat{p}_j(1) = \Phi_1(p_M, N_M, T, D)$$
$$\hat{p}_M'(1) = p_M[1 - \hat{p}_j(1)].$$

Thus

$$\hat{p}_{ns}(1) = \Phi_2(p_1, \ldots, p_{M-1}, p_M[1 - \hat{p}_j(1)])$$
$$N_1, \ldots, N_{M-1} < 1 = \hat{p}_{ns}(0)$$
$$\hat{p}_M'(2) = p_M\hat{p}_{ns}(1) < \hat{p}_M(1)$$

and, from Hypothesis 1, we have

$$\hat{p}_j(2) < \hat{p}_j(1).$$

Assume now that for $n > 0$,

$$\hat{p}_{ns}(n) < \hat{p}_{ns}(n - 1).$$

Then

$$\hat{p}_j(n + 1) < \hat{p}_j(n)$$
$$\hat{p}_M'(n + 1) > \hat{p}_M'(n).$$

Due to monotonicity of operator $\Phi_2$ (see [22]), we have

$$\hat{p}_{ns}(n + 1) < \hat{p}_{ns}(n)$$

which implies that

$$\hat{p}_j(n + 2) < \hat{p}_j(n + 1).$$

Therefore, the sequences $\hat{p}_{ns}(n)$ and $\hat{p}_j(n)$ are monotonically decreasing. From [22] and (20), it follows that $\hat{p}_{ns}(n)$ and
\( \hat{P}(n) \) are bounded from above and below. Therefore, they are convergent.

The proof for the assembly line is similar to the above, taking into account that

\[
\hat{P}_{\text{hs}} = \prod_{i=1}^{\delta} \left[ 1 - \hat{X}_i(0) \right]
\]

and that \( \hat{P}_{\text{hs}} \) is monotonically decreasing with respect to \( \hat{P}_{\text{hs}} \) (see [23]). Proposition 1 is proved.

**Acknowledgment**

Helpful comments of the anonymous reviewers are gratefully acknowledged.

**References**


