Transient Analysis of Serial Production Lines With Perishable Products: Bernoulli Reliability Model

Feng Ju, *Member, IEEE*, Jingshan Li, *Senior Member, IEEE*, and John A. Horst, *Member, IEEE*

***Abstract*—Manufacturing systems with perishable products are widely observed in practice (e.g., food industry, biochemical productions, battery and semiconductor manufacturing). In such systems, the quality of the product is highly affected by its exposure time while waiting for the next operation, i.e., the residence time of intermediate parts within the system. Such a time should be strictly limited in order to ensure the product usability. The parts that reach the maximum allowable residence time need to be scrapped, thus impeding the production. To achieve an efficient production, the time-dependent or transient analysis is important to uncover the underlying principles governing production operations. In this paper, a serial production line model with two Bernoulli reliability machines, a finite buffer and perishable products is presented to analyze the transient behavior of such systems. The analytical formulas are derived to evaluate transient performance, and structural properties are investigated to study the effect of system parameters. In addition, using the model, we address problems of settling time estimation and production control to demonstrate the importance of the proposed method for transient analysis.**

***Index Terms*—Bernoulli machine, perishable part, production control, residence time, settling time, transient analysis.**

I. Introduction

PERISHABILITY refers to the fact that the products have a fixed maximum allowable waiting time, exceeding which the item will lose its utility completely and will be scrapped. Perishability has great impact on every section of a product's life cycle. During a production process, parts may be forced to wait in the buffer due to demand uncertainty, technical failures, and other types of production disruptions, which may lead to excess waiting and scrap. In this paper, we focus on the issue of perishability during production processes and analyze the behavior of production systems subject to perishability constraint.

Perishability is vastly seen in many production systems, ranging from process industries to mass production of discrete products. For example in yogurt production, finished products as well as their intermediate products are highly perishable, thus imposing strict time requirements for the products' flow time in the system. The yogurt can only stay in the buffer within a fixed amount of time, otherwise it will be scrapped. Such an issue is encountered in many other food industries (e.g., pizza, juice, dairy products, etc.), where certain processes must be controlled in a timely manner and production disruptions may cause scrapping and monetary loss.

In battery production, the manufacturing process starts from powder formulation by mixing raw materials. The powders are pressed into tubes and shaped to the designed size through firing, cutting, and grinding operations. Then, chemical materials are filled into cells to form electrodes during the cell assembly process. The cells need to be baked to reach the designed temperature range after adding each material, and the next material must be added within a certain time limit. Otherwise, those cells will fail the inspection and have to be scrapped. Similar cases can be found in stamping, bio-fuel, and chemical production.

As one could see, in addition to the traditional difficulties in analyzing production systems, such as unreliable machines, finite buffers, blockage, and starvation, the unique features introduced above draw more complexities and challenges to production system research. Due to dynamic changes and disruptions in the manufacturing process, machine behaviors, product perishability, and customer demand, the production system operates partially or entirely in the transient regime. The time dynamics and transient behavior have significant manufacturing implications in such systems. Therefore, the traditional steady-state analysis is not applicable in transient settings. Timely decision-making and real-time production control are needed to better operate the production system in order to obtain superior and sustainable performance. The rapid development of information technology enables us to collect more and more real-time production data on the factory floor, which provides the opportunity to implement rigorous methods for analyzing the transient behavior of production systems with perishable products. This relies on full understanding of the system dynamics both in the long run and during an arbitrary time period. Therefore, transient analysis in perishable products manufacturing is critically needed.

Manuscript received July 16, 2015; revised December 28, 2015 and January 6, 2016; accepted April 22, 2016. Date of publication May 24, 2016; date of current version January 26, 2017. This work was supported in part by the National Sceince Foundation under Grant CMMI-1063656 and in part by the National Institute of Standards and Technology under Grant 70NANB14H260. Recommended by Associate Editor L. H. Lee.

F. Ju is with the School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ 85287 USA (e-mail: fengju@asu.edu).

J. Li is with the Department of Industrial and Systems Engineering, University of Wisconsin-Madison, Madison, WI 53706 USA (e-mail: jingshan@engr.wisc.edu).

J. A. Horst is with the Engineering Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899-8230 USA (e-mail: john.horst@nist.gov).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2016.2572119

Although production systems have been studied extensively during the last five decades, most of the studies emphasize steady state analysis (see monographs [1]–[3]). The basic idea is to look at the behavior of production systems in the long run considering the inherent system uncertainty. However, such analyses ignore the substantial amount of production loss due to transients if the steady state is reached after a relatively long period of time. The problem could be even worse when the intermediate products are subject to quality deterioration and perishability, making the production system function at a highly dynamic regime with more losses. Unfortunately, the transient analysis of production system remains largely unexplored in the current literature, and among the limited work addressing transient behavior, the perishability issue is not investigated.

Therefore, in this paper, we propose a novel method to analyze the transient behaviors of a two-machine Bernoulli production line accounting for scrapping due to perishability in the production process. Based on that, we address two important problems, settling time estimation and production control, to demonstrate the importance of the proposed method for transient analysis.

The rest of this paper is organized as follows: Section II reviews the related literature. Section III introduces assumptions and formulates the problem. Based on the modeling approach presented in Section IV, performance measures are evaluated and numerical experiments are conducted to justify their accuracy in Section V. Section VI is devoted to investigating structural properties of the system. Using the proposed transient analysis, settling time estimation and production control strategies are presented in Sections VII and VIII, respectively. Finally, Section IX summarizes the work and casts directions on future research. All the proofs are provided in the Appendix.

II. Literature Review

In production literature, both issues of perishability and quality deterioration have been studied. The quality deterioration is regarded as a continuous process of quality decay such that products gradually become unacceptable while being stored. So the lifetime of products is typically random. In contrast, perishability implies that the maximum allowable exposure time is a fixed number. A product processed within that period will be considered as good as a new one without any quality depreciation or utility loss, while the part exceeding the period will be scrapped. This is also referred to *memoryless perishability* [4].

In production-inventory systems, both issues have been studied in the lot-sizing and inventory control problems (see reviews [5]–[7]). Constraints on the lifetime of products force organizations to thoughtfully plan production and manage inventory incorporating both the inbound material flow and the outbound supply chain ([8], [9]). The general goal is to optimally control the production-inventory system subject to the constraint of perishability and demand ([10], [11]). For example, paper [12] applies the linear quadratic regulator technique to determine the optimal production policy that minimizes the cost associated with inventory and production rate. A three-dimensional matching problem is formulated in [13] to find an optimal assignment of the number of products in launching and completion, subject to perishability of products and deterministic launching and completion time constraints. Paper [14] brings the perspectives of the manufacturer and the retailer together to analyze the integrated production and inventory control decision in the context of perishability, inflation and multiple deliveries. In [15], the discrete lot-sizing and scheduling models are utilized to emphasize the importance of including deterioration and perishability in production planning.

All the above research in production-inventory systems typically ignores the disturbance within production systems. Indeed, due to unanticipated production disruption, the frequent and long waiting times in the intermediate buffers not only increase lead times, but also, more importantly, decrease the quality of products and lead to depreciation or scrapping. Paper [16] evaluates the production performance of a two-stage automatic transfer line with the part being scrapped with a certain probability when the processing machine fails. Such a study is extended to multiple-stage lines in [17]. In [4], issues of both deterioration and perishability have been addressed in a transfer line without intermediate storage space. Recently, paper [18] studies the issue of deterioration due to residence time and develops an aggregation-based procedure to approximate performance measures of general serial production lines with deteriorating products.

The above-mentioned studies are carried out in steady state environments. In production systems research, only a few studies have addressed the transient behavior. Paper [19] discusses the method to calculate the distribution of time to absorption in Markov models in manufacturing systems with deadlocks and failures, and shows the relevance of transient analysis to a multiclass manufacturing system with significant setup times. Paper [20] investigates the transient behavior of two-machine Bernoulli lines. It shows that the transients of production rate and work-in-process are characterized by the second largest eigenvalue (SLE) of the transition matrix of the Markov chain and the pre-exponential factor. The settling time and production loss due to transients are also addressed. Paper [21] extends the work to geometric machine lines, and an aggregation method is introduced to study transients in long Bernoulli lines in [22]. However, all of them ignore the issue of perishability.

In spite of these efforts, an analytical model of transient behavior of production systems subject to perishability has not been studied explicitly. The objective of this paper is intended to contribute to this end.

III. System Description

Consider a two-machine production line shown in Fig. 1. Materials or raw parts flow into machine m_1 and wait in buffer B_1 to be processed by machine m_2. Parts whose waiting time in the buffer exceeds a certain threshold need to be scrapped. The following assumptions define the machines, the buffer, and their interactions.

i) The production system consists of two machines (m_1 and m_2) and a buffer B_1 separating them.
ii) Both machines have constant and identical processing time. Such a time is slotted with cycle time.

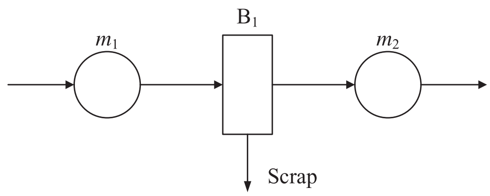


Fig. 1. Bernoulli line with perishable products.

iii) The machines follow the Bernoulli reliability model. In each cycle, machine m_i, $i = 1, 2$, is capable of producing a part with probability p_i and fails to do so with probability $1 - p_i$. The machine status changes at the beginning of each cycle.

Remark 1: The Bernoulli model of machine reliability leads to a faithful model of production systems, where the downtime is relatively short and comparable with the cycle time. It has been successfully used in many production system studies and applications (see [3]). It is also applicable in the production system operating in the transient regime, where the production disruption is typically short and frequent. For instance, the downtime of food package systems is often due to material jam on a machine or conveyor. The sealing process of battery production is often stopped by the operator to manually double check the alignment of cells. A more generalized model for machine reliability (such as geometric and exponential distribution) will be addressed in future work.

iv) Buffer B_1 has finite capacity N $(1 \le N < \infty)$. First-in-first-out (FIFO) is assumed regarding the buffer outflow process. The buffer contents change at the end of each cycle. The time a part resides in the buffer is referred to as residence time. At the beginning of the operation, assume n_0 parts in the buffer with the residence time of the first one being τ_0.

v) The maximum allowable waiting cycle in buffer B_1 is characterized by $T_{\max}$. When a part's waiting time exceeds $T_{\max}$, it has to be scrapped directly from the buffer.

vi) Machine m_1 is blocked during a time slot if it is up, buffer B_1 is full at the beginning of the time slot, and machine m_2 does not take a part and there is no part scrapped from buffer B_1. Machine m_2 is never blocked.

vii) Machine m_2 is starved during a time slot if it is up, buffer B_1 is empty at the beginning of the time slot or there is only one part in B_1 and it has to be scrapped during the time slot. Machine m_1 is never starved.

The problem to be studied in this paper is: *Under assumptions i)–vii), develop a method to evaluate the transient behavior of the production system and investigate system properties*. The solution to the problem is developed in Sections IV–VI.

IV. Modeling

A. Transient States and Transition Equations Formulation

The above assumptions describe a standard model for a two-machine production line subject to scrap due to the residence time limitation. In such production systems, the residence time of parts in the buffer needs to be known in order to determine their usability. Thus, the system state is defined by the number of parts in the system, and the associated residence time of each part. As one can see, when the buffer size is large, the number of states will grow substantially and the computation will also be intensive. Thus, in order to make the analysis tractable, rather than record the residence time of all parts in the buffer, we consider an approximation approach to estimate the residence times of the second to the last parts in the buffer. Then the system state can be defined by (n, τ), representing that the system has n parts in the buffer at the end of the cycle and the part at the head of the buffer has resided for τ cycles, $0 \le n \le N$, $0 \le \tau \le T_{\max} - 1$. The rationale is due to that in each cycle, only the first part (the one waiting at the head of the buffer) may be consumed or scrapped, based on the FIFO assumption. We expect that using the results from the approximation approach can lead to accurate evaluation of system performance. For a two-machine system with buffer size N and maximum residence time $T_{\max}$, the feasible states at arbitrary time t are shown in Table I.

TABLE I
System States at Arbitrary Time t

State	$\tau=0$	$\tau=1$	$\tau=2$	$\dots$	$\tau=N-1$	$\dots$	$\tau=T_{max}-1$
$n=0$	$(0,0)$	–	–	$\dots$	–	$\dots$	–
$n=1$	$(1,0)$	$(1,1)$	$(1,2)$	$\dots$	$(1,N-1)$	$\dots$	$(1,T_{max}-1)$
$n=2$	–	$(2,1)$	$(2,2)$	$\dots$	$(2,N-1)$	$\dots$	$(2,T_{max}-1)$
$n=3$	–	–	$(3,2)$	$\dots$	$(3,N-1)$	$\dots$	$(3,T_{max}-1)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$n=N$	–	–	–	$\dots$	$(N,N-1)$	$\dots$	$(N,T_{max}-1)$

First, consider state (0,0) at time $t+1$. It could be transferred from:

a) state $(0, 0)$ at time t, if no part is produced (machine m_1 is down) during the $(t+1)$th cycle;

b) state $(1, \tau)$ at time t, $\tau = 0, 1, \dots, T_{\max} - 2$, if machine m_1 is down and the part in the buffer is consumed (machine m_2 is up) during the $(t+1)$th cycle;

c) state $(1, T_{\max} - 1)$ at time t, if machine m_1 is down during the tth cycle. In this case, the part at the head of the buffer needs to be either consumed or scrapped since it will reach the maximum allowable residence time $T_{\max}$ at the end of the $(t+1)$th cycle, regardless of the status of machine m_2. Then there is zero part in the buffer at the end of the $(t+1)$th cycle.

Denote $x(n, \tau; t)$ as the probability for state (n, τ) at the end of the tth cycle, $0 \le n \le N$, $0 \le \tau \le T_{\max} - 1$, $0 \le t \le T$. The evolution of the system for state (0, 0) at time $t+1$ can be represented as

$$x(0,0;t+1) = x(0,0;t)(1-p_1) + \sum_{\tau=0}^{T_{\max}-2} x(1,\tau;t)(1-p_1)p_2 + x(1, T_{\max}-1; t)(1-p_1). \quad (1)$$

Second, state (1, 0) at time $t+1$ implies the buffer contains only one part which is produced during the $(t+1)$th cycle. Hence, the state transition is similar to (0, 0) at time $t+1$ except that machine m_1 needs to be up in the $(t+1)$th cycle in order to ensure that a new part flows into the buffer. The mathematical expression is as follows:

$$x(1,0;t+1) = x(0,0;t)p_1 + \sum_{\tau=0}^{T_{\max}-2} x(1,\tau;t)p_1p_2 + x(1,T_{\max}-1;t)p_1. \quad (2)$$

To determine the value for $x(1,\tau;t+1)$, $\tau = 1,2,\ldots, T_{\max}-2$, the following transitions need to be considered:

a) from $(1,\tau-1)$ at time t, if both machines are down during the $(t+1)$th cycle;
b) from $(2,k)$ at time t, $k=\tau,\tau+1,\ldots,T_{\max}-1$, if machine m_1 is down and m_2 is up during the $(t+1)$th cycle, and the second part in the buffer has residence time $\tau-1$;
c) from $(2,T_{\max}-1)$ at time t, if both machines are down during the $(t+1)$th cycle and the second part in the buffer has residence time $\tau-1$.

The key issue here is to determine the residence time for the second part in the buffer given the fact that the state does not specify that information. Here we introduce an operator $\Phi(n,\tau_1,\tau_2)$, which characterizes the probability that the second part in the buffer has residence time τ_2 given that there are n parts in the buffer and residence time of the first one is τ_1 (detailed explanation will be provided in the next subsection), i.e.,

$$\Phi(n,\tau_1;\tau_2) = P\left(T_{\text{res}}^2 = \tau_2 | N_{\text{oc}} = n, T_{\text{res}}^1 = \tau_1\right) \quad (3)$$

where T_{res}^1 and T_{res}^2 present the residence time for the first and second part in the buffer, respectively, and N_{oc} is the number of parts in the buffer.

Using such an operator, the transition for state $(1,\tau)$ at time $t+1$ can be expressed as

$$\begin{aligned} x(1,\tau;t+1) = {} & x(1,\tau-1;t)(1-p_1)(1-p_2) \\ & + \sum_{k=\tau}^{T_{\max}-2} x(2,k;t)(1-p_1)p_2\Phi(2,k,\tau-1) \\ & + x(2,T_{\max}-1;t)(1-p_1)\cdot\Phi(2,T_{\max}-1,\tau-1) \\ & \tau = 1,2,\ldots,T_{\max}-2. \end{aligned} \quad (4)$$

Similarly, when the residence time of the first part is close to its limit (i.e., at $T_{\max}-1$), the transition for state $(1,T_{\max}-1)$ at the end of the $(t+1)$th cycle can also be expressed using operator $\Phi(\cdot)$, which includes transitions from

a) $(1,T_{\max}-2)$ at time t, if both machines are down during the $(t+1)$th cycle;
b) $(2,T_{\max}-1)$ at time t, if machine m_1 is down during the $(t+1)$th cycle, and the second part in the buffer has residence time $T_{\max}-2$. Note that in this case no matter if m_2 is up or down, the transition is the same. If m_2 is up then the first part is processed by m_2. If m_2 is down, the first part will be scrap due to excessive residence time

$$\begin{aligned} x(1,T_{\max}-1;t+1) = {} & x(1,T_{\max}-2;t)(1-p_1)(1-p_2) \\ & + x(2,T_{\max}-1;t)(1-p_1) \\ & \cdot\Phi(2,T_{\max}-1,T_{\max}-2). \end{aligned} \quad (5)$$

Following similar analysis, the transitions for the rest of states are obtained in linear (6)–(10), shown at the bottom of the next page.

Equations (1)–(10) characterize the system evolution given model assumptions and state definitions. They describe how the system transfers from one state to another state along with time, thus establishing a foundation for rigorous transient analysis in such systems.

B. Approximation of Residence Time

Given the states defined in the above model, for the simplest case that $N=2$ and $T_{\max}=2$, the second part can only have residence time 0, i.e., $\Phi(2,1,0)=1$ and all other $\Phi(\cdot)=0$. For all other cases, there is no information indicating the residence time for the second part if the buffer occupancy is more than one. To avoid tracking each part's residence time in the buffer, we could still assume the state only includes the residence time of the first part and introduce an approximation method to estimate the residence time for the second part by utilizing the estimation of probability distribution of parts' residence time in the steady state.

Specifically, denote $P(T_{\text{res}}=\tau)$ as the steady-state probability distribution of part residence time in the buffer, $\tau = 0,1,\ldots,T_{\max}-1$. Its value has been investigated in [18] and is given as follows:

$$P(T_{\text{res}}=\tau) = \sum_{i=0}^{\min(\tau,N-1)} C_\tau^i p_2^{i+1}(1-p_2)^{\tau-i}\widetilde{P}_i \qquad \tau = 0,1,\ldots,T_{\max}-1 \quad (11)$$

where

$$C_\tau^i = \frac{\tau!}{i!(\tau-i)!}, \quad 0 \le i \le \tau$$

$$\widetilde{P}_0 = \frac{Q(p_1,p_2,N)}{(1-p_1)\left[1-Q(p_2,p_1,N)\right]}$$

$$\widetilde{P}_i = \alpha^i(p_1,p_2)\widetilde{P}_0, \quad i=1,\ldots,N-1$$

$$Q(p_1,p_2,N) = \begin{cases} \frac{(1-p_1)(1-\alpha(p_1,p_2))}{1-\frac{p_1}{p_2}\alpha^N(p_1,p_2)}, & \text{if } p_1 \neq p_2 \\ \frac{1-p}{N+1-p}, & \text{if } p_1 = p_2 = p \end{cases}$$

$$\alpha(p_1,p_2) = \frac{p_1(1-p_1)}{p_2(1-p_2)}.$$

In the above derivations, $\widetilde{P}_i$ denotes the probability that the buffer has i parts when machine m_1 produces a part into the buffer, $i=0,1,\ldots,N-1$. Functions $Q(p_1,p_2,N)$ and $\alpha(p_1,p_2)$ are derived in monograph [3].

Suppose there are n parts in the buffer with the first one having residence time τ_1. Then the residence time for the second part, could take values from $n-2, n-1, \ldots$, to $\tau_1 - 1$. Operator $\Phi(n, \tau_1, \tau_2)$ can be estimated as the percentage that the second part has residence time τ_2 among all possible scenarios, which is given by $P(T_{\text{res}} = \tau_2)/(\sum_{i=n-2}^{\tau_1 - 1} P(T_{\text{res}} = i))$.

C. State Transition Matrix

To characterize the transition matrix of the Markov chain described above, the two-dimension state space for an arbitrary time t needs to be transformed into a vector, which contains all the states. According to the feasible states shown in Table I, there is only one state (0, 0) at time t when the buffer occupancy is 0. For a given value of buffer occupancy n $(n \geq 1)$, the number of states is $T_{\max} - n + 1$. Therefore, the total number of states S can be calculated as

$$S = \sum_{i=1}^{N} (T_{\max} - i + 1) + 1 = N \cdot T_{\max} - \frac{N(N-1)}{2} + 1.$$

We denote $X(t)$ as the $S \times 1$ vector at time t, where $X(t) = [x_1(t)\ x_2(t)\ x_3(t) \ldots x_S(t)]^\top$, defined in a finite state space $\mathcal{X}_{S\times 1}$. To transform the system states into a vector, we rank state (n, τ) at time t based on buffer occupancy n first, and then the residence time τ for the first part in the buffer, following an increasing order. In other words, the state (n, τ)'s position $I(n, \tau)$ at cycle t in vector $X(t)$ could be presented as

$$I(n,\tau) = \begin{cases} 1, & \text{if } n = 0 \\ (n-1)T_{\max} + \tau - \frac{n(n-1)}{2} + 2, & \text{if } n \geq 1. \end{cases} \tag{12}$$

Using the above mapping policy, we rearrange the $S \times 1$ vector as $X(t) = [x_1(t), x_2(t), x_3(t), \ldots, x_S(t)]^\top$, with $x_{I(n,\tau)}(t) = x(n, \tau; t)$. Therefore, $x_i(t)$, $i = I(n, \tau)$, represents the probability that the buffer has n parts with the first one staying for τ cycles at the end of time slot t. Denote P as the $S \times S$ transition probability matrix of the Markov chain described in (1)–(10). Assume that buffer has n_0 parts and the first part has residence time τ_0 at the beginning of the first cycle. The mathematical model could be simplified as the following matrix form:

$$X(t+1) = P \cdot X(t), \quad t = 0, 1, \ldots, T \tag{13}$$

with initial condition

$$X(0) = [0 \ldots x_i(0) = 1 \ldots 0]^\top$$
$$i = \begin{cases} 1, & \text{if } n_0 = 0 \\ (n_0 - 1)T_{\max} + \tau_0 - \frac{n_0(n_0-1)}{2} + 2, & \text{if } n_0 \geq 1. \end{cases} \tag{14}$$

$$\begin{aligned} x(n, n-1; t+1) &= x(n-1, n-2; t)p_1(1-p_2) + \sum_{k=n-1}^{T_{\max}-2} x(n, k; t)p_1 p_2 \Phi(n, k, n-2) \\ &\quad + x(n, T_{\max}-1; t)p_1 \Phi(n, T_{\max}-1, n-2), n = \ 2, 3, \ldots, N \end{aligned} \tag{6}$$

$$\begin{aligned} x(n, \tau; t+1) &= x(n-1, \tau-1; t)p_1(1-p_2) + x(n, \tau-1; t)(1-p_1)(1-p_2) + \sum_{k=\tau}^{T_{\max}-2} x(n, k; t)p_1 p_2 \Phi(n, k, \tau-1) \\ &\quad + x(n, T_{\max}-1; t)p_1 \Phi(n, T_{\max}-1, \tau-1) + \sum_{k=m}^{T_{\max}-2} x(n+1, k; t)(1-p_1)p_2 \Phi(n+1, k, \tau-1) \\ &\quad + x(n+1, T_{\max}-1; t)(1-p_1)\Phi(n+1, T_{\max}-1, \tau-1) \\ n &= 2, 3, \ldots, N-1, \quad \tau = n, n+1, \ldots, T_{\max}-2 \end{aligned} \tag{7}$$

$$\begin{aligned} x(n, T_{\max}-1; t+1) &= x(n-1, T_{\max}-2; t)p_1(1-p_2) + x(n, T_{\max}-2; t)(1-p_1)(1-p_2) \\ &\quad + x(n, T_{\max}-1; t)p_1 \Phi(n, T_{\max}-1, T_{\max}-2) \\ &\quad + x(n+1, T_{\max}-1; t)(1-p_1)\Phi(n+1, T_{\max}-1, T_{\max}-2) \\ n &= 2, 3, \ldots, N-1 \end{aligned} \tag{8}$$

$$\begin{aligned} x(N, \tau; t+1) &= x(N-1, \tau-1; t)p_1(1-p_2) + \sum_{k=\tau}^{T_{\max}-2} x(N, k; t)p_1 p_2 \Phi(N, k, \tau-1) \\ &\quad + x(N, \tau-1; t)(1-p_2) + x(N, T_{\max}-1; t)p_1 \Phi(N, T_{\max}-1, \tau-1) \\ \tau &= N, N+1, \ldots, T_{\max}-2 \end{aligned} \tag{9}$$

$$\begin{aligned} x(N, T_{\max}-1; t+1) &= x(N-1, T_{\max}-2; t)p_1(1-p_2) + x(N, T_{\max}-2; t)(1-p_2) \\ &\quad + x(N, T_{\max}-1; t)p_1 \Phi(N, T_{\max}-1; T_{\max}-2) \end{aligned} \tag{10}$$

For instance, transition matrices for systems with $N = 2, T_{\max} = 2$ and $N = 2, T_{\max} = 3$ are 4×4 and 6×6 dimensions, and are shown in (15) and (16), respectively, shown at the bottom of the page.

For the simplest case where $N = 2$ and $T_{\max} = 2$, since a closed form transition matrix $P(N = 2, T_{\max} = 2)$ can be derived, a closed form expression of vector $X(t)$ is obtained.

Proposition 1: Under assumptions i)–vii) when $N = 2$, $T_{\max} = 2$

$$X(t) = \left(L + (p_1 p_2)^t M\right) X(0), \quad t > 0$$

where L and M are presented in (17) and (18), respectively, as shown at the bottom of the page. ■

Proof: See the Appendix.

For all other cases, the closed form solution of $X(t)$ is all but impossible. Thus, (13) will be used to evaluate $X(t)$ iteratively.

The above ranking policy and transition matrix formulation are the basis for the analysis reported in this paper. We will utilize the transition matrix and vectorized state probabilities in the rest of the paper.

V. Transient Performance Evaluation

In this section, we use the expressions developed in Section IV to investigate the system's transient performance. Performance measures that are of particular interests are as follows:

- *Production Rate* $PR(t)$: the expected number of parts produced by machine m_2 in the tth cycle;
- *Consumption Rate* $CR(t)$: the expected number of parts consumed by machine m_1 in the tth cycle;
- *Scrap Rate* $SR(t)$: the expected number of scrapped parts in the tth cycle;
- *Work-In-Process* $WIP(t)$: the expected number of parts in buffer B_1 at the end of the tth cycle.

Such measures can be evaluated using the model and transition matrix obtained from the previous section. Given a production system defined by assumptions i)–vii), the performance measures can be estimated in the following way:

$$\widehat{PR}(t) = p_2 \left(1 - x_1(t-1)\right) \tag{19}$$

$$\widehat{CR}(t) = p_1 \left[1 - (1 - p_2) \sum_{i=N-1}^{T_{\max}-2} x_{I(N,i)}(t-1)\right] \tag{20}$$

$$\widehat{SR}(t) = (1 - p_2) \sum_{i=1}^{N} x_{I(i, T_{\max}-1)}(t-1) \tag{21}$$

$$\widehat{WIP}(t) = \sum_{i=1}^{N} \sum_{j=i-1}^{T_{\max}-1} i \cdot x_{I(i,j)}(t) \tag{22}$$

where $x_i(t-1)$ and $x_i(t)$ are obtained from (1)–(14).

Specifically, $\widehat{PR}(t)$ is the average number of parts produced by the last machine at the tth cycle. It could be calculated based on the probability that machine m_2 is up (i.e., p_2) and the buffer is not empty (i.e., $1 - x_1(t-1)$) during the tth cycle.

$$P(N=2, T_{\max}=2) = \begin{bmatrix} 1-p_1 & (1-p_1)p_2 & 1-p_2 & 0 \\ p_1 & p_1 p_2 & p_1 & 0 \\ 0 & (1-p_1)(1-p_2) & 0 & 1-p_1 \\ 0 & p_1(1-p_2) & 0 & p_1 \end{bmatrix} \tag{15}$$

$$P(N=2, T_{\max}=3) = \begin{bmatrix} 1-p_1 & (1-p_1)p_2 & (1-p_1)p_2 & 1-p_1 & 0 & 0 \\ p_1 & p_1 p_2 & p_1 p_2 & p_1 & 0 & 0 \\ 0 & (1-p_1)(1-p_2) & 0 & 0 & (1-p_1)p_2 & (1-p_1)\Phi(2,2,0) \\ 0 & 0 & (1-p_1)(1-p_2) & 0 & 0 & (1-p_1)\Phi(2,2,1) \\ 0 & p_1(1-p_2) & 0 & 0 & p_1 p_2 & p_1\Phi(2,2,0) \\ 0 & 0 & p_1(1-p_2) & 0 & 1-p_2 & p_1\Phi(2,2,1) \end{bmatrix} \tag{16}$$

$$L = \begin{bmatrix} \frac{p_1^2-2p_1+1}{1-p_1p_2} & \frac{p_1^2-2p_1+1}{1-p_1p_2} & \frac{p_1^2-2p_1+1}{1-p_1p_2} & \frac{p_1^2-2p_1+1}{1-p_1p_2} \\ \frac{p_1-p_1^2}{1-p_1p_2} & \frac{p_1-p_1^2}{1-p_1p_2} & \frac{p_1-p_1^2}{1-p_1p_2} & \frac{p_1-p_1^2}{1-p_1p_2} \\ \frac{p_2p_1^2-p_1^2-p_2p_1+p_1}{1-p_1p_2} & \frac{p_1(p_2p_1-p_1-p_2+1)}{1-p_1p_2} & \frac{p_2p_1^2-p_1^2-p_2p_1+p_1}{1-p_1p_2} & \frac{p_1(p_2p_1-p_1-p_2+1)}{1-p_1p_2} \\ \frac{p_1^2-p_1^2p_2}{1-p_1p_2} & \frac{p_1^2-p_1^2p_2}{1-p_1p_2} & \frac{p_1^2-p_1^2p_2}{1-p_1p_2} & \frac{p_1^2-p_1^2p_2}{1-p_1p_2} \end{bmatrix} \tag{17}$$

$$M = \begin{bmatrix} \frac{p_2p_1-p_1-p_2+1}{p_2(1-p_1p_2)} & \frac{p_2^2p_1^2-p_1^2-p_2^2p_1-p_2p_1+2p_1+p_2-1}{p_1p_2(1-p_1p_2)} & \frac{p_2p_1-p_1-p_2+1}{p_2(1-p_1p_2)} & \frac{-p_1^2+2p_1-1}{p_1p_2(1-p_1p_2)} \\ \frac{p_1(1-p_2)}{p_2(1-p_1p_2)} & \frac{-p_1p_2^2+p_2+p_1-1}{p_2(1-p_1p_2)} & \frac{p_1(1-p_2)}{p_2(1-p_1p_2)} & \frac{p_1-1}{p_2(1-p_1p_2)} \\ \frac{-p_2p_1+p_1+p_2-1}{p_2(1-p_1p_2)} & \frac{-p_2^2p_1^2+p_1^2+p_2^2p_1+p_2p_1-2p_1-p_2+1}{p_1p_2(1-p_1p_2)} & \frac{-p_2p_1+p_1+p_2-1}{p_2(1-p_1p_2)} & \frac{p_1^2-2p_1+1}{p_1p_2(1-p_1p_2)} \\ \frac{p_1(p_2-1)}{p_2(1-p_1p_2)} & \frac{p_1p_2^2-p_2-p_1+1}{p_2(1-p_1p_2)} & \frac{p_1(p_2-1)}{p_2(1-p_1p_2)} & \frac{1-p_1}{p_2(1-p_1p_2)} \end{bmatrix} \tag{18}$$

$\widehat{CR}(t)$ presents the average number of parts consumed by the first machine at the tth cycle and it is obtained by multiplying the probabilities that machine m_1 is functioning (i.e., p_1) and there will be at least one spot available in the buffer by the end of the next cycle (i.e., $1-(1-p_2)\sum_{i=N-1}^{T_{\max}-2} x_{I(N,i)}(t-1)$). Similarly, we could calculate $\widehat{SR}(t)$ by using the probabilities that machine m_2 is down (i.e., $1-p_2$) and the fist part in the buffer has residence time $T_{\max}-1$, which is about to be scrapped during the tth cycle (i.e., $\sum_{i=1}^{N}\widetilde{x}_{I(i,T_{\max}-1)}$). Finally, $\widehat{WIP}(t)$ estimates the expected buffer occupancy at the end of the tth cycle.

To evaluate the accuracy of the above estimates, we first define the steady-state performance measures $PR(\infty)$, $CR(\infty)$, $SR(\infty)$, and $WIP(\infty)$, given as follows:

$$PR(\infty) = p_2(1-\widetilde{x}_1) \tag{23}$$

$$CR(\infty) = p_1\left[1-(1-p_2)\sum_{i=N-1}^{T_{\max}-2}\widetilde{x}_{I(N,i)}\right] \tag{24}$$

$$SR(\infty) = (1-p_2)\sum_{i=1}^{N}\widetilde{x}_{I(i,T_{\max}-1)} \tag{25}$$

$$WIP(\infty) = \sum_{i=1}^{N}\sum_{j=i-1}^{T_{\max}-1} i\cdot\widetilde{x}_{I(i,j)} \tag{26}$$

where $\widetilde{x}_i$ represents the steady-state value of the state probability x_i, $i=1,\ldots,S$.

Then we introduce the following error metrics:

$$\bar{\delta}_{PR} = \frac{1}{T}\sum_{t=1}^{T}\frac{\left|PR(t)-\widehat{PR}(t)\right|}{PR(\infty)}\times 100\%$$

$$\bar{\delta}_{CR} = \frac{1}{T}\sum_{t=1}^{T}\frac{\left|CR(t)-\widehat{CR}(t)\right|}{CR(\infty)}\times 100\%$$

$$\bar{\delta}_{SR} = \frac{1}{T}\sum_{t=1}^{T}\frac{\left|SR(t)-\widehat{SR}(t)\right|}{SR(\infty)}\times 100\%$$

$$\bar{\delta}_{WIP} = \frac{1}{T}\sum_{t=1}^{T}\frac{\left|WIP(t)-\widehat{WIP}(t)\right|}{WIP(\infty)}\times 100\%$$

where $PR(t)$, $SR(t)$, $CR(t)$, and $WIP(t)$ are obtained using numerical simulation. For each parameter setting in simulation, 2000 replications are carried out, resulting in the 95% confidence interval less than 0.005 for all performance measures. Each replication follows the procedure below.

Procedure 1:

1) Randomly generate parameters from the following sets:

$$\begin{aligned} p_1 &\in (0.7, 0.99)\\ p_2 &\in (p_1-0.3, p_1)\\ N &\in \{2,3,\ldots,10\}\\ T_{\max} &\in \{N-1, N, \ldots, N+5\}\\ n_0 &\in \{0,1,\ldots,N\}\\ \tau_0 &\in \{n_0-1, n_0, \ldots, T_{\max}-1\}. \end{aligned} \tag{27}$$

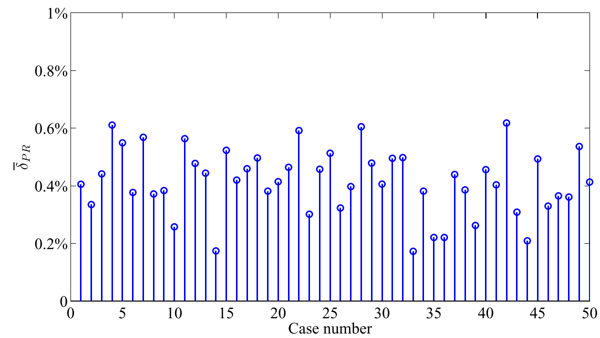


Fig. 2. Accuracy for PR.

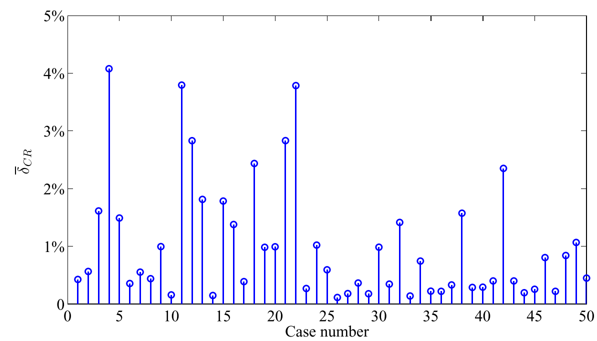


Fig. 3. Accuracy for CR.

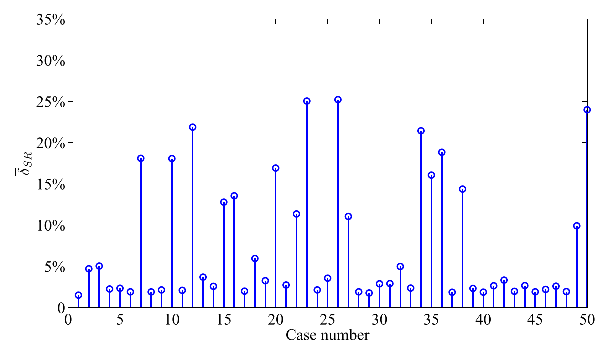


Fig. 4. Accuracy for SR.

2) Set the initial buffer occupancy as n_0 and the residence time of the first part as τ_0. If $n_0>1$, assume the residence time of the rest of the parts to be $\tau_0-1, \tau_0-2, \ldots, \tau_0-n_0+1$.
3) Run the simulation code for a total of 200 time slots to ensure the system reaches the steady state.
4) Take the average of performance measures of the last 50 time slots as the simulated performance in the steady state.

To investigate the accuracy of the proposed estimates, 50 different configurations are tested using both the analytical model and simulation. The resulting accuracies are summarized in Figs. 2–5. As one could see, $\bar{\delta}_{PR}$, $\bar{\delta}_{CR}$, $\bar{\delta}_{SR}$, and $\bar{\delta}_{WIP}$ are typically within 0.5%, 2%, 5%, and 2%, respectively. Note that $\bar{\delta}_{SR}$ is relatively larger. However, since SR is small, typically less than 0.1, its absolute error is generally less than 0.02. Therefore, we conclude that performance estimates obtained by (19)–(22) are effective in analyzing the transient behavior of a two-machine production system defined by

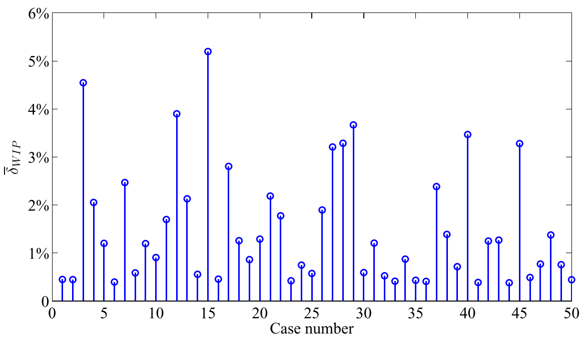


Fig. 5. Accuracy for WIP.

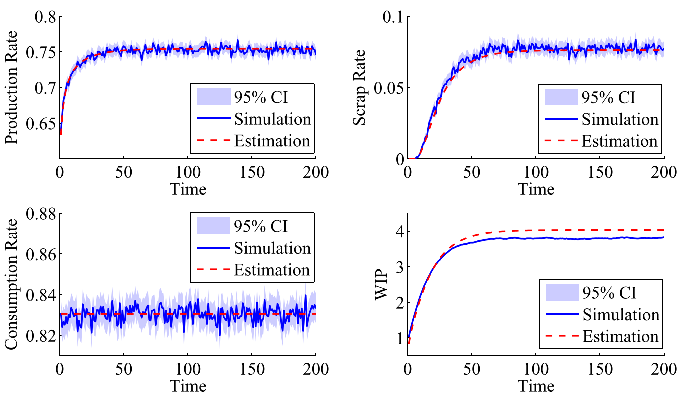


Fig. 6. Case 1: $p_1 = 0.8305$, $p_2 = 0.7623$, $N = 6$, $T_{\max} = 6$, $n_0 = 0$, $\tau_0 = 0$.

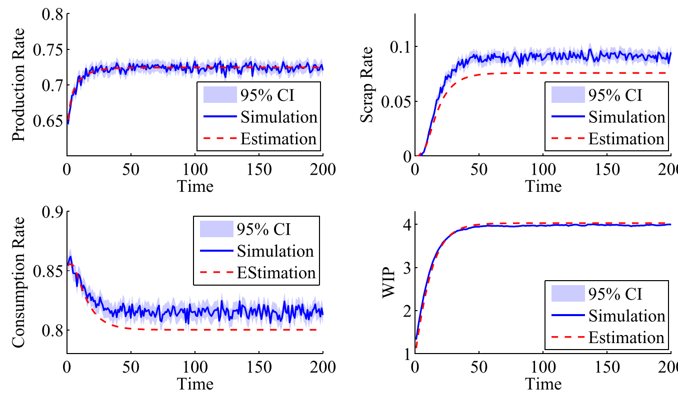


Fig. 7. Case 2: $p_1 = 0.8554$, $p_2 = 0.7273$, $N = 5$, $T_{\max} = 6$, $n_0 = 1$, $\tau_0 = 1$.

assumptions i)–vii) with sufficient accuracy. In addition, since no other approximation is involved except the estimation of residence time, the accuracy of the performance measure also reflects the accuracy of residence time approximation.

To illustrate how the analytical model captures the transient behavior of the system, we select two representative cases from the simulation campaign and report them in Figs. 6 and 7. Simulated performance measures are plotted in solid lines, with shaded area indicating the 95% confidence interval. The dashed lines represent estimation from (19)–(22). As one can see, estimation and simulation results are close in the whole period. Specifically, during the transient period, the estimated performance measures can clearly capture the system dynamics. These results indicate that although the approximation does not fully reveal the system information and may lead to error in predicting the probability of each state, the performance measures are essentially the aggregated values of a set of related state probabilities. Thus, the estimation can still provide enough accuracy for transient analysis in two-machine Bernoulli lines with perishable products.

VI. Structural Properties

In this section, we use the mathematical model and analytical expressions derived in the previous sections to investigate the system-theoretic properties of the transient behavior for two-machine Bernoulli line with perishable products. Specifically, we study how system parameters p_1, p_2, N, and $T_{\max}$ affect the system performance with an emphasis on PR and SR.

To investigate transient properties, numerical experiments are carried out based on the analytical model shown in Section IV. A program is created to emulate the system transition and evaluate performance measures. Specifically, the numerical approach follows the procedure below:

Procedure 2:

1) Set the line configuration with parameters randomly generated from sets (27) in Procedure 1.
2) Define the initial state probability $X(0)$ according to (14), and calculate the transition matrix P based on (1)–(10).
3) Let the system run 100 time slots using (19)–(22), and record $PR(t)$ and $SR(t)$, $t = 1, 2, \ldots, 100$.
4) Change only one parameter (e.g., p_1, p_2, N, $T_{\max}$) by a small amount each time and redo Steps 2 and 3.

Following the above procedure, we randomly generate 1000 cases, and without a single exception we find the results shown below, formulated as numerical facts.

Numerical Fact 1: Consider a production system defined by assumptions i)–vii), at an arbitrary time t, $1 \leq t \leq T$,

i) both $PR(t)$ and $SR(t)$ are monotonically increasing in p_1 and N;
ii) $PR(t)$ is monotonically increasing in p_2 and $T_{\max}$, while $SR(t)$ is monotonically decreasing with respect to p_2 and $T_{\max}$.

To illustrate these numerical facts, four representative line configurations are selected from the numerical study. Without loss of generality, we pick the performance measures at time $t = 10, 20,$ and 70, representing the beginning of the operation, the middle of the transient, and the steady state, respectively. The results are shown in Figs. 8–11. From these figures, we can observe how the system parameters p_1, p_2, N, and $T_{\max}$ affect the system's performance as their values vary. First, as p_1 increases, machine m_1 fails less frequently, which will lead to more parts stacked in buffer B_1. This causes a drop on the starvation probability of machine m_2 and increases the residence time of parts in buffer B_1. As a consequence, production rate and scrap rate will both increase as shown in Fig. 8.

The effect of p_2 on PR and SR is as straightforward as p_1. Intuitively, the increase of p_2 will improve the line efficiency

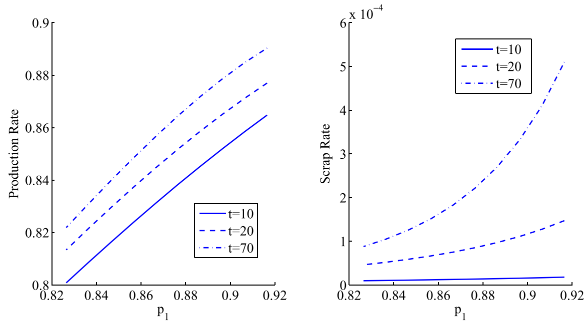


Fig. 8. Performance measures as functions of machine m_1's reliability p_1, where $p_2 = 0.907$, $N = 4$, $T_{\max} = 6$, $n_0 = 0$, $\tau_0 = 0$.

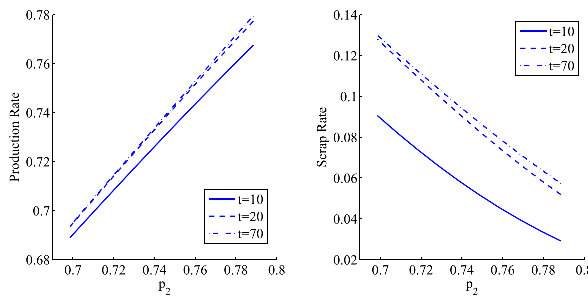


Fig. 9. Performance measures as functions of machine m_2's reliability p_2, where $p_1 = 0.9$, $N = 3$, $T_{\max} = 4$, $n_0 = 0$, $\tau_0 = 0$.

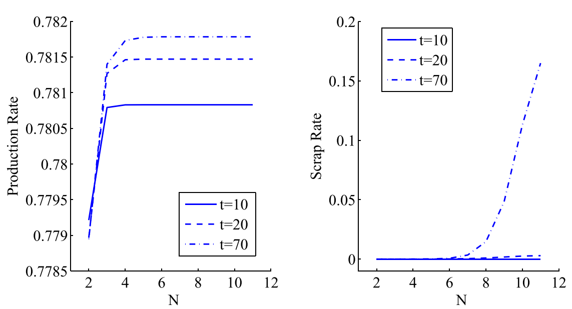


Fig. 10. Performance measures as functions of buffer capacity N ($p_1 = 0.9629$, $p_2 = 0.7818$, $T_{\max} = 11$, $n_0 = 2$, $\tau_0 = 1$).

thus contributing to the improvement of PR. From (21), one could see that SR only depends on p_2 and the probability that the first part in buffer B_1 reaches maximum allowable residence time. When we increase p_2, less stoppage will occur on machine m_2, causing parts in the buffer with less residence time before flowing out. Thus, the increase of p_2 can naturally reduce the probability that the first waiting part reaches maximum allowable residence time. Eventually, SR will decrease in p_2 (see Fig. 9).

For buffer size N, qualitatively similar results are obtained as p_1 (shown in Fig. 10). When N is increased, both PR and SR will increase because there is more holding space between machines, which makes machine m_2 less likely being starved, but also may result in longer residence time of parts in the buffer. Clearly, there is no need to have a buffer larger than $T_{\max}$. A noteworthy observation is that when N is increased to a certain level (for example, $N = 4$ in Fig. 10), the marginal gain in PR is minimal while SR still rises nearly exponentially. Therefore, to achieve a certain increase in PR by adding buffer capacity, one needs to be aware of the increase of SR as the side effect and consider the monetary tradeoff between PR and SR.

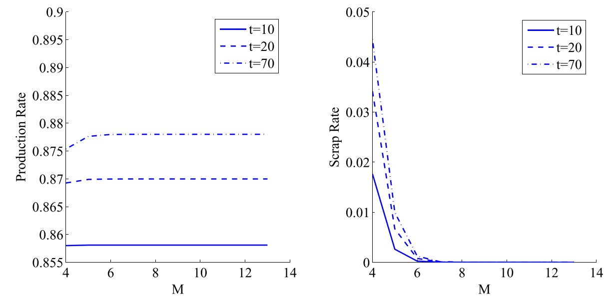


Fig. 11. Performance measures as functions of maximum residence time $T_{\max}$ ($p_1 = 0.92$, $p_2 = 0.8944$, $N = 3$, $n_0 = 0$, $\tau_0 = 0$).

To increase the maximum allowable residence time $T_{\max}$ is not about changing the line configuration, but rather is achieved by product redesign or the manufacturing process improvement to allow the part staying in the buffer for some extra time without losing utility. Clearly from Fig. 11, the monotonic property holds. In addition, the increase of $T_{\max}$ does not help increase PR too much, but could reduce SR tremendously (since SR is typically small).

Numerical Fact 1 qualitatively characterizes the effect of system parameters on performance measures. It not only reveals system dynamics but also provides directions for performance improvement.

VII. Settling Time Estimation

In addition to transient performance evaluation and structural properties, there are many other aspects of manufacturing systems that can be effectively addressed using the proposed analytical model. In this and next sections, we introduce two problems: settling time estimation and real-time production control. We show how the transient model can be effective in evaluating the transient behavior and determining a real-time control policy to achieve the optimal production objective of the system described in this paper.

To describe the transient behavior quantitatively, one important measure is the settling time, denoted as t_s, which is the time that elapses from the beginning of the production operation to the time that the corresponding performance measure reaches and stays within a range of certain tolerance of the final value, such as $\pm 5\%$ ([3], [20]). In the production system illustrated in this paper, the notion of settling time is adopted to characterize the necessary time for the system to reach the steady state given the initial buffer condition. To determine whether the steady state is reached, a small value ϵ (e.g., $\epsilon = 5\%$) is introduced to define the range of tolerance around the steady state value. Since PR and SR are considered as the most critical performance measures in production practice, especially in the system described by assumptions i)–vii), we study the settling times for PR and SR, denoted as $t_s^{PR}(\epsilon)$ and $t_s^{SR}(\epsilon)$ for a given ϵ, respectively.

Let $X(0) = [x_1(0), x_2(0), \ldots, x_S(0)]^\top$ be the initial system condition. Let $X(t|X(0)) = [x_1(t|X(0)), x_2(t|X(0)), \ldots, x_S(t|X(0))]^\top$ be the state probability vector at time t given the initial condition $X(0)$, and denote $\widetilde{X} = [\widetilde{x}_1, \widetilde{x}_2, \ldots, \widetilde{x}_S]^\top$ as the steady state probability vector.

According to (19) and (21), performance measures at time t can be obtained as follows:

$$\widehat{PR}(t) = p_2 \left(1 - x_1 \left(t - 1 | X(0)\right)\right) \tag{28}$$

$$\widehat{SR}(t) = (1 - p_2) \sum_{i=1}^{N} x_{I(i, T_{\max}-1)} \left(t - 1 | X(0)\right). \tag{29}$$

The performance measures at the steady state $PR(\infty)$ and $SR(\infty)$ can be obtained similarly from (23) and (25). To estimate the settling time for PR and SR, the following theorem is obtained.

Theorem 1: Given a production system defined by assumptions i)–vii) with finite state space $\mathcal{X}_{S\times 1}$ and transition matrix $P_{S\times S}$, for an arbitrary small value ϵ

$$\left|\widehat{PR}\left(t_s^{PR}(\epsilon)\right) - PR(\infty)\right| \le \epsilon, \text{ if}$$

$$t_s^{PR}(\epsilon) = \left\lceil \frac{4}{k^2} \log\left(\frac{p_2}{2\epsilon\sqrt{\widetilde{x}_i}}\right)\right\rceil \tag{30}$$

$$\left|\widehat{SR}\left(t_s^{SR}(\epsilon)\right) - SR(\infty)\right| \le \epsilon, \text{ if}$$

$$t_s^{SR}(\epsilon) = \left\lceil \frac{4}{k^2} \log\left(\frac{N(1-p_2)}{2\epsilon\sqrt{\widetilde{x}_i}}\right)\right\rceil \tag{31}$$

where i is obtained such that $x_i(0) = 1$ and k is a Cheeger constant defined by

$$k = \min_{A: \sum_{i\in A} \widetilde{x}_i \le \frac{1}{2}} \frac{\sum_{i\in A}\sum_{j\in A^C} \widetilde{x}_i P_{i,j}}{\sum_{i\in A} \widetilde{x}_i} \tag{32}$$

where A is a subset of state space $\mathcal{X}$ and A^c is the complement of A in $\mathcal{X}$. $P_{i,j}$ represents the transition probability from state x_i to state x_j.

Proof: See the Appendix. ■

In this theorem, we adopt the Cheeger constant [23] that identifies the "bottleneck" of the system that impedes the system's transition to steady state. This implies that within the system's state space, there exist a subset A and its complement set A^c, which have the fewest transitions between them. Such a distribution is viewed as a "bottleneck," which determines the speed of converging to steady state. From (32), the value of the Cheeger constant only depends on state space $\mathcal{X}_{S\times 1}$ and transition matrix $P_{S\times S}$, and is irrelevant to the initial condition.

The interpretation of settling time is straightforward. As one can see, from (30) and (31), the Cheeger constant and system parameters are fixed for a given system configuration. Then the settling time is a function of the range of tolerance ϵ and the initial condition $X(0)$. For a certain ϵ, $t_s^{PR}(\epsilon)$ and $t_s^{SR}(\epsilon)$ are slightly different, which implies that different performance measures may converge to the steady state with different speeds. In addition, it is worth to mention that the calculated settling time is a sufficient but not necessary condition to ensure the steady state being reached. In other words, $t_s^{PR}(\epsilon)$ and $t_s^{SR}(\epsilon)$ provide estimated upper bounds for the real settling times. Such upper bounds are of practical significance to predict the necessary "warmup" time at the beginning of each shift and to determine which initial buffer condition delivers the best settling time performance.

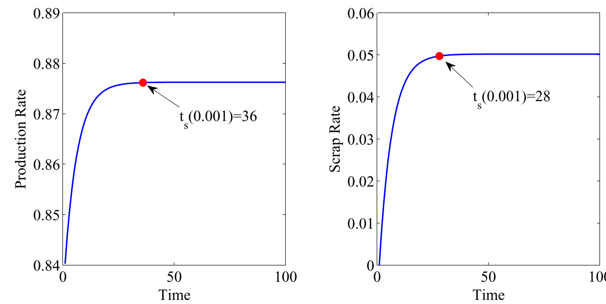


Fig. 12. Settling time for PR and SR ($p_1 = 0.9265$, $p_2 = 0.9070$, $N = 2$, $T_{\max} = 2$, $n_0 = 0$, $\tau_0 = 0$).

To illustrate the applicability of Theorem 1, we randomly generate system parameters as $p_1 = 0.9265$, $p_2 = 0.9070$, $N = 2$, $T_{\max} = 2$, $n_0 = 0$, $\tau_0 = 0$, and set ϵ to be 0.001. The evolution of production rate and scrap rate is shown in Fig. 12. Clearly, both PR and SR encounter a transient period before reaching the steady state. The transient period is roughly 30 cycles for PR and 25 cycles for SR. Based on Theorem 1, the estimated upper bounds $t_s^{PR}(0.001)$ and $t_s^{SR}(0.001)$ are 36 and 28, respectively, which are close to the observation. Therefore, the upper bounds based on the transient model are efficient in estimating the settling time for performance measures.

VIII. Real-Time Production Control

In addition to settling time, another significant implication of transient characteristics is production loss. As shown in Fig. 12, the production rate needs some time (the settling time) to the reach steady state, resulting substantial production loss during the transient period. In fact, as suggested in [20], the production loss due to transients can be up to 10% in an 8-hour production shift. Therefore, how to control the production system to maximize the production rate during the whole observation period is of practical importance.

Moreover, the production system of perishable products is subject to scrap, due to parts' excessive waiting and exposure in the buffer. Such waiting time can be practically reduced by controlling the machine behavior according to the real-time information of the buffer. For instance, when the residence time of the part in the buffer is close to the maximum allowable time because of the frequent stoppage downstream, the upstream machine can be switched off or put on hold for a certain period even though there are still available spots in the buffer. This will potentially avoid increasing the residence time of the parts in the buffer and consequently reduce the scrap rate. The questions are, according to what policy/indicators such a control procedure should be activated, and what is the potential profitability improvement or cost reduction.

To answer these questions, a method for generating production control policies is introduced based on the transient model described in Section IV. The objective is to maximize the total expected reward over a finite production operation horizon. By incorporating both scrap and production loss due to the transient of PR, we establish the control objective function as follows:

$$\max_{\pi} E^{\pi}\left\{\sum_{t=1}^{T}\left(\widehat{PR}(t)-\lambda\cdot\widehat{SR}(t)\right)\right\} \tag{33}$$

where π denotes the control policy, $\widehat{PR}(t)$ and $\widehat{SR}(t)$ are defined in (19) and (21), and λ $(0<\lambda<1)$ represents the discount factor for the scrap parts. To find the best policy, a Markov decision process (MDP) formulation of the real-time production control problem is given as follows:

- Decision epochs: $E=\{1,2,\ldots,T\}$.
- States: $X=\{x_1,x_2,\ldots,x_S\}$.
- Actions: $A=\{1,0\}$, $\forall\, x_i\in X$, i.e., $a_t(x_i)\in\{1,0\}$, $i=1,2,\ldots,S$, where 1 means turning machine m_1 on while 0 means off.
- Rewards: $\forall\, t\in E$

$$r_t(x_i)=\begin{cases}0, & \text{if } i=1\\ p_2, & \text{if } i=I(n,\tau)\\ & n=1,2,\ldots,N\\ & \tau=n-1,n,\ldots,T_{\max}-2\\ p_2-\lambda(1-p_2), & \text{if } i=I(n,T_{\max}-1)\\ & n=1,2,\ldots,N\end{cases}$$

  where $r_t(x_i)$ represents the total expected production reward at time t when the system is state x_i. When $i=0$, the buffer is empty, and no production or scrap will be generated. States with $n>0$ and $\tau<T_{\max}-1$ refer to the scenarios where the buffer is not empty and no part in the buffer will be scrapped. If machine m_2 is up, a product will be produced. Therefore, the corresponding reward is p_2. Similarly for states where the first part has residence time $T_{\max}-1$, its probability being consumed is p_2 while the scrap probability is $1-p_2$. Combining the effect with discounted factor λ, the reward for the last scenario in the reward function is obtained as $p_2-\lambda(1-p_2)$.
- Transition Probabilities:

$$P_{a_t}=\begin{cases}P, & \text{if } a_t=1\\ P(p_1=0), & \text{if } a_t=0,\end{cases}\qquad \forall\, t\in E$$

  where $P(p_1=0)$ is obtained by setting p_1 to be 0 for the original transition probability matrix P. Each item in the transition probability matrix is denoted as $p(j|i,a_t)$, representing the probability from state i to state j given action a_t.
- Objective:

$$\max_{\pi} E^{\pi}\left\{\sum_{t=1}^{T} r_t(X)\right\}. \tag{34}$$

According to the above formulation, we establish the optimality equation as a cost-to-go function shown as follows:

$$u_t(x_i)=\sup_{a_t(x_i)\in A}\left\{r_t(x_i)+\sum_{j\in X}p\left(j|x_i,a_t(x_i)\right)u_{t+1}(j)\right\}$$
$$i=1,\ldots,S;\quad t=1,\ldots,T-1 \tag{35}$$

TABLE II
CONTROL POLICY ($p_1=0.8324$, $p_2=0.6807$, $N=5$, $T_{\max}=6$, $\lambda=0.8$)

State	$\tau=0$	$\tau=1$	$\tau=2$	$\tau=3$	$\tau=4$	$\tau=5$
$n=0$	1	–	–	–	–	–
$n=1$	1	1	1	1	1	1
$n=2$	–	1	1	1	1	0
$n=3$	–	–	1	1	0	0
$n=4$	–	–	–	0	0	0
$n=5$	–	–	–	–	0	0

with boundary condition

$$u_T(x_i)=r_T(x_i),\quad i=1,\ldots,S.$$

Since $a_t(x_i)$ is compact and $r_t(x_i)$ and P_{a_t} are continuous functions on $a_t(x_i)$, there exists an optimal deterministic Markov policy. Thus, the control procedure is time invariant and only depends on the system state. Therefore, the optimality equation can be rewritten as

$$u_t(x_i)=\sup_{a(x_i)\in A}\left\{r_t(x_i)+\sum_{j\in X}p\left(j|x_i,a(x_i)\right)u_{t+1}(j)\right\}$$
$$i=1,\ldots,S;\quad t=1,\ldots,T-1 \tag{36}$$

with the boundary condition unchanged.

To solve the problem, we first need to construct reward functions and transition probability matrices according to the transient model described in Section IV. Then using the backward induction algorithm, we could compare policies by inductively evaluating the cost-to-go function in (36), which yields an optimal Markovian deterministic policy [24].

To illustrate the proposed method, consider a two-machine manufacturing system with configuration of $p_1=0.8324$, $p_2=0.6807$, $N=5$, $T_{\max}=6$, and $\lambda=0.8$. By applying the above procedures, the optimal control policy is generated and shown in Table II. As one can see, the structure of the optimal policy is a lookup table. At the beginning of each cycle, the controller for machine m_1 will determine its functionality. As long as the system reaches $(2,5),(3,4),(3,5),(4,i),i=3,4,5,(5,i),i=4,5$, machine m_1 will be switched off for one cycle. Otherwise, it can still function normally. Such a control policy is easy to implement in production systems with programmable logic controllers embedded.

To illustrate the advantage of employing the proposed real-time control methodology, we use simulation to generate the performance measures with and without implementation of the control policy in Table II, and the results are shown in Fig. 13. The solid blue line represents the case where no control is employed, while the red line with round markers characterizes the control case. The shaded areas stand for the 95% confidence interval. As one could see, SR, CR, and WIP are tremendously reduced after implementing the control policy, almost 70% reduction for SR, 10% for CR, and 40% for WIP. At the same time, the drop for PR is only about 1.5%. In addition, the settling times for performance measures are reduced as well. Therefore, we could conclude that by applying the real-time control policy generated by the proposed methodology, considerable reduction on production loss can be achieved without sacrificing too much of production gain.

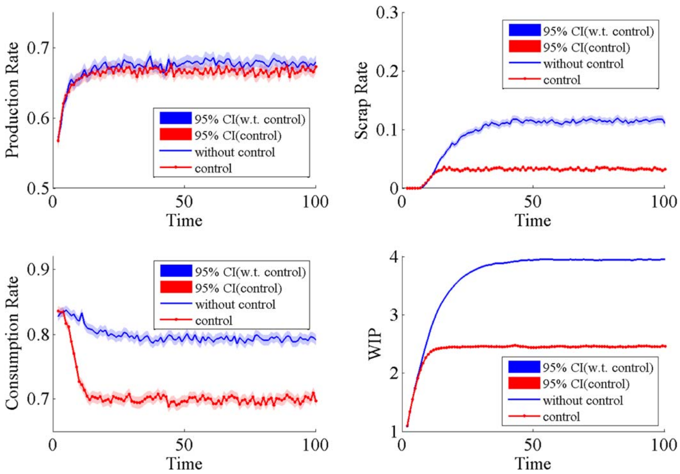


Fig. 13. Comparison of performance measures with and without control.

IX. Conclusion

In this paper, a serial production line model with two Bernoulli reliability machines, a finite buffer, and perishable products is presented to analyze the transient behavior. Analytical formulas to evaluate the transient performance are derived. The structural properties are provided to investigate the effect of system parameters. In addition, we utilize the proposed transient model to address two important problems: settling time estimation and real-time production control. It has been shown that the analytical model is effective in determining the transient period of the system and generating the optimal production control policy. Such a method can provide production engineers and managers a quantitative tool for real-time operation management.

To extend the study, the following topics can be addressed in future work.

- Extend the model to address longer lines and assembly systems.
- Expand the study to more general reliability models (e.g., exponential, Weibull, and general distributions).
- Detailed discussion of the applicability of settling time estimation and the effect of initial conditions will be conducted.
- Incorporate the continuous data thread into real-time production control strategy and investigate the optimal control policy.
- Apply the results on the factory floor to validate the model and improve production operations.

Appendix: Proofs

Proof of Proposition 1: For the simplest case where $N = 2$, $T_{\max} = 2$, to derive the closed form solution, from equation

$$X(t+1) = PX(t)$$

taking the z-transform, we have

$$z^{-1}\left[\chi(z) - X(0)\right] = P \cdot \chi(z)$$
$$\chi(z) = (1 - zP)^{-1} X(0)$$

where $\chi(z) = \mathcal{Z}\{X(t)\}$ is the z-transform of $X(t)$, and $X(0)$ is the initial state vector. From $\chi(z)$, taking the inverse z-transform we can obtain $X(t)$, for $t \geq 0$, which provides the transient probability vector. Specifically, given the transition matrix P, we have

$$(I - zP) = \begin{bmatrix} 1-(1-p_1)z & -(1-p_1)p_2 z & -(1-p_1)z & 0 \\ -p_1 z & 1-p_1p_2 z & -p_1 z & 0 \\ 0 & -(1-p_1)(1-p_2)z & 1 & -(1-p_1)z \\ 0 & -p_1(1-p_2)z & 0 & 1-p_1 z \end{bmatrix}.$$

The closed form expression of $(I - zP)^{-1}$ can be obtained as shown at the bottom of the page. By partial fraction expansion, $(I - zP)^{-1}$ can be expressed as

$$(I - zP)^{-1} = C + \frac{1}{1-z}L + \frac{1}{1-p_1p_2 z}M$$

where C, L, and M are also shown at the top of the next page. Let $H(t)$ be the inverse z-transform of $(I - zP)^{-1}$

$$H(t) = \delta(t)C + L + (p_1p_2)^t S$$

where $\delta(t)$ is a discrete delta function, i.e.,

$$\delta(t) = \begin{cases} 1, & t = 0 \\ 0, & t \neq 0. \end{cases}$$

Therefore, we obtain

$$X(t) = H(t)X(0).$$

■

Note that for a general case, transition matrix P cannot be written in a closed form since $\Phi(\cdot)$ has to be estimated through steady state distribution. In principle, P can be evaluated numerically when $\Phi(\cdot)$ is estimated. However, deriving the inverse and carrying out partial fraction expansion still require substantial efforts.

To prove Theorem 1, the following lemma is needed, which is described in [23] that investigates the convergence speed of

$$(I - zP)^{-1}$$
$$= \begin{bmatrix} \frac{p_1^2z^2-p_1z^2+p_1p_2(z^2-z)-p_1z+1}{p_1p_2z^2-p_1p_2z-z+1} & \frac{p_1^2z^2-2p_1z^2+p_1p_2(z^2-z)-p_2z^2+z^2+p_2z}{p_1p_2z^2-p_1p_2z-z+1} & \frac{p_1^2z^2-p_1z^2-p_1z+z}{p_1p_2z^2-p_1p_2z-z+1} & \frac{p_1^2z^2-2p_1z^2+z^2}{p_1p_2z^2-p_1p_2z-z+1} \\ \frac{p_1z-p_1^2z^2}{p_1p_2z^2-p_1p_2z-z+1} & \frac{-p_1^2z^2+p_1z^2-z+1}{p_1p_2z^2-p_1p_2z-z+1} & \frac{p_1z-p_1^2z^2}{p_1p_2z^2-p_1p_2z-z+1} & \frac{p_1z^2-p_1^2z^2}{p_1p_2z^2-p_1p_2z-z+1} \\ \frac{-p_1^2z^2+p_1z^2+p_1^2p_2z^2-p_1p_2z^2}{p_1p_2z^2-p_1p_2z-z+1} & -\frac{(1-(1-p_1)z)(p_1z-p_1p_2z+p_2z-z)}{p_1p_2z^2-p_1p_2z-z+1} & \frac{-p_1^2z^2+p_1z^2+p_1^2p_2(z^2-z)-z+1}{p_1p_2z^2-p_1p_2z-z+1} & \frac{(1-p_1)z(p_1z-p_1p_2z-z+1)}{p_1p_2z^2-p_1p_2z-z+1} \\ \frac{p_1^2z^2-p_1^2p_2z^2}{p_1p_2z^2-p_1p_2z-z+1} & \frac{p_1(1-p_2)z(p_1z-z+1)}{p_1p_2z^2-p_1p_2z-z+1} & \frac{p_1^2z^2-p_1^2p_2z^2}{p_1p_2z^2-p_1p_2z-z+1} & \frac{p_1^2z^2-p_1z^2-p_1^2p_2z^2+p_1p_2(z^2-z)+p_1z-z+1}{p_1p_2z^2-p_1p_2z-z+1} \end{bmatrix}$$

$$C = \begin{bmatrix} \frac{p_1+p_2-1}{p_2} & \frac{(p_1-1)(p_1+p_2-1)}{p_1p_2} & \frac{p_1-1}{p_2} & \frac{(p_1-1)^2}{p_1p_2} \\ -\frac{p_1}{p_2} & -\frac{p_1-1}{p_2} & -\frac{p_1}{p_2} & \frac{1-p_1}{p_2} \\ \frac{(p_1-1)(p_2-1)}{p_2} & \frac{(p_1-1)^2(p_2-1)}{p_1p_2} & \frac{p_2p_1-p_1+1}{p_2} & \frac{(p_1-1)(p_2p_1-p_1+1)}{p_1p_2} \\ \frac{p_1(1-p_2)}{p_2} & -\frac{(1-p_1)(1-p_2)}{p_2} & \frac{p_1(1-p_2)}{p_2} & -\frac{(1-p_1)(1-p_2)}{p_2} \end{bmatrix}$$

$$L = \begin{bmatrix} \frac{p_1^2-2p_1+1}{1-p_1p_2} & \frac{p_1^2-2p_1+1}{1-p_1p_2} & \frac{p_1^2-2p_1+1}{1-p_1p_2} & \frac{p_1^2-2p_1+1}{1-p_1p_2} \\ \frac{p_1-p_1^2}{1-p_1p_2} & \frac{p_1-p_1^2}{1-p_1p_2} & \frac{p_1-p_1^2}{1-p_1p_2} & \frac{p_1-p_1^2}{1-p_1p_2} \\ \frac{p_2p_1^2-p_1^2-p_2p_1+p_1}{1-p_1p_2} & \frac{p_1(p_2p_1-p_1-p_2+1)}{1-p_1p_2} & \frac{p_2p_1^2-p_1^2-p_2p_1+p_1}{1-p_1p_2} & \frac{p_1(p_2p_1-p_1-p_2+1)}{1-p_1p_2} \\ \frac{p_1^2-p_1^2p_2}{1-p_1p_2} & \frac{p_1^2-p_1^2p_2}{1-p_1p_2} & \frac{p_1^2-p_1^2p_2}{1-p_1p_2} & \frac{p_1^2-p_1^2p_2}{1-p_1p_2} \end{bmatrix}$$

$$M = \begin{bmatrix} \frac{p_2p_1-p_1-p_2+1}{p_2(1-p_1p_2)} & \frac{p_2^2p_1^2-p_1^2-p_2^2p_1-p_2p_1+2p_1+p_2-1}{p_1p_2(1-p_1p_2)} & \frac{p_2p_1-p_1-p_2+1}{p_2(1-p_1p_2)} & \frac{-p_1^2+2p_1-1}{p_1p_2(1-p_1p_2)} \\ \frac{p_1(1-p_2)}{p_2(1-p_1p_2)} & \frac{-p_1p_2^2+p_2+p_1-1}{p_2(1-p_1p_2)} & \frac{p_1(1-p_2)}{p_2(1-p_1p_2)} & \frac{p_1-1}{p_2(1-p_1p_2)} \\ \frac{-p_2p_1+p_1+p_2-1}{p_2(1-p_1p_2)} & \frac{-p_2^2p_1^2+p_1^2+p_2^2p_1+p_2p_1-2p_1-p_2+1}{p_1p_2(1-p_1p_2)} & \frac{-p_2p_1+p_1+p_2-1}{p_2(1-p_1p_2)} & \frac{p_1^2-2p_1+1}{p_1p_2(1-p_1p_2)} \\ \frac{p_1(p_2-1)}{p_2(1-p_1p_2)} & \frac{p_1p_2^2-p_2-p_1+1}{p_2(1-p_1p_2)} & \frac{p_1(p_2-1)}{p_2(1-p_1p_2)} & \frac{1-p_1}{p_2(1-p_1p_2)} \end{bmatrix}$$

non-reversible, aperiodic, and irreducible Markov chain on the finite state space.

Lemma A.1: For a nonreversible, aperiodic and irreducible Markov chain on the finite state space $\mathcal{X}_{S\times 1}$ with transition matrix $P_{S\times S}$, we have

$$d\left[X\left(t_s(\epsilon)|X(0)\right), \widetilde{X}\right] \leq \epsilon, \text{for}$$

$$t_s(\epsilon) = \frac{4}{k^2}\log\left(\frac{1}{2\epsilon\sqrt{\widetilde{x}_i}}\right) \tag{A.1}$$

where i is obtained such that $x_i(0) = 1$ and k is a Cheeger constant defined in (32).

Proof of Theorem 1: Given the transient model and performance measures introduced in Sections IV and V, respectively, the following facts can be obtained:

1) The Markov chain described by (1)–(10) is nonreversible, aperiodic, and irreducible.
2) From (19)–(22), all the performance measures are linear functions of system states.

Therefore, the problem of calculating settling time for the system performance could be transformed into investigating the number of cycles required for the values of system states to converge. The distance between two probability measures $d[X(t|X(0)), \widetilde{X}]$ is given by the *total variation norm*:

$$d\left[X\left(t|X(0)\right), \widetilde{X}\right] = \max_{1\leq i\leq S} |x_i\left(t|X(0)\right) - \widetilde{x}_i|.$$

Then from Lemma A.1 and (23) and (28), we have

$$\begin{aligned}\left|\widehat{PR}\left(t_s^{PR}(\epsilon)\right) - PR(\infty)\right| &\leq p_2\left|x_1\left(t_s^{PR}|X(0)\right) - \widetilde{x}_1\right| \\ &\leq p_2 d\left[X\left(t_s^{PR}(\epsilon)|X(0)\right), \widetilde{X}\right].\end{aligned}$$

Thus, when

$$t_s\left(\frac{\epsilon}{p_2}\right) = \frac{4}{k^2}\log\left(\frac{p_2}{2\epsilon\sqrt{\widetilde{x}_i}}\right)$$

it follows that:

$$\begin{aligned}\left|\widehat{PR}\left(t_s^{PR}(\epsilon)\right) - PR(\infty)\right| &\leq p_2 d\left[X\left(t_s^{PR}(\epsilon)|X(0)\right), \widetilde{X}\right] \\ &\leq \epsilon.\end{aligned}$$

Therefore, $t_s^{PR}(\epsilon) = t_s(\epsilon/p_2)$. Since $t_s^{PR}(\epsilon)$ is a positive integer, we round up the result to be

$$t_s^{PR}(\epsilon) = \left\lceil t_s\left(\frac{\epsilon}{p_2}\right)\right\rceil.$$

Similarly, for t_s^{SR}, according to (25) and (29)

$$\begin{aligned}&\left|\widehat{SR}\left(t_s^{SR}(\epsilon)\right) - SR(\infty)\right| \\ &\leq (1-p_2)\sum_{i=1}^{N}\left|x_{I(i,M-1)}\left(t_s^{SR}(\epsilon)|X(0)\right) - \widetilde{x}_{I(i,M-1)}\right| \\ &\leq N(1-p_2)d\left[X\left(t_s^{SR}(\epsilon)|X(0)\right), \widetilde{X}\right].\end{aligned}$$

By replacing ϵ with $\epsilon/(N(1-p_2))$ in (A.1) and taking the ceiling of the result, it is easy to obtain

$$t_s^{SR}(\epsilon) = \left\lceil t_s\left(\frac{\epsilon}{N(1-p_2)}\right)\right\rceil$$

which concludes the proof. ■

References

[1] J. A. Buzacott and J. G. Shantikumar, *Stochastic Models of Manufacturing Systems*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[2] S. B. Gershwin, *Manufacturing Systems Engineering*. Englewood Cliffs, NJ, USA: PTR Prentice-Hall, 1994.

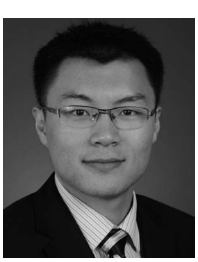
[3] J. Li and S. M. Meerkov, *Production Systems Engineering*. New York, NY, USA: Springer-Verlag, 2009.

[4] G. Liberopoulos, G. Kozanidis, and P. Tsarouhas, "Performance evaluation of an automatic transfer line with WIP scrapping during long failures," *Manuf. Service Operat. Manage.*, vol. 9, no. 1, pp. 62–83, 2007.

[5] S. Nahmias, "Perishable inventory theory: A review," *Operat. Res.*, vol. 30, no. 4, pp. 680–708, 1982.

[6] I. Karaesmen, A. Scheller-Wolf, and B. Deniz, "Managing Perishable and Aging Inventories: Review and Future Research Directions," in *Planning*

Production and Inventories in the Extended Enterprise. New York, NY, USA: Springer-Verlag, 2011, pp. 393–436.
[7] M. Bakker, J. Riezebos, and R. H. Teunter, "Review of inventory systems with deterioration since 2001," *Eur. J. Operat. Res.*, vol. 221, no. 2, pp. 275–284, 2012.
[8] X. Chao, X. Gong, C. Shi, and H. Zhang, "Approximation algorithms for perishable inventory systems," *Operat. Res.*, vol. 63, no. 3, pp. 585–601, 2015.
[9] W. L. Cooper, "Pathwise properties and performance bounds for a perishable inventory system," *Operat. Res.*, vol. 49, no. 3, pp. 455–466, 2001.
[10] Z. Lian and L. Liu, "Continuous review perishable inventory systems: Models and heuristics," *IIE Trans.*, vol. 33, no. 9, pp. 809–822, 2001.
[11] C. L. Williams and B. E. Patuwo, "A perishable inventory model with positive order lead times," *Eur. J. Operat. Res.*, vol. 116, no. 2, pp. 352–373, 1999.
[12] A. Andijani and M. Al-Dajani, "Analysis of deteriorating inventory/production systems using a linear quadratic regulator," *Eur. J. Operat. Res.*, vol. 106, no. 1, pp. 82–89, 1998.
[13] C. Arbib, D. Pacciarelli, and S. Smriglio, "A three-dimensional matching model for perishable production scheduling," *Discrete Appl. Math.*, vol. 92, no. 1, pp. 1–15, 1999.
[14] S. Lo, H. Wee, and W. Huang, "An integrated production-inventory model with imperfect production processes and Weibull distribution deterioration under inflation," *Int. J. Product. Econ.*, vol. 106, no. 1, pp. 248–260, 2007.
[15] J. Pahl and S. Voß, "Discrete lot-sizing and scheduling including deterioration and perishability constraints," *Adv. Manuf. Sustain. Logist.*, pp. 345–357, 2010.
[16] J. G. Shanthikumar and C. C. Tien, "An algorithmic solution to two-stage transfer lines with possible scrapping of units," *Manage. Sci.*, vol. 29, no. 9, pp. 1069–1086, 1983.
[17] M. A. Jafari and J. G. Shanthikumar, "An approximate model of multistage automatic transfer lines with possible scrapping of workpieces," *IIE Trans.*, vol. 19, no. 3, pp. 252–265, 1987.
[18] R. Naebulharam and L. Zhang, "Bernoulli serial lines with deteriorating product quality: Performance evaluation and system-theoretic properties," *Int. J. Product. Res.*, vol. 52, no. 5, pp. 1479–1494, 2014.
[19] Y. Narahari and N. Viswanadham, "Transient analysis of manufacturing systems performance," *IEEE Trans. Robot. Autom.*, vol. 10, no. 2, pp. 230–244, 1994.
[20] S. M. Meerkov and L. Zhang, "Transient behavior of serial production lines with Bernoulli machines," *IIE Trans.*, vol. 40, no. 3, pp. 297–312, 2008.
[21] S. M. Meerkov, N. Shimkin, and L. Zhang, "Transient behavior of two-machine geometric production lines," *IEEE Trans. Autom. Control*, vol. 55, no. 2, pp. 453–458, Feb. 2010.
[22] L. Zhang, C. Wang, J. Arinez, and S. Biller, "Transient analysis of Bernoulli serial lines: Performance evaluation and system-theoretic properties," *IIE Trans.*, vol. 45, no. 5, pp. 528–543, 2013.
[23] J. A. Fill, "Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process," *Ann. Appl. Probab.*, vol. 1, no. 1, pp. 62–87, 1991.
[24] M. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Hoboken, NJ, USA: Wiley, 2005.



Feng Ju (S'11–M'15) received the B.S. degree from the Department of Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China, in 2010 and the M.S. degree in electrical and computer engineering and the Ph.D. degree in industrial and systems engineering from the University of Wisconsin, Madison, WI, USA, in 2011 and 2015, respectively.

He is an Assistant Professor in the School of Computing, Informatics, and Decision Systems Engineering at Arizona State University, Tempe, AZ, USA. His research interests include modeling, analysis, continuous improvement, and optimization of manufacturing systems.

Dr. Ju received multiple awards including the best paper award in IFAC conference on Manufacturing Modeling, Management and Control (MIM). He is a member of the Institute for Operations Research and the Management Sciences (INFORMS) and Institute of Industrial Engineers (IIE).

Jingshan Li (SM'97–M'00–SM'06) received the B.S. degree from the Department of Automation, Tsinghua University, Beijing, China, the M.S. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, and the Ph.D. degree in electrical engineering-systems from University of Michigan, Ann Arbor, MI, USA, in 1989, 1992, and 2000, respectively.

He was a Staff Research Engineer at the Manufacturing Systems Research Laboratory, General Motors Research and Development Center, Warren, MI, USA, from 2000 to 2006, and was with Department of Electrical and Computer Engineering and Center for Manufacturing, University of Kentucky, Lexington, KY, USA, from 2006 to 2010. He is now a Professor in Department of Industrial and Systems Engineering, University of Wisconsin, Madison, WI, USA. His primary research interests are in modeling, analysis and control of manufacturing and healthcare systems.

Dr. Li received the 2010 NSF Career Award, the 2009 IIE Transactions Best Application Paper Award, the 2005 IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING Best Paper Award, the 2006 IEEE Early Industry/Government Career Award in Robotics and Automation, and multiple awards in international conferences. He is a Senior Editor of the IEEE ROBOTICS AND AUTOMATION LETTERS, Department Editor of IIE Transactions and Associate Editor of IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING, *International Journal Of Production Research*, *Flexible Service and Manufacturing*, and *International Journal of Automation Technology*, and was an Associate Editor of *Mathematical Problems in Engineering*.

John A. Horst (M'80) received the B.A. degree in music, the B.S.E.E. degree (cum laude with honors), and the M.S.E.E. degree all from the University of Maryland, College Park, MD, USA, in 1976, 1985, and 1991, respectively.

He is currently with the Engineering Laboratory, National Institute of Standards and Technology, Gaithersburg, MD, USA. He has been conducting research in manufacturing performance measurement to address the real-time control problem at the manufacturing production level. He has focused on Key Performance Indicator (KPI) effectiveness for manufacturing. Several KPI effectiveness methods and models are currently under development. He is also working to define, validate, and verify information exchange standards and their implementations for all interfaces within the manufacturing quality process. NIST facilitates correct, complete, and unambiguous interface standards, enables public interoperability testing events, and helps define and conduct interoperability certifications. This work has culminated in several ANSI standards including the Quality Information Framework (QIF) suite of standards. QIF supplies digital representations for interface information throughout the entire quality measurement process.