Jing Zou
Department of Mechanical Engineering,
Stony Brook University,
156A Light Engineering,
Stony Brook, NY 11794
e-mail: jing.zou@stonybrook.edu

Jorge Arinez
Manufacturing Systems Research Lab,
General Motors Research
and Development Center,
30500 Mound Road,
Warren, MI 48090
e-mail: jorge.arinez@gm.com

Qing Chang[1]
Department of Mechanical Engineering,
Stony Brook University,
163 Light Engineering,
Stony Brook, NY 11794
e-mail: qing.chang@stonybrook.edu

Yong Lei
State Key Laboratory of Fluid Power
Transmission and Control,
Department of Mechanical Engineering,
Zhejiang University,
Hangzhou 310013, China
e-mail: ylei@zju.edu.cn

Opportunity Window for Energy Saving and Maintenance in Stochastic Production Systems

Energy efficiency improvement and timely preventive maintenance (PM) are critical in manufacturing industry due to the rising energy cost, environmental concerns, and increasing requirements on system reliability. By strategically turning appropriate machines in down state, the corresponding energy consumption can be reduced, and at the meantime, the necessary PM works can be carried out to increase PM completion rate and reduce potential extra expense on PM during nonproduction shifts. However, there is usually a tradeoff between time dedicated to production and time available for energy saving and PM. In this paper, a systematic method is developed to identify opportunity windows (OWs) during which certain machines can be shut down to save energy and PM tasks can be performed while maintaining a desired production throughput. The method is based on stochastic serial production lines and real-time production data. A profit function is formulated to illustrate the tradeoff between energy cost savings and potential throughput loss. The profit function is used to justify the cost savings by utilizing the proposed OWs during production operation. [DOI: 10.1115/1.4033757]

Keywords: energy saving opportunity, maintenance opportunity, discrete-time Markov chain, stochastic production systems

1 Introduction

Energy consumption, used to receive relatively little attention, is now an important factor to evaluate a production system. With rising energy costs, increasing global competitiveness, environmental concerns, and more government regulations, energy consumption plays a more critical role in modern manufacturing operation. However, most of the studies in manufacturing system modeling and analysis have mainly focused on improving the production efficiency, flexibility, and responsiveness [1–3]. In most current manufacturing execution systems, there is no module or function to deal with energy management during production operation. Researchers have discovered that 85% of energy consumption in the production plant is used on functions not related to the production of parts [4], and a huge amount of energy is wasted during machines blockages and starvations without production [5]. This emphasizes the importance of identifying energy saving opportunities through the analysis of stochastic production systems.

Another significant factor to be considered in manufacturing systems is PM. For manufacturing systems with unreliable machines and finite internal buffers, PM is critical to reduce machine failures and increase system reliability and efficiency [6]. Usually, PM tasks require the equipment to be stopped in order for work to be safely performed. Consequently, there is a tradeoff between time dedicated to production and time available for PM. This is a particularly important issue in three-shift operations, where any downtime, planned or unplanned, directly impacts the amount of time available for production. In such a situation, planned downtime for maintenance, i.e., PM, must be balanced against the avoidance of unplanned downtime resulting from equipment failures. Some PM tasks are often performed late or not at all, resulting in a long-term decline in equipment reliability and system capability [6]. Even in plants operating on a one-shift or two-shift schedule, where planned downtime is readily available, the cost of staffing nonproduction shifts and/or scheduled overtime to perform PM may be prohibitive [6]. Therefore, it is desirable to find opportunities during the production shifts to do more PM works.

It is noted that to acquire opportunities for energy saving and maintenance, real-time dynamic analysis of manufacturing systems is needed. However, traditional methods in modeling manufacturing systems have focused on steady-state performance analysis. The studies can be grouped into analytical methods and simulation methods. While simulation models are widely adopted to evaluate performance of complex manufacturing systems [7–10], simulation modeling and analysis can be time-consuming and expensive, and simulation results can be difficult to interpret [11]. For analytical methods, decomposition and aggregation methods are utilized with Markov chain models to estimate the important performance metrics [12,13]. For example, Xia et al. [14] developed a decomposition method for the performance evaluation of serial production lines with unreliable machines and finite transfer-delay buffers. These methods are all based on long-term steady-state assumptions and are no longer applicable to the real-time dynamic analysis.

For the past decade, there is an increasing interest in the study of the transient behavior of production systems [15–18]. Chen et al. derived the mathematical model for transient performance evaluation for synchronized serial production lines with geometric machines [16]. Transient system performance, system-theoretic properties, workforce allocation, and bottlenecks of serial production lines with Bernoulli machines and finite buffers are studied in Ref. [18]. However, the method only provides some basic

[1]Corresponding author.

Manuscript received August 17, 2015; final manuscript received May 19, 2016; published online August 5, 2016. Assoc. Editor: Jianjun Shi.

properties such as settling time monotonicity, while the transient throughput and work-in-process are actually derived based on approximation.

The OW concepts and evaluation methods have been developed based on deterministic analysis [5,6,19–21]. The methods bring explicit and systematic solutions to estimate the energy saving potential and maintenance opportunity. However, the deterministic-based analysis only provides a loose bound solution of OW. A real production system is stochastic. Therefore, it is necessary to develop a methodology to evaluate energy saving and maintenance opportunities in stochastic production systems. This paper is dedicated to this end.

The contribution of this research is in twofold. First, an innovative approach is developed to combine the stochastic modeling and data-driven method. Markov property is adopted and real-time data (e.g., machine up–down status and buffers' levels) are used to evaluate system status and OW. This is fundamentally different from traditional stochastic modeling techniques. The advantage of this modeling approach is that it enables a real-time prediction capability to more accurately estimate system performance by utilizing real-time information. Such a method was not explored before. Second, the OW concepts are extended from deterministic analysis to stochastic analysis.

The remainder of this paper is organized as follows. In Sec. 2, we provide manufacturing system assumptions and notations. Section 3 estimates the status of a discrete-time stochastic production system and evaluates the corresponding OW and recovery time. Case studies are given in Sec. 4 to validate the performance of the estimation and evaluation results. Conclusions and future work are summarized in Sec. 5.

2 Assumptions and Notations

We consider the discrete-time (or discrete event system) continuous flow models to analyze the impact of OW on the production process. The continuous flow model is used in this paper because the production dynamics can be conveniently described by integral or differential equations [13,19]. The continuous flow model assumes that the quantity of jobs in the buffer varies continuously from zero to its capacity. For a serial production line with M machines (represented by the rectangles) and $M-1$ buffers (represented by the circles) as shown in Fig. 1, the following notations are used in this paper:

- M_i denotes the ith machine, where $1 \leq i \leq M$.
- B_i denotes the ith buffer; each buffer $B_2,\ B_3, \ldots, B_M$ has a finite capacity. With abuse of notation, $B_2,\ B_3, \ldots, B_M$ are also used to denote the capacity of the buffer.
- $b_i(n)$ denotes buffer level of $B_i,\ i=2,\ \ldots,\ M$, at the nth unit cycle.
- $s_i(n)$ denotes actual processing speed measured in unit cycles for machine M_i at the nth unit cycle.
- The production process is described in the discrete-time model with unit cycle duration as Δt. Each machine M_i has a constant rated speed $1/(T_i \Delta t),\ i=1,\ 2,\ \ldots,\ M$, where $T_i \Delta t$ is the cycle time for machine M_i, and T_i denotes the number of unit cycles in one work cycle of machine M_i. Therefore, in the discrete-time model, the rated speed measured in unit cycles for machine M_i is $1/T_i$. An operational machine runs at its rated speed when it is neither starved nor blocked.
- $M_{M_k^*}$ denotes the slowest machines, i.e., $M_k^* = \arg\max (T_i)_{i=1,\ \ldots,\ M},\ 1 \leq k \leq M$.

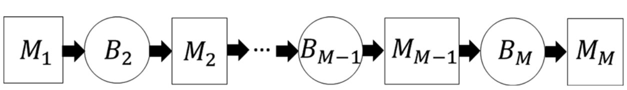

Fig. 1 A serial production line with M machines and $M-1$ buffers

- M_{M^*} denotes the slowest machine that is closest to the end-of-line machine S_M, since there might be one or multiple slowest machines in a system as described in Eq. (3). It is also denoted as the last slowest machine.
- $\vec{e}_i = (j, n_i, d_i),\ i=1,\ 2,\ \ldots,\ j=1,\ 2,\ \ldots,\ M$, denotes a downtime event that machine M_j is down at the n_ith unit cycle for d_i unit cycles.
- $\boldsymbol{E} = \left\{\vec{e}_1, \vec{e}_2, \ldots,\ \vec{e}_k\right\}$ denotes a sequence of downtime events of the line.
- PL denotes permanent production loss of the whole line.
- $pi_{10},\ i=1,\ 2,\ \ldots,\ M$ denotes the transition probability that machine M_i transit from up to down state at the nth unit cycle.
- $pi_{01},\ i=1,\ 2,\ \ldots,\ M$ denotes the transition probability that machine M_i transit from down to up state at the nth unit cycle.
- $Pi_1(n),\ i=1,\ 2,\ \ldots,\ M,\ n=1,\ 2,\ 3,\ \ldots,$ denotes the probability that machine M_i is up at the nth unit cycle.
- $Pi_0(n),\ i=1,\ 2,\ \ldots,\ M, n=1,\ 2,\ 3,\ \ldots,$ denotes the probability that machine M_i is down at the nth unit cycle.
- $\theta_i(n),\ i=1,\ 2,\ \ldots,\ M,\ n=1,\ 2,\ 3,\ \ldots,$ denotes the processing probability of machine M_i at the nth unit cycle.

We make the following assumptions:

(1) A machine is blocked if it is up and its downstream buffer is full. The last machine is never blocked.
(2) A machine is starved if it is up and its upstream buffer is empty. The first machine is never starved.
(3) Machine failures are time-dependent. This means that the machine breakdowns may occur even while it is blocked or starved.
(4) MCBF_i and MCTR_i denote the mean-cycle-between-failure and the mean-cycle-to-repair of machine M_i, respectively. They are assumed to be geometric random variables [22]. Let $\lambda_i = 1/\mathrm{MCBF}_i$ and $\mu_i = 1/\mathrm{MCTR}_i$ [12].
(5) The power consumption of machine M_i is reduced to a certain level if it is turned off.
(6) Each machine will run at its power rating P_m when up and will consume no power when turned off.
(7) The machine warm up time is not considered for the ease of mathematical expression.

3 Discrete-Time Markov Chain Stochastic Serial Production Line Model

The relationships between pi_{10}, pi_{01} and $\lambda_i, \mu_i, i=1,\ 2,\ \ldots, M$ are shown in the following equations based on assumption (4):

$$pi_{10} = \frac{1}{\mathrm{MCBF}_i} = \lambda_i \tag{1}$$

$$pi_{01} = \frac{1}{\mathrm{MCTR}_i} = \mu_i \tag{2}$$

We will first consider a stochastic two-machine-one-buffer system. Then, we extend the analysis to the general stochastic serial production systems.

3.1 Stochastic Two-Machine-One-Buffer System. For a two-machine-one-buffer stochastic system, let $P1(n) = \begin{bmatrix} P1_1(n) \\ P1_0(n) \end{bmatrix}$ and $P2(n) = \begin{bmatrix} P2_1(n) \\ P2_0(n) \end{bmatrix}$ denote the probability distribution for the states of M_1 and M_2 at the nth unit cycle, respectively. The system possesses Markov property due to geometric reliability assumption [13,23–27]. Therefore, the relationship between $Pi(n)$ and $Pi(n+1)$ for M_1 and M_2 is

$$P1(n+1) = \begin{bmatrix} 1-p1_{10} & p1_{01} \\ p1_{10} & 1-p1_{01} \end{bmatrix} \begin{bmatrix} P1_1(n) \\ P1_0(n) \end{bmatrix} = A_1 P1(n) \quad (3)$$

$$P2(n+1) = \begin{bmatrix} 1-p2_{10} & p2_{01} \\ p2_{10} & 1-p2_{01} \end{bmatrix} \begin{bmatrix} P2_1(n) \\ P2_0(n) \end{bmatrix} = A_2 P2(n) \quad (4)$$

where $A_1 = \begin{bmatrix} 1-p1_{10} & p1_{01} \\ p1_{10} & 1-p1_{01} \end{bmatrix}$ and $A_2 = \begin{bmatrix} 1-p2_{10} & p2_{01} \\ p2_{10} & 1-p2_{01} \end{bmatrix}$ are the probability transition matrices for M_1 and M_2.

We denote $u_i(n)$ as the control input at M_i at the nth unit cycle. In this paper, a control of switching on or off a machine is adopted. $u_i(n)$ is denoted by the binary variable as

$$u_i(t) = \begin{cases} 0, & \text{turn off } M_i \text{ at the } n\text{th unit cycle} \\ 1, & \text{turn on } M_i \text{ at the } n\text{th unit cycle} \end{cases}$$

The processing probabilities of M_1 and M_2 at the nth unit cycle are defined as

$$\theta_1(n) = \begin{cases} P1_1(n), & b_2(n-1) < B_2 \text{ and } u_1(n) = 1 \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

$$\theta_2(n) = \begin{cases} P2_1(n), & b_2(n-1) > 0 \text{ and } u_2(n) = 1 \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

When M_i is not being blocked or starved at the nth unit cycle and $u_i(n) = 1$, its processing probability equals the probability that it is up. Otherwise, the processing probability is zero.

Based on the conservation of flow, at the nth unit cycle, we can estimate the value of $b_2(n+1)$ by its expected value as

$$b_2(n+1) = b_2(n) + \frac{\theta_1(n)}{T_1} - \frac{\theta_2(n)}{T_2} \quad (7)$$

Iteratively applying Eqs. (5)–(7), we can estimate the approximate buffer level in future time. At the n_1th unit cycle, $\forall n_2 > n_1$, the estimated value of $b_2(n_2)$ is

$$b_2(n_2) = b_2(n_1) + \sum_{k=n_1}^{n_2-1} \frac{\theta_1(k)}{T_1} - \sum_{k=n_1}^{n_2-1} \frac{\theta_2(k)}{T_2} \quad (8)$$

3.2 Stochastic OW and Recovery Time. The OW of machine M_i is defined as the longest possible downtime of M_i that does not result in permanent production loss at the end-of-line machine [6,12,13].

Definition 1. *The OW $W_i(n_d)$ for machine M_i at the n_dth unit cycle is*

$$W_i(n_d) = \sup\left\{ d \ge 0\text{: s.t. } \exists N^*(d),\ \sum_{n=1}^{N} s_M(n) = \sum_{n=1}^{N} \tilde{s}_M(n; \vec{e}),\ \forall N \ge N^*(d) \right\}$$

where $\sum_{n=1}^{N} s_M(n)$ and $\sum_{n=1}^{N} \tilde{s}_M(n;\vec{e})$ are the production counts of the end-of-line station M_M within $[1, N]$, with and without inserted downtime event $\vec{e} = (i, n_d, d)$, respectively. $N^*(d)$ signifies the potential dependency of N^* on d.

The definition only requires that the production count trajectory of the end-of-line machine can eventually recover as if there were no inserted downtime event $\vec{e}$. For the two-machine-one-buffer line, M_1 will not be blocked if $b_2(n) < B_2$, and M_2 will not be starved if $b_2(n) > 0$. The OW of M_i equals the amount of time it takes for the buffers between M_i and the slowest machine to become empty (if M_i is upstream of the slowest machine) or full (if M_i is downstream of the slowest machine). T_1 and T_2 are the cycle times of M_1 and M_2 measured in unit cycles. Therefore, we have

$$W_1(n) = \begin{cases} T_2 * b_2(n), & \text{if } T_1 < T_2 \\ 0, & \text{if } T_1 > T_2 \end{cases}$$

$$W_2(n) = \begin{cases} T_1 * (B_2 - b_2(n)), & \text{if } T_1 > T_2 \\ 0, & \text{if } T_1 < T_2 \end{cases}$$

The calculation methods for OW in Refs. [5], [19], and [20] are designed for systems under deterministic scenario and may not be suitable in stochastic scenario. However, Definition 1, i.e., the definition of OW, is still applicable for the stochastic system. Let $WS_i(n)$ denote the OW of M_i in stochastic scenario at the nth unit cycle. Based on the estimation results from Sec. 3.1 and Definition 1, $WS_i(n)$ can be evaluated as

$$WS_1(n) = \begin{cases} \inf\{WS \ge 1\text{: s.t. } b_2(n+WS) = 0\}, & 1/T_1 > 1/T_2 \\ 0, & 1/T_1 < 1/T_2 \end{cases} \quad (9)$$

$$WS_2(n) = \begin{cases} \inf\{WS \ge 1\text{: s.t. } b_2(n+WS) = B_2\}, & 1/T_1 < 1/T_2 \\ 0, & 1/T_1 > 1/T_2 \end{cases} \quad (10)$$

where $b_2(n+WS)$ is estimated using Eqs. (5)–(8).

Another quantity of practical interest is how long it takes for the system to recover after a downtime event [19,28,29]. We can define the recovery time of each machine in the discrete-time case as follows.

Definition 2. *The recovery time of machine M_i after an isolated downtime event $\vec{e} = (i, T_d, d)$ at M_i is*

$$T_r(i, \vec{e}) = \inf\left\{ \Delta N \ge 0\text{: s.t. } \sum_{k=1}^{N} s_i(k) = \sum_{k=1}^{N} \tilde{s}_i(k, \vec{e}),\ \forall N \ge \Delta N + T_d + d \right\}$$

where $\sum_{k=1}^{N} s_i(k)$ and $\sum_{k=1}^{N} \tilde{s}_i(k, \vec{e})$ are the production counts of M_i within $[1, N]$, with and without downtime event $\vec{e} = (i, T_d, d)$.

Since the performance measures of the end-of-line machine are commonly considered as performance measures of the whole line, it is reasonable to consider the recovery time of the whole line as the time it takes for the end-of-line station to recover.

Definition 3. *The recovery time of the whole line after an isolated downtime event $\vec{e} = (i, T_d, d)$ at machine M_i is*

$$T_r(\vec{e}) = T_r(M, \vec{e})$$

where $T_r(M, \vec{e})$ is the recovery time of the end-of-line machine M_M after the downtime event $\vec{e} = (i, T_d, d)$.

The recovery time for M_1 and M_2 in the deterministic two-machine-one-buffer discrete-time system after downtime event $\vec{e} = (j, T_d, d)$ is the time for the buffers between M_i and the slowest machine M_{M^*} to become full (if M_i is an upstream machine of M_{M^*}) or empty (if M_i is a downstream machine of M_{M^*}). Then, for $1/T_1 > 1/T_2$, the recovery time T_r for M_1 after $\vec{e} = (1, T_d, d)$ is

$$T_r(1, (1, T_d, d)) = \begin{cases} \dfrac{B_2 - b_2(T_d) + d/T_2}{1/T_1 - 1/T_2}, & d \le W_1(T_d) \\ \infty, & d \ge W_1(T_d) \end{cases}$$

and the recovery time for M_2 after $\vec{e} = (1, T_d, d)$ is

$$T_r(2,(1,T_d,d)) = \begin{cases} 0, & d \le W_1(T_d) \\ \infty, & d \ge W_1(T_d) \end{cases}$$

For $1/T_1 < 1/T_2$, the recovery time for M_2 after $\vec{e} = (2, T_d, d)$ is

$$T_r(2,(2,T_d,d)) = \begin{cases} \dfrac{b_2(T_d) + d/T_1}{1/T_2 - 1/T_1}, & d \le W_2(T_d) \\ \infty, & d \ge W_2(T_d) \end{cases}$$

and the recovery time for M_1 after $\vec{e} = (2, T_d, d)$ is

$$T_r(1,(2,T_d,d)) = \begin{cases} 0, & d \le W_2(T_d) \\ \infty, & d \ge W_2(T_d) \end{cases}$$

where $W_1(T_d)$ and $W_2(T_d)$ denote the OWs in deterministic analysis for M_1 and M_2, respectively.

In stochastic systems, the above calculation needs to be revised. For a two-machine-one-buffer production system, let $T_{\rm re}(i, (i, T_d, d))$ denote the recovery time for M_i after an inserted downtime event $\vec{e} = (i, T_d, d)$ in stochastic scenario, by definition we have

For $1/T_1 > 1/T_2$

$$T_{\rm re}(1,(1,T_d,d)) = \begin{cases} \inf\{T \ge d: \ \text{s.t. } b_2(T_d+T) = b_2(T_d)\}, \text{for } d \le WS_1(T_d) \\ \infty, \text{for } d > WS_1(T_d) \end{cases}$$

$$T_{\rm re}(2,(1,T_d,d)) = \infty$$

For $1/T_1 < 1/T_2$

$$T_{\rm re}(2,(2,T_d,d)) = \begin{cases} \inf\{T \ge d: \text{s.t. } b_2(T_d+T) = b_2(T_d)\}, \text{for } d \le WS_2(T_d) \\ \infty, \text{for } d > WS_2(T_d) \end{cases}$$

$$T_{\rm re}(1,(2,T_d,d)) = \infty$$

3.3 Stochastic Serial Production Line Model With Multimachine Multibuffer. We can extend the analysis on the two-machine-one-buffer line to the multimachine multibuffer production line. Again $Pi(n) = \begin{bmatrix} Pi_1(n) \\ Pi_0(n) \end{bmatrix}$ denotes the probability distribution for the states of M_i at the nth unit cycle, and the relationship between $Pi(n)$ and $Pi(n+1)$ is

$$Pi(n+1) = \begin{bmatrix} 1 - pi_{10} & pi_{01} \\ pi_{10} & 1 - pi_{01} \end{bmatrix} \begin{bmatrix} Pi_1(n) \\ Pi_0(n) \end{bmatrix} = A_i Pi(n) \quad (13)$$

Similar to the analysis for Eqs. (5) and (6) in Sec. 3.1, we have the processing probability of $M_i, i = 1, \ldots, M$ as follows:

For $i = 1$

$$\theta_1(n) = \begin{cases} P1_1(n), & b_2(n-1) < B_2 \text{ and } u_1(n) = 1 \\ 0, & \text{otherwise} \end{cases} \quad (14)$$

For $i = M$

$$\theta_M(n) = \begin{cases} PM_1(n), & b_M(n-1) > 0 \text{ and } u_M(n) = 1 \\ 0, & \text{otherwise} \end{cases} \quad (15)$$

For $2 \le i \le M-1$

$$\theta_i(n) = \begin{cases} Pi_1(n), b_{i+1}(n-1) < B_{i+1} \text{ and } b_i(n-1) > 0 \text{ and } u_i(n) = 1 \\ 0, \quad \text{otherwise} \end{cases} \quad (16)$$

At the n_1th unit cycle, $\forall n_2 > n_1$, the estimated value of $b_i(n_2)$, $i = 2, \ldots, M$ is

$$b_i(n_2) = b_i(n_1) + \sum_{k=n_1}^{n_2-1} \frac{\theta_{i-1}(k)}{T_{i-1}} + \sum_{k=n_1}^{n_2-1} \frac{\theta_i(k)}{T_i} \quad (17)$$

The relationship between the production loss of the entire production line and the production loss of the slowest machine for the discrete-time model is presented in Lemma 1.

Lemma 1. *Let machine M_{M^*} be the unique slowest station in a serial transfer line consisting of M machines, as shown in Fig. 1. Suppose a single isolated downtime event $\vec{e}' = (j, n'_d, d')$ at M_j leads to stoppage event $\vec{e} = (M^*, n_d, d)$ at M_{M^*}. Then for any given machine M_i in the line, $\exists n^* \ge n_d + d$, which may depend on the location of M_i such that [19]*

$$\sum_{n=1}^{N} s_i(n) - \sum_{n=1}^{N} \tilde{s}_i(t; \vec{e}') = d/T_{M^*}, \ \forall N > n^*$$

where $\sum_{n=1}^{N} s_i(n)$ and $\sum_{n=1}^{N} \tilde{s}_i(t; \vec{e}')$ are the production counts of M_i within $[1, N]$, with and without inserted downtime event $\vec{e}' = (j, n'_d, d')$.

Proof. Due to the fact that the slowest machine can only operate at two different speed, we have $\tilde{s}_{M^*}(n; \vec{e}') = 1/T_{M^*}, \ n \notin [n_d, \ n_d + d - 1]$ and $\tilde{s}_{M^*}(n; \vec{e}') = 0, \ n \in [n_d, \ n_d + d - 1]$. Clearly, $n'_d + d' = n_d + d$ since the stoppage event $(M^*, \ n_d, \ d)$ results from $\vec{e}' = (j, \ n'_d, d')$. If the slowest machine is M_i, i.e., $M^* = i$, let $n^* = n_d + d$, then $\forall N > n^*$

$$\begin{aligned} \sum_{n=1}^{N} s_{M^*}(n) - \sum_{n=1}^{N} \tilde{s}_{M^*}(n; \vec{e}) &= \sum_{n=1}^{n_d-1} \left(s_{M^*}(n) - \tilde{s}_{M^*}(n; \vec{e}) \right) \\ &+ \sum_{n=n_d}^{n_d+d-1} \left(s_{M^*}(n) - \tilde{s}_{M^*}(n; \vec{e}) \right) \\ &+ \sum_{n=n_d+d}^{N} \left(s_{M^*}(n) - \tilde{s}_{M^*}(n; \vec{e}) \right) = d/T_{M^*} \end{aligned}$$

In the case when $i > M^*$, given the line segment from the slowest machine M_{M^*} to M_i, by conservation of flow, we have

$$\sum_{n=1}^{N} s_i(n) = \sum_{n=1}^{N} s_{M^*}(n) - \sum_{k=M^*+1}^{i} [b_k(N) - b_k(1)]$$

$$\sum_{n=1}^{N} \tilde{s}_i(n;\vec{e}') = \sum_{n=1}^{N} \tilde{s}_{M^*}(n;\vec{e}') - \sum_{k=M^*+1}^{i} [b_k(N;\vec{e}') - b_k(1;\vec{e}')]$$

Because the slowest machine has the slowest rated processing speed and is unique per our assumption, we have $1/T_{M^*} < \min\{1/T_{M^*+1}, \ldots, 1/T_M\}$. When there is no inserted downtime event, the total buffer level between M_{M^*} and M_i decreases at a rate at least as fast as $\min\{1/T_{M^*+1}, \ldots, 1/T_M\} - 1/T_{M^*}$. Therefore, if we choose $n_1^* = \sum_{k=M^*+1}^{i} b_k(1)/[\min\{1/T_{M^*+1}, \ldots, 1/T_M\} - 1/T_{M^*}]$, we will have

$$\sum_{n=1}^{N} s_i(n) = \sum_{n=1}^{N} s_{M^*}(n) + \sum_{k=M^*+1}^{i} b_k(1), \ \forall N > n_1^*$$

Similarly, when there is an inserted downtime event $\vec{e}'$, by choosing $n_2^* = n_d + d - 1 + \sum_{k=M^*+1}^{i} b \ \ k(n_d + d - 1;\vec{e}')/[\min\{1/T_{M^*+1}, \ldots, 1/T_M\} - 1/T_{M^*}]$, we have

$$\sum_{n=1}^{N} \tilde{s}_i(n;\vec{e}') = \sum_{n=1}^{N} \tilde{s}_{M^*}(n;\vec{e}') + \sum_{k=M^*+1}^{i} b_k(1;\vec{e}'), \ \forall N > n_2^*$$

The system starts from exactly the same initial conditions, i.e., $b_k(1) = b_k(1;\vec{e}'), \ \forall k = M^* + 1, \ldots, M. \ \ \forall N > n^* = \max\{n_1^*, n_2^*\}$, we have

$$\begin{aligned}\sum_{n=1}^{N} s_i(n) - \sum_{n=1}^{N} \tilde{s}_i(n;\vec{e}) &= \sum_{n=1}^{n_d-1} \left(s_{M^*}(n) - \tilde{s}_{M^*}(n;\vec{e})\right) \\ &+ \sum_{n=n_d}^{n_d+d-1} \left(s_{M^*}(n) - \tilde{s}_{M^*}(n;\vec{e})\right) \\ &+ \sum_{n=n_d+d}^{N} \left(s_{M^*}(n) - \tilde{s}_{M^*}(n;\vec{e})\right) = d/T_{M^*}\end{aligned}$$

In the case when $m < M^*$, considering the line segment between machine M_i and the slowest machine M_{M^*}, by choosing

$$n^* = \max\left\{ \frac{\sum_{k=i+1}^{M^*} [B_k - b_k(1)]}{[\min\{1/T_{M^*+1}, \ldots, 1/T_M\} - 1/T_{M^*}]}, \frac{\sum_{k=i+1}^{M^*} [B_k - b_k(n_d + d - 1;\vec{e}')]}{[\min\{1/T_{M^*+1}, \ldots, 1/T_M\} - 1/T_{M^*}]} \right\}$$

we can show that $\forall N > n^*$

$$\sum_{n=1}^{N} s_i(n) = \sum_{n=1}^{N} s_{M^*}(n) + \sum_{k=i+1}^{M^*} [B_k - b_k(1)]$$

$$\sum_{n=1}^{N} \tilde{s}_i(t;\vec{e}') = \sum_{n=1}^{N} \tilde{s}_{M^*}(n;\vec{e}) + \sum_{k=i+1}^{M^*} [B_k - b_k(1;\vec{e}')]$$

Taking the difference between the two equations yields

$$\sum_{n=1}^{N} s_i(n) - \sum_{n=1}^{N} \tilde{s}_i(t;\vec{e}') = d/T_{M^*}$$

By definition, the OW for machine M_i in the discrete-time model is the time for the buffers between M_i and the slowest machine to become empty (if M_i is upstream of the slowest machine) or full (if M_i is downstream of the slowest machine), which can be calculated in deterministic scenario as [5]

$$W_i(n) = \begin{cases} T_{M^*} \sum_{k=i+1}^{M^*} b_k(n), i < M^* \\ 0, \ \ i = M^* \\ T_{M^*} \sum_{k=M^*+1}^{i} (B_k - b_k(n)), i > M^* \end{cases} \quad (18)$$

Similar to the analysis for the two-machine-one-buffer system case, Definitions 1 and 2 are still applicable for the general stochastic systems. Let $WS_i(n)$ denote the OW for M_i at the nth unit cycle in stochastic scenario, we have the following theorem:

THEOREM 1. *Let machine M_{M^*} be the unique slowest machine in a serial transfer line consisting of M machines, as shown in Fig. 1, M_i stop from the nth unit cycle, and all other machines operate without inserted downtime; the OW of M_i at the nth unit cycle is*

$$WS_i(n) = \begin{cases} \inf\left\{ WS \geq 0 : \text{s.t.} \sum_{k=i+1}^{M^*} b_k(n+WS) = 0 \right\}, & i < M^* \\ 0, & i = M^* \\ \inf\left\{ WS \geq 0 : \text{s.t.} \sum_{k=M^*+1}^{i} (B_k - b_k(n+WS)) = 0 \right\}, & i > M^* \end{cases} \quad (19)$$

Proof. The case $i = M^*$ is proved by contradiction. Suppose the OW of the slowest machine M_{M^*} at the nth unit cycle is not zero, i.e., $WS_{M^*}(n) > 0$. Then, for any downtime event (M^*, n, d) with $0 < d \leq WS_{M^*}(n)$, the difference between the undisturbed and disturbed production count trajectories of the end-of-line machine M_M is nonzero, which contradicts the definition of OW.

We will prove the case when $i < M^*$. An equivalent condition of Definition 1 requires that the duration of the stoppage event at the slowest machine M_{M^*} be zero since any nontrivial stoppage event (M^*, n_s, s) resulted from an inserted downtime event $\vec{e} = (i, n, d)$ at another machine M_i eventually leads to nontrivial discrepancy between $\sum_{k=1}^{N} s_M(k)$ and $\sum_{k=1}^{N} \tilde{s}_M(k;\vec{e}')$.

Considering the line segment between M_i and M_{M^*}, as shown in Fig. 2, we insert a downtime event $\vec{e} = (i, n, d)$ at M_i with duration d at the nth unit cycle. Immediately after M_i is down, there is no flow into this line segment until the $(n + d)$th unit cycle. Applying the conservation of flow during the interval $[n, \ n + d - 1]$ yields

$$\begin{aligned}\sum_{k=n}^{n+d-1} \tilde{s}_i(k;\vec{e}) - \sum_{k=n}^{n+d-1} \tilde{s}_{M^*}(k;\vec{e}) &= \sum_{m=i+1}^{M^*} b_m(n + d - 1;\vec{e}) \\ &- \sum_{m=i+1}^{M^*} b_m(n;\vec{e})\end{aligned}$$

where $s_i(k;\vec{e}) = 0, \ k \in [n, \ n + d - 1]$. In the case when $i < M^*$, the slowest machine will not be starved by M_i until $\sum_{m=i+1}^{M^*} b_m(n;\vec{e})$ becomes zero. In the stochastic model, the buffer content between M_i and the slowest machine M_{M^*} will be gradually drained. Therefore, the number of unit cycles it takes for all the buffers between M_i and the slowest machine M_{M^*} to become empty is $d^* = \inf\left\{WS \geq 0\text{: s.t. } \sum_{k=i+1}^{M^*} b_k(n+WS) = 0\right\}$. If the downtime duration d is greater than d^*, the slowest machine M_{M^*} will be starved by M_i. In light of Lemma 1, this will eventually lead to a permanent production loss between $\sum_{k=1}^{N} s_M(k)$ and $\sum_{k=1}^{N} \tilde{s}_M\left(k;\vec{e}'\right)$. Therefore, in the case when $i < M^*$, the OW of M_i at the nth unit cycle is $WS_i(n) = d^*$. Analogously, one can also prove the case when $i > M^*$.

The upper bound of recovery time for the deterministic serial production line can be derived based on Definition 2 as

$$T_{ir}(i,\ (i,n,d)) = \begin{cases} \dfrac{\sum_{k=i+1}^{M^*}[B_k - b_k(n)] + \dfrac{d}{T_{M^*}}}{\min\left\{\dfrac{1}{T_1},\dots,\dfrac{1}{T_{M^*}}\right\} - \dfrac{1}{T_{M^*}}}, & i < M^* \\ \infty, & i = M^* \\ \dfrac{\sum_{k=M^*+1}^{i} b_k(n) + \dfrac{d}{T_{M^*}}}{\min\left\{\dfrac{1}{T_{M^*+1}},\dots,\dfrac{1}{T_M}\right\} - \dfrac{1}{T_{M^*}}}, & i > M^* \end{cases} \tag{20}$$

For the convenience of evaluating the upper bound of recovery time in the stochastic scenario, we first introduce Lemmas 2 and 3.

Lemma 2. *Let machine M_{M^*} be the unique slowest machine in a serial transfer line consisting of M machines as shown in Fig. 1, and then when M_i and M_{M^*} are up at the nth unit cycle*

(1) *for $i > M^*$, if $s_i(n) = s_{M^*}(n) = 1/T_{M^*}$, then all the buffers between M_{M^*} and M_i are empty, i.e., $\sum_{j=M^*+1}^{i} b_j(n) = 0$*
(2) *for $i < M^*$, if $s_i(n) = s_{M^*}(n) = 1/T_{M^*}$, then all the buffers between the M_{M^*} and M_i are full, i.e., $\sum_{j=i+1}^{M^*} b_j(n) = \sum_{j=i+1}^{M^*} B_j$*

Proof. (1) For $i > M^*$. If $s_i(n) = s_{M^*}(n) = 1/T_{M^*} = \min\{1/T_m, m = 1,\dots,M\}$, machine M_i is partially starved by M_{M^*}. This means that all the buffers between M_{M^*} and M_i are empty, i.e., $\sum_{j=M^*+1}^{i} b_j(n) = 0$. Similarly one can also prove case (2).

Lemma 3. *Let machine M_{M^*} be the unique slowest machine in a serial transfer line consisting of M machines as shown in Fig. 1. The recovery time of machine $M_i, i \neq M^*$ after a single isolated downtime event $\vec{e} = (i,\ T_d, d)$ with $d \leq WS_i(T_d)$ is bounded from above by the amount of time for all buffers between M_{M^*} and M_i to become empty (when $k > M^*$) or full (when $k < M^*$).*

Proof. In the case when $k > M^*$, for a single isolated downtime event $\vec{e} = (i,\ T_d, d)$ with $d \leq WS_i(T_d)$, consider the following sets:

$$A = \left\{\Delta N \geq 0\text{: s.t., } \sum_{k=1}^{N} s_i(k) = \sum_{k=1}^{N} \tilde{s}_i\left(k;\vec{e}\right), \forall N \geq \Delta N + T_d + d\right\}$$

$$B = \left\{\Delta N \geq 0\text{: s.t., } s_i(N) = \tilde{s}_i\left(N;\vec{e}\right), \forall N \geq \Delta N + T_d + d\right\}$$

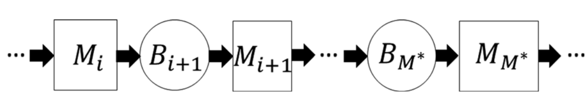

Fig. 2 The line segment between the machines M_i and M^* when $i < M^*$

$$C = \left\{\Delta N \geq 0\text{: s.t., } s_i(N) = s_{M^*}(N), \tilde{s}_{M^*}\left(N;\vec{e}\right) = \tilde{s}_i\left(N;\vec{e}\right), \forall N \geq \Delta N + T_d + d\right\}$$

$$D = \left\{\Delta N \geq 0\text{: s.t., } \sum_{j=M^*+1}^{i} b_j(N) = 0, \forall N \geq \Delta N + T_d + d\right\}$$

Since $d \leq WS_i(T_d)$, from the proof of Theorem 1, we know that the equality $s_{M^*}(N) = \tilde{s}_{M^*}\left(N;\vec{e}\right)$ is always true. Therefore, we have $A \subseteq B \subseteq C$. And Lemma 2 indicates that $C \subseteq D$. Together with the Definition 2, we have

$$T_r\left(i,\vec{e}\right) = \sup\{A\} \leq \sup\{B\} \leq \sup\{C\} \leq \sup\{D\}$$

This completes the proof for the case when $k > M^*$. Similarly, one can also prove the case when $k < M^*$.

Let $T_{\text{re}}(i,\ (i,\ n,\ d))$ denote the recovery time in stochastic scenario for M_i after an inserted downtime event $\vec{e} = (i,\ n,\ d)$, we have the following theorem:

Theorem 2. *Let machine M_{M^*} be the unique slowest machine in a serial transfer line consisting of M machines, as shown in Fig. 1, and let $\vec{e} = (i,n,d), d \leq WS_i(n)$ be the only inserted downtime; the upper bound of recovery time for M_i is*

$$T_{\text{re}}(i,(i,n,d)) = \begin{cases} \inf\left\{T \geq d\text{: s.t.} \sum_{j=i+1}^{M^*}\left[B_j - b_j(n+T)\right] = 0\right\}, & i < M^* \\ \infty, & i = M^* \\ \inf\left\{T \geq d\text{: s.t.} \sum_{j=M^*+1}^{i} b_j(n+T) = 0\right\}, & i > M^* \end{cases} \tag{21}$$

Proof. As discussed previously, if the downtime event is larger than the OW of any individual machine, then the recovery time is infinite. Similar to the analysis on OW, the recovery time of each machine depends on the location of M_i in relation to the slowest machine in the line. We start with $i = M^*$. Based on the statement in Theorem 2, we have $d \leq WS_{M^*}(n) = 0$, namely, no downtime is allowed to insert on M_{M^*}. Therefore, the recovery time $T_{\text{re}}(i,(i,n,d)) = \infty$.

For the case $i > M^*$, by the conservation of flow, we have

$$\sum_{k=n+d}^{N} \tilde{s}_i\left(k;\vec{e}\right) = \sum_{k=n+d}^{N} \tilde{s}_{M^*}\left(k;\vec{e}\right) - \sum_{k=M^*+1}^{i}\left(b_k\left(N;\vec{e}\right) - b_k\left(n+d;\vec{e}\right)\right)$$

From the above equation, we get

$$N^* = \inf\left\{N \geq d\text{: s.t.} \sum_{k=n+d}^{n+d+N}\left[\tilde{s}_i\left(k;\vec{e}\right) - \tilde{s}_{M^*}\left(k;\vec{e}\right)\right] = \sum_{j=M^*+1}^{i} b_j\left(n+d;\vec{e}\right)\right\}$$

which gives the amount of unit cycles required for all buffers between M_{M^*} and M_i to first become empty.

According to Lemma 3, N^* is an upper bound for the recovery time of M_i after a downtime event $\vec{e} = (i,n,d)$ with $d \leq WS_i(n)$. And it is apparent that $T_{2i} = N_1$. Analogously, one can also prove the case when $i < M^*$.

Remark. The relationships between the estimated OWs/recovery time and the system parameters such as machines'

Fig. 3 The serial production line in case study

MCBFs and MCTRs can be explicitly examined. According to the machine probability distribution analysis and Eq. (13), the probability, $Pi_1(n), \forall i \neq M^*$, becomes larger if MCBF_i increases or MCTR_i decreases. Therefore, from the OW and recovery time evaluation analysis in Eqs. (19) and (21), under the same initial conditions, the corresponding OW of machine $M_i, \forall i \neq M^*$, will remain unchanged, while the corresponding recovery time will decrease. This means that the OW evaluation is relatively insensitive to machines' MCBFs and MCTRs.

3.4 Profit Analysis.

The profit of the production system is considered as follows:

$$\text{Profit} = \text{total revenue} - \text{total cost} = (\text{PC})(c_p) - \text{CE} \tag{22}$$

where PC is the system production count, CE is the total cost of energy used by the production line, and c_p is the profit per part.

We will assume that the expenses to produce the part in terms of material and labor are already considered in the profit per part calculation. In this research, inventory cost is ignored, since the tradeoff between throughput and energy savings is the main concern.

Note that OWs provide opportunities for energy saving and opportunities to perform more PM works during production hours, so that more PM can be accomplished, and after-production shift PM can be reduced for cost savings. This scenario of maintenance cost is out of the scope of this paper, since we focus on the real-time estimation of OWs and their potential impact on throughput. The maintenance analysis and simulation can be found in previous research [6,19].

The cost of power consumption of the production system is calculated based on a time of use basis. This charge is dictated by the energy demand of the grid at the current time. If we assume that the energy price per kWh is c_e during the production hours, then CE in Eq. (22) can be evaluated as

$$\text{CE} = c_e \sum_{m=1}^{M} P_m \sigma_m \tag{23}$$

where σ_m is the total operation time of machine M_m.

4 Case Studies

Extensive numerical experiments are executed to verify the performance of the proposed method. The system analyzed is composed of 15 machines and 14 buffers as shown in Fig. 3, which is based on a portion of an engine block line. By changing the system parameters, hundreds of different lines can be obtained. We evaluate the profit through applying a control scheme using the OW and recovery time calculation. Compared to the deterministic-based method as described in Eqs. (18) and (20), the stochastic-based method is numerically proved to be more effective, which results in an average of 10–20% profit increase depending on the energy price and profit of each part as described in the profit analysis in Sec. 3.4. As an illustration, one case study is presented using the system parameters shown in Tables 1 and 2. The profit of each part is assumed to be $c_p = \$300$, and the energy price per kWh is assumed to be $c_e = \$0.2$.

The unit cycle length is set as $\Delta t = 1$ min, and the total simulation length is set as $25,000$ min, which is approximately two and half weeks. For demonstration purpose, all the OWs are taken at M_6, which is the machine with the highest speed. Whenever M_6 takes an OW, it has to wait for a period of recovery time to take the next OW. Three scenarios are compared: (1) deterministic-based method as in Eqs. (18) and (20), (2) stochastic analysis as described in Eqs. (19) and (21), and (3) baseline production system with no OWs. The corresponding OW schedules, i.e., the time to insert OWs and the durations of OWs, are shown in Tables 3 and 4. Using Eqs. (22) and (23) in Sec. 3.4, we can calculate the corresponding profits for those three scenarios. Multiple simulation iterations are carried out. The profit results are shown in Table 5 with 95% confidence interval included.

It can be observed that profit increment can be achieved by applying the OWs based on either deterministic method or stochastic analysis. More importantly, the profit is more effectively increased by utilizing the stochastic analysis which provides a more accurate OW estimation. Note that there might be a minor production throughput impact by applying OWs due to the stochastic nature of production systems and estimation errors.

Table 1 Parameters of the 15 machines in the production system

	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8
MCBF	7200	7200	7800	6000	3840	14,400	5000	4500
MCTR	2400	3600	3600	1500	2640	2400	2315	1840
Cycle time (min)	30	18	30	48	24	12	20	32
pi_{10}	$\frac{1}{7200}$	$\frac{1}{7200}$	$\frac{1}{7800}$	$\frac{1}{6000}$	$\frac{1}{3840}$	$\frac{1}{14,400}$	$\frac{1}{5000}$	$\frac{1}{4500}$
Pi_{01}	$\frac{1}{2400}$	$\frac{1}{3600}$	$\frac{1}{3600}$	$\frac{1}{1500}$	$\frac{1}{2640}$	$\frac{1}{2400}$	$\frac{1}{2315}$	$\frac{1}{1840}$
Energy consumption rate c_i (kW)	10	20	30	20	10	40	50	40
	M_9	M_{10}	M_{11}	M_{12}	M_{13}	M_{14}	M_{15}	
MCBF	6000	5600	8000	6150	4580	5970	7400	
MCTR	2300	2000	3200	2600	1600	2150	2730	
Cycle time (min)	15	25	30	35	24	36	28	
pi_{10}	$\frac{1}{6000}$	$\frac{1}{5600}$	$\frac{1}{8000}$	$\frac{1}{6150}$	$\frac{1}{4580}$	$\frac{1}{5970}$	$\frac{1}{7400}$	
Pi_{01}	$\frac{1}{2300}$	$\frac{1}{2000}$	$\frac{1}{3200}$	$\frac{1}{2600}$	$\frac{1}{1600}$	$\frac{1}{2150}$	$\frac{1}{2730}$	
Energy consumption rate c_i (kW)	20	50	70	40	30	10	50	

Table 2 Parameters of the buffers in the production line

	B_2	B_3	B_4	B_5	B_6	B_7	B_8
Buffer capacity	12	15	16	5	15	18	13
Initial buffer level	7	3	5	4	5	6	4
	B_9	B_{10}	B_{11}	B_{12}	B_{13}	B_{14}	B_{15}
Buffer capacity	12	15	16	10	14	9	16
Initial buffer level	5	2	7	3	4	5	8

Table 3 OW schedule on M_6 based on deterministic method

Time to insert OW (min)	1000	2920	4840	6760	8680	10,600	12,520
OW duration (min)	960	960	960	960	960	960	960
Time to insert OW (min)	14,440	16,360	18,280	20,200	22,120	24,040	
OW duration (min)	960	960	960	960	960	960	

Table 4 OW schedule on M_6 Based on stochastic method

Time to insert OW (min)	1000	2389	3899	5434	7028	8625	10,221
OW duration (min)	1063	1115	1132	1138	1139	1139	1139
Time to insert OW (min)	11,819	13,477	15,076	16,674	18,262	19,861	21,519
OW duration (min)	1139	1139	1139	1139	1139	1139	1139
Time to insert OW (min)	23,118	24,716					
OW duration (min)	1139	286					

Table 5 Profit

	Average profit ($)	95% CI for profit ($)
Deterministic method	77517.2	[77341.3, 77693.1]
Stochastic method	80307.9	[80136.4, 80479.3]
Baseline	7259.5	[7206.9, 7312.0]

However, based on the concepts of the OW, the throughput impact is minimal and the overall profit increments are significant.

To further illustrate the sensitivity of estimated OW with respects to machines' MCBFs and MCTRs, we decrease the value of MCBF_6 from 14,400 to 10,000 for demonstration purpose. The results show that the estimated OW of machine M_6 remains unchanged, which is exactly as expected from the discussion in the aforementioned remark.

5 Discussion and Conclusion

This paper investigates the energy saving and maintenance opportunities in the stochastic serial production systems. The geometric reliability model and Markov property are assumed, and the real-time data are utilized in the analysis. A systematic method is developed to calculate the OW and recovery time in stochastic scenario. It is concluded that the OW concepts are valid in various scenarios. More importantly, the revised calculation methods based on the stochastic analysis can lead to more accurate prediction for production control schedule, which can result in bigger profit increment. In the future, we will further extend the application of OW and recovery time to more complicated systems such as parallel structures and develop optimal control scheme for more efficient production control.

Acknowledgment

This work was supported by the U.S. National Science Foundation (NSF) Grant No. 1351160. The authors would like to thank Dr. Michael Brundage for his assistance to this work.

References

[1] Alden, J. M., Burns, L. D., Costy, T., Hutton, R. D., Jackson, C. A., Kim, D. S., Kohls, K. A., Owen, J. H., Turnquist, M. A., and Vander Veen, D. J., 2006, "General Motors Increases Its Production Throughput," Interfaces, **36**(1), pp. 6–25.

[2] Blumenfeld, D. E., and Li, J., 2005, "An Analytical Formula for Throughput of a Production Line With Identical Stations and Random Failures," Math. Probl. Eng., **3**, pp. 293–308.

[3] Chang, Q., Ni, J., Bandyopadhyay, P., Biller, S., and Xiao, G., 2007, "Supervisory Factory Control Based on Real-Time Production Feedback," ASME J. Manuf. Sci. Eng., **129**(3), pp. 653–660.

[4] Gutowski, T., Murphy, C., Allen, D., Bauer, D., Bras, B., Piwonka, T., Sheng, P., Sutherland, J., Thurston, D., and Wolff, E., 2005, "Environmentally Benign Manufacturing: Observations From Japan, Europe and the United States," J. Cleaner Prod., **13**(1), pp. 1–17.

[5] Chang, Q., Xiao, G., Biller, S., and Lin, L., 2012, "Energy Saving Opportunity Analysis of Automotive Serial Production Systems," IEEE Trans. Autom. Sci. Eng., **10**(2), pp. 334–342.

[6] Chang, Q., Ni, J., Bandyopadhyay, P., Biller, S., and Xiao, G., 2007, "Maintenance Opportunity Planning System," ASME J. Manuf. Sci. Eng., **129**(3), pp. 661–668.

[7] Ding, Y., Elsayed, E. A., Kummara, S., Lu, J. C., Niu, F., and Shi, J. J., 2006, "Distributed Sensing for Quality and Productivity Improvements," IEEE Trans. Autom. Sci. Eng., **3**(4), pp. 344–359.

[8] Košturiak, J., and Gregor, M., 1999, "Simulation in Production System Life Cycle," Comput. Ind., **38**(2), pp. 159–172.

[9] Fowler, J. W., and Rose, O., 2004, "Grand Challenges in Modeling and Simulation of Complex Manufacturing Systems," Simulation, **80**(9), pp. 469–476.

[10] Lee, S., Cho, H., and Jung, M., 2000, "A Conceptual Framework for the Generation of Simulation Models From Process Plans and Resource Configuration," Int. J. Prod. Res., **38**(4), pp. 811–828.

[11] Banks, J., Carson, J. S., Nelson, B. L., and Nicol, D. M., 2010, *Discrete-Event System Simulation*, Prentice-Hall, Englewood Cliffs, NJ.

[12] Gershwin, S. B., 1994, *Manufacturing Systems Engineering*, Prentice-Hall, Englewood Cliffs, NJ.

[13] Li, J., and Meerkov, S. M., 2008, *Production Systems Engineering*, 3rd ed., WingSpan Press, Livermore, CA.

[14] Xia, B., Xi, L., Zhou, B., and Du, S., 2013, "An Efficient Analytical Method for Performance Evaluation of Transfer Lines With Unreliable Machines and Finite Transfer-Delay Buffers," Int. J. Prod. Res., **51**(6), pp. 1799–1819.

[15] Jia, Z., Zhang, L., Arinez, J., and Xiao, G., 2015, "Finite Production Run-Based Serial Lines With Bernoulli Machines: Performance Analysis, Bottleneck, and Case study," IEEE Trans. Autom. Sci. Eng., **99**, pp. 1–15.

[16] Chen, G., Wang, C., Zhang, L., Arinez, J., and Xiao, G., 2015, "Transient Performance Analysis of Serial Production Lines With Geometric Machines," IEEE Trans. Autom. Control, **61**(4), pp. 877–891.

[17] Shi, C., and Gershwin, S. B., 2014, "Improvement of the Evaluation of Closed-Loop Production Systems With Unreliable Machines and Finite Buffers," Comput. Ind. Eng., **75**, pp. 239–256.

[18] Shi, C., and Gershwin, S. B., 2009, "An Efficient Buffer Design Algorithm for Production Line Profit Maximization," Int. J. Prod. Econ., **122**(2), pp. 725–740.
[19] Chang, Q., Biller, S., Xiao, G., and Liu, J., 2010, "Transient Analysis of Downtimes and Bottleneck Dynamics in Serial Manufacturing Systems," ASME J. Manuf. Sci. Eng., **132**(5), p. 051015.
[20] Brundage, M. P., Chang, Q., Li, Y., Xiao, G., and Arinez, J., 2014, "Energy Efficiency Management of an Integrated Serial Production Line and HVAC System," IEEE Trans. Autom. Sci. Eng., **11**(3), pp. 789–797.
[21] Gu, X., Jin, X., and Ni, J., 2015, "Prediction of Passive Maintenance Opportunity Windows on Bottleneck Machines in Complex Manufacturing Systems," ASME J. Manuf. Sci. Eng., **137**(3), p. 031017.
[22] Tolio, T., Gershwin, S. B., and Matta, A., 2011, "Analysis of Two-Machine Lines With Multiple Failure Modes," IIE Trans., **34**(1), pp. 51–62.
[23] Hassett, T. F., Dietrich, D. L., and Szidarovszky, F., 1995, "Time-Varying Failure Rates in the Availability & Reliability Analysis of Repairable Systems," IEEE Trans. Reliab., **44**(1), pp. 155–160.
[24] Cui, L. R., Li, H. J., and Li, J. L., 2007, "Markov Repairable Systems With History-Dependent Up and Down States," Stochastic Models, **23**(4), pp. 665–681.
[25] Zhang, T. L., and Horigome, M., 2001, "Availability and Reliability of System With Dependent Components and Time-Varying Failure and Repair Rates," IEEE Trans. Reliab., **50**(2), pp. 151–158.
[26] Wijngaard, J., 1979, "The Effect of Inter Stage Buffer Storage on Output of Two Unreliable Production Units in Series With Different Production Rates," AIIE Trans., **11**(1), pp. 42–47.
[27] Gershwin, S. B., and Schick, I., 1980, "Continuous Model of an Unreliable Two-Stage Material Flow System With a Finite Inter-Stage Buffer," Massachusetts Institute of Technology, Cambridge, MA, Technical Report No. 1039.
[28] Liu, J., Chang, Q., Xiao, G., and Biller, S., 2012, "The Costs of Downtime Incidents in Serial Multi-Stage Manufacturing Systems," ASME J. Manuf. Sci. Eng., **134**(2), p. 021016.
[29] Li, Y., Chang, Q., Brundage, M., Biller, S., Arinez, J., and Xiao, G., 2014, "Market Demand Oriented Data-Driven Modeling for Dynamic Manufacturing System Control," IEEE Trans. Syst., Man., Cybern., Syst., **45**(1), pp. 109–121.