REAL-TIME ANALYSIS AND CONTROL OF SERIAL PRODUCTION LINES FOR ENERGY EFFICIENT MANUFACTURING

by

Guorong Chen

A dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy
in Engineering

at

The University of Wisconsin - Milwaukee

May, 2014

ABSTRACT

REAL-TIME ANALYSIS AND CONTROL OF SERIAL PRODUCTION LINES FOR ENERGY EFFICIENT MANUFACTURING

by

Guorong Chen

The University of Wisconsin-Milwaukee, 2014
Under the Supervision of Professor Liang Zhang

Productivity analysis, operation control and energy consumption reduction have been the central topics in manufacturing research and practice. They are closely related to each other. Control of production operations is considered as one of the most economical methods to improve energy efficiency in manufacturing systems, while system performance analysis serves as the base of production control. On the other hand, effective operation control can result to energy efficiency in manufacturing.

Steady state analysis has been investigated extensively; however, transient analysis remained largely unexplored. Our research focuses on system modeling, performance analysis, and real-time operation control of serial production lines with unreliable machines and finite buffers, especially in transient period, with Bernoulli or geometric reliability. Analytical results, practical case studies and applications for energy efficient manufacturing are provided. A simulation using ARENA software to reproduce and analyze brewery production line is performed.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my research advisor Professor Liang Zhang for his patient guidance, monitoring and constant encouragement of this research work. It has been a great pleasure and honor to work with him during my years at the University of Wisconsin-Milwaukee.

I also would like to express my appreciation to the other members in my committee: Professors Matthew Petering, Vishnuteja Nanduri, Yingchun Yuan, and Zhijian Huang.

I would like to extend my thanks to Drs. S. Biller, J. Arinez and G. Xiao from General Motors, and the personnel from MillerCoors, for their help with the case studies.

I wish to thank my wife, Guannan Sun, for her constant love, support, and sacrifice, my cutest son Jason, my father Guantian Chen, and my mother Guimei Lin. Without your love and support, this work would not be possible.

Finally, I would like to acknowledge the financial support from National Science Foundation, Department of Industrial/Manufacturing Engineering, and General Motors Corporation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Improving productivity and quality has been the central topics in manufacturing research and practice (see, for instance, monographs [1]-[4]). In recent years, with increasing energy costs and global environmental concerns, reducing energy consumption and greenhouse gas emissions has become a critical issue for the manufacturing industry. In the U.S. alone, hundreds of millions of dollars are spent every year on energy in automotive assembly plants (see [5]). In 2009, the manufacturing industry in the U.S. as a whole spent over $96 billion on fuel and electricity [6]. The energy expenditure in the transportation equipment manufacturing industry (NACIS code 336) is estimated at over $4 billion, in which $791 million are spent in vehicle assembly plants (SIC 3711) [6]. To identify potential energy-saving opportunities, an inside-out approach is proposed in [7], while indicators for benchmarking energy use in manufacturing plants are developed in [8] and [9]. Moreover, studies have shown that approximately 60% of the total energy consumed in an automotive assembly plant is used during the painting process (see [10]). Further investigation indicates that most of this energy is used for ventilation, heating, and air-conditioning in the booths and ovens of a paint shop (see [11]). In addition to enormous amount of energy consumption, the painting process is an indirect source of greenhouse gas emissions as well, due to the carbon dioxide (CO2) emitted at the electric utility that generates the electricity used by the painting process, and by the emission during the painting process of volatile organic compounds (VOCs), a precursor to ozone. Indeed, dur-

ing the painting process, considerable amount of VOC is generated and needs to be destructed before being released to the environment [12].

In contrast to the vast studies on steady state behavior of production lines, very few results are available in the literature on the transient properties of production systems. It has been shown that production losses during transient period can be as much as 12% (see [13]). On the other hand, many production systems need to deplete parts in the buffers at the end of a shift, such as perishable product, or the system suffers a major breakdown, both of which will result to another transient production. The current literature does offer a number of publications on the transients of queueing systems. These studies, however, are mostly focused on singer-server or single-stage queueing systems. One of the early papers in this direction is by [14], who studied the approximation of state probability distribution in Markovian queueing systems using Runge-Kutta, Liou's method, and randomization. [15] discussed several issues related to transient and steady state behavior of $M/G/1$ queues. In addition, [16] and [17] studied the transients of $GI/G/1$ queues and $M/M/1$ queues, respectively. [18] proposed an empirical formula to approximate the expected queue length during transients of a single-server Markovian queue with infinite capacity and empty initial occupancy. [19] discussed the transients in one-machine production systems using the idea of Markov process absorption time. [20] derived closed-form expressions for analytical evaluation of state probability distribution during transients for a $M/D/1/N$ queue. [21] considered a pull production system as a closed $M/M/1/0$ queue and analyzed the average and variance of its transient throughput. [22] developed an approach to estimate the performance of a $G(t)/G/1/K$ loss-waiting queueing systems. Other representative research in this area includes, but not limited to, [23], [24], [25], [26], [27], [28], [29], [30], [31], as well as the references therein.

For multi-stage production systems, very few analytical results have been reported. Among the limited number of publications available on this topic, [32] studied the

transient behavior of a Markovian two-stage tandem queue with no in-process buffer. [33] discussed the transients of a two-stage production system, in which the first stage contains two parallel machines, the second stage is a merge machine, and the system has no in-process buffer but a finished goods buffer. It is assumed in both papers that the machines are reliable and have exponentially distributed processing times. Also, [34] developed an algorithm for solving the partial differential equation, which describes the evolution of the probability density function of a buffer with Markov-modulated input and output flows. In addition to these analytical studies, computer simulation has been used in the investigation of production system transients. For instance, [35] and [36] proposed to approximate the transient performance in multi-stage-multi-server production systems using exponential functions of time. However, no insight on how to determine the parameters of the approximation formulas was provided. This method was later applied by [37] to model the work-in-process in a production system. In a recent paper, [38] applied time series analysis (ARMA models) to approximate the transient performance of production systems. However, this method requires a given simulation model for the system under consideration in order to "train" the approximate time series model. If system parameters are modified (e.g., through improvements), then the model must be re-trained. Finally, using simulations, [39] studied the transient behavior of serial production lines with reliable machines.

The productivity, quality, and energy performances of manufacturing systems are closely related. Indeed, continuous improvement and efficient design for productivity and quality usually lead to energy performance improvements as well. Although this qualitative conclusion is agreed upon in numerous publications (see survey papers [40] and [41]), very few quantitative studies on this issue have been carried out.

Due to the importance of painting in automotive assembly and its critical role in energy consumption and greenhouse gas emissions, significant research efforts have

been dedicated to improve its productivity, quality, energy, and environmental performances. Specifically, an effective performance evaluation method is developed in [42] and [43] to calculate the throughput in a paint shop with rework. In addition, paper [44] investigates the effect of system design on the quality buy rate of a paint shop. The effect of quality-quantity coupling in painting operations is studied in [45], in which an improvement project is designed that results in significant quality and productivity improvements. From the energy perspective, [46] develops a control strategy in order to reduce the HVAC utility in the paint booth of an automotive assembly plant, while [47] presents several case studies and guidelines to improve energy efficiency in automotive manufacturing industry.

Research efforts for energy consumption reduction in manufacturing systems have been centered at technology and process innovation. These projects, however, often involve major capital investment of new equipment and material. Control of production operations is considered as one of the most economical methods to improve energy efficiency in manufacturing systems, since it usually does not require major capital investment and has a relatively short payback period. However, very few quantitative studies on this issue have been carried out. In two recent papers, using simplified painting process models, [48] and [49] study the problem of optimal vehicle batching and the design of spot repair capacity to reduce energy consumption for fixed production volume, respectively, where it is assumed that the energy consumption is a linear function of the number of jobs processed.

We explore energy saving opportunities through improvement in factory floor operations. Shortening machines' running time is one of the simplest way to reduce energy consumption in manufacturing systems. Specifically, in the framework of Bernoulli serial lines, we consider production systems with stripping operations. In such systems, the in-process buffers have to be depleted at the end of each shift to avoid quality deterioration during off-shift periods.

Specifically, for Bernoulli production lines *without* buffer stripping, mathematical models have been developed and several important properties of transient characteristics have been studied in [13] and [50]. This analysis is later extended to serial lines with two geometric machines in [51]. Finally, float-based systems with two Bernoulli machines are studied in [52], and an application in dairy filling and packing production systems is carried out in [53]. Transient analysis of the systems are carried out and formulas to calculate the performance measures are derived. In addition, we investigate the effect of machine startup schedule on the system performances and develop optimal startup schedule which can lead to significant improvement in energy utilization efficiency.

On the other hand, order completion time estimation is one of the most important considerations in production planning and control, which involves the order type and size as well as production capacity for the specific type of order. Unlike planning, control is a deterministic process based on short-term detailed technical data to optimize production scheduling under resource capacity constraints, a mature research area of operational research [54]. Well-controlled operation schedules may not only increase productivity but also reduce energy consumption [55]. On the other hand, research has primarily focused on completion time problems with serial systems in which either intermediate storage is unlimited or the number of machines are not more than three. For example, [56] addressed the determination of completion times of serial multiproduct processes with three buffer types. An optimal scheduling to minimize makespan and reduce energy and environmental effects with different types of jobs and various machine speeds was proposed in [57] using multi-objective mixed integer programming method, which required significant computation time. Paper [58] proposed an improved mixed integer linear formulation and lower bounds for minimizing makespan with batch processing machines. Paper [59] investigated the distribution of completion time in a two-machine serial system with two kinds of

jobs. An algorithm based on semi-Markovian decision processes and graph theory to find a path which is a lower bound for expected completion time in the dynamic PERT network was developed in [60]. Note that the simulation technique can be applied to validate production process [61] and to determine order completion time [62]. However, it is far from being satisfactory to the requirement of fast assessment for real-time planning and control. Therefore, our research developed an analytical approximation procedure to efficiently evaluate the order completion time as well as important system structural properties.

The remainder chapters are organized as follows: Chapter II describes model and performance measures of serial multi-machine production lines with Bernoulli and geometric reliability and finite buffers. A recursive aggregation technique proposed by [63] is included since the methods proposed in later chapters are an extension of this method. Chapter III investigates energy consumption reduction in two-machine production systems through effective scheduling of machine startup and shutdown. Chapter IV extends the startup and shutdown scheduling discussion to production lines with more than two machines. Chapter V investigates the order completion time of a required demand size in serial lines. Chapter VI estimates geometric transient performance by transforming to Bernoulli lines with two methods. Chapter VII derives closed-form formulas as well as approximation method to evaluate transient performance of serial production lines with geometric machines. Chapter VIII constructs a simulation model using ARENA to perform productivity capacity analysis in manufacturing. Chapter IX gives the conclusions and future work.

CHAPTER II

BERNOULLI AND GEOMETRIC SERIAL PRODUCTION SYSTEMS

2.1 Modeling and Performance Measures

2.1.1 Bernoulli serial production lines

The assumptions for a serial production line in Figure 2.1 with Bernoulli reliability are as follows:

Figure 2.1: Serial production line

(i) The system consists of M machines and $M-1$ in-process buffers.

(ii) The time axis is slotted with duration τ. The machines, m_i, $i = 1, \ldots, M$, have identical cycle time τ.

(iii) Each in-process buffer, b_i, $i = 1, \ldots, M-1$, is characterized by its capacity, $0 < N_i < \infty$.

(iv) Machine m_i, $i = 1, \ldots, M$, when it is neither blocked nor starved, produces a part during time slot n with probability $p_i(n)$ and fails to do so with probability $1 - p_i(n)$.

(v) Machine m_i, $i = 2, \ldots, M$ is starved during a time slot if it is up and buffer b_{i-1} is empty at the beginning of the time slot. It is assumed that machine m_1 is never starved for raw material.

(vi) Machine m_i, $i = 1, \ldots, M-1$, is blocked during a time slot if it is up, buffer b_i has N_i parts at the beginning of the time slot and machine m_{i+1} fails to take a part during that time slot. It is assumed that m_M is never blocked.

(vii) All machines operate independently. If a machine is shut down during a shift, it cannot be started again during the same shift.

Note that in large volume production systems, machine cycle time is practically always constant or close to being constant. Note also that Bernoulli reliability model is applicable to operations where the unscheduled downtime is, on the average, close to the machine cycle time. This often happens in automotive painting and assembly operations, where the downtime is primarily due to quality problems rather than mechanical breakdowns. For example, a vehicle needs to be processed through the following operations: phosphate coating, electrocoating, sealing, sanding, prime booth and oven, cleaning and preparation, paint booth and oven, and, finally, quality inspection. During this process, phosphate, electrocoating, prime oven & booth, and paint booth & oven usually require buffer stripping.. In addition, the paint booth and oven are the largest energy consumers in the painting process [11].

For assumption (iv), if the probability $p_i(n)$ is not time dependent, i.e., $p_i(n) = p_i$, for all n, then p_i could be used instead and is referred to as machine's *efficiency*.

2.1.2 Geometric serial production lines

For a geometric serial line, the assumptions are all the same as Bernoulli lines except that assumption (iv) is replaced by (iv'):

(iv') The machines obey the geometric reliability model, i.e., machine m_i, $i = 1, \ldots, M$,

will be up with probability R_i and down with probability $1-R_i$ when it is down in the previous time slot, and will be up with probability $1-P_i$ and down with probability P_i when it is up in the previous time slot. Machine will produce a part during a time slot when it is up and neither blocked nor starved. R_i is referred to repair rate and P breakdown rate.

Geometric lines have one-step memory and can be described by an ergodic Markov chain. The states transition of geometric machines is illustrated by Figure 2.2. Note also that the geometric reliability model is usually used when the machine's average downtime is much longer than its cycle time (see, for instance, [2, 1, 3, 64]). These models are applicable in many manufacturing systems, such as machining, heat treatment, washing, etc.

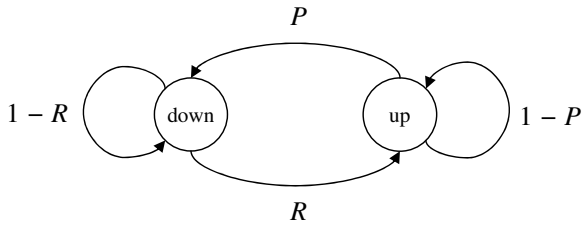


Figure 2.2: States transition diagram of geometric machine

The states transition can be described mathematically by equation 2.1 and teh uptime and downtime probability mass functions of machine m_i are given by 2.2.

$$\begin{cases} P[s_i(n+1)=0|s_i(n)=1] & = P_i, \\ P[s_i(n+1)=1|s_i(n)=0] & = 1-P_i, \\ P[s_i(n+1)=1|s_i(n)=1] & = R_i, \\ P[s_i(n+1)=0|s_i(n)=0] & = 1-R_i, \end{cases} \tag{2.1}$$

where $s(n)$ denotes status in time slot n, 0 is down and 1 is up.

$$\begin{cases} P[t_{up} = t] \quad = P_i(1-P_i)^{t-1},\ t = 1, 2, \ldots, \\ P[t_{down} = t] = R_i(1-R_i)^{t-1},\ t = 1, 2, \ldots, . \end{cases} \tag{2.2}$$

The average up- and downtime of machine m_i are $T_{i,up} = 1/P_i$ and $T_{i,down} = 1/R_i$ and machine efficiency is given by

$$e_i = \frac{R_i}{R_i + P_i} = \frac{T_{i,up}}{T_{i,up} + T_{i,down}}.$$

2.1.3 Performance measures

The productivity performance measures of interest are:

- Production rate, $PR(n)$: the expected number of finished parts produced by m_M during time slot $n+1$;
- Consumption rate, $CR(n)$: the expected number of raw parts consumed by m_1 during time slot $n+1$;
- Work-in-process, $WIP_i(n)$: the expected number of parts in buffer b_i, $i = 1, \ldots, M-1$, at the beginning of time slot $n+1$;
- Machine starvation $ST_i(n)$: the probability that machine m_i, $i = 2, \ldots, M$, is starved during time slot $n+1$;
- Machine blockage $BL_i(n)$: the probability that machine m_i, $i = 1, \ldots, M-1$, is blocked during time slot $n+1$.

In addition, to quantify the total production in a shift, we define the shift production volume, SPV, as follows:

$$SPV \quad = \quad \sum_{n=0}^{T_{\text{shift}}-1} PR(n).$$

Among these performance measures, $PR(n)$, $CR(n)$ and $WIP(n)$ are called the *transient performance measures*, since they are functions of time, while SPV is referred to as *terminal performance measures*. It should be noted that a machine may be starved and blocked during the same time slot.

2.2 Approximation of Performance Measures of Bernoulli Lines

2.2.1 Recursive aggregation procedure

There are closed-form formulas to calculate some performance measures in steady state (see [4]) and/or transient period, some of which will be proposed in the following chapters; however, the calculation requires significant time as the dimension of the system grows. To resolve this issue, an aggregation-based recursive procedure is proposed in [65] to approximate the transient performance of Bernoulli serial lines with constant machine efficiencies. In fact, the idea of machine aggregation is originated in [63] to evaluate the steady state performance of Bernoulli serial lines. We extend the aggregation procedure to evaluate the performance measures of Bernoulli lines defined by (i)-(vii):

$$\begin{aligned} p_i^b(s+1;n) &= \frac{p_i(n)}{p_i^f(s;n)} \cdot CR(n-1;\mathbf{p}_i^f(s;n),\mathbf{p}_{i+1}^b(s;n),N_i,h_i(0)), \\ &\quad i=1,\ldots,M-1,\ s=0,1,\ldots,\ n=1,2,\ldots, \\ p_i^f(s+1;n) &= \frac{p_i(n)}{p_i^b(s+1;n)} \cdot PR(n-1;\mathbf{p}_{i-1}^f(s+1;n),\mathbf{p}_i^b(s+1;n),N_{i-1},h_{i-1}(0)), \\ &\quad i=2,\ldots,M,\ s=0,1,\ldots,\ n=1,2,\ldots, \end{aligned} \tag{2.3}$$

with initial condition

$$p_i^f(0;n) = p_i(n), \quad i=1,\ldots,M, \quad n=1,2,\ldots, \tag{2.4}$$

and boundary condition

$$p_1^f(s;n) = p_1(n),\ p_1^b(s;n) = p_M(n),\ \forall s = 0, 1, \ldots, \tag{2.5}$$

where

$$\begin{aligned} \mathbf{p}_i^b(s;n) &= [p_i^b(s;1)\ \ldots\ p_i^b(s;n)],\ i = 1, \ldots, M, \\ \mathbf{p}_i^f(s;n) &= [p_i^f(s;1)\ \ldots\ p_i^f(s;n)],\ i = 1, \ldots, M, \end{aligned}$$

and $PR(n-1; \mathbf{v}_1, \mathbf{v}_2, v_3, v_4)$ and $CR(n-1; \mathbf{v}_1, \mathbf{v}_2, v_3, v_4)$ denote the production rate and consumption rate, respectively, during time slot n, of a two-machine Bernoulli line with buffer capacity v_3, initial buffer occupancy v_4 and time-dependent efficiencies of the first and the second machine given by vectors $\mathbf{v}_1$ and $\mathbf{v}_2$. Formulas for calculating these two functions are given in (4.5) and (4.6) with $M = 2$.

2.2.2 Properties and accuracy

Numerical Fact 2.1. *For any $n = 1, 2 \ldots$, sequences $p_i^b(s;n)$, $i = 1, \ldots, M-1$, and $p_i^f(s;n)$, $i = 2, \ldots, M$, are practically always convergent with respect to s with probability 1, i.e.:*

$$\begin{aligned} P[p_i^b(n) = \lim_{s\to\infty} p_i^b(s;n)] &= 1,\ i = 1, \ldots, M-1, \\ P[p_i^f(n) = \lim_{s\to\infty} p_i^f(s;n)] &= 1,\ i = 2, \ldots, M, \\ & \quad n = 1, 2, \ldots. \end{aligned}$$

Justification: To justify this numerical fact, a total of 1,800,000 production lines are generated with 100,000 for each $M \in \{3, 4, \ldots, 20\}$. The max time duration is selected as $T = 10,000$ time slots. In addition, the machine efficiency, buffer capacity, and initial buffer occupancy of the lines were selected randomly and equiprobably from

the following sets:

$$\begin{aligned} p_i(n) &\in (0.5, 1), \quad i = 1, \ldots, M, \quad n = 1, \ldots, T, \\ N_i &= \{1, 2, \ldots, 10\}, \quad i = 1, \ldots, M-1, \\ h_i(0) &\in \{0, 1, \ldots, N_i\}, \quad i = 1, \ldots, M-1. \end{aligned}$$

For each line, thus constructed, Recursive Procedure 1 was performed. To determine the convergency of the procedure, introduce

$$\begin{aligned} \Delta_1(s) = \sum_{i=1}^{M} \sum_{n=1}^{T} \Big[& |p_i^f(s;n) - p_i^f(s-1;n)| + \\ & |p_i^b(s;n) - p_i^b(s-1;n)| \Big], \quad s = 1, 2, \ldots. \end{aligned}$$

During the justification, it was observed that for each of the 1.8 million lines generated above, there always exists an s_0 such that $\Delta_1(s) \leq 10^{-8}$ for $s \geq s_0$. Therefore, we conclude that Numerical Fact 5.1 holds, i.e., Recursive Procedure 1 is convergent. ■

The implications of these parameter limits can be interpreted as follows: From the point of view of buffer b_i, $i = 1, \ldots, M-1$, the upstream of the line, i.e., the capability of *producing* parts into buffer b_i during time slot n, is represented by "virtual" Bernoulli machine m_i^f with time-dependent efficiency $p_i^f(n)$. Similarly, the downstream of the line, i.e., the capability of *consuming* parts from buffer b_i during time slot n, is represented by virtual Bernoulli machine m_{i+1}^b with time-dependent efficiency $p_{i+1}^b(n)$. Finally, the raw material consumption is represented by $p_1^b(n)$, while the finished part production is represented by $p_M^f(n)$.

Therefore, based on these interpretations, the following estimates for the produc-

tivity performance measures are defined:

$$\begin{aligned}
\widehat{PR}(n) &= p_M^f(n+1), && (2.6)\\
\widehat{CR}(n) &= p_1^b(n+1), && (2.7)\\
\widehat{WIP}_i(n) &= WIP(n; \mathbf{p}_i^f(n), \mathbf{p}_{i+1}^b(n), N_i, h_i(0)), \\
&\qquad i = 1, \ldots, M-1, && (2.8)\\
\widehat{ST}_i(n) &= ST(n; \mathbf{p}_i^f(n), \mathbf{p}_{i+1}^b(n), N_i, h_i(0)), \\
&\qquad i = 2, \ldots, M, && (2.9)\\
\widehat{BL}_i(n) &= BL(n; \mathbf{p}_i^f(n), \mathbf{p}_{i+1}^b(n), N_i, h_i(0)), \\
&\qquad i = 1, \ldots, M-1, && (2.10)
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{p}_i^b(n) &= [p_i^b(1) \ \ldots \ p_i^b(n)], \ i = 1, \ldots, M,\\
\mathbf{p}_i^f(n) &= [p_i^f(1) \ \ldots \ p_i^f(n)], \ i = 1, \ldots, M,
\end{aligned}$$

and $WIP(n-1; \mathbf{v}_1, \mathbf{v}_2, v_3, v_4)$, $ST(n-1; \mathbf{v}_1, \mathbf{v}_2, v_3, v_4)$, and $BL(n-1; \mathbf{v}_1, \mathbf{v}_2, v_3, v_4)$ denote the work-in-process, second machine starvation and first machine blockage, respectively, during time slot n, in a two-machine Bernoulli line with buffer capacity v_3, initial buffer occupancy v_4 and time-dependent efficiencies of the first and the second machine given by vectors $\mathbf{v}_1$ and $\mathbf{v}_2$. Formulas for calculating these functions are given in (4.7)-(4.9) with $M = 2$.

To evaluate the accuracy of these estimates, the following metrics are used:

$$\begin{aligned}
\delta_{PR}(n) &= \frac{|PR(n) - \widehat{PR}(n)|}{p_M(n+1)} \cdot 100\%,\\
\delta_{CR}(n) &= \frac{|CR(n) - \widehat{CR}(n)|}{p_1(n+1)} \cdot 100\%,\\
\delta_{WIP}(n) &= \frac{1}{M-1} \sum_{i=1}^{M-1} \frac{|WIP_i(n) - \widehat{WIP}_i(n)|}{N_i} \cdot 100\%,
\end{aligned}$$

$$\delta_{ST}(n) = \frac{1}{M-1}\sum_{i=2}^{M}|ST_i(n) - \widehat{ST}_i(n)|,$$
$$\delta_{BL}(n) = \frac{1}{M-1}\sum_{i=1}^{M-1}|BL_i(n) - \widehat{BL}_i(n)|.$$

Note that $p_M(n+1)$ and $p_1(n+1)$ are used in $\delta_{PR}(n)$ and $\delta_{CR}(n)$ because $PR(n)$ and $CR(n)$ are upper-bounded by $p_M(n+1)$ and $p_1(n+1)$, respectively. Then, we calculated these accuracy metrics for the 1,800,000 lines generated in the justification of Numerical Fact 5.1. As a result, among all systems studied, $\delta_{PR}(n)$ and $\delta_{CR}(n)$ are typically within 2.5%, $\delta_{WIP}(n)$ is typically within 4%, and $\delta_{ST}(n)$ and $\delta_{BL}(n)$ are typically within 0.015. Therefore, we claim that Recursive Procedure 1 can be used to estimate the productivity performance measures in Bernoulli serial lines with time-dependent machine efficiencies, and, thus, can be applied to the systems defined by assumptions (i)-(viii) with startup/shutdown times.

As an illustration, consider a five-machine Bernoulli line with time-dependent machine efficiencies defined as follows (see also Figure 2.3):

$$p_1(n) = 0.8 + 0.15\sin\left(\frac{\pi}{15}\cdot n\right),$$
$$p_2(n) = 0.75 + 0.2\sin\left(\frac{\pi}{10}\cdot n\right),$$
$$p_3(n) = \begin{cases} 0.9, & \text{if } 30k \le n < 30k+15,\ k = 0, 1, \ldots, \\ 0.7, & \text{otherwise,} \end{cases}$$
$$p_4(n) = \begin{cases} 0.83, & \text{if } 11 \le n \le 55, \\ 0, & \text{otherwise,} \end{cases}$$
$$p_5(n) = 0.55 \cdot e^{\frac{n}{120}}, \quad n \le 60.$$

In addition, let the shift duration be $T_{\text{shift}} = 60$ and assume the buffer capacity and initial buffer occupancy are given by the following vectors:

$$\mathbf{N} = [3,\ 5,\ 4,\ 2,\ 4,\ 5,\ 2],$$

$$\mathbf{h}(0) \quad = \quad [1,\, 3,\, 1,\, 0,\, 2,\, 0,\, 1].$$

The performance measure estimates for PR, CR and total WIP, obtained using (2.6)-(2.8) and those obtained by simulations, are plotted in Figure 2.4. As one can see, the performance estimates derived based on Recursive Procedure 1 can effectively track the transients of the performance measures with high accuracy.

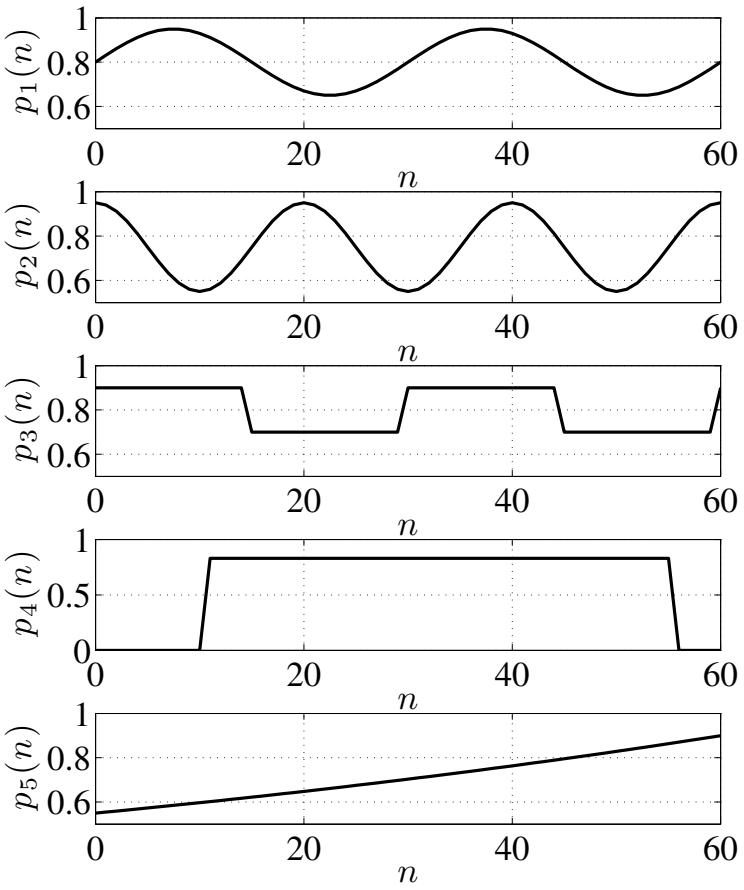
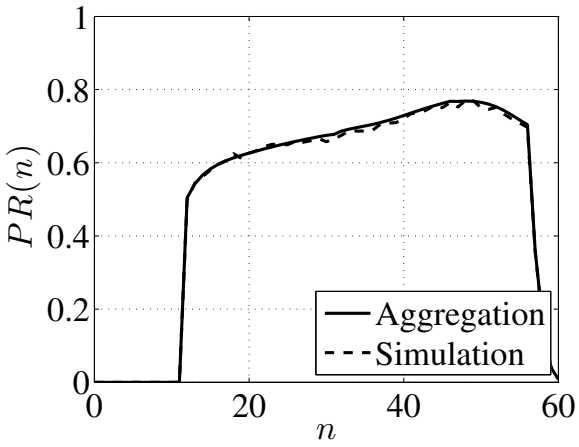
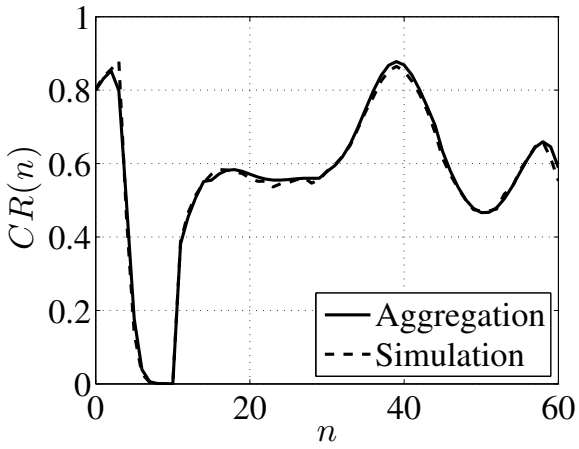


Figure 2.3: Machine efficiencies as functions of time n

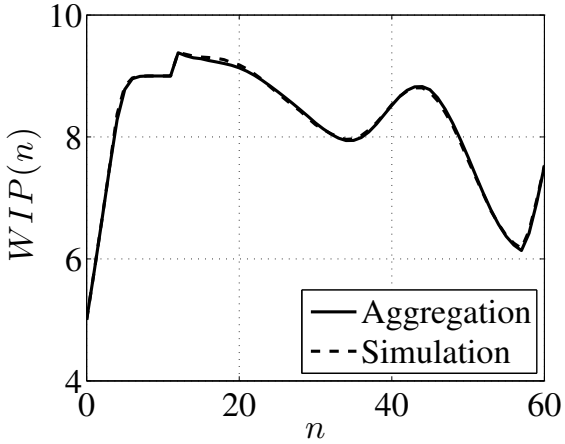
Clearly, this method enables analytical study of production lines with time-dependent parameters, and can be applied to, for example, production systems with equipment deterioration, operator fatigue, etc. It also provides an effective tool for other real-time analysis and control problems in production systems.



(a) Production rate



(b) Consumption rate



(c) Work-in-process

Figure 2.4: Illustration of performance estimation using Recursive Procedure

CHAPTER III

FEEDBACK CONTROL OF OPERATION SCHEDULE IN TWO-MACHINE BERNOULLI SERIAL LINES

3.1 Introduction

In automotive paint shop systems, due to process and quality considerations, some of the in-process buffers (for instance, those within the painting booths and ovens) must be depleted at the end of each shift to maintain job condition and prevent quality deterioration. Such activity is referred to as *buffer stripping*. In addition to painting systems, buffer stripping is widely used in manufacturing systems with perishable products (e.g., food production). Because of these depleted buffers, the downstream machines will be starved for a certain period of time at the beginning of each shift. It is intuitive that these machines should not be turned on immediately when a shift starts but rather wait until some work-in-process has been accumulated in front of them. Therefore, a number of questions arise. For example, how to evaluate the performance of a production system with stripping? How does the startup schedule affect the overall system performance? How to formulate an optimal startup schedule such that the energy consumption is minimized while maintaining desired throughput level of the system? How much energy could be saved using the optimal schedule? The goal of this chapter is to provide answers to these questions.

This chapter investigates energy consumption reduction in production systems through effective scheduling of machine startup and shutdown. Specifically, we con-

sider serial production lines with finite buffers and machines having Bernoulli reliability model. Using transient analysis of the systems at hand, an analytical performance evaluation technique is developed for Bernoulli serial lines with time-dependent machine efficiencies. In addition, trade-off between productivity and energy-efficiency in production systems is discussed and the energy-efficient production problem is formulated as a constrained optimization problem. The effects and practical implications of operations schedule are demonstrated using a numerical study on automotive paint shop operations.

This chapter develops an effective analytical tool to evaluate the performance of production systems with time-varying parameters of machine reliability. Using this tool, production engineers and managers can predict the performance of the production systems in real-time with high accuracy. In addition, based on this tool, production operators can determine the machine startup and shutdown schedule based on the current status of the line and production requirement. Numerical experiments show that significant energy savings can be obtained by applying effective machine operations schedule.

3.2 Model and Analysis

In this chapter we only study the case of two-machine lines. Extensions to longer lines will be carried out in future work.

Consider a serial production line shown in Figure 3.1 defined by the assumptions described in Chapter II. Additoinal assumptions are as follows:

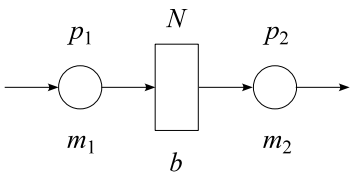


Figure 3.1: Two-machine Bernoulli line

(i) The system operates on a shift-basis, with shift duration T_{shift} time slots. At the end of each shift, i.e., after the $T_{\text{shift}}{}^{\text{th}}$ time slot, machine m_1 is shut down immediately, while machine m_2 continues operating until the buffer is empty.

(ii) At the beginning of a shift, both machines are turned on immediately. In addition, it is assumed that once turned on, the machines cannot be shut down during the shift.

Note that assumption (i) in this chapter indicates that the system operates over a finite period of time. Therefore, the steady-state analysis for serial production lines, strictly speaking, is not applicable, and transient analysis of the system must be used. Since no feedback control on the startup schedule of m_2 is used, we refer to the system as *open-loop*.

Closed-loop system: Closed-loop system is an open-loop system with the following assumptions:

(iii) At the beginning of each shift, m_1 is turned on immediately, while machine m_2 starts up at the $t_s{}^{\text{th}}$ time slot, $0 \leq t_s \leq T_S$, where $T_S \leq T_{\text{shift}}$ is the latest startup time.

To determine t_s, introduce a feedback controller defined as follows:

$$u(n) = \begin{cases} 0, \ (m_2 \text{ cannot start}), & \text{if } h(t) < h^*, \forall t \leq n, \\ 1, \ (m_2 \text{ authorized to start}), & \text{otherwise.} \end{cases} \tag{3.1}$$

where $h(t)$ is the buffer occupancy at the end of time slot t. In other words, in a closed-loop system, m_2 cannot start operating until enough (h^*) parts have been accumulated in the buffer. Therefore, t_s is the smallest n to have $u(n) = 1$. Clearly, for a closed-loop system, the control parameter h^* should be selected appropriately in order to achieve desired performance. In addition, as one can see, the open-loop system can be regarded as a special case of the closed-loop system with $h^* = 0$. (Note

that since the Bernoulli machines do not possess memory, feedback controller (3.1) only depends on the buffer occupancy.)

To construct the mathematical model of the system under consideration, note that the open-loop system is actually a special case of the closed-loop system (when $h^* = 0$). Thus, the mathematical model and the results reported in subsequent sections are directly derived for the closed-loop system. To accomplish this, we first divide the system operation time into three stages:

- *Stage 1*: Only machine m_1 is operating, i.e., from time slot $n = 0$ till $n = t_s - 1$;
- *Stage 2*: Both machines are operating, i.e., from time slot $n = t_s$ till $n = T_{\text{shift}}$;
- *Stage 3*: Only machine m_2 is operating, i.e., from time slot $n = T_{\text{shift}} + 1$ till $n = t_d$, where t_d is the last time slot of m_2 operation in a shift.

Stage 1: The operation of the system at hand during Stage 1 is described by a Markov chain. The state of this Markov chain is the buffer occupancy at the end of time slot n, $h(n) \in \{0, 1, \ldots, N\}$, where state h^* is an absorbing state for this stage. Therefore, the startup time for m_2, t_s, is actually the time to absorption for this Markov process. Let $x_i(n) = P[h(n) = i]$, $i = 0, \ldots, N$, and $\mathbf{x}(n) = [x_0(n) \ldots x_N(n)]^T$. Then, the evolution of the system states can be expressed as:

$$\mathbf{x}(n) = \mathbf{A_1}\mathbf{x}(n-1), \quad n = 1, 2, \ldots, \tag{3.2}$$

with initial condition $h(0) = 0$, i.e.,

$$\mathbf{x}(0) = [1 \quad 0 \ \ldots \ 0]^T. \tag{3.3}$$

The transition probability matrix $\mathbf{A_1}$ is given by

$$\mathbf{A_1} = \begin{bmatrix} \mathbf{B_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{(N+1)\times(N+1)}, \tag{3.4}$$

$$\mathbf{A_2} = \begin{bmatrix} 1-p_1 & p_2(1-p_1) & 0 & \cdots & 0 \\ p_1 & 1-p_1-p_2+2p_1p_2 & p_2(1-p_1) & \cdots & 0 \\ 0 & p_1(1-p_2) & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & 1-p_1-p_2+2p_1p_2 & p_2(1-p_1) \\ 0 & 0 & \cdots & p_1(1-p_2) & p_1p_2+1-p_2 \end{bmatrix}. \tag{3.6}$$

$$\mathbf{B_1} = \begin{bmatrix} 1-p_1 & 0 & \cdots & 0 \\ p_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1-p_1 & 0 \\ 0 & \cdots & p_1 & 1 \end{bmatrix}_{(h^*+1)\times(h^*+1)}. \tag{3.5}$$

Stage 2: The operation of the system during Stage 2 is also characterized by a Markov chain with the state being the buffer occupancy. The $(N+1)\times(N+1)$-dimension transition probability matrix $\mathbf{A_2}$ is given by (3.6), and the evolution of the system states can be expressed as:

$$\mathbf{x}(n) = \mathbf{A_2}\mathbf{x}(n-1), \quad n = t_s, \ldots, T_{\text{shift}}, \tag{3.7}$$

with "initial" condition $h(t_s-1) = h^*$, i.e.,

$$\mathbf{x}(t_s-1) = [0 \ \ldots \ 0 \quad x_{h^*}(t_s-1) = 1 \quad 0 \ \ldots \ 0]^T. \tag{3.8}$$

Stage 3: Similarly to Stages 1 and 2, the operation of the system during Stage 3 is also defined by a Markov chain but with the state 0 being an absorption state. The transition probability matrix $\mathbf{A_3}$ is given by:

$$\mathbf{A_3} = \begin{bmatrix} 1 & p_2 & \cdots & 0 \\ 0 & 1-p_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & p_2 \\ 0 & \cdots & 0 & 1-p_2 \end{bmatrix}_{(N+1)\times(N+1)}. \tag{3.9}$$

The evolution of the system states can be expressed as:

$$\mathbf{x}(n) = \mathbf{A_3}\mathbf{x}(n-1), \quad n = T_{\text{shift}} + 1, T_{\text{shift}} + 2, \ldots, t_d, \tag{3.10}$$

where t_d is the time to absorption.

Therefore, the dynamics of the system is characterized by a series of linear systems (3.2)-(3.10). These descriptions are the basis for the analysis carried out in this chapter.

3.3 Performance Analysis

3.3.1 Performance measures

In the framework of the serial production line, the additional performance measures of interest are:

- Running time, RT_i: the expected number of time slots that machine m_i, $i = 1, 2$, is on;
- Overtime, OT_i: the expected number of time slots that machine m_i, $i = 1, 2$, works after $n = T_{\text{shift}}$;
- Total energy, E: the expected total energy consumed by the production system during a shift;
- Energy consumed per part, E_v: the average energy consumed to produce one part.

Within these performances, it can be obtained immediately that $RT_1 = T_{\text{shift}}$ and $OT_1 = 0$. Below, we derive the formulas to calculate the other non-trivial performance measures.

3.3.2 Performance evaluation

Since the dynamics of the system defined by (i)-(vi) and (viii) are given by (3.2)-(3.10), the transients of production rate and work-in-process during the shift can be expressed as follows:

$$\begin{aligned} PR(n) &= P[m_2 \text{ up and } b \text{ non-empty during the time slot}] \\ &= \mathbf{C_1}(n)\mathbf{x}(n), \qquad (3.11) \\ WIP(n) &= \sum_{i=0}^{N} i \cdot P[\text{buffer } b \text{ has } i \text{ parts during the time slot}] \\ &= \mathbf{C_2}\mathbf{x}(n), \qquad (3.12) \end{aligned}$$

where

$$\begin{aligned} \mathbf{C_1}(n) &= \widetilde{p}_2(n)[0 \quad 1 \ \ldots \ 1], \qquad (3.13) \\ \mathbf{C_2} &= [0 \quad 1 \ \ldots \ N], \qquad (3.14) \\ \widetilde{p}_2(n) &= \begin{cases} 0, & n = 0, 1, \ldots, t_s - 2, \\ p_2, & n = t_s - 1, t_s, \ldots, t_d. \end{cases} \qquad (3.15) \end{aligned}$$

To calculate the terminal performance measures, the information about the startup time of machine m_2 must be known. Since t_s is a random variable due to random breakdowns of m_1, we need to derive its probability distribution function. For systems defined by assumptions (i)-(vi) and (viii), it is easy to show that for any given $h^* \leq N$, the startup time for m_2 is a "bounded" negative binomial random variable with probability mass function defined as follows:

$$P[t_s = t] = \begin{cases} \begin{pmatrix} t-2 \\ h^*-1 \end{pmatrix} p_1^{h^*}(1-p_1)^{t-h^*-1}, & \text{if } h^* < t < T_S, \\ 1 - \sum_{i=h^*+1}^{T_S-1} \begin{pmatrix} i-2 \\ h^*-1 \end{pmatrix} p_1^{h^*}(1-p_1)^{i-h^*-1}, & \text{if } t = T_S, \\ 0, & \text{otherwise.} \end{cases} \qquad (3.16)$$

An illustration of the distribution of t_s is shown in Figure 3.2.

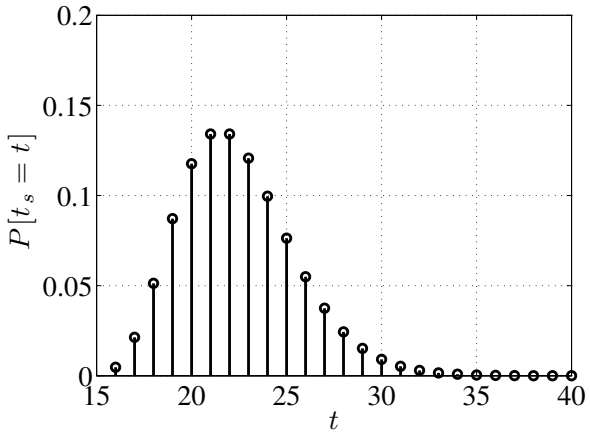


Figure 3.2: Probability distribution of m_2 startup time t_s under control (3.1) for $p_1 = 0.7$, $p_2 = 0.9$, $N = 20$, $T_{\text{shift}} = 100$, $T_S = 40$, $h^* = 15$

In addition, for sufficiently large T_s, the average startup time $\bar{t}_s$ can be approximated as $\bar{t}_s \approx h^*/p_1 + 1$.

Next, we derive the expressions for evaluate the terminal performance measures.

For shift production volume, due to stripping operation, the following equation holds:

$$\begin{aligned} SPV &= \sum_{n=0}^{T_{\text{shift}}-1} PR(n) + WIP(T_{\text{shift}}) \\ &= E_{t_s}\left\{ \sum_{n=t-1}^{T_{\text{shift}}-1} PR(n|t_s = t) + WIP(T_{\text{shift}}|t_s = t) \right\} \\ &= \sum_{t=h^*+1}^{T_S} \left(\sum_{n=t-1}^{T_{\text{shift}}-1} \mathbf{C_1}(t_s - 1)\mathbf{A_2}^{n-t+1}\mathbf{x}(t_s - 1) + \right. \\ &\qquad \left. \mathbf{C_2}\mathbf{A_2}^{T_{\text{shift}}-t+1}\mathbf{x}(t_s - 1) \right) P[t_s = t], \end{aligned} \tag{3.17}$$

where $P[t_s = t]$, $t = h^* + 1, \ldots, T_S$, is calculated in (3.16). Similarly, it can be obtained that

$$RT_2 = \sum_{t=h^*+1}^{T_S} \left(\frac{\mathbf{C_2}\mathbf{A_2}^{T_{\text{shift}}-t+1}\mathbf{x}(t_s - 1)}{p_2} + T_{\text{shift}} - t + 1 \right) P[t_s = t].$$

(3.18)

In practice, the energy consumption of a machine is generally proportional to its running time. Thus, in this paper, we evaluate the total energy consumed by the system during a shift as follows:

$$E = \alpha_1 RT_1 + \alpha_2 RT_2, \tag{3.19}$$

where α_i, $i = 1, 2$, is the energy consumption of machine m_i per cycle time. In practice, these constants can be evaluated based on the energy consumption mechanism of the equipment (see [11] for example). Clearly, the energy consumed per part is given by:

$$E_v = \frac{E}{SPV}. \tag{3.20}$$

Although these formulas provide exact evaluation of the performance measures, the computational burden can become intolerable as the dimension of the system grows. To resolve this issue, an aggregation-based recursive procedure is proposed in [65] and it is extended in Chapter II.

3.4 Monotonicity and Optimal Startup Schedule

3.4.1 Monotonicity

Numerical Fact 3.1. *Consider a production line defined by assumptions (i)-(vi) and (viii).*

- *OT_2 is monotonically increasing as a function of h^*;*
- *RT_2 is monotonically decreasing as a function of h^*;*
- *E is monotonically decreasing as a function of h^*.*

Thus, as the control parameter h^* increases, i.e., as more work-in-process is required to start m_2, more time is needed for machine m_2 to work "overtime." This is because later start will result in more work-in-process when m_1 is shut down after the ${T_{\text{shift}}}^{\text{th}}$ time slot. On the other hand, later start will lead to shorter total running

time for m_2. Finally, since RT_1 is always T_{shift} time slots, decreasing RT_2 leads to reduced total energy consumption E. An illustration of Proposition 1 is given in Figures 3.3-3.5.

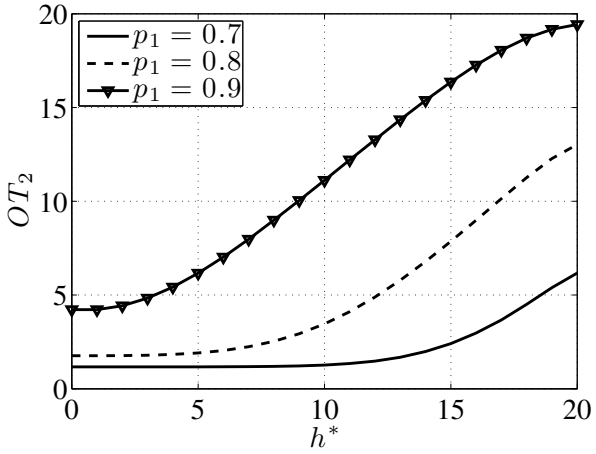


Figure 3.3: OT_2 as a function of h^* ($p_2 = 0.8$, $N = 20$, $T_s = 30$, $T_{\text{shift}} = 100$)

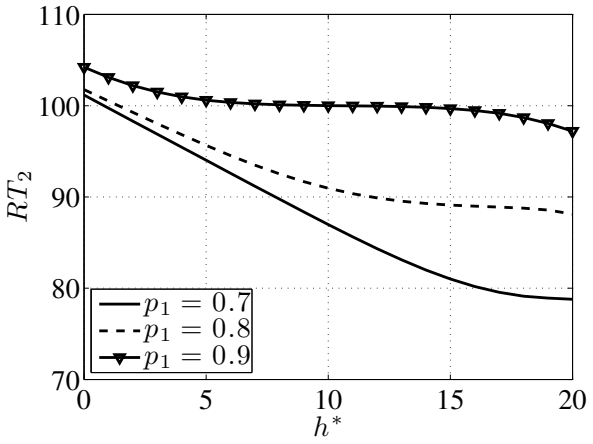


Figure 3.4: RT_2 as a function of h^* ($p_2 = 0.8$, $N = 20$, $T_s = 30$, $T_{\text{shift}} = 100$)

The monotonicity property of SPV, however, is more complicated, where SPV could be monotonically increasing in h^*, monotonically decreasing in h^*, first increasing then decreasing in h^*, etc. (see Figure 3.6 for an illustration).

Despite the complicated nature of SPV caused by the interactions among multiple system parameters, the monotonicity property E_v is relatively simple. Among

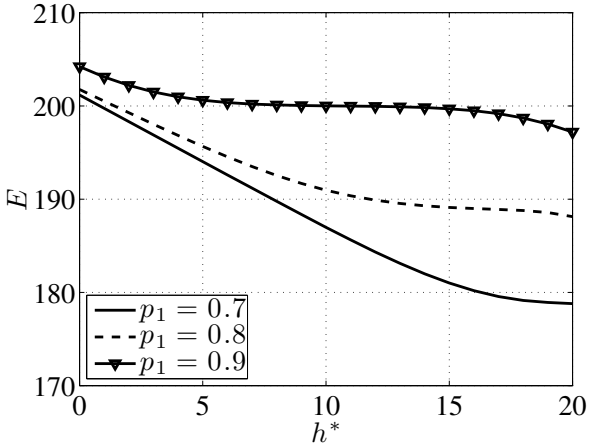


Figure 3.5: E as a function of h^* for $\alpha_1 = \alpha_2 = 1$ ($p_2 = 0.8$, $N = 20$, $T_s = 30$, $T_{\text{shift}} = 100$)

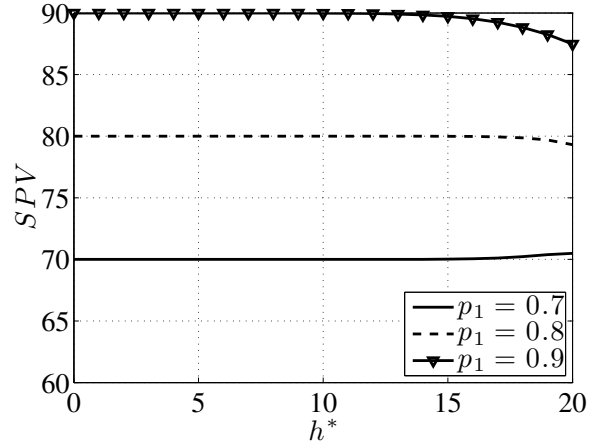


Figure 3.6: SPV as a function of h^* ($p_2 = 0.8$, $N = 20$, $T_s = 30$, $T_{\text{shift}} = 100$)

numerous 10,000 randomly generated cases studied, only two patterns are observed (see Figure 3.7). Therefore, we obtain the following result:

Numerical Fact 3.2. *Consider a production line defined by assumptions (i)-(vi) and (viii). Expected energy consumed per part, E_v, is either monotonically decreasing in h^*, or non-monotonic convex (thus, first decreasing then increasing) in h^*.*

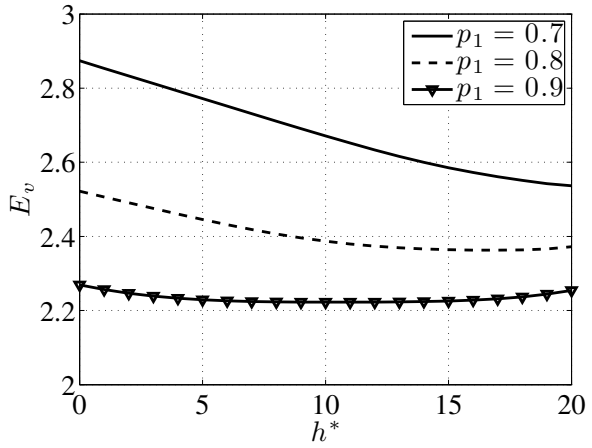


Figure 3.7: E_v as a function of h^* ($p_2 = 0.8$, $N = 20$, $T_s = 30$, $T_{\text{shift}} = 100$)

3.4.2 Optimal startup schedule

To design an optimal startup schedule, we must first specify the objective. Since improving the energy utilization is often the goal in energy efficient manufacturing, we use E_v as the objective for optimal schedule. Based on the monotonicity property of E_v outlined in Proposition 2, the following procedure is proposed:

Optimal schedule design procedure:

(i) For $s = 1$, select $h^*(s) = 0$ and let $E_v(s = 0) = \infty$.

(ii) Evaluate $E_v(s)$ using (4.2).

(iii) If $E_v(s) < E_v(s-1)$ and $h^*(s) < T_s$, then let $h^*(s+1) = h^*(s) + 1$ and return to (b); otherwise, select $h^* = h^*(s)$ and terminate the procedure.

Example: Consider a production line defined by assumptions (i)-(vi) and (viii) with parameters defined as follows:

$$p_1 = 0.8, \quad p_2 = 0.95, \quad N = 40, \quad T_s = 60, \quad T_{\text{shift}} = 450.$$

Assume that $\alpha_1 = \alpha_2 = 1$. Using the optimal schedule design procedure described above, it can be obtained that the optimal control parameter for this system is $h^* = 40$, i.e., machine m_2 will not start until the buffer is full. The performance measures of the open-loop system (i.e., when $h^* = 0$) and the optimal controlled closed-loop system is shown in Table 3.1. As one can see, using the optimal control of startup schedule, the running time of m_2 can be reduced by 11.1%, while the total shift production volume remains practically the same. From the energy point of view, the total energy consumption in a shift is reduced by 5.1% and the energy consumption per part is reduced by 5.0%.

Table 3.1: Performance measure comparisons

	Open-loop	Optimal closed-loop
SPV	360	359.75
RT_2	451.12	401.17
OT_2	1.12	1.14
E	901.12	851.17
E_v	2.50	2.37

Practical implications: The example above illustrates the significance in energy savings for a production system. In fact, the parameters are selected to reflect part of the an automotive paint shop, where m_2 is used to represent the painting booth, while m_1 the cleaning and preparation. In practice, the painting booth usually consists of several lines in parallel, and, thus, has a higher efficiency than the upstream operation. In addition, since the typical cycle time of an automotive assembly line is approximately 1 min per job, we select $T_{\text{shift}} = 450$, which is roughly the duration of an 8-hour shift. Note that during 2000 about 60% of the $700 million energy expenditures in 37 U.S. automotive assembly plants are spent in painting processes. Thus,

a 5% reduction in energy consumption will result in a saving of more than half million dollars per year for each plant. Considering the rapidly increasing energy prices during the past 10 years, the saving would be even more remarkable in present days. It should be pointed out that this significant amount of savings is achieved without any major equipment investment.

3.5 Summary

In this chapter, we study energy reduction problem in Bernoulli serial lines with stripping operations through optimal control of machine startup schedule. Specifically, using transient analysis, analytical mathematical model is developed, which describes the dynamics of the system. In addition, closed-form expressions are derived to evaluate the productivity and energy performances. Based on these expressions, the effects of startup schedule on system performances are discussed and procedures for developing optimal machine startup schedules are formulated. Numerical results show that the optimal schedule can lead to significant improvements in energy utilization efficiency.

CHAPTER IV

CONTROL OF OPERATION SCHEDULE IN MULTI-MACHINE SERIAL BERNOULLI LINES

4.1 Introduction

In this chapter, we extend the research in Chapter III and study the operations schedule problem in the framework of serial production lines with multiple Bernoulli machines and finite buffers.

4.2 Model and Performance Measures

Additional assumptions along with those for Bernoulli serial lines in Chapter II are as follows:

(viii) The system operates on a shift-basis, with shift duration T_{shift} time slots. The startup and shutdown times for machine m_i, $i = 1, \ldots, M$, are $T_{\text{on},i}$ and $T_{\text{off},i}$, respectively. That is, machine m_i is started at the beginning of time slot $T_{\text{on},i}$, where $1 \leq T_{\text{on},i} \leq T_{\text{shift}}$, and shut down at the end of time slot $T_{\text{off},i}$, where $T_{\text{on},i} \leq T_{\text{off},i} \leq T_{\text{shift}}$. In addition, assume that the warm-up time for m_i, $i = 1, \ldots, M$, is $T_{\text{warm},i} \geq 0$, i.e., after m_i is turned on at $T_{\text{on},i}$, it must wait for additional $T_{\text{warm},i}$ time slots to start processing parts.

(ix) Once started, machine m_i, $i = 1, \ldots, M$, consumes α_i units of energy per time slot until it is shut down. Finally, we assume that the system consumes α_0 units

of energy per time slot from the beginning of the shift until the shift is over and all machines are shut down.

Remark: The energy performance of the machines are taken into account in assumption (ix). While α_i's, $i = 1, \ldots, M$, characterize the energy consumption of individual machines, α_0 is used to quantify the energy shared by all operations, for instance, heating and air-conditioning in the plant, etc. Production systems with more complicated energy consumption models will be addressed in future work.

Below, we use the terms "startup/shutdown" and "on/off" interchangeably.

Let RT_i denote the running time of machine m_i, $i = 1, \ldots, M$, i.e., the total time that m_i is on (including warm-up):

$$RT_i = T_{\text{off},i} - T_{\text{on},i} + 1.$$

Then, E and E_v can be evaluated as follows:

$$\begin{aligned} E &= \sum_{i=1}^{M} \alpha_i RT_i + \alpha_0 T_{\text{shift}} \\ &= \sum_{i=1}^{M} \alpha_i (T_{\text{off},i} - T_{\text{on},i} + 1) + \alpha_0 T_{\text{shift}}, \qquad (4.1) \\ E_v &= \frac{E}{SPV}. \qquad (4.2) \end{aligned}$$

In this chapter, we develop methods to evaluate the performance measures of the production system defined above and discuss the effects of machine startup/shutdown schedule on these performances.

4.3 Performance analysis

To analyze the performance of a production line defined by assumptions (i)-(ix), note that the efficiency of machine m_i, $i = 1, \ldots, M$, during time slot n is given by

$$p_i(n) = \begin{cases} p_i, & T_{\text{on},i} + T_{\text{warm},i} \leq n \leq T_{\text{off},i}, \\ 0, & \text{otherwise.} \end{cases}$$

Thus, the production system under consideration can be viewed as a Bernoulli serial line with time-dependent machine efficiencies.

In the current literature, performance analysis of serial production lines with Bernoulli machines during transients have been discussed in [13] and [65]. Specifically, closed-form formulas have been derived to evaluate the productivity performance measures of two-machine (see [13]) and $M > 2$-machine Bernoulli lines with constant machine efficiencies (see [65]). For Bernoulli lines with time-dependent machine efficiencies, only formulas for two-machine lines are available (see [65]). In this section, we develop techniques to evaluate the performance in $M > 2$-machine Bernoulli lines with time-dependent machine efficiencies.

Clearly, the system described above is characterized by a *time-inhomogeneous Markov chain* with $S = \prod_{i=1}^{M-1}(N_i + 1)$ states defined by the occupancy of buffer b_i at the end of time slot n:

$$\mathbf{h}(n) = [h_1(n)\ h_2(n)\ \ldots\ h_{M-1}(n)], \quad h_i(n) \in \{0, 1 \ldots, N_i\},\ i = 1, \ldots, M-1.$$

The dynamics of the production system are defined by the following expressions:

$$\begin{aligned} h_i(n+1) &= h'_i(n+1) + s_i(n+1) \cdot \min\{h_{i-1}(n), N_i - h'_i(n+1), 1\}, \\ &\quad i = 2, \ldots, M-1, \\ h_1(n+1) &= h'_1(n+1) + s_1(n+1) \min\{N_1 - h'_1(n+1), 1\}, \end{aligned} \tag{4.3}$$

where $s_i(n) \in \{1{=}\text{up}, 0{=}\text{down}\}$, $i = 1, \ldots, M$, denotes the status of m_i during time

slot n, and

$$\begin{aligned} h'_{M-1}(n+1) &= h_{M-1}(n) - s_M(n+1) \cdot \min\{h_{M-1}(n), 1\}, \\ h'_i(n+1) &= h_i(n) - s_{i+1}(n+1) \cdot \min\{h_i(n), N_{i+1} - h'_{i+1}(n+1), 1\}, \\ & \quad i = 1, \ldots, M-2. \end{aligned}$$

To calculate the transition probability matrix of this Markov chain, we first rank the system states based on the buffer occupancy, as illustrated in Table 7.1. Thus, given any buffer occupancy status $[h_1 \quad h_2 \quad \ldots \quad h_{M-1}]$, the ranking of the state under this arrangement can be calculated as

$$\text{State ranking} = \sum_{i=1}^{M-1} h_i \beta_{i+1} + 1,$$

where

$$\beta_i = \begin{cases} \prod_{j=i}^{M-1}(N_j + 1), & \text{for } i = 1, \ldots, M-1, \\ 1, & \text{for } i = M. \end{cases}$$

Table 4.1: Ranking the system states

State #	h_1	h_2	$\ldots$	h_{M-2}	h_{M-1}
1	0	0	$\ldots$	0	0
2	0	0	$\ldots$	0	1
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\vdots$
$N_{M-1}+1$	0	0	$\ldots$	0	N_{M-1}
$N_{M-1}+2$	0	0	$\ldots$	1	0
$N_{M-1}+3$	0	0	$\ldots$	1	1
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\vdots$
$S-1$	N_1	N_2	$\ldots$	N_{M-2}	$N_{M-1}-1$
S	N_1	N_2	$\ldots$	N_{M-2}	N_{M-1}

Next, note that for each time slot, the sample space is comprised of a total of 2^M combinations of machine status. The probability of each of these combinations is given by

$$P[s_1(n) = \alpha_1, \ldots, s_M(n) = \alpha_M] = \prod_{i=1}^{M} [p_i(n)]^{\alpha_i} [1 - p_i(n)]^{1-\alpha_i}, \quad \alpha_i \in \{0, 1\}.$$

Thus, for state i, $i = 1, \ldots, S$, we can enumerate all 2^M combinations of machine status and determine the corresponding outcome states using (7.14). Then, the combinations of machine status that lead to the same outcome state j, $j = 1 \ldots, S$, are identified and the probabilities of these combinations are summed up to obtain the transition probability from the original state i to this particular outcome state j. Repeat this procedure for all S states and all transition probabilities can be obtained.

Using the state transition probabilities calculated above, the transition probability matrix of the Markov chain that describes the M-machine Bernoulli line at hand can be obtained. Let $\mathbf{A}_M(n)$ denote this transition probability matrix during time slot n, and $\mathbf{x}(n) = [x_1(n) \ \ldots \ x_S(n)]^T$ denote the probability distribution of the states at the end of time slot n. Thus, the evolution of the system is given by

$$\mathbf{x}(n+1) = \mathbf{A}_M(n)\mathbf{x}(n), \quad \sum_{i=1}^{S} x_i(n) = 1, \quad n = 0, 1, \ldots. \tag{4.4}$$

In Control Theory, vector $\mathbf{x}(n)$ is often referred to as the *state* of linear system (4.4).

Using the mathematical model developed above, it can be shown that the performance of the production line can be calculated as:

$$\begin{aligned}
PR(n) &= \left[\begin{array}{cccc} \mathbf{D}_1(n) & \mathbf{D}_1(n) & \cdots & \mathbf{D}_1(n) \end{array} \right] \mathbf{x}(n), && (4.5)\\
CR(n) &= \mathbf{D}_2(n)\mathbf{x}(n), && (4.6)\\
WIP_i(n) &= \left[\begin{array}{cccc} \mathbf{D}_{3,i} & \mathbf{D}_{3,i} & \cdots & \mathbf{D}_{3,i} \end{array} \right] \mathbf{x}(n), \quad i = 1, \ldots, M-1, && (4.7)\\
ST_i(n) &= \left[\begin{array}{cccc} \mathbf{D}_{4,i}(n) & \mathbf{D}_{4,i}(n) & \cdots & \mathbf{D}_{4,i}(n) \end{array} \right] \mathbf{x}(n), \quad i = 2, \ldots, M, && (4.8)\\
BL_i(n) &= \mathbf{D}_{5,i}(n)\,\mathbf{x}(n), \quad i = 1, \ldots, M-1, && (4.9)
\end{aligned}$$

where $\mathbf{D}_1(n)$, $\mathbf{D}_2(n)$, $\mathbf{D}_{3,i}(n)$, $\mathbf{D}_{4,i}(n)$ and $\mathbf{D}_{5,i}(n)$ are row vectors with β_{M-1}, S, β_i, β_{i-1} and S entries, respectively, given by:

$$\mathbf{D}_1(n) = \left[\begin{array}{cc} 0 & p_M(n+1)\mathbf{J}_{\beta_{M-1}-1} \end{array} \right],$$

$$
\begin{aligned}
\mathbf{D}_2(n) &= p_1(n+1)\mathbf{J}_S - \mathbf{D}_{5,1}(n), \\
\mathbf{D}_{3,i} &= \begin{bmatrix} 0\mathbf{J}_{\beta_{i+1}} & 1\mathbf{J}_{\beta_{i+1}} & \cdots & N_i\mathbf{J}_{\beta_{i+1}} \end{bmatrix}, \\
\mathbf{D}_{4,i}(n) &= \begin{bmatrix} p_i(n+1)\mathbf{J}_{\beta_i} & 0\mathbf{J}_{N_{i-1}\beta_i} \end{bmatrix}, \\
\mathbf{D}_{5,i}(n) &= \begin{bmatrix} d_{5,i,1}(n) & d_{5,i,2}(n) & \cdots & d_{5,i,S}(n) \end{bmatrix}, \\
d_{5,i,j}(n) &= \begin{cases} p_i(n+1)(1 - p_{i+1}(n+1) + d_{5,i+1,j}(n)), \\ \qquad j = (r-1)\beta_i + N_i\beta_{i+1} + 1, \ldots, r\beta_i, \\ \qquad r = 1, \ldots, S/\beta_{i+1}, \\ 0, \text{otherwise}, \end{cases}
\end{aligned}
$$

with $d_{5,M,j}(n) = 0$, $j = 1, \ldots, S$, and $\mathbf{J}_K$ denoting the 1-by-K matrix of ones, i.e., $\mathbf{J}_K = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}$.

4.4 Optimal Operation Schedule

4.4.1 Productivity-energy tradeoff

In a production system defined by assumptions (i)-(ix), it is intuitive and easy to show that

- SPV is monotonically decreasing in $T_{\text{on},i}$, and monotonically increasing in $T_{\text{off},i}$, $i = 1, \ldots, M$;
- E is monotonically decreasing in $T_{\text{on},i}$, and monotonically increasing in $T_{\text{off},i}$, $i = 1, \ldots, M$.

However, in general, average energy consumption per part, E_v, does not possess monotonic properties with respect to $T_{\text{on},i}$ or $T_{\text{off},i}$.

4.4.2 Constrained optimal schedule

In manufacturing practice, although reducing energy consumption is desirable, meeting certain production demand is usually a hard requirement. Therefore, in this

subsection, we use production volume as a constraint and have energy efficiency as the optimization objective.

Consider a production line defined by assumptions (i)-(ix). Introduce the following notions:

Production requirement (R): the shift production volume of the line, SPV, in units of the largest possible shift production volume of the system for any machine startup/shutdown schedule.

Clearly, due to the monotonicity properties of SPV described above, the largest SPV is achieved when all machines are started at the beginning of the shift and shut down at the end of the shift, i.e., when $T_{\text{on},i} = 0$ and $T_{\text{off},i} = T_{\text{shift}}$, for all $i = 1, \ldots, M$. We denote this shift production volume as SPV_{max}. Then, the production requirement can be expressed as

$$R = \frac{SPV}{SPV_{max}}.$$

Optimal operation schedule under production requirement R: the machine startup times $\mathbf{T}_{\text{on},R} = [T_{\text{on},1,R}, \ldots, T_{\text{on},M,R}]$ and shutdown times $\mathbf{T}_{\text{off},R} = [T_{\text{off},1,R}, \ldots, T_{\text{off},M,R}]$ such that the production requirement R is achieved while the resulting E_v is minimized.

Exact formulas for optimal operation schedule are all but impossible to derive. Therefore, we propose the following iterative procedure based on greedy algorithm to obtain a sub-optimal one:

Algorithm 1: Given $R \in (0, 1]$,

1) Set $k = 0$, $T_{\text{on},i,R}(k) = 0$ and $T_{\text{off},i,R}(k) = T_{\text{shift}}$, $i = 1, \ldots, M$. Define

$$\begin{aligned} \mathbf{T}_{\text{on},R}(k) &= [T_{\text{on},1,R}(k), \ldots, T_{\text{on},M,R}(k)], \\ \mathbf{T}_{\text{off},R}(k) &= [T_{\text{off},1,R}(k), \ldots, T_{\text{off},M,R}(k)]. \end{aligned}$$

2) Using Recursive Procedure 1, evaluate the energy consumption per part and

shift production volume under machine startup schedule $\mathbf{T}_{\text{on},R}(0)$ and shutdown schedule $\mathbf{T}_{\text{off},R}(0)$, denoted as

$$\widehat{E}_v(\mathbf{T}_{\text{on},R}(0), \mathbf{T}_{\text{off},R}(0)),$$
$$\widehat{SPV}(\mathbf{T}_{\text{on},R}(0), \mathbf{T}_{\text{off},R}(0)),$$

and, therefore, $\widehat{SPV}_{max} = \widehat{SPV}(\mathbf{T}_{\text{on},R}(0), \mathbf{T}_{\text{off},R}(0))$. In addition, denote the "best-so-far" operations schedule and the corresponding performance measures as:

$$\begin{aligned}
\mathbf{T}^{best}_{\text{on},R} &= \mathbf{T}_{\text{on},R}(0), \\
\mathbf{T}^{best}_{\text{off},R} &= \mathbf{T}_{\text{off},R}(0), \\
\widehat{SPV}_{best} &= \widehat{SPV}(\mathbf{T}_{\text{on},R}(0), \mathbf{T}_{\text{off},R}(0)), \\
E_{best} &= E(\mathbf{T}_{\text{on},R}(0), \mathbf{T}_{\text{off},R}(0)), \\
\widehat{E}^{best}_v &= \widehat{E}_v(\mathbf{T}_{\text{on},R}(0), \mathbf{T}_{\text{off},R}(0)).
\end{aligned}$$

3) For all $i \in \{1, \ldots, M\}$, calculate

$$\widehat{E}_v(\mathbf{T}_{\text{on},R}(k) + \mathbf{e}_i, \mathbf{T}_{\text{off},R}(k)),$$
$$\widehat{SPV}(\mathbf{T}_{\text{on},R}(k) + \mathbf{e}_i, \mathbf{T}_{\text{off},R}(k)),$$

where $\mathbf{e}_i$ is the 1-by-M row vector with the ith element equal to 1 and all others 0.

4) Determine the set, $S_{\text{on}}(k) \subseteq \{1, \ldots, M\}$, of all machines that satisfy the following condition:

$$\widehat{SPV}(\mathbf{T}_{\text{on},R}(k) + \mathbf{e}_i, \mathbf{T}_{\text{off},R}(k)) \geq R \cdot \widehat{SPV}_{max}, \quad \forall i \in S_{\text{on}}(k).$$

If $S_{\text{on}}(k)$ is empty, then go to Step 7); otherwise, identify machine m_j within

$S_{\text{on}}(k)$ such that

$$\widehat{E}_v(\mathbf{T}_{\text{on},R}(k)+\mathbf{e}_j, \mathbf{T}_{\text{off},R}(k)) < \widehat{E}_v(\mathbf{T}_{\text{on},R}(k)+\mathbf{e}_l, \mathbf{T}_{\text{off},R}(k)),\ \forall l \neq j,\ l \in S_{\text{on}}(k).$$

In addition, if $\widehat{E}_v^{best} > \widehat{E}_v(\mathbf{T}_{\text{on},R}(k)+\mathbf{e}_j, \mathbf{T}_{\text{off},R}(k))$, then update the "best-so-far" solution:

$$\begin{aligned}
\widehat{SPV}_{best} &= \widehat{SPV}(\mathbf{T}_{\text{on},R}(k)+\mathbf{e}_j, \mathbf{T}_{\text{off},R}(k)),\\
E_{best} &= E(\mathbf{T}_{\text{on},R}(k)+\mathbf{e}_j, \mathbf{T}_{\text{off},R}(k)),\\
\widehat{E}_v^{best} &= \widehat{E}_v(\mathbf{T}_{\text{on},R}(k)+\mathbf{e}_j, \mathbf{T}_{\text{off},R}(k)),\\
\mathbf{T}_{\text{on},R}^{best} &= \mathbf{T}_{\text{on},R}(k)+\mathbf{e}_j,\\
\mathbf{T}_{\text{off},R}^{best} &= \mathbf{T}_{\text{off},R}(k).
\end{aligned}$$

Finally, let $\mathbf{T}_{\text{on},R}(k) = \mathbf{T}_{\text{on},R}(k)+\mathbf{e}_j$ and continue to Step 5).

5) For all $i \in \{1,\ldots,M\}$, calculate

$$\widehat{E}_v(\mathbf{T}_{\text{on},R}(k), \mathbf{T}_{\text{off},R}(k)-\mathbf{e}_i),$$
$$\widehat{SPV}(\mathbf{T}_{\text{on},R}(k), \mathbf{T}_{\text{off},R}(k)-\mathbf{e}_i).$$

6) Determine the set, $S_{\text{off}}(k) \subseteq \{1,\ldots,M\}$, of all machines that satisfy the following condition:

$$\widehat{SPV}(\mathbf{T}_{\text{on},R}(k), \mathbf{T}_{\text{off},R}(k)-\mathbf{e}_i) \geq R \cdot \widehat{SPV}_{max}, \quad \forall i \in S_{\text{off}}(k).$$

If $S_{\text{off}}(k)$ is empty, then go to Step 7); otherwise, identify machine m_j within $S_{\text{off}}(k)$ such that

$$\widehat{E}_v(\mathbf{T}_{\text{on},R}(k), \mathbf{T}_{\text{off},R}(k)-\mathbf{e}_j) < \widehat{E}_v(\mathbf{T}_{\text{on},R}(k), \mathbf{T}_{\text{off},R}(k)-\mathbf{e}_l),\ \forall l \neq j,\ l \in S_{\text{off}}(k).$$

In addition, if $\widehat{E}_v^{best} > \widehat{E}_v(\mathbf{T}_{\text{on},R}(k), \mathbf{T}_{\text{off},R}(k)-\mathbf{e}_j)$, then update the "best-so-far"

solution:

$$\begin{aligned}
\widehat{SPV}_{best} &= \widehat{SPV}(\mathbf{T}_{\text{on},R}(k), \mathbf{T}_{\text{off},R}(k) - \mathbf{e}_j), \\
E_{best} &= E(\mathbf{T}_{\text{on},R}(k), \mathbf{T}_{\text{off},R}(k) - \mathbf{e}_j), \\
\widehat{E}_v^{best} &= \widehat{E}_v(\mathbf{T}_{\text{on},R}(k), \mathbf{T}_{\text{off},R}(k) - \mathbf{e}_j), \\
\mathbf{T}_{\text{on},R}^{best} &= \mathbf{T}_{\text{on},R}(k), \\
\mathbf{T}_{\text{off},R}^{best} &= \mathbf{T}_{\text{off},R}(k) - \mathbf{e}_j.
\end{aligned}$$

Finally, let $\mathbf{T}_{\text{off},R}(k) = \mathbf{T}_{\text{on},R}(k) - \mathbf{e}_j$, update $k = k + 1$, and return to Step 3).

7) Let

$$\begin{aligned}
\widehat{SPV}_{end} &= \widehat{SPV}(\mathbf{T}_{\text{on},R}(k), \mathbf{T}_{\text{off},R}(k)), \\
E_{end} &= E(\mathbf{T}_{\text{on},R}(k), \mathbf{T}_{\text{off},R}(k)), \\
\widehat{E}_v^{end} &= \widehat{E}_v(\mathbf{T}_{\text{on},R}(k), \mathbf{T}_{\text{off},R}(k)), \\
\mathbf{T}_{\text{on},R}^{end} &= \mathbf{T}_{\text{on},R}(k), \\
\mathbf{T}_{\text{off},R}^{end} &= \mathbf{T}_{\text{off},R}(k),
\end{aligned}$$

and terminate the algorithm.

After the algorithm is terminated, machine startup schedule $\mathbf{T}_{\text{on},R}^{best}$ and shutdown schedule $\mathbf{T}_{\text{off},R}^{best}$ represent the operations schedule that has the smallest $\widehat{E}_v$ among all schedules evaluated during the course of the algorithm, and thus, is adopted as the final solution. Note that based on our observation, among all solutions visited in Algorithm 1, operations schedule $[\mathbf{T}_{\text{on},R}^{end} \quad \mathbf{T}_{\text{off},R}^{end}]$ is practically always the one that leads to the $\widehat{SPV}$ closest to the production requirement. However, this schedule does not necessarily have the smallest $\widehat{E}_v$. Below, we refer to these two schedules as the *best solution* and the *ending solution* of Algorithm 1.

It should be noted that in practice real-time feedback control-based scheduling is more desirable than off-line optimization because the "optimized" solutions are

usually sensitive to system parameters, which are often identified with errors. Therefore, in the current paper, we do not pursue the optimality of the greedy algorithm formulated above but rather use it to demonstrate the potential improvement in energy-efficiency one can obtain by appropriately designing machine operation schedule. Investigations of system structural properties and development of effective and efficient feedback controllers as well as robust optimization algorithms are considered as important directions of future work.

4.5 Case Study

4.5.1 Motivation

Studies have shown that approximately 60% of the total energy consumed in an automotive assembly plant is used during the painting process (see [10]). Further investigation indicates that most of this energy is used for heating, ventilation, and air-conditioning (HVAC) in the booths and ovens of the paint shop (see [11]). In addition to enormous amount of energy consumption, the painting process is a huge source of greenhouse gas emissions as well. Indeed, during the painting process, considerable amount of volatile organic compound (VOC) is generated and needs to be destructed before being released to the environment [12]. Due to the importance of painting in automotive assembly and its critical role in energy consumption, significant research efforts have been dedicated to improving its productivity and quality performances (see [42, 43, 44, 45]). From the energy perspective, [46] develops a control strategy in order to reduce the HVAC utility in the paint booth of an automotive assembly plant, while [47] presents several guidelines to improve energy efficiency in automotive manufacturing industry. In addition, [66] uses linear programming to optimize the environmental performance in passenger car coating process. Despite these studies, the effects of operations schedule on paint shop productivity and energy performance have not been investigated, and, thus, are discussed in this section.

4.5.2 System parameters selection

To carry out the analysis, consider an 11-machine Bernoulli serial line defined by assumptions (i)-(viii) with machine efficiencies, warm-up times, energy consumption rates as well as buffer capacities and initial occupancies are given in Table 4.2. In addition, assume that $T_{\text{shift}} = 500$ and $\alpha_0 = 5$. Note that these parameters are chosen so that they reflect typical production scenario in automotive paint shops (see [11], [42, 43, 44, 45]).

Table 4.2: Paint shop production system parameters

Machine	p_i	$T_{\text{warm},i}$	α_i	N_i	$h_i(0)$
1	0.80	20	5	5	4
2	0.82	20	13	8	3
3	0.75	15	7	10	2
4	0.80	5	12	3	0
5	0.75	10	3	3	2
6	0.76	10	12	8	0
7	0.77	30	8	25	12
8	0.70	5	1	6	5
9	0.82	10	12	11	0
10	0.80	30	12	14	0
11	0.78	5	1	-	-

Using Recursive Procedure, it can be obtained that $\widehat{SPV}_{max} = 328.15$ (i.e., when $T_{\text{on},i} = 0$ and $T_{\text{off},i} = T_{\text{shift}}$) and the corresponding total energy consumption and energy consumption per job are $E = 45586$ and $\widehat{E}_v = 138.92$, respectively.

4.5.3 Effects of operations schedule on system performance

To illustrate the consequence of an ill-designed machine startup and shutdown schedule, assume that the following machine operation schedule is used:

$$\begin{aligned} \mathbf{T}_{\text{on}} &= [2,\ 12,\ 16,\ 10,\ 25,\ 13,\ 20,\ 32,\ 5,\ 0,\ 3,\ 25], \\ \mathbf{T}_{\text{off}} &= [450, 463, 446, 497, 500, 477, 490, 483, 472, 490, 497] \end{aligned} \tag{4.10}$$

The resulting transients of PR, CR and WIP are evaluated using Recursive Pro-

cedure and simulations, and shown in Figure 4.1. Note that Recursive Procedure still provides highly accurate estimation for the transients of the performance measures. Note also that under this machine startup/shutdown schedule, both $PR(n)$ and $CR(n)$ exhibit strong oscillations at the beginning of the shift. Such oscillations are due to machine starvations and blockages caused by the inappropriate operations schedule. In addition, these oscillations will lead to unnecessary production losses and wasted energy during transients. In fact, under the machine startup and shutdown schedule above, the system performance can be obtained as follows:

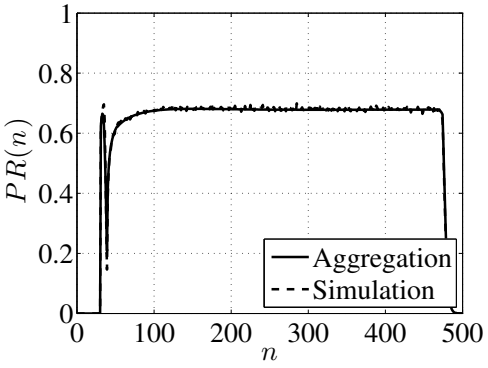
$$\widehat{SPV} = 299.92, \quad E = 42725, \quad \widehat{E}_v = 142.47.$$

As one can see, the inappropriate machine startup and shutdown schedule leads to less production and more energy consumption per job.

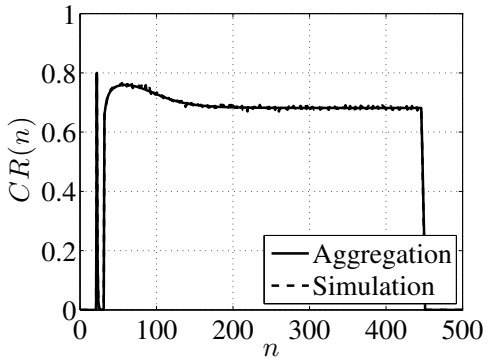
To obtain an energy-efficient operations schedule, Algorithm 1 can be applied. Assume the shift production volume of 299.92 is required, i.e., the production requirement is $R = 0.914$. Then, the best solution from Algorithm 1 for the system under consideration is as follows:

$$\begin{aligned}
\widehat{SPV}_{best} &= 306.31, \\
E_{best} &= 40442, \\
\widehat{E}_v^{best} &= 132.03, \\
\mathbf{T}_{\text{on},R}^{best} &= [16, 14, 17, 28, 24, 25, 13, 46, 43, 37, 75], \\
\mathbf{T}_{\text{off},R}^{best} &= [432, 437, 448, 459, 463, 464, 469, 496, 497, 498, 500].
\end{aligned}$$

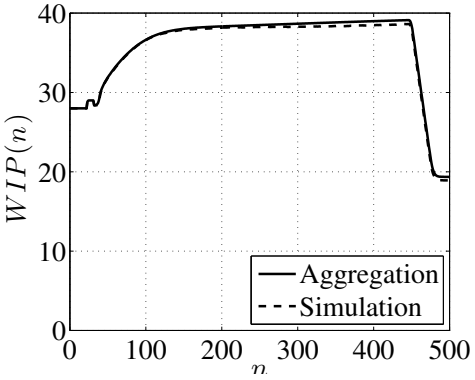
Therefore, the total energy consumption and energy consumption per part are reduced by 5.34% and 7.33%, respectively, from those obtained under operation schedule (4.10). The resulting transients of PR, CR, and WIP are shown in Figure 4.2. As illustrated in the figure, the performance measures exhibit smoother transients compared to those shown in Figure 4.1. Moreover, the best solution from Algorithm



(a) Production rate



(b) Consumption rate

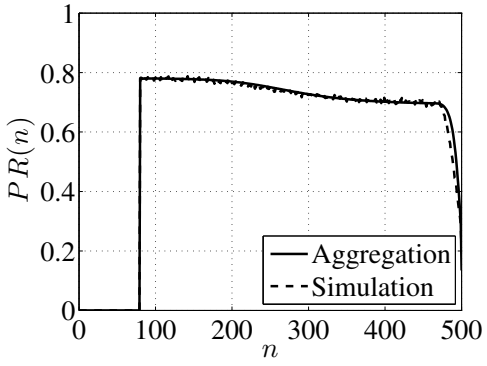


(c) Work-in-process

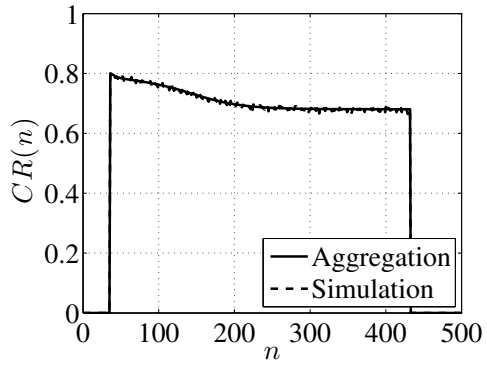
Figure 4.1: System performance under ill-designed operations schedule

1 also reduces E and $\widehat{E}_v$ by 11.3% and 4.96% from operations schedule $T_{\text{on},i} = 0$, $T_{\text{off},i} = T_{\text{shift}}$, $i = 1, \ldots, M$. It should be pointed out that during 2000 about 60% of the $700 million energy expenditures in 37 U.S. automotive assembly plants are spent in painting processes. Thus, a 5% reduction in energy consumption will result in a saving of more than half million dollars per year for each plant. Considering the rapidly increasing energy prices during the past 10 years, the saving would be even more remarkable in present days. Note also that this significant amount of savings is achieved without any major equipment investment.

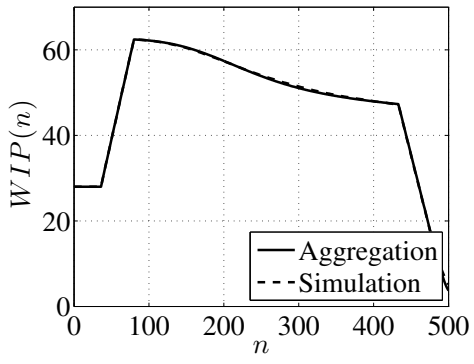
Finally, we investigate the productivity and energy performance measures, obtained under the best and ending solutions of Algorithm 1, as functions of production requirement R in Figure 4.3. As one can see, all three performance measures obtained under the best solution of Algorithm 1 are monotonically increasing in R. For the ending solution of Algorithm 1, however, the resulting $\widehat{SPV}$ and E are monotonically increasing in R, while $\widehat{E}_v$ is non-monotonic. This is because, the best solution focuses on achieving lower energy consumption per job, while the ending solution aims at cutting the production volume as close to the requirement as possible. As a result, although the ending solution can lead to a smaller total energy consumption when R is below a certain level, it may suffer from a lower energy efficiency due to smaller production volume. Thus, in production practice, the decision maker should not simply focus on cutting total energy consumption but should examine the overall energy efficiency in terms of energy consumption per job.



(a) $\widehat{PR}(n)$

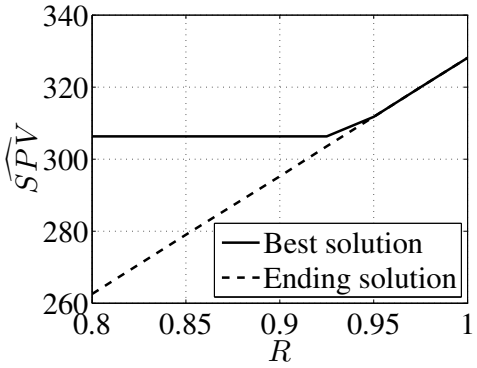


(b) $\widehat{CR}(n)$

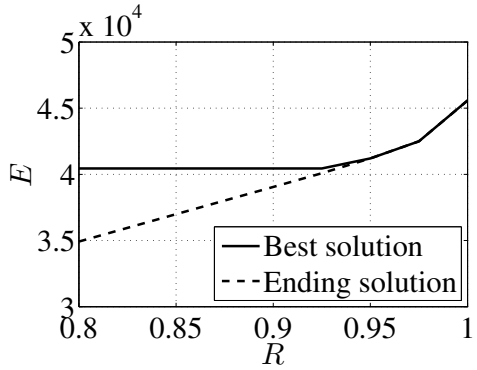


(c) $\widehat{WIP}(n)$

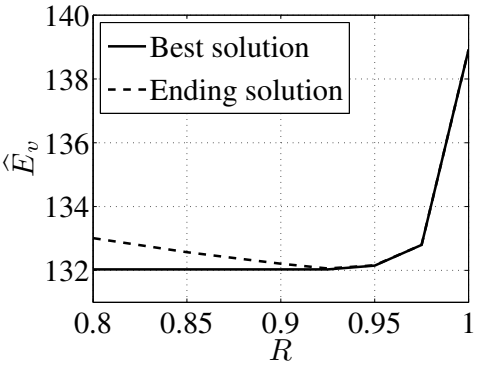
Figure 4.2: Transients of performance measures under improved operations schedule



(a) $\widehat{SPV}$ vs. R



(b) E vs. R



(c) $\widehat{E}_v$ vs. R

Figure 4.3: Productivity and energy performance as functions of R

4.6 Summary

In this chapter, we study productivity and energy performance in Bernoulli serial lines with machine startup and shutdown schedule. Since the system operates in transient regimes, steady state analysis is not applicable. Using Markovian analysis, closed-form expressions are provided to calculate the performance measures of Bernoulli serial lines with time-dependent machine efficiencies, and a recursive procedure based on aggregation is developed. Based on this technique, the effects of machine startup and shutdown schedule on system performances are discussed and a greedy algorithm-based procedure for obtaining sub-optimal operations schedules is formulated. Numerical results show that the operations schedule can lead to significant improvements in energy efficiency.

CHAPTER V

ORDER COMPLETION TIME IN BERNOULLI SERIAL SYSTEMS

5.1 Introduction

This chapter addresses analytical evaluation of order completion time in multi-machine serial production lines with Bernoulli reliability and finite buffer capacity using transient analysis of the systems.

5.2 Model and Performance Measures

5.2.1 Model description

The considered production system is illustrated in Figure 2.1 with assumptions (i)-(vii) in Chapter II. An additional assumption is as follows:

(viii) All buffers are empty initially and all machines begin to work until a batch of D products are produced and no products are left in the buffers.

Batch/lot-based production operations are widely used in various manufacturing systems. However, due to its finite size, these production activities are operated partially or entirely in the transient regime, and, therefore, steady state analysis is, strictly speaking, not applicable and transient analysis of the system must be carried out.

5.2.2 Performance measures

The productivity performance measures of interest are:

- Production rate, $PR(n)$: the expected number of finished parts produced by m_M during time slot $n+1$;
- Consumption rate, $CR(n)$: the expected number of raw parts consumed by m_1 during time slot $n+1$;
- Work-in-process, $WIP_i(n)$: the expected number of parts in buffer b_i, $i = 1, \ldots, M-1$, at the beginning of time slot $n+1$;
- Completion time, CT: the expected completion time of the batch.

5.3 Analysis

5.3.1 Linear approximation

We assume in this section that the Bernoulli line has identical capacity for each buffer and identical efficiency for each machine. There is a special case that can be addressed analytically, where $D = 1$, i.e., the batch contains only one product. Then we have the following theorem:

Theorem 5.1. *The completion time for Bernoulli serial lines with identical machines and buffers if $D = 1$ is*

$$E(CT|D=1) \;=\; \sum_{i=0}^{\infty}(M+i)\binom{M+i-1}{M-1}p^M(1-p)^i. \tag{5.1}$$

Proof. Note that N has no impact on CT in this case and the shortest possible completion time is M. The probability mass function of CT is given by

$$P\{CT = M+i\} = \binom{M+i-1}{M-1}p^M(1-p)^i,\ i = 0, \ldots, \infty.$$

Taking expectation of the probabilities proves the theorem. Besides, according to

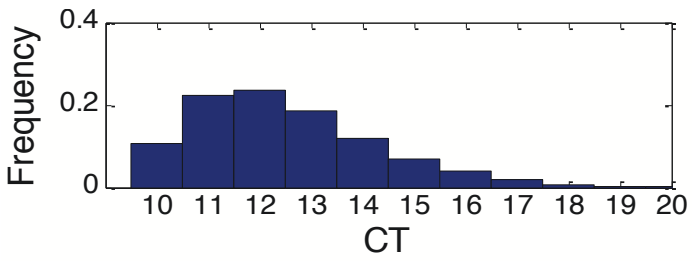


Figure 5.1: Linear relationship between D and CT in a 10-machine Bernoulli line with efficiency of 0.8 and buffer capacity of 4

the fact that

$$\sum_{i=0}^{\infty} \binom{M+i-1}{M-1} (1-p)^i \;=\; \frac{1}{1-(1-p)^M},\ |1-p|<1,$$

it can be proved without much effort that

$$\sum_{i=0}^{\infty} P\{T = M+i\} \;=\; 1.$$

□

A 10-machine Bernoulli line is provided to illustrate the probability mass function of CT obtained from simulation (see Figure. 5.1). As expected, the probability mass function has single mode with a heavy right tail. Systems with distinct machines and buffers maintain similar properties.

For $D > 1$, simulation experiments show that CT is approximately linear in D. As an illustration, Figure. 5.2 displays this relationship in the system as in Figure 5.1. The linear relationship can be explained by PR. Research (see [4]) shows that Bernoulli serial systems have short transient period and PR enters steady state quickly. Therefore, extra operation time needed for extra demands will be approximately a linear function of PR in steady states (PR_{ss}).

Thanks to the linearity between D and CT, a lower bound of CT can be given by a straight line

$$E(CT) = kD + b,$$

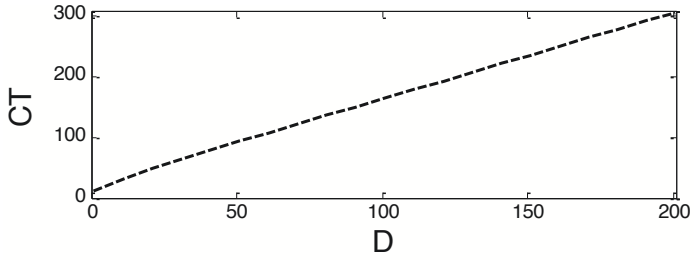


Figure 5.2: Linear relationship between D and CT in the system as in Figure. 5.1

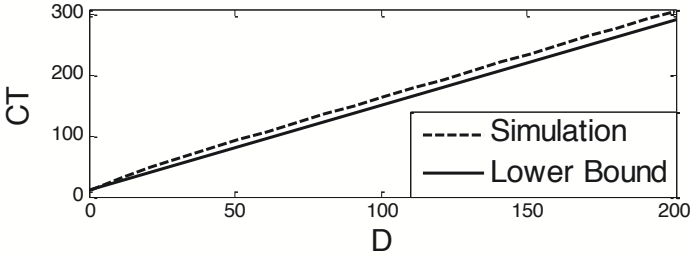


Figure 5.3: Lower bound of CT in the system as in Figure. 5.1

where $k = \frac{1}{PR_{ss}}, b = E(CT|D = 1) - \frac{1}{PR_{ss}}$, and $E(CT|D = 1)$ is from equation 5.1. It is intuitive and can be proved that intercept b converges to $M - 1$ as efficiency approaches 1. Figure. 5.3 is an example of this linear approximation.

The reason that the line is a lower bound of CT is intuitive. At the beginning of the production, PR increases from zero to steady state level due to initial empty buffer occupancy. As a result, the reciprocal of PR, i.e., the slope of the line, is greater than that in actual line at the beginning. The lower bound also indicates production loss during transient period resulting to longer CT.

The difference between actual CT and lower bound of CT worths our attention. Two additional Bernoulli lines are provided for illustrations in Figure. 5.4. The first line (top in Figure. 5.4) has 15 machines with efficiency of 0.7 and buffer capacity of 3. The second line (bottom in Figure. 5.4) has 20 machines with efficiency of 0.9 and buffer capacity of 2. Table 5.1 summarizes differences between actual CT and lower bound of CT for different D in absolute errors and relative errors.

Experiments show that if D is large, or the system enters steady state quickly, i.e., with small M and/or greater p_i, the relative disparity between actual and lower

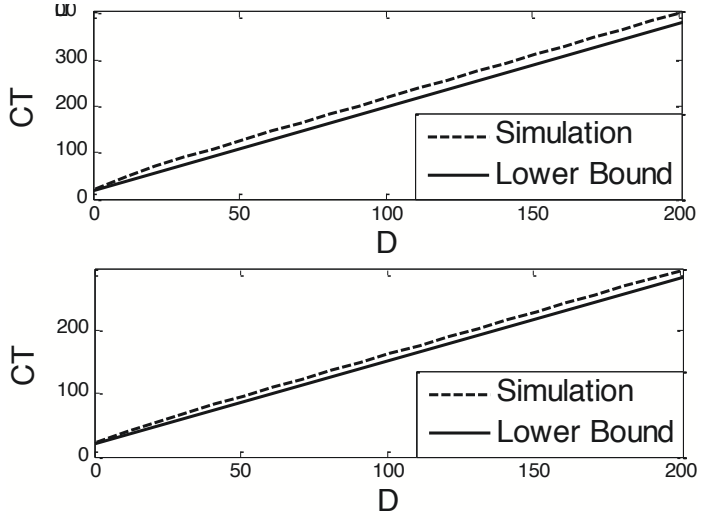


Figure 5.4: Lower bounds in two additional lines

Table 5.1: Difference between actual CT and lower bound of CT (relative errors in parentheses)

Line	$D = 10$	$D = 20$	$D = 50$	$D = 100$	$D = 200$
1	5 (17%)	7 (16%)	10 (11%)	12 (8%)	14 (4%)
2	10 (22%)	13 (19%)	17 (14%)	19 (9%)	23 (6%)
3	4 (11%)	6 (12%)	8 (9%)	10 (6%)	12 (4%)

bound CT will become small and the lower bound could be an acceptable estimate of CT if accuracy is not highly required.

The major advantage of the lower bound is that, since all values can be obtained by formulas, the computation burden is much less than simulation and aggregation technique introduced later, especially for large D.

5.3.2 Recursive aggregation approximation

Obviously CT can be obtained using simulation. However, due to long simulation time, this method cannot satisfy real time analysis, control or scheduling optimization. We propose an estimate method based on recursive aggregation technique. In this method, a virtual machine m_0 and a virtual buffer b_0 are added to the beginning of the Bernoulli line, with $M + 1$ machines and M buffers in total in the system, as shown in Figure. 5.5. The virtual machine has efficiency of 0 all the time and the virtual buffer has initial occupancy of D.

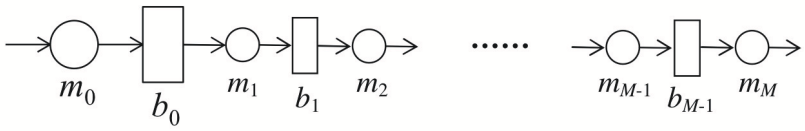


Figure 5.5: Bernoulli serial line with virtual machine and buffer

To approximate the steady state performance of Bernoulli serial lines, a recursive aggregation procedure has been developed in [4], which is extended to evaluate transient period performance of Bernoulli serial lines in [67].

The system as in Figure. 5.1 with $D = 100$ is used to compare performance measures in simulation and aggregation. $PR(n)$ and $CR(n)$ are compared in Figure. 5.6 while their relative errors, defined as follows, are shown in Figure. 5.7:

$$\varepsilon_{PR}(n) = \frac{\widehat{PR}(n) - PR(n)}{PR_{ss}} \cdot 100\%,$$
$$\varepsilon_{CR}(n) = \frac{\widehat{CR}(n) - CR(n)}{CR_{ss}} \cdot 100\%,$$

where $\widehat{PR}(n)$ is estimated from aggregation and $PR(n)$ from simulation. The same notation applies to $CR(n)$ in the definition. $PR_{ss} = CR_{ss} = 0.7106$ in this system. The total work in process $WIP(n)$, defined as follows, and WIP_1, WIP_5, WIP_9 are compared in Figure. 5.8:

$$WIP(n) = \sum_{i=1}^{M-1} WIP_i(n)$$

The aggregation approach works very well before the virtual buffer depletes, but it differs from simulation since then, except $WIP(n)$. However, since we focus on average completion time rather than the performance during the production duration, such distortion does not impact this method significantly.

There is an issue that needs to address: how to cut off the line to obtain CT? Ideally we wish to find a time slot n^* so that the sum of production rate from time slot 0 to n^* is equal to D, which is almost impossible and inappropriate, since PR will approach zero but is almost always greater than zero. Therefore we need to provide

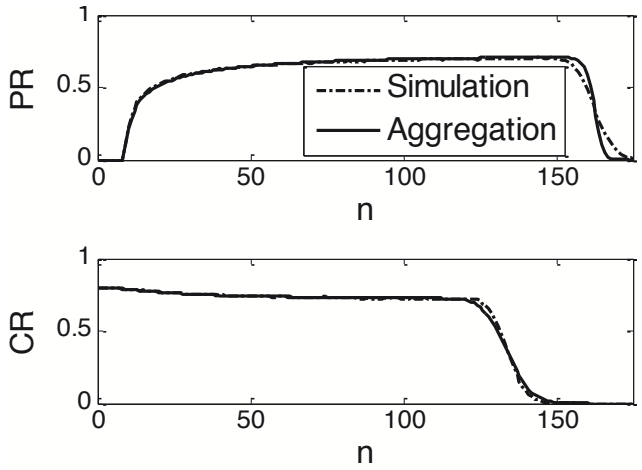


Figure 5.6: PR and CR comparison in simulation and aggregation in the system as in Figure. 5.1

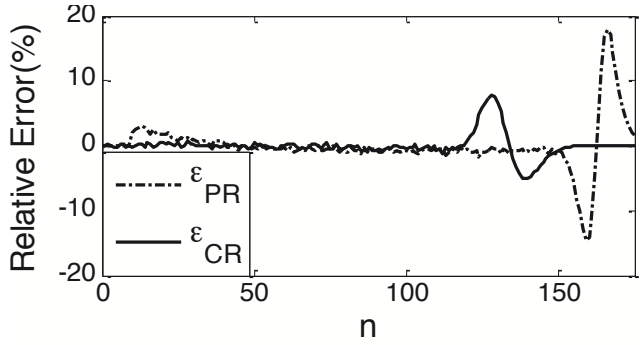


Figure 5.7: Relative errors in simulation and aggregation in the system as in Figure. 5.1

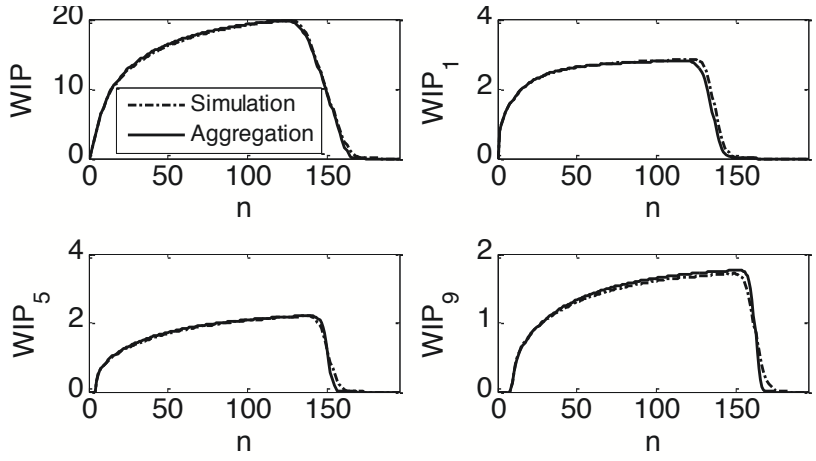


Figure 5.8: WIP comparison in simulation and aggregation in the system as in Figure. 5.1

Table 5.2: Parameters and values in cutoff experiment

Parameter	Value Set
M	$\{5, 10, 15, 20, 25, 30\}$
N	$\{2, 4, 6\}$
p	$\{0.7, 0.8, 0.9\}$
D	$\{30, 60, 100, 150, 200\}$

Table 5.3: Proposed cutoffs

D	(1,50)	[50,100]	(100,∞)
Cutoff	97%	98.5%	99%

cutoff criteria, such as 99%, 99.5%, etc., of D, to obtain expected completion time. Let r denote the cutoff, then the CT is the smallest n^* such that

$$\sum_{n=0}^{n^*} \widehat{PR}(n) \geq D \cdot r.$$

In order to investigate the relation between cutoff criteria and CT, experiment with 3,000 lines was conducted with parameters shown in Table 5.2.

Based on experiment, we propose appropriate cutoffs in Table 5.3, whose feasibility is evaluated by

$$\varepsilon_r = \frac{\mid \widehat{CT} - CT \mid}{CT} \cdot 100\%.$$

The experiment result is summarized in Figure. 5.9. 52% of ε_r are less than 1% and 94% are less than 3%, which is acceptable considering relatively small D (less than 200). Another experiment allowing distinct machines and distinct buffers shows similar accuracy.

Another experiment was conducted to test the proposed cutoff accuracy in more general Bernoulli systems. In these systems, the buffer capacities can be different from each other, and the efficiencies for each machine at each time slot can be different too, i.e., $p_i(n)$ is randomly chosen from 0.7 to 0.99 for any i and any n. Buffer capacity can be different from each other, i.e., N_i is randomly chosen in the integer set from 2

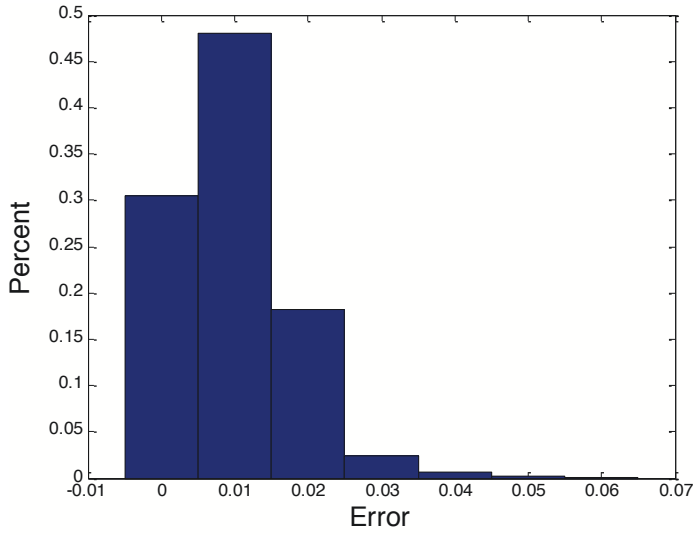


Figure 5.9: Errors in aggregation with identical machines and buffers

Table 5.4: Parameters and values in cutoff test in general cases

Parameter	Value Set
M	[5 40]
N_i	[2 6]
$p_i(n)$	[0.7 0.99]
D	[20 300]

to 6 for any i. System parameters and value set are shown in Table 5.4. 10,000 lines are included in this experiment.

The experiment result is summarized in Figure. 5.10, which shows that 93% of the errors are within 2%. This is an acceptable result since the system parameters obtained from facilities normally contains around 5% of discrepency. The experiment also shows that small D leads to lowest accuracy.

As an illustration, Figure. 5.11 shows the best cutoff, i.e., under which CT is equal to that in simulation, in a 10-machine line. In the figure, best cutoff increases fast from $D = 1$ to around $D = 50$, which slows down while $D \geq 50$ and almost remains flat for $D > 100$. This is why D is broke into three intervals. Of course, the proposed cutoffs do not guarantee the best result and other cutoffs with more breakdown intervals for D will provide better results. However, they are simple and

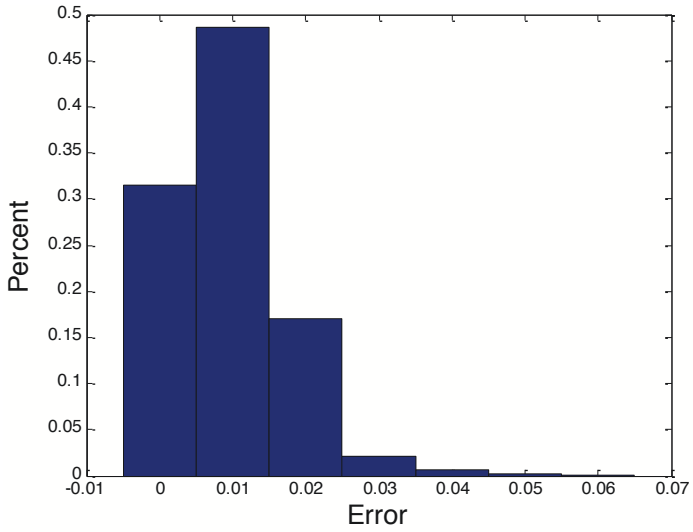


Figure 5.10: Errors in aggregation in general systems

provide acceptable level of accuracy.

There are two advantages in this method. The first one is that it can apply to systems with distinct machines and buffers which are seen more often in practice. Besides, it can also apply to systems with time varying parameters. The second one is that it requires less computation time than simulation for small D, say, less than 100. However, when D is much larger, the computation time is higher than simulation.

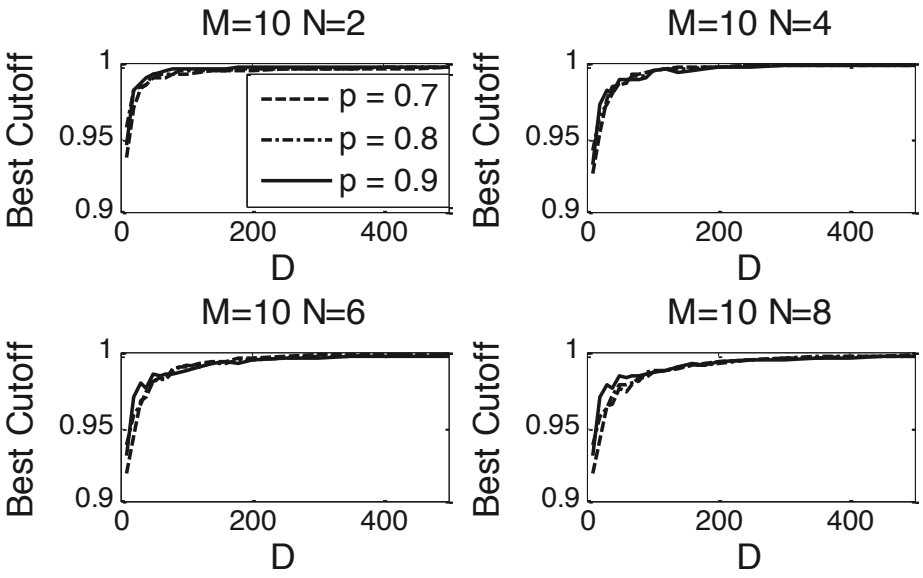


Figure 5.11: Best cutoff in 10-machine systems

5.4 Structural Properties

5.4.1 Reversibility

Consider a Bernoulli serial line L and its reverse L_r (see Figure. 5.12), experiments support the following numerical fact:

Numerical Fact 5.1. *The performance measures of a serial Bernoulli line, L, and its reverse, L_r, are related as follows:*

$$PR^L(n) = PR^{L_r}(n),\ CT^L = CT^{L_r}.$$

Justification: Two experiments, each with 10,000 lines, are conducted to justify the reversibility, one in simulation and the other in aggregation, with parameter values

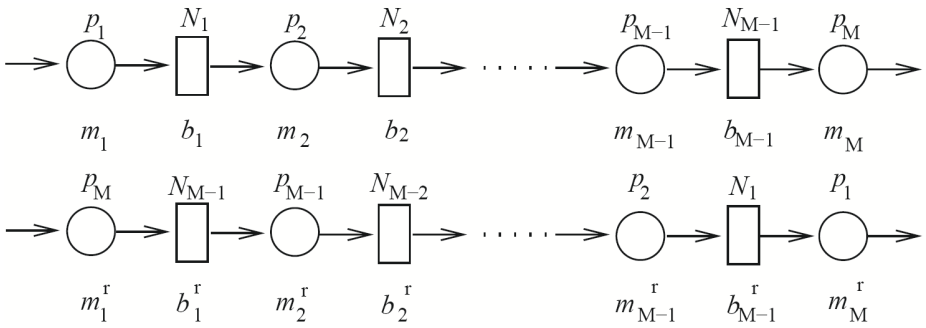


Figure 5.12: M-machine Bernoulli serial line and its reverse

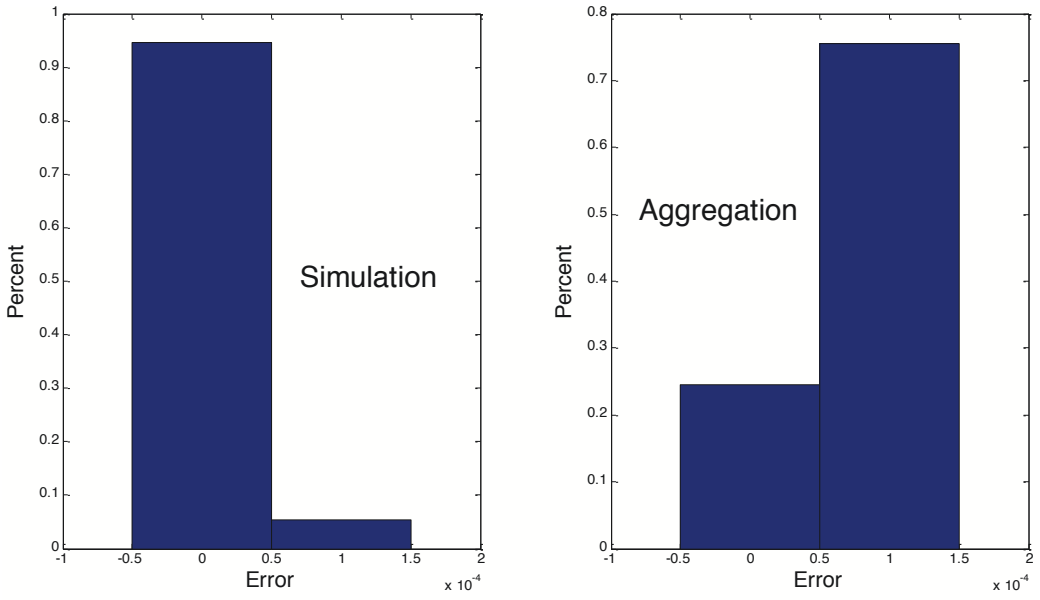


Figure 5.13: Reversibility test in simulation (left) and aggregation (right)

Table 5.5: Parameters and values in reversibility test

Parameter	Value Set
M	[3 40]
N_i	[2 6]
$p_i(n)$	[0.5 0.99]
D	[20 300]

in Table 5.5. Figure. 5.13 summarizes the relative errors in simulation aggregation, whose values are calculated similarly in Figure. 5.10.

Results show that in simulation the error is almost less than 0.00% and the largest error is about 0.06%; while in aggregation, the error is within 0.01% in most cases, the largest error is 6.15% and only 425 lines have errors greater than 2%, which are 4.25% of total lines. Note that other performance measures do not hold this property.

As an illustration, obtained from simulation, performance measures of a 8-machine line with $D = 82$ are shown in Figure. 5.14 and its reverse in Figure. 5.15. Original line's parameters are as follows:

$$N = \{4, 6, 3, 4, 6, 3, 5\},$$

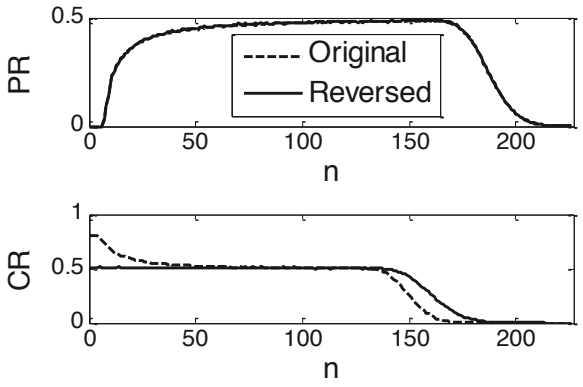


Figure 5.14: PR and CR of a 8-machine line and its reverse in simulation

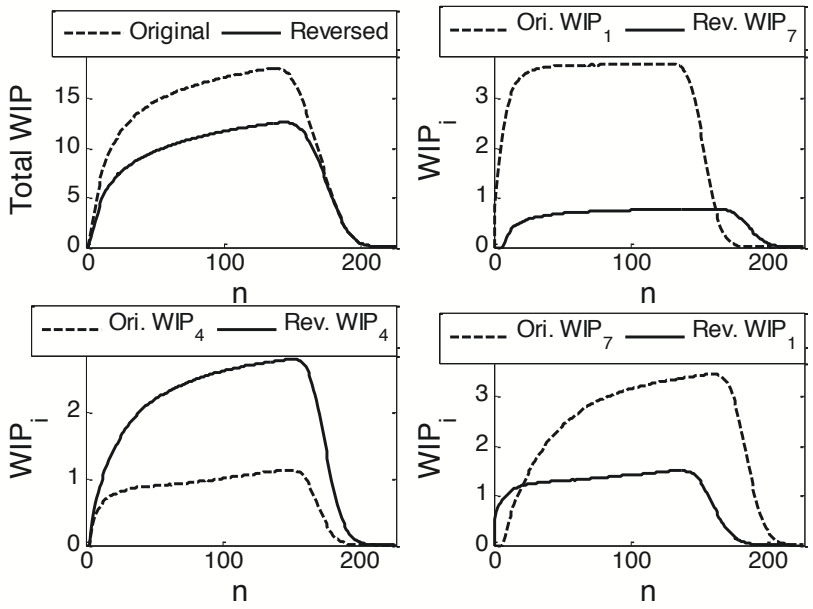


Figure 5.15: WIP of a 8-machine line and its reverse in simulation

$$p = \{0.81, 0.58, 0.60, 0.58, 0.77, 0.71, 0.71, 0.51\}.$$

The illustration verifies that $PR^L(n) = PR^{L_r}(n)$, which leads to $CT^L = CT^{L_r}$. Another serial line using aggregation is shown in Figure. 5.16 and Figure. 5.17 with parameters as follows:

$$M = 10,\ D = 100,\ N = \{2, 6, 6, 5, 2, 2, 2, 5, 2, 4\},$$

$$p = \{0.72, 0.87, 0.97, 0.78, 0.67, 0.85, 0.88, 0.56, 0.57, 0.62\}.$$

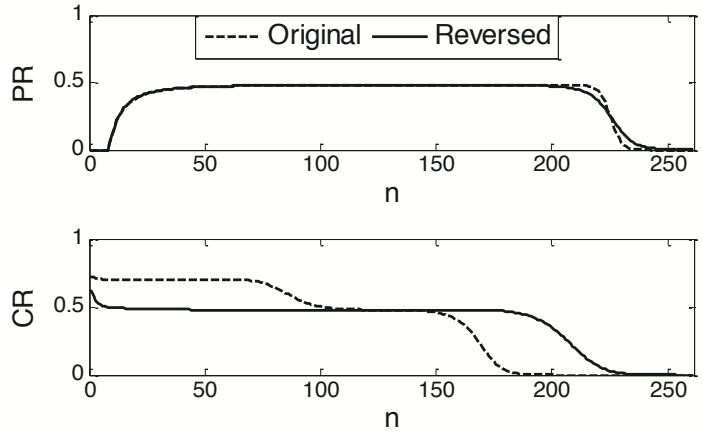


Figure 5.16: PR and CR of a 10-machine line and its reverse in aggregation

The same disparity occurs again when virtual buffer depletes. However, applying the proposed cutoff, original line's CT is 220, while the CT of its reverse is 222, with relative error of $(222-220)/220 = 0.9\%$. Note that original line's CT in Figure. 5.14 is 343.01 and the CT of its reverse is 343.12, with relative error of 0.03%. Also note that in both systems, with CT and system components unchanged, thanks to lower WIP, their reverses are preferable to the original systems.

5.4.2 Monotonicity

In a production system defined by assumptions (i)-(vii), it is intuitive and supported by experiments that:

- CT strictly monotonically decreasing in p;
- CT strictly monotonically decreasing in N and the decreasing rate slows down very quickly.

As an illustration, Figure. 5.18 shows the sensitivity of CT to buffer capacity and machine efficiency in a 10-machine system with $D = 50$.

The monotonicity property provides evaluation information to managers on system improvement. Take the system in Figure. 5.18 as an example and assume that

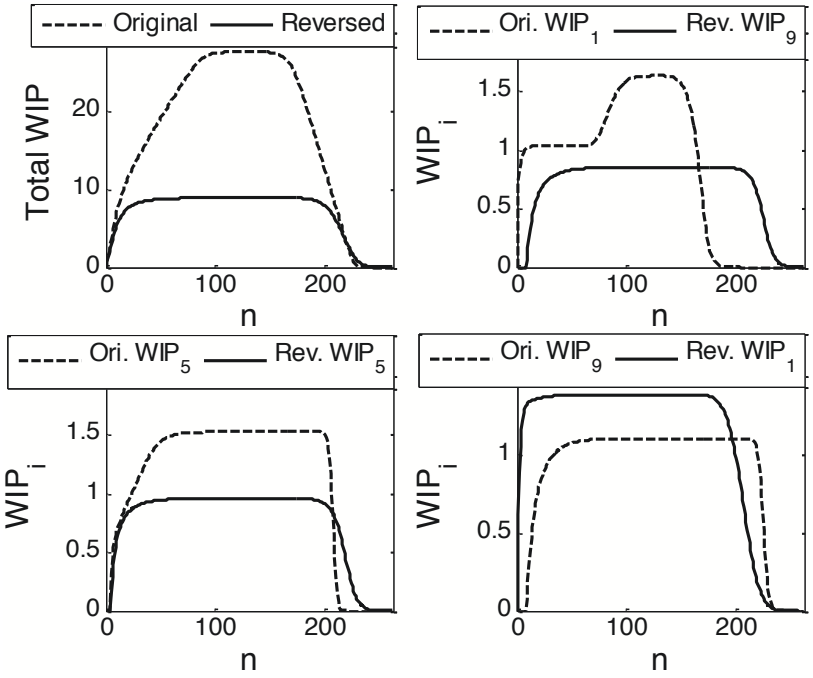


Figure 5.17: WIP of a 10-machine line and its reverse in aggregation

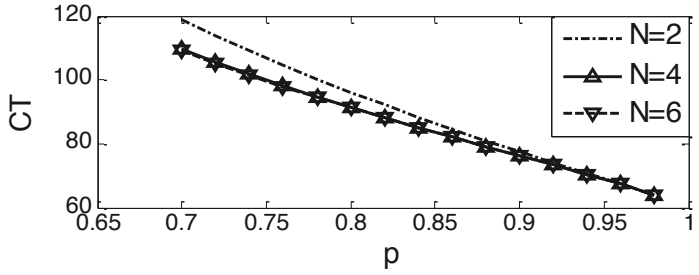


Figure 5.18: CT sensitivity to buffer capacity and machine efficiency

currently $p = 0.8$ and $N = 2$, which results to $CT = 97$. Increasing machine efficiencies to 0.9 can reduce CT to 78; however, increasing buffer size to 4 only reduce CT to 95, and further buffer size increasing does not lead to noticeable CT reduction. Therefore, investment in machines rather than buffers results to higher payback in terms of CT.

5.5 Summary

This chapter investigates the expected order completion time in a multi-machine Bernoulli serial line with unreliable machines and finite buffer capacity.

Special case with $D = 1$ is discussed and closed form formulas are provided. Then a lower bound of CT is provided for $D > 1$ cases. A method based on recursive aggregation technique by adding a virtual machine and a virtual buffer to the system is developed to approximate CT. Appropriate cutoffs for this method are also provided. Experiments show that this method maintains high accuracy and can be extended to Bernoulli lines with distinct machines and buffers. The structural properties, including reversibility and monotonicity, are also discussed at the end.

Therefore, this chapter provides technique to obtain order completion time and insights on its reversibility and monotonicity, with which production operators and managers can control and plan production more effectively to increase productivity, satisfy customer demand, and/or reduce energy consumption.

CHAPTER VI

GEOMETRIC TO BERNOULLI SYSTEMS TRANSFORMATION

6.1 Introduction

It is easy to show that a geometric serial line is also characterized by a Markov chain and similar approach can be applied to derive the performance evaluation formulas. However, since geometric machines have one-step memory, an M-machine geometric serial line has a total of $2^M \prod_{i=1}^{M-1}(N_i+1)$ states, which is larger than a Bernoulli one. Therefore, exact evaluation of geometric serial lines using closed-form expressions is not pursued here. Instead, we approximate its performance measures by transforming the system into a Bernoulli line. If the transformation maintains acceptable accuracy, then we can apply the aggregation-based performance evaluation technique for the transformed Bernoulli line and approximate the transients of the original geometric line.

6.2 Method One

6.2.1 Methodology

For simplicity, we first consider the following M-machine geometric serial line:

- All the $M \geq 2$ geometric machines are identical, i.e., they have the same P and R;
- All the $M-1$ buffers are identical, i.e., they have the same buffer capacity;

- All machines are down and all buffers are empty initially.

We are looking for approximated production rate in Bernoulli lines (PR^B), to estimate production rate in geometric lines (PR). Therefore, the corresponding characteristics in Bernoulli lines, p^B, N^B and τ^B need to be estimated so that the $PR(n)$ and $PR_B(n)$ are sufficiently close to each other.

The idea behind this method is to equalize the average downtime protection for the two systems. Specifically, note that $\frac{1}{R}$ and $\frac{1}{p^B}$ are the average downtime for geometric and Bernoulli models, respectively. Then $\frac{N}{\frac{1}{R}}$ or equivalently $N \times R$ measures how many downtimes can be protected by the buffer on average. In Bernoulli model, this corresponding value is $N^B \times p^B$. To equalize the two systems, this value should be equal. Thus, the corresponding Bernoulli line can be defined by, if $\frac{N \times R}{p^B} = N^*$ is an integer:

$$p^B = \frac{R}{P+R},$$
$$N^B = N^* + 1,$$

or, otherwise,

$$p^B = \frac{R}{P+R},$$
$$N^B = \lceil N^* \rceil + 1,$$

The extra unit buffer capacity is inspired by [4] and experiments show that this transformation receives higher accuracy than that without the extra capacity.

Let e denote the machine efficiency in geometric lines. Then $p^B = e$, i.e., both geometric and Bernoulli machines have identical efficiencies. p^B and e will be used interchangeably in the following.

It is easy to calculate all the steady state performance measures of geometric and Bernoulli lines, PR_{ss}, WIP_{ss}, and PR^B_{ss}, WIP^B_{ss}, respectively, which are assumed to

be known. Then the production rate and total work-in-process during transient can be characterized by normalized values. In the research, we divide the transient PR and WIP by PR_{ss} and WIP_{ss}, respectively, to obtain the corresponding normalized values. The closeness of PR and WIP between these two lines can be proved by the closeness of the normalized values. We will still use PR and WIP to denote the normalized values in the following for convenience.

However, experiments show that this method does not work so well regarding WIP. Therefore we will only discuss PR approximation.

Note that the original and the transformed lines have different scale of the time axis. For all non-zero PR and WIP, one time slot in the Bernoulli line is equivalent to $r = p^B/R$ time slots in the geometric line. For example, assume that we obtain the following PR for a transformed 5-machine Bernoulli line in the first 7 time slots:

$$PR^B = [0\ 0\ 0\ 0\ 0.1\ 0.2\ 0.25]$$

Then we can obtain the PR for a 5-machine geometric line in the first 5 time slots:

$$PR = [0\ 0\ 0\ 0\ 0.1]$$

To obtain other PR in time slots followed, we need to know r to adjust the time axis and the values. Assume that $r = 2.5$. Then the time slot 6 in Bernoulli line will be time slot $5 + (6 - 5) \times 2.5 = 7.5$ in geometric line, and similarly time slot 7 in Bernoulli line will be time slot $5 + (7 - 5) \times 2.5 = 10$ in geometric line. To obtain $PR(6)$, we apply linear interpolation:

$$PR(6) = PR^B(5) \times \frac{7.5 - 6}{2.5} + PR^B(6) \times \frac{6 - 5}{2.5}$$

Similarly we can approximate other $PR(n)$. In this example,

$$PR = [0\ 0\ 0\ 0\ 0.1\ 0.12\ 0.16\ 0.21\ 0.23\ 0.25]$$

Then there are 10 time slots in geometric line now.

6.2.2 Experiment

Case of integer N^*

The following parameters will impact or decide the accuracy of this method:

$$M,\ N,\ p^B,\ r$$

To observe to what extend each parameter/factor will impact the accuracy and which is the key impact factor, we design the following simulation experiments with C++ programming:

$$M \in \{5, 10, 15\},\ N \in \{16, 24, 32\},\ p^B \in \{0.75, 0.85, 0.95\},\ r \in \{2, 4, 8\}$$

Therefore, there are $3 \times 3 \times 3 \times 3 = 81$ lines in total. We will monitor the PR during transient and the corresponding average (absolute value of) difference between the two lines and its standard deviation.

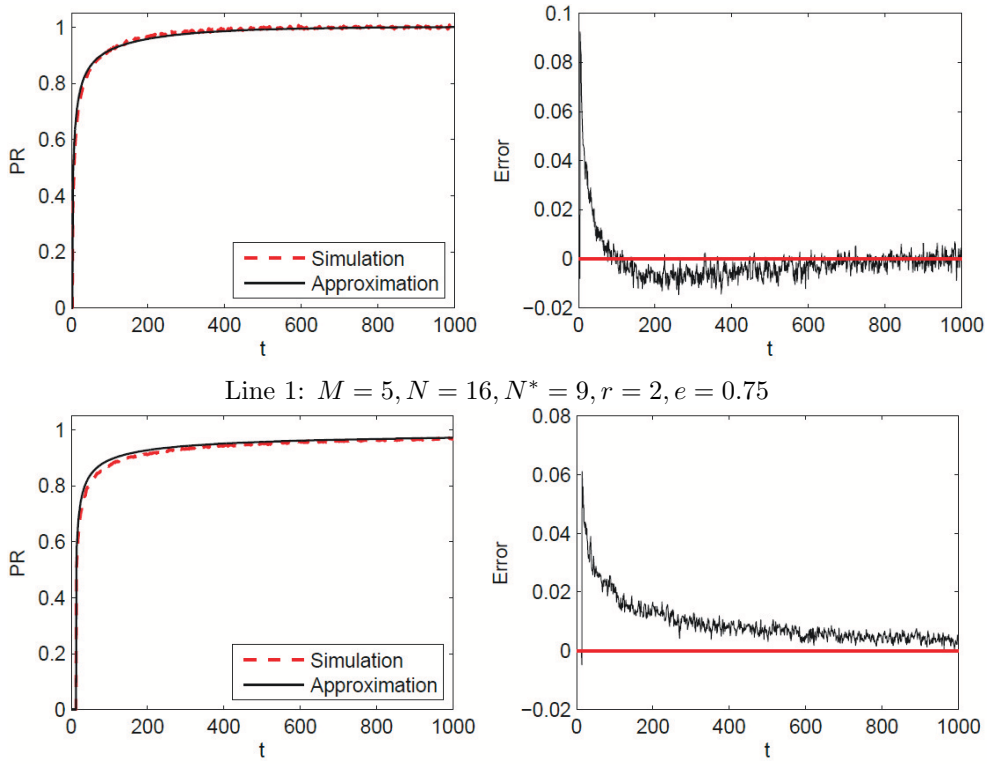
In the experiment, transient performance measures are obtained from time slot 0 to time slot 1,000 with 50,000 iterations, and steady state performance measures are obtained from time slot 200,000 to time slot 300,000 with 20 iterations.

Note that $N^B \in \{3, 4, 5, 7, 9, 13, 16\}$, more than twice of the average downtime of Bernoulli lines, and

$$R \in \{0.0938, 0.1063, 0.1187, 0.1875, 0.2125, 0.2375, 0.3750, 0.4250, 0.4750\},$$
$$P \in \{0.0063, 0.0125, 0.0188, 0.0250, 0.0313, 0.0375, 0.0625, 0.0750, 0.1250\}.$$

Let us visually check the accuracy of the approximation of two individual lines in Figure 6.1. The errors in the figure is the difference between $PR(n)$ and $PR^B(n)$.

The PR approximation works well, normally within 2%. The noticeable fluctuations in the very beginning of the shift are partially due to small absolute PR values.



Line 1: $M = 5, N = 16, N^* = 9, r = 2, e = 0.75$

Line 2: $M = 5, N = 32, N^* = 17, r = 2, e = 0.95$

Figure 6.1: Examples with small disparity for integer N^*

However, in some other systems, the approximation does not produce such acceptable results, two of which are shown in Figure 6.2.

Although there exists more discrepancy, the errors are still within 8% most of the time. Significant disparity may be due to more machines, larger buffer capacities and/or lower efficiencies, i.e., slower transients.

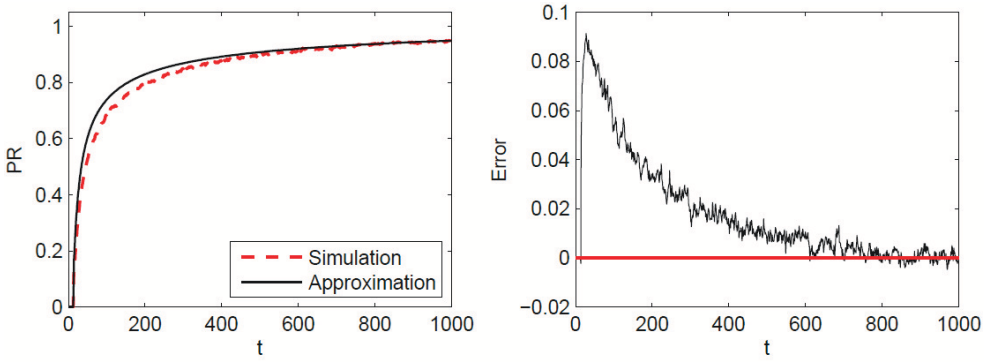
Case of non-integer N^*

The system parameters values in this experiment is as follows:

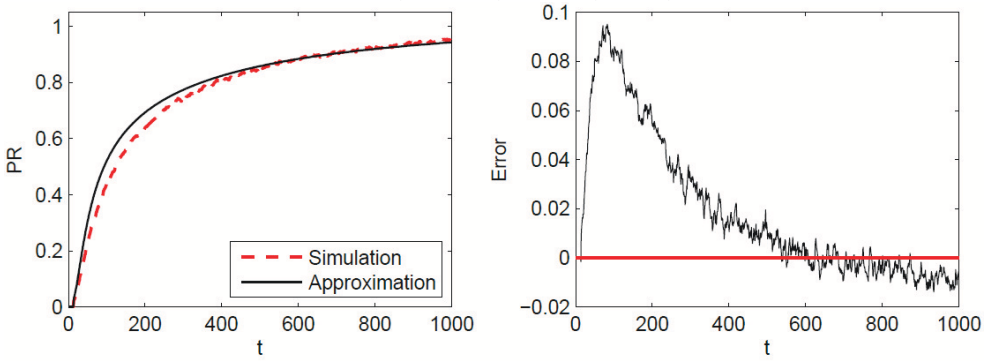
$$M \in \{5, 10, 15\},\ N \in \{10, 20, 40\},\ p^B \in \{0.75, 0.85, 0.95\},\ r \in \{1.2, 5.5, 9.8\}$$

There are $3 \times 3 \times 3 \times 3 = 81$ lines in total as well. Other experiment parameters are the same with integer case, such as the iterations and time slots.

In this case, for simplicity, we denote N^B the smaller Bernoulli buffer capacity, since the other Bernoulli buffer capacity can be obtained easily by adding one to



Line 1: $M = 15, N = 32, N^* = 9, r = 4, e = 0.85$



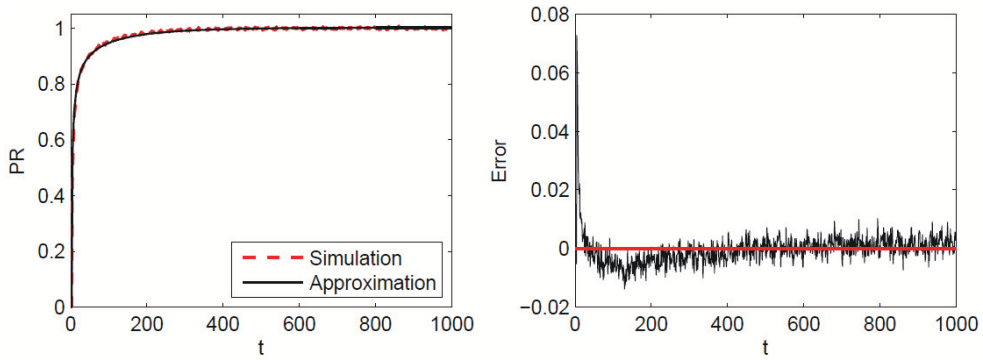
Line 2: $M = 15, N = 32, N^* = 9, r = 8, e = 0.75$

Figure 6.2: Examples with large disparity for integer N^*

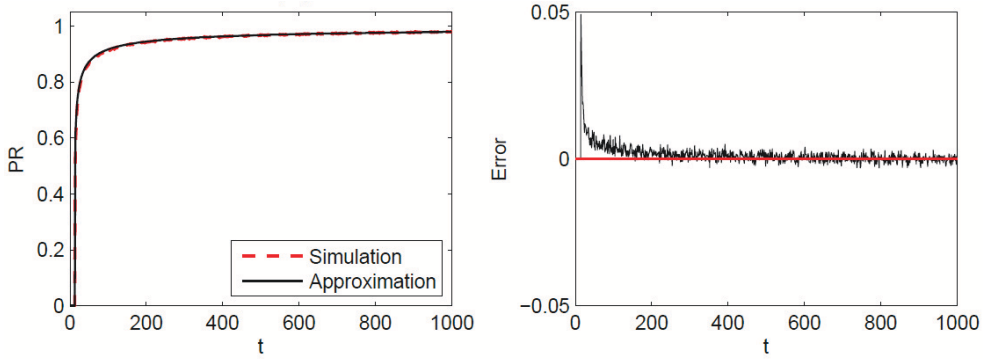
it. Then $N^B \in \{2, 3, 4, 5, 8, 9, 17, 34\}$, more than twice of the average downtime of Bernoulli lines.

$$R \in \{0.0765, 0.0867, 0.0969, 0.1364, 0.1545, 0.1727, 0.6250, 0.7083, 0.7917\},$$
$$P \in \{0.0051, 0.0091, 0.0153, 0.0255, 0.0273, 0.0417, 0.0455, 0.1250, 0.2083\}.$$

Again, let us take a look at two lines (see Figure 6.3) and visually check the accuracy first.



Line 1: $M = 5, N = 10, N^* = 10, r = 1.2, e = 0.75$



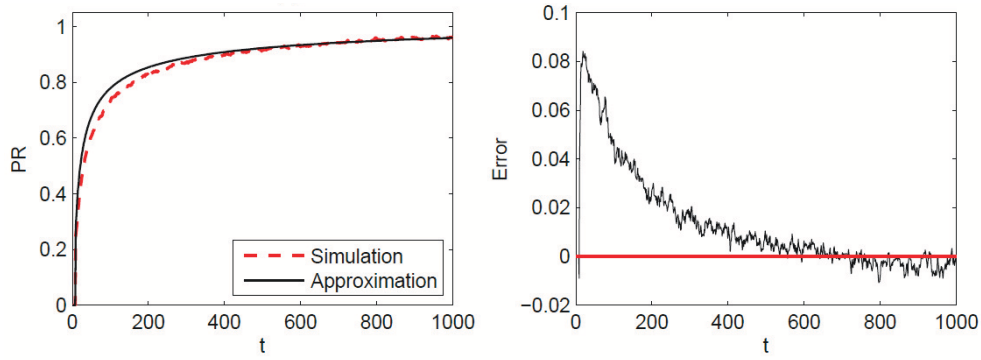
Line 2: $M = 15, N = 20, N^* = 18, r = 1.2, e = 0.95$

Figure 6.3: Examples with small disparity for non-integer N^*

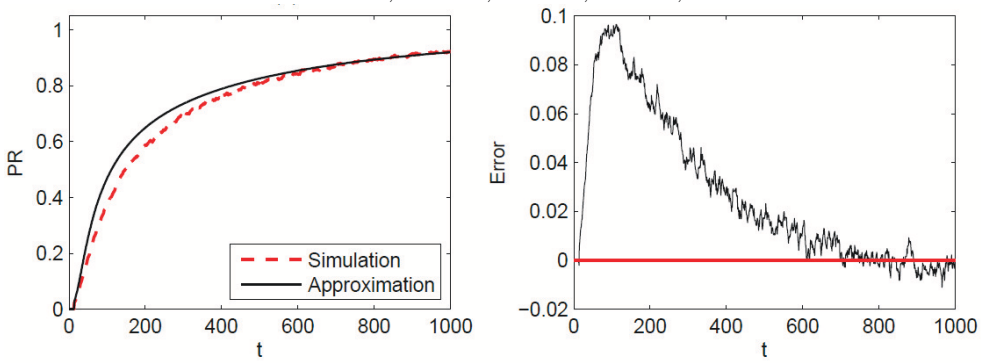
PR approximation works well in these lines, with errors normally within 2%. However, there exist some not so good examples (see Figure 6.4).

Experiments are conducted to check how the approximation responds to the change of each parameter. The results are shown in Figure 6.5 with integer N^* and in Figure 6.6 with non-integer N^*.

Based on the observations above, we have the following conjectures:



Line 1: $M = 10, N = 40, N^* = 9, r = 5.5, e = 0.85$



Line 2: $M = 15, N = 40, N^* = 6, r = 9.8, e = 0.75$

Figure 6.4: Examples with large disparity for non-integer N^*

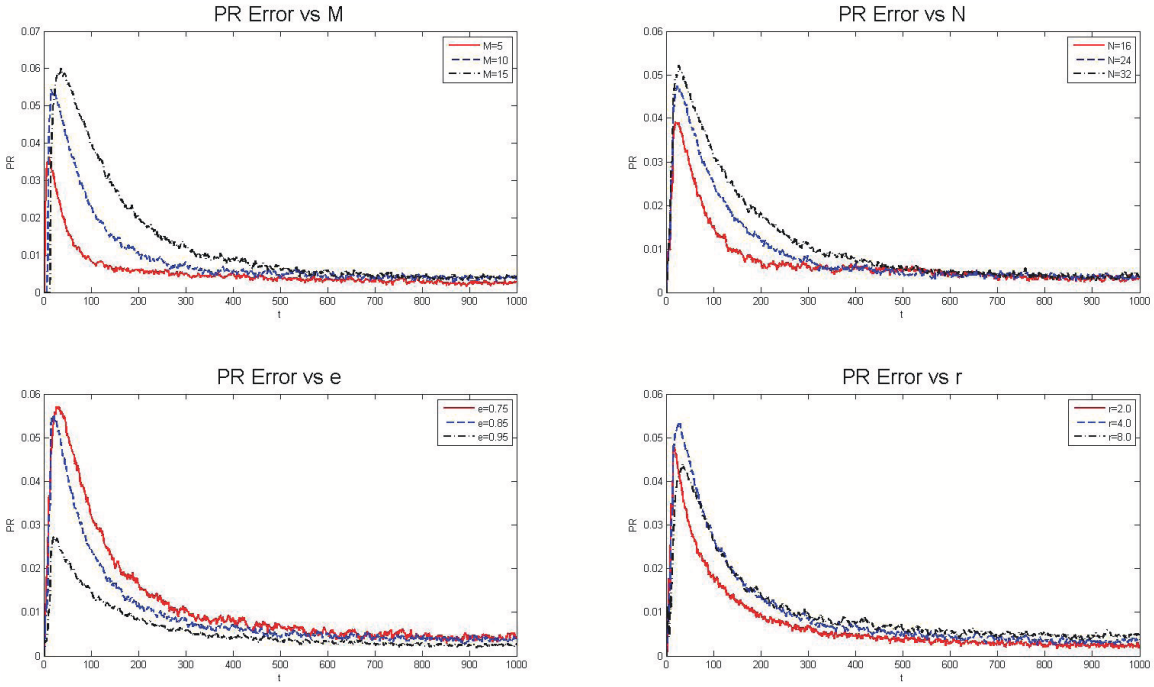


Figure 6.5: Parameters effect in approximation for integer N^*

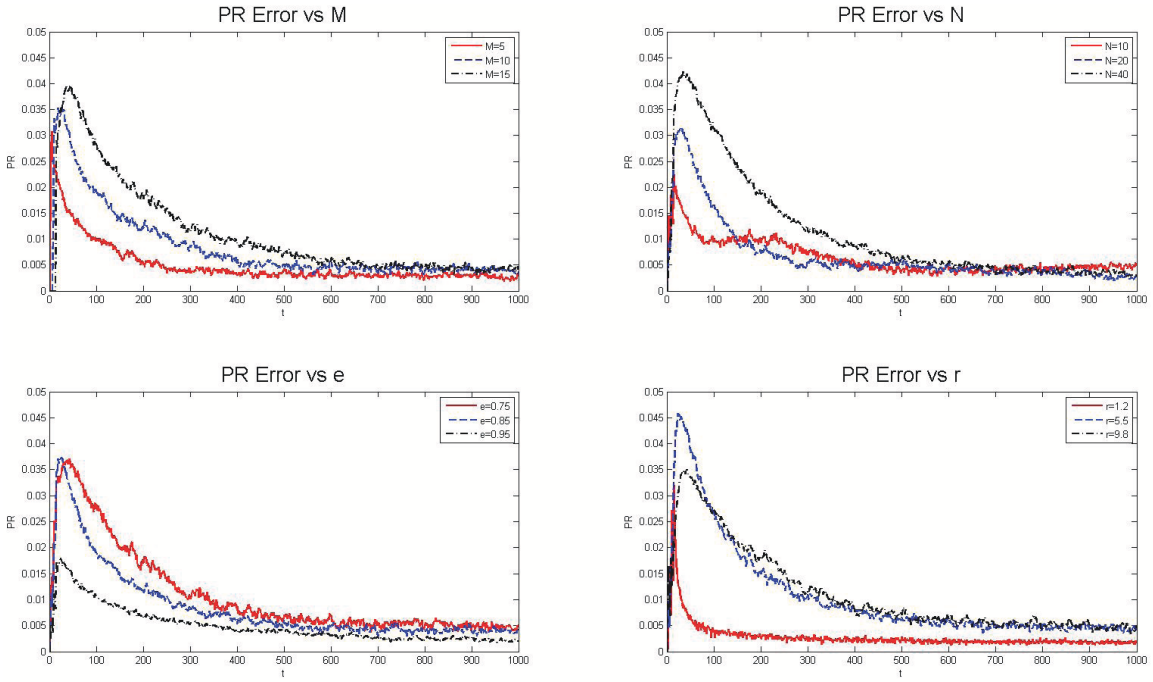


Figure 6.6: Parameter effects in approximation for non-integer N^*

- Approximation performance is monotonically decreasing in M;
- Approximation performance is monotonically decreasing in N;
- Approximation performance is monotonically increasing in e;
- Approximation performance is monotonically decreasing in r, especially for very small r.

The summary for this method is that for small M, small N, large e, and/or large r, i.e., short transient period, this approximate provides acceptable level of accuracy. In those cases the method does not work well, further modification and improvement are needed.

6.3 Method Two

6.3.1 Methodology

This method is inspired by [4]. Assume all machines have identical efficiency and initial (up or down) status, and all baffurs have identical capacity and level initial occupancy h_0. Then for $n = 0, 1, 2, \cdots$,

$$PR(n) = \begin{cases} 0, & \text{if } PR^B(n) = 0; \\ PR^B(n) \times (1 + \frac{\Delta}{e}\lambda^{(n-k+1)})^M, & \text{if } PR^B(n) > 0; \end{cases}$$

where $\lambda = 1-P-R$, the second largest eigenvalue (SLE); k is the smallest number of n such that $PR^B(n) > 0$; $r = e/k$; and

$$\Delta = \begin{cases} 1 - e, & \text{if machines are initially up;} \\ -e, & \text{if machines are initially down.} \end{cases}$$

The initial buffer occupancy in corresponding Bernoulli lines, h_0^B, is estimated in three conjectures:

$$h_0^B = \lfloor \tfrac{h_0}{r} \rfloor, h_0^B = \lceil \tfrac{h_0}{r} \rceil, h_0^B = \lceil \tfrac{h_0}{r} \rceil + 1,$$

where $\lfloor x \rfloor$ calculates the biggest integer that is not greater than x and $\lceil x \rceil$ the smallest integer that is not less than x.

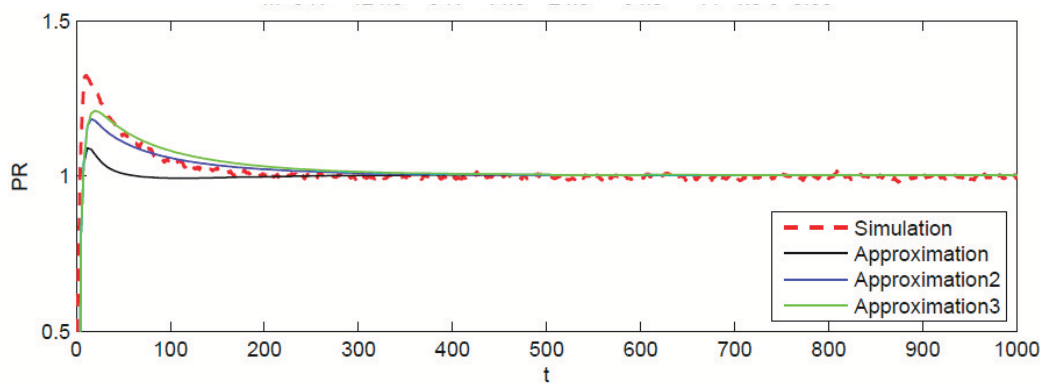
6.3.2 Experiment

Figure 6.7 are some examples of test results if machines are initially down using the three approximations. PR and PR^B are normalized, i.e.,

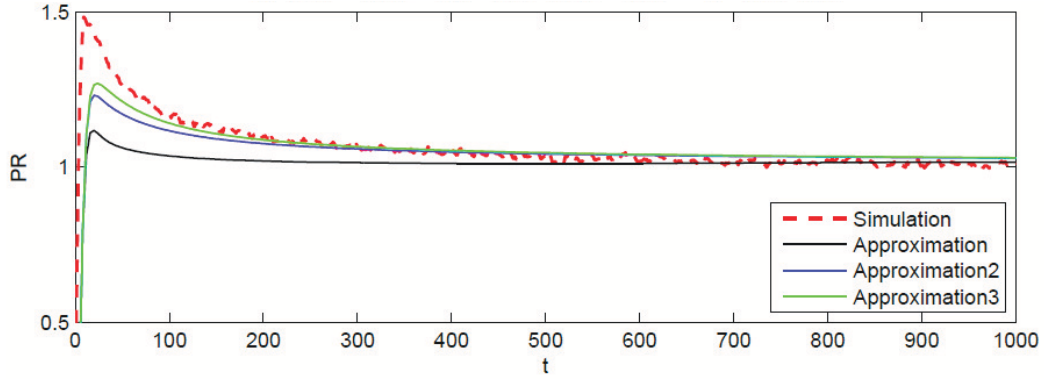
$$PR = \tfrac{PR}{PR_{ss}}, PR^B = \tfrac{PR^B}{PR^B_{ss}}.$$

Note that $PR_{ss} \approx PR^B_{ss}$.

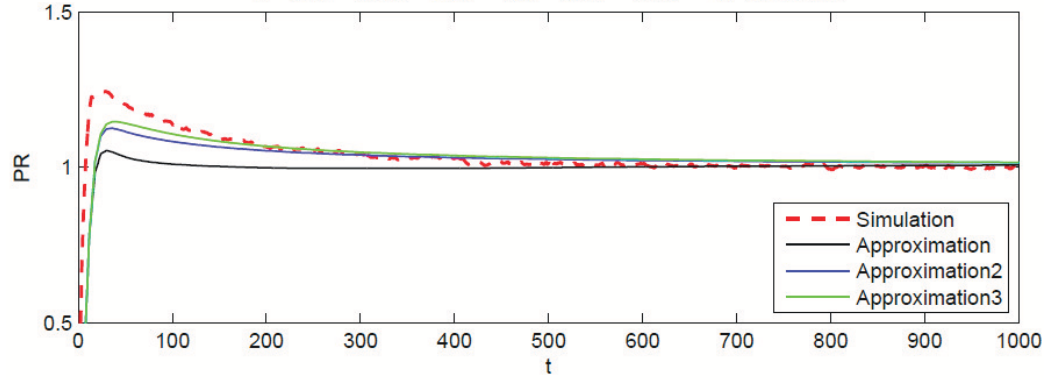
Figure 6.8 are examples if machines are initially up.



Line 1: $M = 5, N = 12, h_0 = 9, r = 4, e = 0.55$

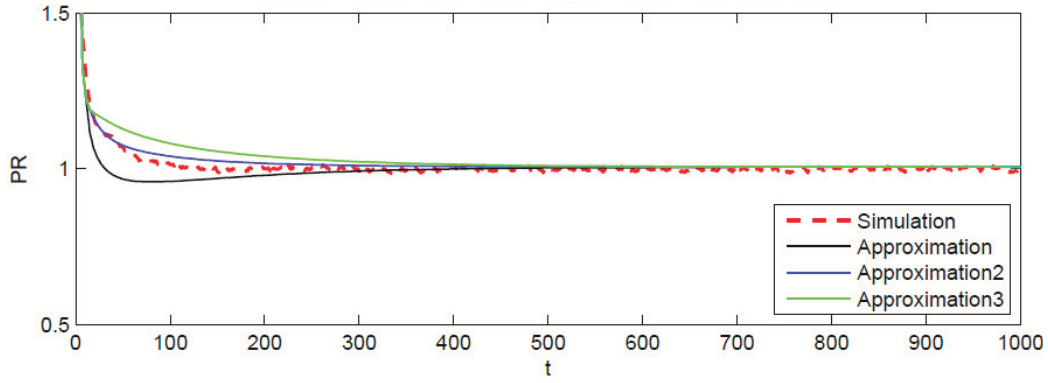


Line 2: $M = 15, N = 12, h_0 = 9, r = 4, e = 0.55$

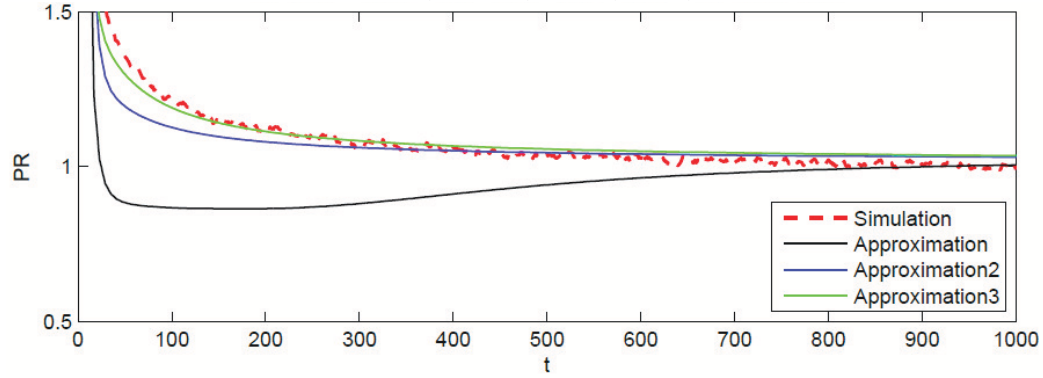


Line 3: $M = 10, N = 18, h_0 = 9, r = 6, e = 0.75$

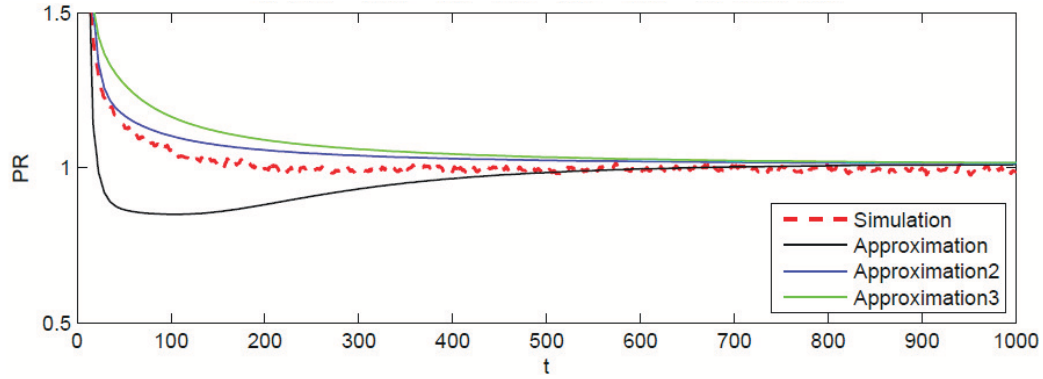
Figure 6.7: Examples for machines initially down



Line 1: $M = 5, N = 18, h_0 = 9, r = 4, e = 0.55$



Line 2: $M = 15, N = 12, h_0 = 9, r = 6, e = 0.55$



Line 3: $M = 10, N = 12, h_0 = 6, r = 6, e = 0.55$

Figure 6.8: Examples for machines initially up

Other examples show similar disparity during transient period. Further investigation shows that the third conjecture works better than the other two.

The summary for this method is that it does not estimate PR with acceptable accuracy. Further modification and improvement are needed and will be future work.

6.4 Summary

This chapter investigates two methods to transform Geometric lines to Bernoulli lines to estimate PR. When the transient period is short, Method One maintains high accuracy; otherwise, the accuracy will be reduced. However, Method Two in general does not maintain high accuracy. Further modifications are needed to improve the approximation.

CHAPTER VII

TRANSIENT ANALYSIS OF SERIAL PRODUCTION LINES WITH GEOMETRIC MACHINES

7.1 Introduction

Transients of serial production lines with finite buffers and unreliable machines have only received some preliminary studies in a few recent publications. Applications of Bernoulli line transient analysis have been reported in [49], [68] and [69]. In particular, [69] extended the algorithm developed in [70] to Bernoulli serial lines with time-varying machine parameters. Despite these important results regarding production system transients, it should be noted that most of the analytical studies described above are only for systems with machines having the Bernoulli reliability model, which is applicable only in the cases where the machine downtime is, on the average, comparable to its cycle time. Although [45] attempted to study the transient behavior of serial lines with machines having the geometric reliability model, the results were limited to the case of two-machine lines only. For longer lines, to the best of our knowledge, no analytical methods have been developed for analysis of their transient behavior, and simulation remains as the only tool for this purpose. Thus, the goal of this chapter is to derive analytical models that describe the transients of serial production lines with finite buffers and machines having the geometric reliability model, and to develop analytical methods for their performance evaluation during transients.

7.2 Transient Performance of Individual Geometric Machines

Individual geometric machine with constant parameters: As it was mentioned above, transient analysis of individual geometric machines with constant parameters has been carried out in [45]. Since the performance evaluation method derived is the foundation of the study in the chapter, we briefly review it below.

The state transition diagram for an individual geometric machine is shown in Figure 7.1. Let $x_i(n)$, $i \in \{0 = \text{down}, 1 = \text{up}\}$, denote the probability that the

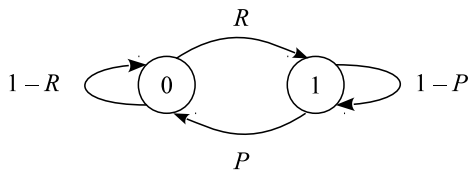


Figure 7.1: State transition diagram of a geometric machine

machine is in state i during time slot n, i.e., $x_i(n) = \text{Prob}[s(n) = i]$. Clearly, the system is characterized by a two-state ergodic Markov chain and the evolution of state vector $\mathbf{x}(n) = [x_0(n)\ \ x_1(n)]^T$ can be described by

$$\mathbf{x}(n+1) = \mathbf{A}_1\mathbf{x}(n), \quad x_0(n) + x_1(n) = 1, \tag{7.1}$$

where

$$\mathbf{A}_1 = \begin{bmatrix} 1-R & P \\ R & 1-P \end{bmatrix}. \tag{7.2}$$

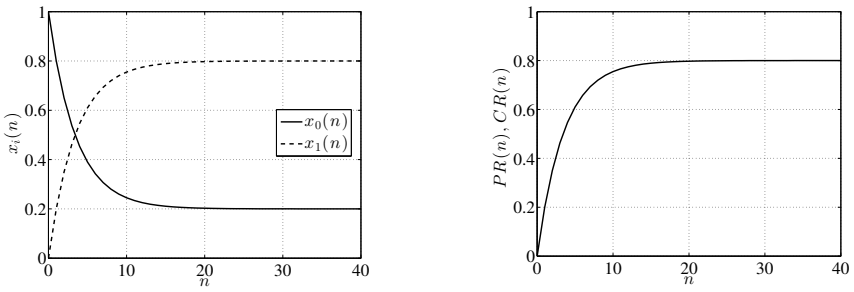
The production rate and consumption rate of an individual machine can be calculated as:

$$PR(n) = CR(n) = x_1(n) = [0 \quad 1]\mathbf{x}(n) = [0 \quad 1]\mathbf{A}_1^n\mathbf{x}(0), \tag{7.3}$$

which are both *linear* in machine state $\mathbf{x}(n)$.

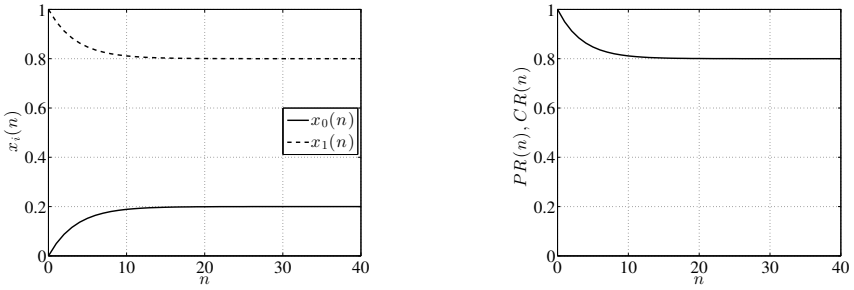
As an illustration, consider a geometric machine with breakdown probability $P =$

0.05 and repair probability $R = 0.2$. The transients of the system state and the performance measures are given in Figures 7.2 and 7.3, assuming the machine is initially down and up, respectively. As one can see, the initial condition of a machine has strong impact on system transients − which may result in production loss (see Figure 7.2) or production gain (see Figure 7.3).



(a) Transients of $x_i(n)$ (b) Transients of $PR(n)$ and $CR(n)$

Figure 7.2: Transients of an individual geometric machine when it is initially down



(a) Transients of $x_i(n)$ (b) Transients of $PR(n)$ and $CR(n)$

Figure 7.3: Transients of an individual geometric machine when it is initially up

Individual geometric machine with time-varying parameters: When the parameters of a geometric machine change over time, the system is characterized by a two-state inhomogeneous Markov chain. Let $P(n)$ and $R(n)$ denote the breakdown and repair probabilities of a geometric machine during time slot n, respectively. Then, the state transition diagram for such an individual geometric machine is shown in Figure 7.4.

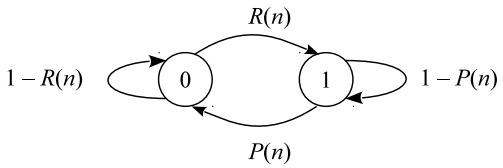


Figure 7.4: State transition diagram of geometric machine with time-varying parameters

Again, let $x_i(n)$, $i \in \{0, 1\}$, be the probability that the machine is in state i during time slot n. Then, the evolution of the state vector $\mathbf{x}(n) = [x_0(n)\ \ x_1(n)]^T$ is given by

$$\mathbf{x}(n+1) = \mathbf{A}_1(n)\mathbf{x}(n), \quad x_0(n) + x_1(n) = 1, \tag{7.4}$$

where

$$\mathbf{A}_1(n) = \begin{bmatrix} 1 - R(n) & P(n) \\ R(n) & 1 - P(n) \end{bmatrix}. \tag{7.5}$$

The production rate and consumption rate of this individual machine can be calculated as:

$$PR(n) = CR(n) = x_1(n) = [0 \quad 1]\mathbf{x}(n) = [0 \quad 1]\mathbf{A}_1(n-1)\cdots\mathbf{A}_1(0)\mathbf{x}(0). \tag{7.6}$$

As an illustration, consider a geometric machine with time-varying breakdown probability and repair probabilities depicted in Figure 7.5. The transients of the system state and its performance measures are provided in Figures 7.6 and 7.7 for different machine initial conditions.

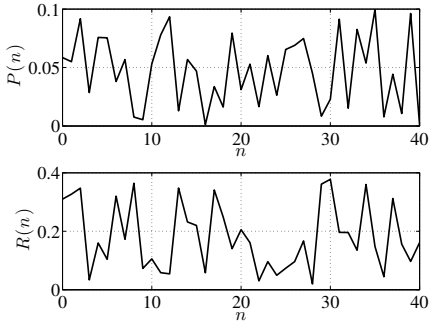
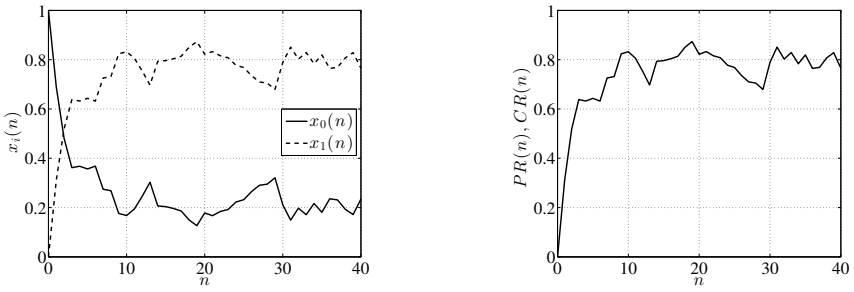
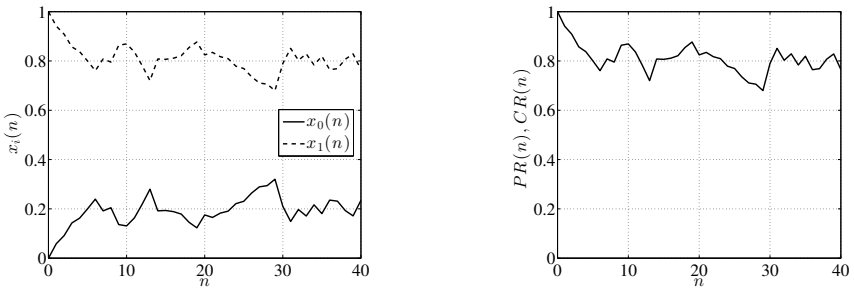


Figure 7.5: Breakdown and repair probabilities of a geometric machine as functions of time n



(a) Transients of $x_i(n)$ (b) Transients of $PR(n)$ and $CR(n)$

Figure 7.6: Transients of an individual geometric machine with time-varying parameters when it is initially down



(a) Transients of $x_i(n)$ (b) Transients of $PR(n)$ and $CR(n)$

Figure 7.7: Transients of an individual geometric machine with time-varying parameters when it is initially up

In the next section, we will use the results on individual geometric machine with time-varying parameters to describe the transient behavior of two-machine geometric lines.

7.3 Transient Performance of Two-Machine Geometric Lines

7.3.1 Mathematical analysis

Consider a two-machine geometric line illustrated in Figure 7.8. It is easy to show

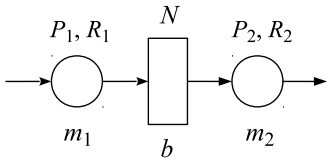


Figure 7.8: Two-machine geometric serial line

that the system is characterized by an ergodic Markov chain. In addition to machine state $s_i(n)$, let $h(n)$ denote the number of parts in the buffer at the beginning of time slot n. Then, the state of the Markov chain is defined by a triple $(s_1(n), s_2(n), h(n))$, where $s_1(n), s_2(n) \in \{0, 1\}$ and $h(n) \in \{0, 1, \ldots, N\}$. Clearly, the system has a total of $4(N + 1)$ states. To calculate the transition probabilities among these states, we arrange the states in the following manner: Let $r(s_1, s_2, h)$ denote the order number of state (s_1, s_2, h), $s_1, s_2 \in \{0, 1\}$, $h \in \{0, 1, \ldots, N\}$. Define:

$$r(s_1, s_2, h) = 4h + 2s_1 + s_2 + 1. \tag{7.7}$$

Then, the arrangement of the $4(N + 1)$ system states can be summarized in Table 7.1. In other words, each system state is assigned a unique number ranging from 1 to $4(N + 1)$. For example, State 1 is when both machines are down and the buffer is empty, while State $4(N + 1)$ is when both machines are up and the buffer is full.

Table 7.1: Arrangement of the system states ($k = 0, 1, \ldots, N$)

State number (r)	s_1	s_2	h
$4k+1$	0	0	k
$4k+2$	0	1	k
$4k+3$	1	0	k
$4k+4$	1	1	k

Note that the transition of $h(n)$ is deterministic given $s_1(n)$ and $s_2(n)$:

$$h(n+1) = h(n) - s_2(n) \min\{h(n), 1\} + s_1(n) \min\{N - h(n) + s_2(n) \min\{h(n), 1\}, 1\}. \tag{7.8}$$

The transitions of $s_i(n)$'s, on the other hand, are probabilistic based on (2.1). Therefore, we can examine each of the $4(N+1)$ states, then, based on (7.8), identify all possible destination states after one time slot by enumerating all four combinations of $s_1(n)$ and $s_2(n)$, and, finally, calculate the corresponding transition probabilities using (2.1). Let $\mathbf{A}_2$ denote the transition probability matrix obtained and let $x_i(n)$, $i \in \{1, 2, \ldots, 4(N+1)\}$, denote the probability that the system, i.e., the Markov chain, is in state i during time slot n. Then, the evolution of the system state, $\mathbf{x}(n) = [x_1(n)\ x_2(n)\ \ldots\ x_{4(N+1)}(n)]^T$, is given by

$$\mathbf{x}(n+1) = \mathbf{A}_2 \mathbf{x}(n), \quad \sum_{i=1}^{4(N+1)} x_i(n) = 1. \tag{7.9}$$

Then the performance measures of the two-machine geometric line system can be calculated as follows:

$$\begin{aligned}
PR(n) &= \text{Prob}[m_2 \text{ is up and buffer } b \text{ is not empty during time slot } n] \\
&= \mathbf{C}_1 \mathbf{x}(n) = [\mathbf{C}_{1,0}\ \mathbf{C}_{1,1}\ \ldots\ \mathbf{C}_{1,N}]\mathbf{x}(n), \\
CR(n) &= \text{Prob}[m_1 \text{ is up and not blocked during time slot } n] \\
&= \mathbf{C}_2 \mathbf{x}(n) = [\mathbf{C}_{2,0}\ \mathbf{C}_{2,1}\ \ldots\ \mathbf{C}_{2,N}]\mathbf{x}(n), \\
WIP(n) &= \mathbf{C}_3 \mathbf{x}(n) = [\mathbf{C}_{3,0}\ \mathbf{C}_{3,1}\ \ldots\ \mathbf{C}_{3,N}]\mathbf{x}(n),
\end{aligned} \tag{7.10}$$

$$
\begin{aligned}
ST_2(n) &= \mathbf{C}_4\mathbf{x}(n) = [\mathbf{C}_{4,0}\ \mathbf{C}_{4,1}\ \ldots\ \mathbf{C}_{4,N}]\mathbf{x}(n), \\
BL_1(n) &= \mathbf{C}_5\mathbf{x}(n) = [\mathbf{C}_{5,0}\ \mathbf{C}_{5,1}\ \ldots\ \mathbf{C}_{5,N}]\mathbf{x}(n),
\end{aligned}
$$

where

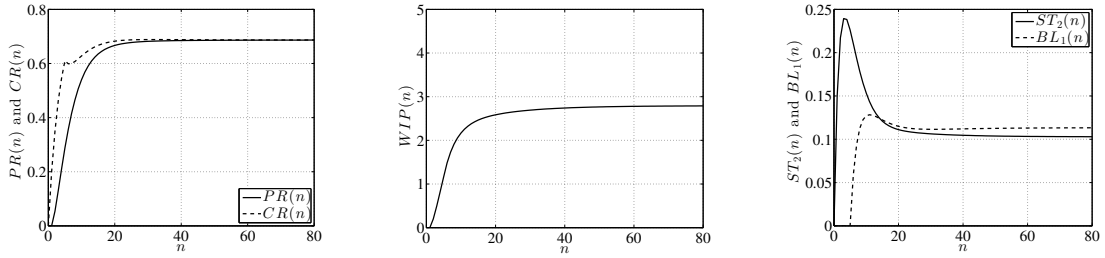
$$
\begin{aligned}
\mathbf{C}_{1,0} &= [0\ 0\ 0\ 0], \quad \mathbf{C}_{1,i} = [0\ 1\ 0\ 1], \quad i = 1, \ldots, N, \\
\mathbf{C}_{2,N} &= [0\ 0\ 0\ 1], \quad \mathbf{C}_{2,i} = [0\ 0\ 1\ 1], \quad i = 0, \ldots, N-1, \\
\mathbf{C}_{3,i} &= [i\ i\ i\ i], \quad i = 0, \ldots, N, \\
\mathbf{C}_{4,0} &= [0\ 1\ 0\ 1], \quad \mathbf{C}_{4,i} = [0\ 0\ 0\ 0], \quad i = 1, \ldots, N, \\
\mathbf{C}_{5,N} &= [0\ 0\ 1\ 0], \quad \mathbf{C}_{5,i} = [0\ 0\ 0\ 0], \quad i = 0, \ldots, N-1.
\end{aligned}
\tag{7.11}
$$

Therefore, all these performance measures are *linear* in system state $\mathbf{x}(n)$.

As an illustration, consider a two-machine geometric line machine and buffer parameters:

$$
P_1 = 0.05, \quad R_1 = 0.2, \quad P_2 = 0.04, \quad R_2 = 0.15, \quad N = 5.
$$

Assume that both machines are initially down and the buffer is initially empty. The transients of the performance measures of this system are illustrated in Figure 7.9.



(a) $PR(n)$ and $CR(n)$ (b) $WIP(n)$ $ST_2(n)$ and $BL_2(n)$

Figure 7.9: Transients of a two-machine geometric line

Finally, for systems with machines having time-varying parameters, the evolution

of the system state, $\mathbf{x}(n)$, is given by

$$\mathbf{x}(n+1) = \mathbf{A}_2(n)\mathbf{x}(n), \quad \sum_{i=1}^{4(N+1)} x_i(n) = 1, \tag{7.12}$$

where transition probability matrix $\mathbf{A}_2(n)$ can be calculated using the same method described above, but with P_i and R_i replaced by $P_i(n)$ and $R_i(n)$, respectively.

7.3.2 Equivalent aggregation

At the input end of a two-machine geometric serial line, the consumption of raw materials is characterized by the joint effect of machine m_1 and its blockage due to downstream failures. In other words, we can view the raw materials as being consumed by a "modified" version of m_1 with the effects of both buffer b and machine m_2 included. We referred to this as *backward aggregation* (since the buffer and machine m_2 are aggregated with and into m_1 in the backward direction of the parts flow), and refer to the modified m_1 as m_1^b, where superscript b stands for "backward" (see Figure 7.10(a)). Similarly, at the output end, the production of finished parts is characterized by machine m_2 and its starvation due to upstream failures. We aggregate machine m_1 and the buffer with m_2 in the *forward aggregation* and obtain a modified version of the second machine, m_2^f, where superscript f stands for "forward" (see Figure 7.10(b)).

(a) Backward aggregation (b) Forward aggregation

Figure 7.10: Equivalent aggregation of a two-machine geometric line

Since both m_1 and m_2 are geometric machines, we assume that the two aggregated machines m_1^b and m_2^f are also geometric. However, due to system transients, the parameters of m_1^b and m_2^f may change over time. Therefore, we let $P_1^b(n)$, $R_1^b(n)$, $P_2^f(n)$, and $R_2^f(n)$ denote "breakdown" and "repair" probabilities of m_1^b and m_2^f during

time slot n, respectively. Then, based on the definition of a geometric machine, the breakdown probability of machine m_2^f is defined by

$$\begin{aligned} P_2^f(n) &= \text{Prob}[m_2^f \text{ is down in time slot } n+1 | m_2^f \text{ is up in time slot } n] \\ &= \text{Prob}[\text{System doesn't produce in time slot } n+1 | \text{System produces a part} \\ &\quad \text{in time slot } n] \\ &= \frac{\sum_{i \in I_{np}} \sum_{j \in I_p} \mathbf{A}_{2,ij} x_j(n)}{PR(n)}, \end{aligned}$$

where $\mathbf{A}_{2,ij}$ is the element in the ith row and jth column of matrix $\mathbf{A}_2$, and I_p and I_{np} are sets containing the indices of states in which the system produces and does not produce a part in a time slot, respectively. Clearly, the indices contained in I_p correspond to the 1-entries in $\mathbf{C}_1$, while the indices in I_{np} correspond to the 0-entries in $\mathbf{C}_1$. Therefore, the numerator of the expression above is equal to the probability that the system produces a part in time slot n *and* does not produce in time slot $n+1$. Similarly, the repair probability of machine m_2^f can be calculated as

$$\begin{aligned} R_2^f(n) &= \text{Prob}[m_2^f \text{ is up in time slot } n+1 | m_2^f \text{ is down in time slot } n] \\ &= \text{Prob}[\text{System produces a part in time slot } n+1 | \text{System doesn't produce} \\ &\quad \text{in time slot } n] \\ &= \frac{\sum_{i \in I_p} \sum_{j \in I_{np}} \mathbf{A}_{2,ij} x_j(n)}{1 - PR(n)}. \end{aligned}$$

The same approach also applies to the breakdown and repair probabilities of m_1^b, which can be calculated as

$$\begin{aligned} P_1^b(n) &= \text{Prob}[m_1^b \text{ is down in time slot } n+1 | m_1^b \text{ is up in time slot } n] \\ &= \text{Prob}[\text{System doesn't consume in time slot } n+1 | \text{System consumes a part} \\ &\quad \text{in time slot } n] \\ &= \frac{\sum_{i \in I_{nc}} \sum_{j \in I_c} \mathbf{A}_{2,ij} x_j(n)}{CR(n)}, \end{aligned}$$

$$
\begin{aligned}
R_1^b(n) &= \text{Prob}[m_1^b \text{ is up in time slot } n+1 | m_1^b \text{ is down in time slot } n] \\
&= \text{Prob}[\text{System consumes a part in time slot } n+1 | \text{System doesn't consume} \\
&\quad \text{in time slot } n] \\
&= \frac{\sum_{i \in I_c} \sum_{j \in I_{nc}} \mathbf{A}_{2,ij} x_j(n)}{1 - CR(n)},
\end{aligned}
$$

where I_c and I_{nc} denote the indices of states in which the system consumes and does not consume a raw part in a time slot, respectively. In other words, the indices contained in I_c correspond to the 1-entries in $\mathbf{C}_2$, while the indices in I_{nc} correspond to the 0-entries in $\mathbf{C}_2$.

With the breakdown and repair probabilities of m_1^b and m_2^f calculated, we are now able to study their transient behavior using the results described in Section 7.2. Let $x_i^f(n)$, $i \in \{0 = \text{down}, 1 = \text{up}\}$, denote the probability that the aggregated machine m_2^f is in state i during time slot n. Thus, m_2^f being up implies that the two-machine line produces a finished part during this time slot. Then, the evolution of the vector $\mathbf{x}^f(n) = [x_0^f(n)\ x_1^f(n)]^T$ can be expressed as

$$\mathbf{x}^f(n+1) = \mathbf{A}_1^f(n)\mathbf{x}^f(n),$$

where

$$
\begin{aligned}
\mathbf{A}_1^f(n) &= \begin{bmatrix} 1 - R_2^f(n) & P_2^f(n) \\ R_2^f(n) & 1 - P_2^f(n) \end{bmatrix}, \\
\mathbf{x}^f(0) &= \begin{bmatrix} 1 - PR(0) & PR(0) \end{bmatrix}^T.
\end{aligned}
$$

Similarly, let $x_i^b(n)$, $i \in \{0 = \text{down}, 1 = \text{up}\}$, denote the probability that aggregated machine m_1^b is in state i during time slot n. Thus, m_1^b being up implies that the two-machine line consumes a raw part during this time slot. Then, the evolution of the vector $\mathbf{x}^b(n) = [x_0^b(n)\ x_1^b(n)]^T$ can be expressed as

$$\mathbf{x}^b(n+1) = \mathbf{A}_1^b(n)\mathbf{x}^b(n),$$

where

$$\begin{aligned}\mathbf{A}_1^b(n) &= \begin{bmatrix} 1-R_2^b(n) & P_2^b(n) \\ R_2^b(n) & 1-P_2^b(n) \end{bmatrix}, \\ \mathbf{x}^b(0) &= \begin{bmatrix} 1-CR(0) & CR(0) \end{bmatrix}^T.\end{aligned}$$

Theorem 7.1. *Consider a two-machine geometric line and its equivalent representations resulted from backward and forward aggregations. Then,*

$$PR(n) = x_1^f(n), \quad CR(n) = x_1^b(n).$$

Proof. According to initial condition, $x_1^f(0) = PR(0)$. Assume that $x_1^f(n) = PR(n)$. Then,

$$\begin{aligned}PR(n+1) &= \sum_{i\in I_p} x_i(n+1) \\ &= \sum_{i\in I_p}\sum_{j} \mathbf{A}_{2,ij}x_j(n) \\ &= \sum_{i\in I_p}\sum_{j\in I_{np}} \mathbf{A}_{2,ij}x_j(n) + \sum_{i\in I_p}\sum_{j\in I_p} \mathbf{A}_{2,ij}x_j(n) \\ &= R_2^f(n)[1-PR(n)] + \sum_{i}\sum_{j\in I_p} \mathbf{A}_{2,ij}x_j(n) - \sum_{i\in I_{np}}\sum_{j\in I_p} \mathbf{A}_{2,ij}x_j(n) \\ &= R_2^f(n)[1-PR(n)] + \sum_{j\in I_p} x_j(n)\sum_{i} \mathbf{A}_{2,ij} - \sum_{i\in I_{np}}\sum_{j\in I_p} \mathbf{A}_{2,ij}x_j(n) \\ &= R_2^f(n)[1-PR(n)] + \sum_{j\in I_p} x_j(n) - \sum_{i\in I_{np}}\sum_{j\in I_p} \mathbf{A}_{2,ij}x_j(n) \\ &= R_2^f(n)[1-PR(n)] + PR(n)[1-P_2^f(n)] \\ &= R_2^f(n)x_0^f(n) + [1-P_2^f(n)]x_1^f(n) \\ &= x_1^f(n+1).\end{aligned}$$

Similarly, we can prove that $CR(n) = x_1^b(n)$ for $n = 0, 1, \ldots$. □

It should be noted that, although $CR(n)$ and $PR(n)$ provide the probability, i.e., efficiency, of raw material consumption and finished part production, respectively, neither of them measures the dynamics of the consumption and production of the system. Parameters $P_1^b(n)$, $R_1^b(n)$, $P_2^f(n)$, and $R_2^f(n)$, on the other hand, characterize the dynamics of the two-machine system during transients by viewing the entire system as individual geometric machines. In addition, it can be shown that as the

system approaches steady state, the parameters of the two aggregated machines also converge:

$$P_1^b = \lim_{n\to\infty} P_1^b(n), \quad R_1^b = \lim_{n\to\infty} R_1^b(n), \quad P_2^f = \lim_{n\to\infty} P_2^f(n), \quad R_2^f = \lim_{n\to\infty} R_2^f(n),$$

and,

$$CR_{ss} = \lim_{n\to\infty} \frac{R_1^b(n)}{P_1^b(n) + R_1^b(n)} = \lim_{n\to\infty} \frac{R_2^f(n)}{P_2^f(n) + R_2^f(n)} = PR_{ss},$$

where CR_{ss} and PR_{ss} are the steady state values of the system consumption rate and production rate, respectively. Finally, it should be noted that the production and consumption activities of the original two-machine line are not "Markovian" per se and internal system state must be known in order to predict the system production and consumption in the future. Indeed, during the equivalent aggregation, this information has been incorporated in the calculation of parameters $P_1^b(n)$, $R_1^b(n)$, $P_2^f(n)$, and $R_2^f(n)$ through $x_j(n)$, $PR(n)$ and $CR(n)$. As a result, the up probabilities of the aggregated machines coincide with the production rate and consumption rate of the original system.

As an illustration, the transients of $P_1^b(n)$, $R_1^b(n)$, $P_2^f(n)$, and $R_2^f(n)$ of the system discussed in the previous subsection is given in Figure 7.11. It can be observed in the figure and should also be noted that although the steady state efficiency of the two aggregated machines are identical, their corresponding steady state "uptimes" (i.e., $1/P_1^b$ and $1/P_2^f$) and "downtimes" (i.e., $1/R_1^b$ and $1/R_2^f$) are not necessarily equal.

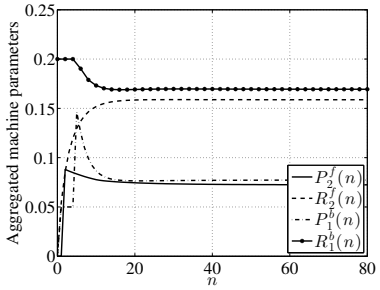


Figure 7.11: Parameters of aggregated machines in a two-machine geometric serial line

Finally, the equivalent aggregation described above also applies to systems with time-varying parameters with the elements in $\mathbf{A}_2$ replaced by the corresponding ones in $\mathbf{A}_2(n)$, i.e.,

$$P_1^b(n) = \frac{\sum_{i\in I_{nc}} \sum_{j\in I_c} \mathbf{A}_{2,ij}(n)x_j(n)}{CR(n)}, \quad R_1^b(n) = \frac{\sum_{i\in I_c} \sum_{j\in I_{nc}} \mathbf{A}_{2,ij}(n)x_j(n)}{1-CR(n)},$$
$$P_2^f(n) = \frac{\sum_{i\in I_{np}} \sum_{j\in I_p} \mathbf{A}_{2,ij}(n)x_j(n)}{PR(n)}, \quad R_2^f(n) = \frac{\sum_{i\in I_p} \sum_{j\in I_{np}} \mathbf{A}_{2,ij}(n)x_j(n)}{1-PR(n)}.$$

In the next section, we will use the results on equivalent aggregation of two-machine geometric lines to develop a method for performance evaluation in multi-machine geometric lines.

7.4 Transient Performance of Multi-Machine Geometric Lines

7.4.1 Exact analysis: Markovian approach

Consider an M-machine geometric line. Due to the memoryless property of geometric random variables, the system is still characterized by a Markov chain. Let $h_i(n)$ denote the number of parts in buffer b_i at the beginning of time slot n. Then, the state of the Markov chain is defined by vector $(s_1(n), \ldots, s_M(n), h_1(n), \ldots, h_{M-1}(n))$, where $s_i(n) \in \{0,1\}$, $i = 1, \ldots, M$, and $h_i(n) \in \{0, 1, \ldots, N_i\}$, $i = 1, \ldots, M-1$. Clearly, the system has a total of $S = 2^M \prod_{i=1}^{M-1}(N_i + 1)$ states. To calculate the transition probabilities among these states, we use the same approach described in Subsection 7.3.1 to linearize the state space. Specifically, arrange the states from State 1 to State S in the following manner: Let $r(s_1, \ldots, s_M, h_1, \ldots, h_{M-1})$ denote the order number of state $(s_1, \ldots, s_M, h_1, \ldots, h_{M-1})$. Define:

$$r(s_1, \ldots, s_M, h_1, \ldots, h_{M-1}) = 1 + \sum_{i=1}^{M} s_i\alpha_i + \sum_{i=1}^{M-1} h_i\beta_i, \tag{7.13}$$

where

$$\begin{aligned}
\alpha_i &= 2^{M-1}, \quad i = 1, \ldots, M, \\
\beta_i &= \begin{cases} 2^M \prod_{j=i+1}^{M-1}(N_j + 1), & i = 1, \ldots, M-2, \\ 2^M, & i = M-1. \end{cases}
\end{aligned}$$

In this manner, each state is assigned a unique number between 1 and S. On the other hand, given the order number r of a system state, $r \in \{1, \ldots, S\}$, the corresponding machine state $s_i^{(r)}$, $i = 1, \ldots, M$, and buffer state, $h_i^{(r)}$, $i = 1, \ldots, M-1$, can be obtained as follows:

$$\begin{aligned}
h_i^{(r)} &= \begin{cases} \left\lfloor \frac{r}{\beta_1} \right\rfloor, & i = 1, \\ \left\lfloor \frac{r - \sum_{j=1}^{i-1} h_j^{(r)} \beta_j}{\beta_i} \right\rfloor, & i = 2, \ldots, M-1, \end{cases} \\
s_i^{(r)} &= \begin{cases} \left\lfloor \frac{r - \sum_{j=1}^{M-1} h_j^{(r)} \beta_j}{\alpha_1} \right\rfloor, & i = 1, \\ \left\lfloor \frac{r - \sum_{j=1}^{M-1} h_j^{(r)} \beta_j - \sum_{j=1}^{i-1} s_j^{(r)} \alpha_j}{\alpha_i} \right\rfloor, & i = 2, \ldots, M, \end{cases}
\end{aligned}$$

where $\lfloor x \rfloor$ represents the largest integer not greater than x.

Similar to the two-machine case, the transitions of $h_i(n)$'s are deterministic given machine states $s_1(n), \ldots, s_M(n)$, and can be calculated based on the following equations:

$$\begin{aligned}
h_i(n+1) &= h_i'(n) + s_i(n) \min\{h_{i-1}(n), N_i - h_i'(n), 1\}, \quad i = 2, \ldots, M-1, \\
h_1(n+1) &= h_1'(n) + s_1(n) \min\{N_1 - h_1'(n), 1\},
\end{aligned} \tag{7.14}$$

where

$$\begin{aligned}
h_{M-1}'(n) &= h_{M-1}(n) - s_M(n) \min\{h_{M-1}(n), 1\}, \\
h_i'(n) &= h_i(n) - s_{i+1}(n) \min\{h_i(n), N_{i+1} - h_{i+1}'(n), 1\}, \quad i = 1, \ldots, M-2.
\end{aligned} \tag{7.15}$$

Next, we arrange all S system states from 1 to S based on the order number calculated in (7.13) and define $x_i(n) = \text{Prob}[\text{System in state } i \text{ in time slot } n]$. Then, the evolution of the state of the Markov chain, $\mathbf{x}(n) = [x_1(n)\ x_2(n)\ \ldots\ x_S(n)]^T$, is given by

$$\mathbf{x}(n+1) = \mathbf{A}_M \mathbf{x}(n), \quad \sum_{i=1}^{S} x_i(n) = 1. \tag{7.16}$$

The performance measures of the system can be calculated as:

$$\begin{aligned}
PR(n) &= \mathbf{D}_1 \mathbf{x}(n) = [d_{1,1} \quad d_{1,2} \quad \cdots \quad d_{1,S}]\, \mathbf{x}(n), \\
CR(n) &= \mathbf{D}_2 \mathbf{x}(n) = [d_{2,1} \quad d_{2,2} \quad \cdots \quad d_{2,S}]\, \mathbf{x}(n), \\
WIP_i(n) &= \mathbf{D}_{3,i} \mathbf{x}(n) = [d_{3,i,1} \quad d_{3,i,2} \quad \cdots \quad d_{3,i,S}]\, \mathbf{x}(n), \quad i = 1, \ldots, M-1, \\
ST_i(n) &= \mathbf{D}_{4,i} \mathbf{x}(n) = [d_{4,i,1} \quad d_{4,i,2} \quad \cdots \quad d_{4,i,S}]\, \mathbf{x}(n), \quad i = 2, \ldots, M, \\
BL_i(n) &= \mathbf{D}_{5,i} \mathbf{x}(n) = [d_{5,i,1} \quad d_{5,i,2} \quad \cdots \quad d_{5,i,S}]\, \mathbf{x}(n), \quad i = 1, \ldots, M-1,
\end{aligned} \tag{7.17}$$

where

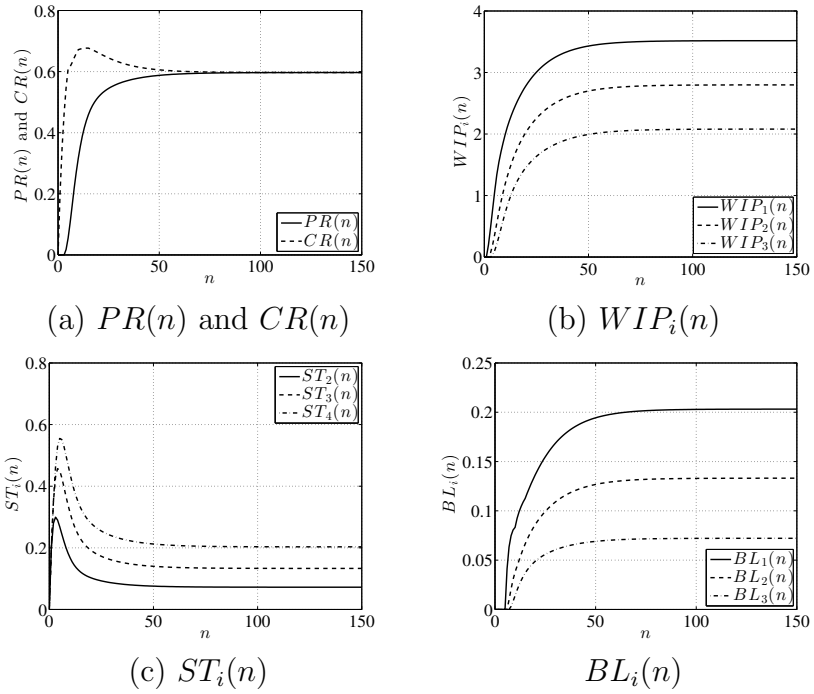
$$\begin{aligned}
d_{1,r} &= \begin{cases} 1, & \text{if } s_M^{(r)} h_{M-1}^{(r)} > 0, \\ 0, & \text{otherwise,} \end{cases} \\
d_{2,r} &= 1 - d_{5,1,r}, \\
d_{3,i,r} &= h_i^{(r)}, \quad i = 1, \ldots, M-1, \quad r = 1, \ldots, S, \\
d_{4,i,r} &= \begin{cases} s_i^{(r)}, & \text{if } i \neq 1 \text{ and } h_{i-1}^{(r)} = 0, \\ 0, & \text{otherwise,} \end{cases} \\
d_{5,i,r} &= \begin{cases} s_i^{(r)} \left[1 - s_{i+1}^{(r)} + d_{5,i+1,r}\right], & \text{if } i \neq M \text{ and } h_i^{(r)} = N_i, \\ 0, & \text{otherwise,} \end{cases}
\end{aligned}$$

i.e., $d_{1,r}$, $d_{2,r}$, $d_{3,i,r}$, $d_{4,i,r}$ and $d_{5,i,r}$ denote the rth element in row vectors $\mathbf{D}_1$, $\mathbf{D}_2$, $\mathbf{D}_{3,i}$, $\mathbf{D}_{4,i}$ and $\mathbf{D}_{5,i}$, respectively. Clearly, all these performance measures are still *linear* in system state $\mathbf{x}(n)$.

As an illustration, consider a four-machine geometric line with machine and buffer parameters:

$$P_i = 0.05, \quad R_i = 0.2, \quad i = 1, \ldots, 4; \quad N_i = 5, \quad i = 1, \ldots, 4.$$

Assume that all machines are initially down and the buffers are initially empty. The transients of the performance measures of this system are illustrated in Figure 7.12.



(a) $PR(n)$ and $CR(n)$ (b) $WIP_i(n)$

(c) $ST_i(n)$ $BL_i(n)$

Figure 7.12: Transients of four-machine geometric line

Similar to the two-machine case, the mathematical model and the performance evaluation formulas obtained above can be extended to systems with time-varying machine parameters as well − by replacing transition probability matrix $\mathbf{A}_2$ with $\mathbf{A}_2(n)$, i.e., by replacing P_i and R_i with $P_i(n)$ and $R_i(n)$, respectively.

7.4.2 Approximate analysis: Recursive aggregation

Idea of the approach

Although equation (7.17) provides closed-form expressions to calculate the exact values of the performance measures in an M-machine serial line with geometric machines

during transients, the cost of using this method may be huge: The calculation requires constructing the S-by-S transition probability matrix $\mathbf{A}_M$, with S growing exponentially in the number of machines and buffer capacities. For instance, for a 10-machine geometric line with all buffer capacities equal to 5, matrix $\mathbf{A}_M$ has a total of 1.0649×10^{20} elements, which requires enormous memory space and computation time to calculate even though the matrix is sparse. To overcome this issue, a computationally efficient procedure is necessary, and, thus, is developed in this subsection by recursively applying the two-machine equivalent aggregation technique described in Subsection 7.3.2 to M-machine lines so as to approximate the transient performance measures.

To approximate the transient performance measures of Bernoulli serial lines, a recursive aggregation procedure has been developed in [70]. The idea of this method is to view the serial line from the perspective of buffer b_i, $i = 1, \ldots, M-1$, and represent the upstream and downstream of the line by two virtual Bernoulli machines with time-varying parameters. The parameters of these virtual Bernoulli machines are obtained by aggregating two Bernoulli machines at a time based on the expressions for the two-machine case. We extend this method to develop a procedure for transient performance estimation in geometric lines.

The idea of this procedure is as follows: Consider a serial line with M geometric machines. Assume that the parts flow in and out of buffer b_i, $i = 1, \ldots, M-1$, can be represented by a two-machine geometric line shown in Figure 7.13. In this representation, machine m_i^f is a modified version of the original machine m_i, which is intended to represent the *aggregated* behavior of all machines and buffers upstream of b_i *producing* parts into buffer b_i. Similarly, machine m_{i+1}^b is a modified version of the original machine m_{i+1}, which is intended to represent the *aggregated* behavior of all machines and buffers downstream of b_i *consuming* parts from buffer b_i. We assume that the two aggregated machines are still geometric but with time-varying

parameters during transients. Next, we apply the backward and forward aggregation described in Subsection 7.3.2 to the two-machine lines shown in Figure 7.14 to form machines m_i^b and m_{i+1}^f, and repeat this procedure to obtain all aggregated machines in these virtual two-machine lines. Note that, during this process, the two machines on the boundary, m_1^f and m_M^b, remain the original machines m_1 and m_M since they are the only system components upstream of b_1 and downstream of b_{M-1}, respectively. Finally, this process may need to be repeated in several iterations to obtain satisfactory parameters for all aggregated machines.

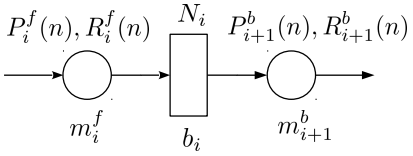
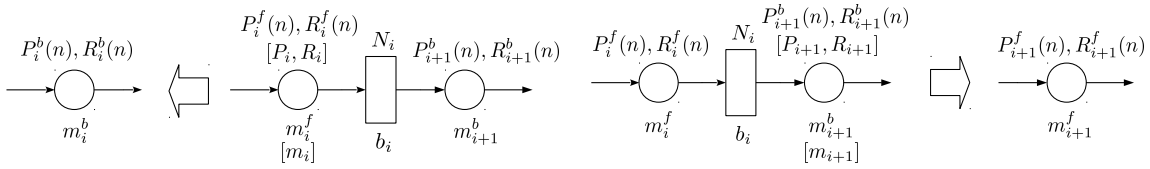


Figure 7.13: Two-machine line representation at buffer b_i in an M-machine geometric serial line



(a) Backward aggregation (b) Forward aggregation

Figure 7.14: Aggregation in an M-machine geometric serial line

Mathematical realization

The mathematical realization of this procedure is given as follows: Assume we are currently at the kth iteration of the procedure. Consider the two-machine line shown in Figure 7.15, where $P_i^f(k;n)$ and $R_i^f(k;n)$ denote the breakdown and repair probabilities of machine m_i^f at time n during the kth iteration of the procedure, and $P_{i+1}^b(k;n)$ and $R_{i+1}^b(k;n)$ denote the breakdown and repair probabilities of machine m_{i+1}^b at time n during the kth iteration of the procedure. The initial buffer occupancy is $h_i(0)$, while the initial state of the two virtual machines, $s_i^f(0)$ and $s_{i+1}^b(0)$, can be

determined by

$$s_i^f(0) = \begin{cases} 1, & \text{if the original machine } m_i \text{ is up and not starved at time 0,} \\ 0, & \text{otherwise,} \end{cases} \tag{7.18}$$

$$s_{i+1}^b(0) = \begin{cases} 1, & \text{if the original machine } m_{i+1} \text{ is up and not blocked at time 0,} \\ 0, & \text{otherwise.} \end{cases} \tag{7.19}$$

Let $\mathbf{x}^{(i)}(k;n) = [x_1^{(i)}(k;n)\ x_2^{(i)}(k;n)\ \ldots\ x_{4(N_i+1)}^{(i)}(k;n)]^T$ and $\mathbf{A}_2^{(i)}(k;n)$ denote the state of this two-machine system and its transition probability matrix at time n during the kth iteration of the procedure, respectively. Then, the evolution of the two-machine line can be calculated as

$$\mathbf{x}^{(i)}(k;n+1) = \mathbf{A}_2^{(i)}(k;n)\mathbf{x}^{(i)}(k;n). \tag{7.20}$$

In addition, the performance measures of the two-machine line can be calculated using equation (7.10). Here, we use $PR^{(i)}(k;n)$ and $CR^{(i)}(k;n)$ to denote the production rate and consumption rate of the system at time n during the kth iteration of the procedure, respectively. Next, consider the two-machine line formed by buffer b_i in the middle but with original machines m_i and m_{i+1} upstream and downstream, respectively (see Figure 7.16). Let $\widetilde{\mathbf{A}}_2^{(i,b)}(k;n)$ and $\widetilde{\mathbf{A}}_2^{(i,f)}(k;n)$ denote the transition probability matrices of the two systems (i.e., with original machines, see Figure 7.16(a) and (b), respectively) at time n during the kth iteration of the procedure. Then, the following equations for the recursive aggregation are formulated:

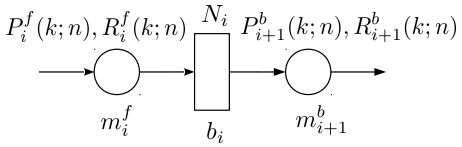
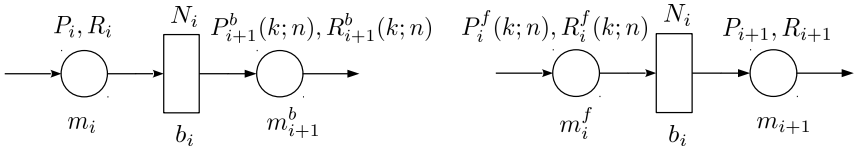


Figure 7.15: Two-machine line representation at buffer b_i at the kth iteration



(a) Backward aggregation (b) Forward aggregation

Figure 7.16: Two-machine lines at buffer b_i with original and aggregated machines

Recursive Procedure 1:

$$
\begin{aligned}
P^b_i(k+1;n) &= \frac{\sum_{j\in I_{nc}}\sum_{l\in I_c}\widetilde{\mathbf{A}}^{(i,b)}_{2,jl}(k;n)x^{(i)}_l(k;n)}{CR^{(i)}(k;n)},\\
R^b_i(k+1;n) &= \frac{\sum_{j\in I_{c}}\sum_{l\in I_{nc}}\widetilde{\mathbf{A}}^{(i,b)}_{2,jl}(k;n)x^{(i)}_l(k;n)}{1-CR^{(i)}(k;n)},\\
&\qquad i=1,\dots,M-1, \quad k=0,1,\dots, \quad n=0,1,\dots,\\
P^f_i(k+1;n) &= \frac{\sum_{j\in I_{np}}\sum_{l\in I_p}\widetilde{\mathbf{A}}^{(i,f)}_{2,jl}(k;n)x^{(i)}_l(k;n)}{PR^{(i)}(k+1;n)},\\
R^f_i(k+1;n) &= \frac{\sum_{j\in I_{p}}\sum_{l\in I_{np}}\widetilde{\mathbf{A}}^{(i,f)}_{2,jl}(k;n)x^{(i)}_l(k;n)}{1-PR^{(i)}(k+1;n)},\\
&\qquad i=2,\dots,M, \quad k=0,1,\dots, \quad n=0,1,\dots,
\end{aligned}
$$

with initial condition

$$P^f_i(0;n)=P^b_i(0;n)=P_i, \quad R^f_i(0;n)=R^b_i(0;n)=R_i, \quad i=1,\dots,M, \quad n=0,1,2\dots.$$

and boundary condition

$$
\begin{aligned}
&P^f_1(k;n)=P_1, \quad R^f_1(k;n)=R_1, \quad P^b_M(k;n)=P_M, \quad R^b_M(k;n)=R_M,\\
&\qquad k=0,1,\dots, \quad n=0,1,2\dots.
\end{aligned}
$$

Numerical Fact 7.1. *For any $n\in\{1,2,\dots,\}$, sequences $P^b_i(k;n)$, $R^b_i(k;n)$, $P^f_i(k;n)$, and $R^f_i(k;n)$, $i=1,\dots,M$, are convergent with respect to k with probability 1, i.e., there exist limits $P^b_i(n)$, $R^b_i(n)$, $P^f_i(n)$, and $R^f_i(n)$, $n=1,2,\dots,$ such that*

$$
\begin{aligned}
Prob\left[\lim_{k\to\infty}P^b_i(k;n)=P^b_i(n)\right]=1, &\qquad Prob\left[\lim_{k\to\infty}R^b_i(k;n)=R^b_i(n)\right]=1,\\
Prob\left[\lim_{k\to\infty}P^f_i(k;n)=P^f_i(n)\right]=1, &\qquad Prob\left[\lim_{k\to\infty}R^f_i(k;n)=R^f_i(n)\right]=1,\\
i=1,\dots,M, &\qquad n=1,2,\dots.
\end{aligned}
$$

Justification: To justify this numerical fact, a total of 180,000 production lines were generated randomly with 10,000 for each $M \in \{3, 4, \ldots, 20\}$. For each line, the machine repair probability and machine efficiency were selected randomly and equiprobably from the following sets:

$$R_i \in (0.05, 0.5), \quad e_i \in (0.6, 0.99).$$

In other words, the average downtime of a machine is randomly selected from a range of 2 to 20 cycle times, with efficiency from 60% to 99%. These parameter ranges were used so that they could reflect typical production situations on the factory floor. Then, the breakdown probability can be calculated based on the relationship that $e_i = T_{up,i}/(T_{up,i} + T_{down,i}) = R_i/(R_i + P_i)$. Next, the capacity of each buffer was randomly selected from

$$N_i \in \{\lceil T_{down,i} \rceil, \lceil T_{down,i} \rceil + 1, \ldots, 5\lceil T_{down,i} \rceil\}.$$

Each machine was initially up or down with probability 0.5, and the initial occupancy of each buffer was randomly and uniformly selected from $\{0, 1, \ldots, N_i\}$. The total time duration T of transient study was selected as 1,000 time slots, since preliminary experiments showed that for most production lines with parameters selected from the above ranges, $PR(n)$ entered steady state before 1000 time slots. For each line, thus constructed, Recursive Procedure 1 was performed. To determine the convergence of the procedure, introduce

$$\begin{aligned}\Delta(k) \;=\; \sum_{i=1}^{M}\sum_{n=1}^{T} \Big[\big|P_i^b(k;n) - P_i^b(k-1;n)\big| + \big|R_i^b(k;n) - R_i^b(k-1;n)\big| + \\ \Big|P_i^f(k;n) - P_i^f(k-1;n)\Big| + \Big|R_i^f(k;n) - R_i^f(k-1;n)\Big| \Big].\end{aligned}$$

During the justification, it was observed that for each of the 180,000 lines generated above, there always exists an k_0 such that $\Delta(k) \leq 10^{-6}$ for $k \geq k_0$ (in most cases, $k_0 < 20$). Therefore, we conclude that Numerical Fact 7.1 holds, i.e., Recursive

Procedure 1 converges with probability 1.

■

Performance estimation

Based on Recursive Procedure 1 and Numerical Fact 7.1, the estimates of transient performance measures for multi-machine geometric lines can be formulated as follows:

$$\widehat{PR}(n) = PR(n; \mathbf{P}^f_{M-1}(n), \mathbf{R}^f_{M-1}(n), \mathbf{P}^b_M(n), \mathbf{R}^b_M(n), N_{M-1}), \tag{7.21}$$

$$\widehat{CR}(n) = CR(n; \mathbf{P}^f_1(n), \mathbf{R}^f_1(n), \mathbf{P}^b_2(n), \mathbf{R}^b_2(n), N_1), \tag{7.22}$$

$$\widehat{WIP}_i(n) = WIP(n; \mathbf{P}^f_i(n), \mathbf{R}^f_i(n), \mathbf{P}^b_{i+1}(n), \mathbf{R}^b_{i+1}(n), N_i), \tag{7.23}$$

$$\widehat{ST}_i(n) = \frac{\text{Prob}[m_i \text{ is up at time } n]}{\text{Prob}[m^f_i \text{ is up at time } n]} \cdot ST(n; \mathbf{P}^f_{i-1}(n), \mathbf{R}^f_{i-1}(n), \mathbf{P}^b_i(n), \mathbf{R}^b_i(n), N_{i-1}) \tag{7.24}$$

$$\widehat{BL}_i(n) = \frac{\text{Prob}[m_i \text{ is up at time } n]}{\text{Prob}[m^b_i \text{ is up at time } n]} \cdot BL(n; \mathbf{P}^f_i(n), \mathbf{R}^f_i(n), \mathbf{P}^b_{i+1}(n), \mathbf{R}^b_{i+1}(n), N_i), \tag{7.25}$$

where

$$\begin{aligned}
\mathbf{P}^f_i(n) &= \left[P^f_i(0)\ P^f_i(1)\ \ldots\ P^f_i(n)\right], \\
\mathbf{R}^f_i(n) &= \left[R^f_i(0)\ R^f_i(1)\ \ldots\ R^f_i(n)\right], \\
\mathbf{P}^b_i(n) &= \left[P^b_i(0)\ P^b_i(1)\ \ldots\ P^b_i(n)\right], \\
\mathbf{R}^b_i(n) &= \left[R^b_i(0)\ R^b_i(1)\ \ldots\ R^b_i(n)\right],
\end{aligned}$$

and $PR(n; \mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}, \mathbf{v_4}, N)$, $CR(n; \mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}, \mathbf{v_4}, N)$, $WIP(n; \mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}, \mathbf{v_4}, N)$, $ST(n; \mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}, \mathbf{v_4}, N)$ and $BL(n; \mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}, \mathbf{v_4}, N)$ denote the production rate, consumption rate, work-in-process, second machine starvation and first machine blockage of a two-machine geometric line with buffer capacity N and machine parameters specified by vectors $\mathbf{v_1}$ (first machine breakdown probabilities from time 0 to n), $\mathbf{v_2}$ (first machine repair probabilities from time 0 to n), $\mathbf{v_3}$ (second machine breakdown probabilities from time 0 to n) and $\mathbf{v_4}$ (second machine repair probabilities from time 0 to n).

To evaluate the accuracy of these estimates, the following metrics are defined:

$$\begin{aligned}
\delta_{PR}(n) &= \frac{|PR(n) - \widehat{PR}(n)|}{PR_{ss}} \cdot 100\%, \\
\delta_{CR}(n) &= \frac{|CR(n) - \widehat{CR}(n)|}{CR_{ss}} \cdot 100\%, \\
\delta_{WIP_i}(n) &= \frac{|WIP_i(n) - \widehat{WIP}_i(n)|}{N_i} \cdot 100\%, \\
\delta_{ST_i}(n) &= |ST_i(n) - \widehat{ST}_i(n)|, \\
\delta_{BL_i}(n) &= |BL_i(n) - \widehat{BL_i}(n)|.
\end{aligned} \tag{7.26}$$

Then, we calculated these accuracy metrics for all the 180,000 lines generated in the justification of Numerical Fact 7.1. Specifically, performance measure estimate $\widehat{PR}(n)$, $\widehat{CR}(n)$, $\widehat{WIP_i}(n)$, $\widehat{ST_i}(n)$ and $\widehat{BL_i}(n)$ were evaluated using (7.21)-(7.25) with Recursive Procedure 1 terminated at $\Delta(k) \leq 10^{-6}$. On the other hand, due to long calculation time required by exact evaluation of the performance measures, the "true" transient performance involved in (7.26) as well as the steady state performance PR_{ss} were evaluated using simulations based on the following:

Simulation Procedure 1:

(1) Set the initial state of machines and the initial occupancy of the buffers the same as the line generated in the justification of Numerical Fact 7.1.

(2) For transient performance evaluation, carry out 10,000 runs of the simulation code for each line and calculate the average performance during each time slot.

(3) For steady state performance evaluation, carry out 20 runs of the simulation code for each line. In each run, use the first 20,000 time slots as a warm-up period and the subsequent 400,000 time slots to statistically calculate the average performance.

(4) This results 95% confidence intervals of less than 0.001 for PR_{ss} and 0.005 for $PR(n)$.

The resulting means and standard deviations of the accuracy metrics defined in (7.26) are summarized in Figure 7.17. As one can see from the figure, although the accuracy of the aggregation procedure varies significantly with the configuration of the production lines considered (see the standard deviations of the errors illustrated in the right column of Figure 7.17), the average of $\delta_{PR}(n)$ and $\delta_{CR}(n)$ is typically within 3-4%. Due to the higher complexity of the production system and the recursive procedure, the level of accuracy is slightly lower than that of a similar technique developed for Bernoulli serial lines reported in [70].

It should be noted that numerical experiments showed that the accuracy of the recursive procedure is low when the buffer capacity in the system is small. This observation is illustrated below for a five-machine line with identical machines $P_i = 0.01$ and $R_i = 0.05$ and identical buffers $N_i = N$. In this illustration, the buffers are initially empty and the machines are initially up. Then, we increased the buffer capacity N from 1 to 20 and evaluate the average accuracy of the recursive procedure based on:

$$\bar{\delta}_{PR} = \frac{1}{T}\sum_{n=1}^{T}\delta_{PR}(n), \quad \bar{\delta}_{CR} = \frac{1}{T}\sum_{n=1}^{T}\delta_{CR}(n), \quad \bar{\delta}_{WIP} = \frac{1}{T}\sum_{n=1}^{T}\delta_{WIP}(n),$$
$$\bar{\delta}_{ST} = \frac{1}{T}\sum_{n=1}^{T}\delta_{ST}(n), \quad \bar{\delta}_{BL} = \frac{1}{T}\sum_{n=1}^{T}\delta_{BL}(n).$$

The results are summarized in Figure 7.18 for $T = 800$. As one can see, the averages of δ_{PR}, δ_{CR}, δ_{WIP}, δ_{ST} and δ_{BL} are all monotonically decreasing in N. As N becomes larger than 10, i.e., half of the machine average downtime, $\bar{\delta}_{PR}$ and $\bar{\delta}_{CR}$ are below 5%, $\bar{\delta}_{WIP}$ is less than 2.5%, and $\bar{\delta}_{ST}$ and $\bar{\delta}_{BL}$ are less than 0.02. Similar observations have been made for lines with non-identical machines and non-identical buffers as well that *the accuracy of the recursive procedure is satisfactory when the capacity of each buffer in the line is at least half of the longer average downtime of its two surrounding machines.*

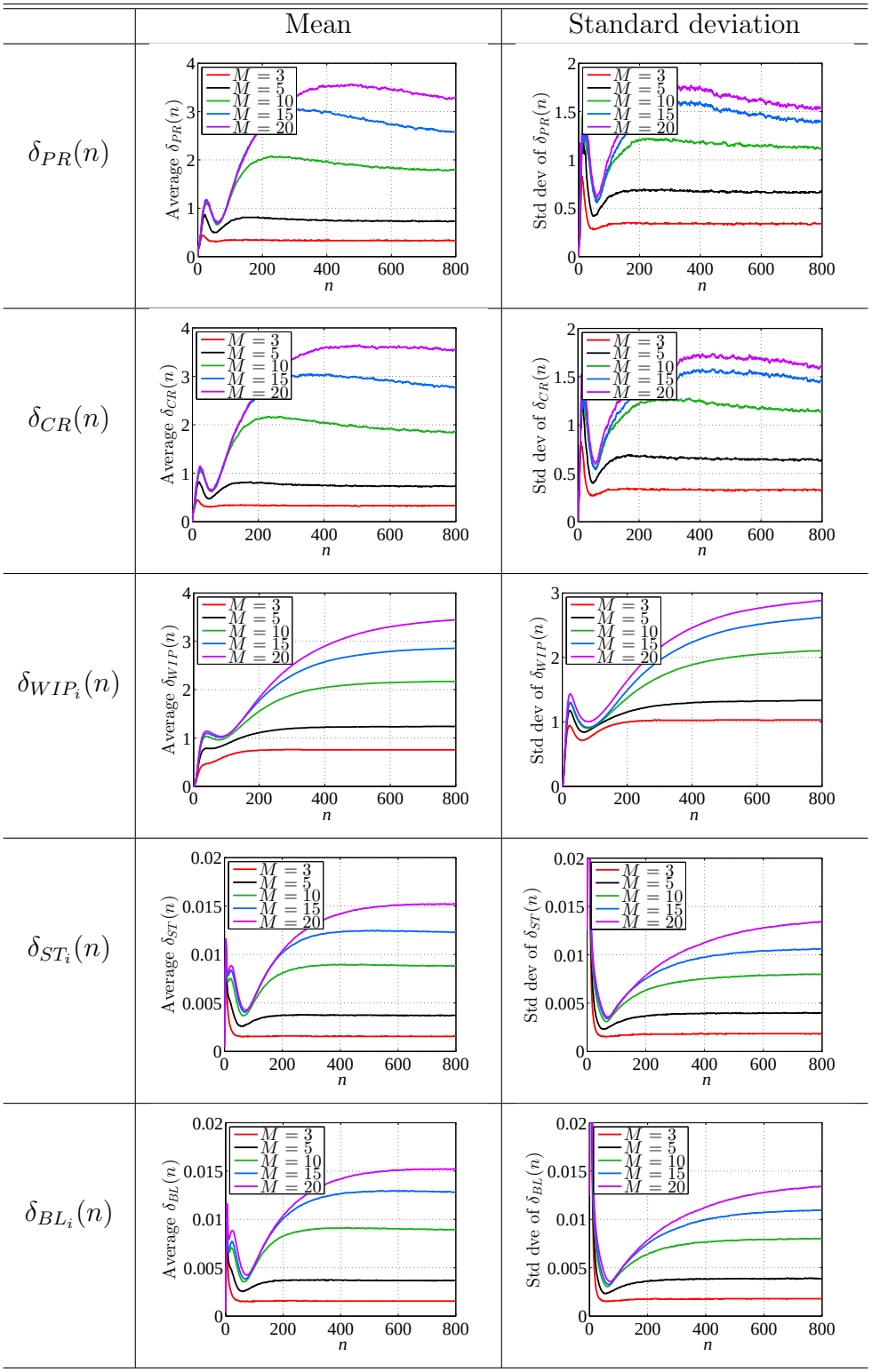


Figure 7.17: Accuracy of transient performance estimates

Thus, taking into account that the parameters of the machines and buffers are rarely known on the factory floor with accuracy better than 5%-10% and that in-process buffers are usually selected to accommodate at least one average downtime in practice, we claim that Recursive Procedure 1 and equations (7.21)-(7.25) can be used to effectively approximate the performance measures of serial lines with geometric machines during transients. In addition, since all calculations involved in Recursive Procedure 1 are based on two-machine line formulas, the computational effort is greatly reduced compared to the exact calculations. Finally, since no other analytical methods are available for transient performance evaluation of the production systems considered in this chapter, no comparisons can be carried out at this point.

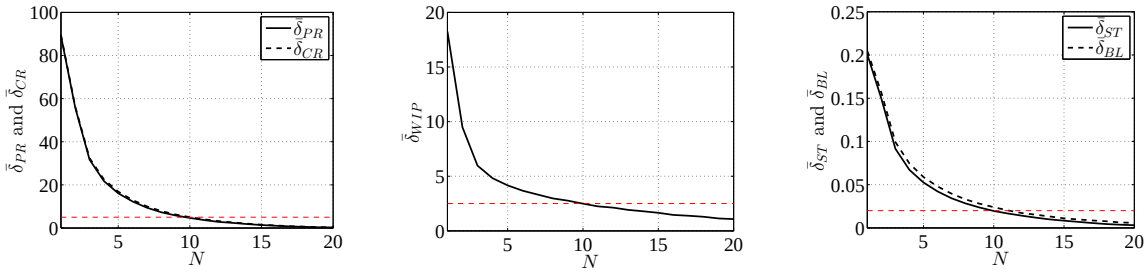


Figure 7.18: Accuracy of transient performance as functions of buffer capacity N

As an illustration, consider a randomly generated six-machine geometric serial line with machine and buffer parameters and initial condition given in Table 7.3. The transient performance measures of this system, obtained using Simulation Procedure 1 and their approximate evaluation obtained using (7.21)-(7.25), are plotted in the left column of Figure 7.19. The accuracy metrics calculated using (7.26) are also plotted in the right column of Figure 7.19. As one can see from the figure, the estimates track the real transient performance closely with very small errors.

Extension to systems with time-varying parameters

The recursive aggregation procedure developed in the above subsection can be extended to geometric lines with machines having time-varying parameters as well. In

Table 7.2: System parameters and initial condition

i	1	2	3	4	5	6
P_i	0.013	0.049	0.019	0.030	0.003	0.011
R_i	0.100	0.132	0.127	0.112	0.126	0.112
N_i	13	8	19	21	18	–
$s_i(0)$	1	1	1	1	1	1
$h_i(0)$	4	4	16	17	17	–

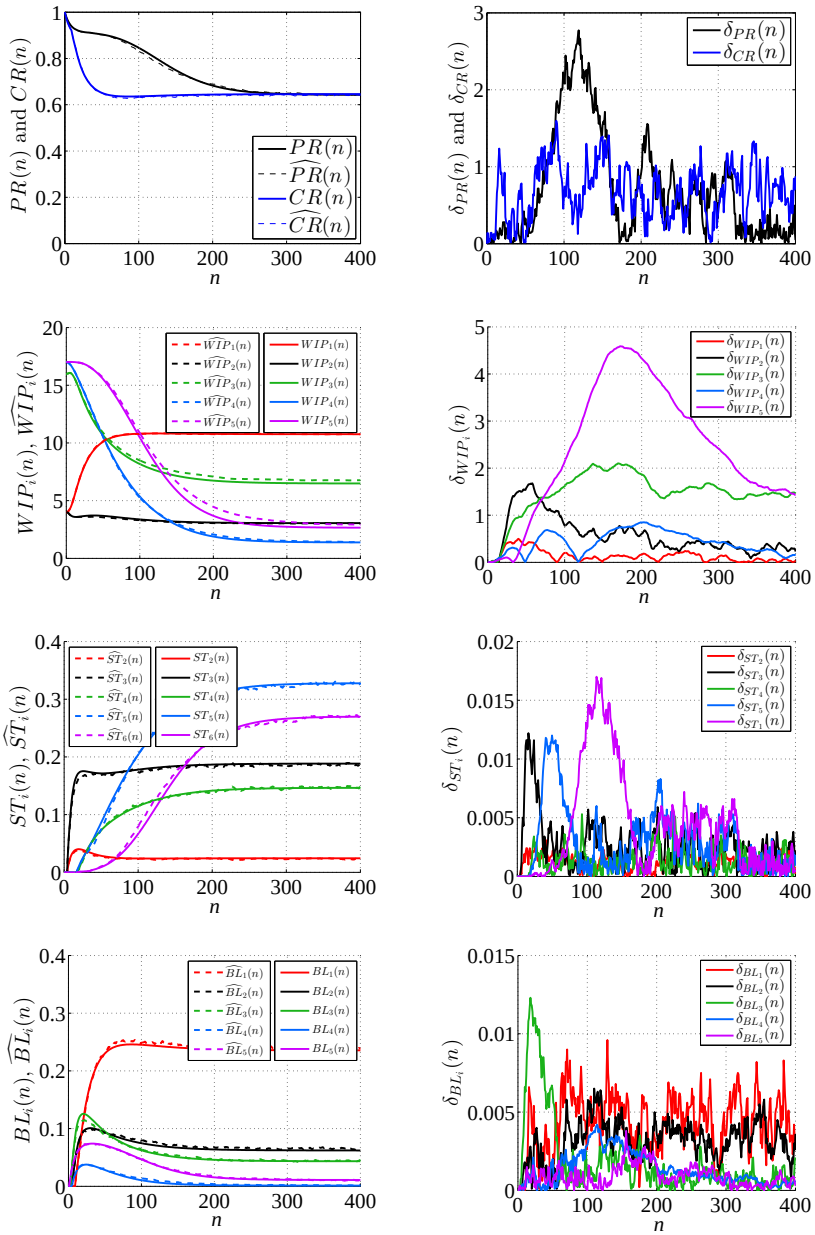


Figure 7.19: Comparison of simulation and approximate evaluation of transient performance measures of a geometric serial line

this case, the aggregation formulas remain the same but the transition probability matrices of the two-machine lines with original machines, $\widetilde{\mathbf{A}}_2^{(i,b)}$ and $\widetilde{\mathbf{A}}_2^{(i,f)}$, must be modified to take into account the time-varying machine parameters, i.e., with P_i and R_i replaced by $P_i(n)$ and $R_i(n)$, respectively. Similar modifications in the initial condition and boundary condition also must be made:

Initial condition:

$$P_i^f(0;n) = P_i^b(0;n) = P_i(n), \quad R_i^f(0;n) = R_i^b(0;n) = R_i(n), \quad i = 1,\ldots,M,\ n = 0,1,2\ldots.$$

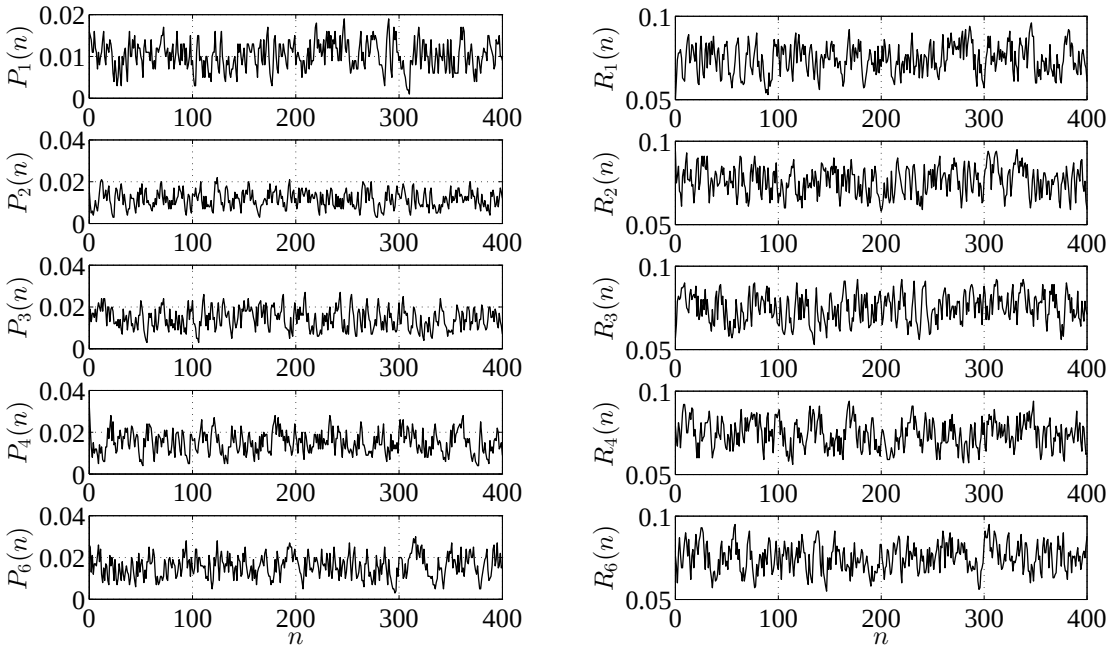
Boundary condition:

$$P_1^f(k;n) = P_1(n), \quad R_1^f(k;n) = R_1(n), \quad P_M^b(k;n) = P_M(n), \quad R_M^b(k;n) = R_M(n),$$
$$k = 0,1,\ldots, \quad n = 0,1,2\ldots.$$

Using similar numerical experiments, we obtained that the recursive procedure is also convergent with probability 1. Moreover, equations (7.21)-(7.25) can still be used to estimate the system transient performance with similar accuracy compared to the results described in the previous subsection. As an illustration, consider a six-machine geometric line with machines' time-varying breakdown and repair probabilities plotted in Figure 7.20. The capacities of the buffers as well as the system initial condition are given in Table 7.3. For this system, the recursive procedure converges at the 6th iteration. The transients of the system's performance measures, obtained using Simulation Procedure 1 and formulas (7.21)-(7.25), are plotted in Figure 7.21. As one can see, even with highly oscillating system parameters, the recursive procedure and formulas (7.21)-(7.25) can still track the real transient performance very closely.

Table 7.3: Buffer capacities and system initial condition

i	1	2	3	4	5	6
N_i	43	33	22	20	17	–
$s_i(0)$	1	1	0	1	0	0
$h_i(0)$	16	11	5	19	8	–



(a) $P_i(n)$ (b) $R_i(n)$

Figure 7.20: Time-varying breakdown and repair probabilities of machines

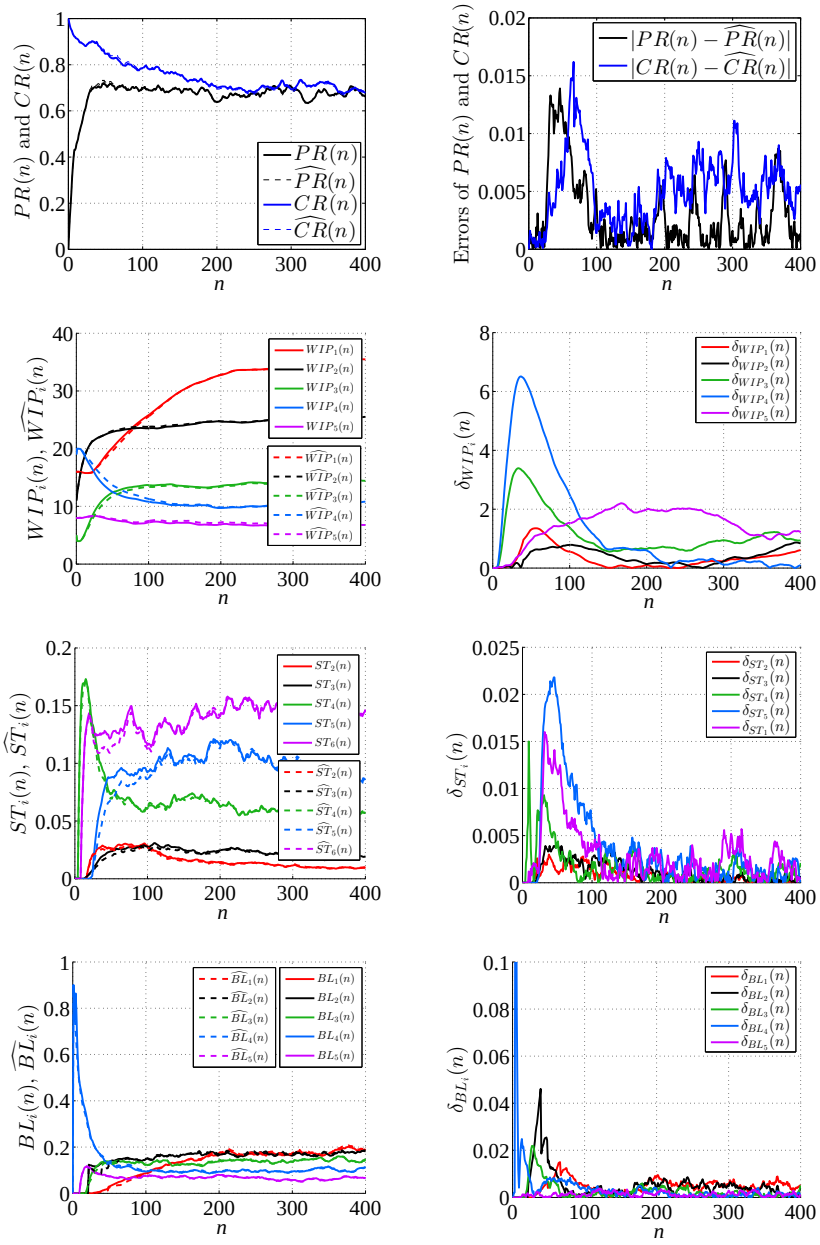


Figure 7.21: Comparison of simulation and approximate evaluation of transient performance in a geometric serial line with time-varying parameters

7.5 Summary

In this chapter, we study the problem of performance evaluation in serial production lines with finite buffers and machines having the geometric reliability model. Specifically, we derive mathematical model for two-machine geometric lines and closed-form formulas for transient performance evaluation. In addition, we develop the equivalent aggregation technique to represent the dynamics of a two-machine geometric line as a single geometric machine with time-varying parameters. Then, for $M > 2$-machine lines, we, again, derive mathematical model and closed-form expressions for transient performance evaluation using Markovian analysis. The resulting formulas, however, requires extremely large amount of computing effort. To resolve this issue, a computationally efficient algorithm based on recursive applications of the two-machine equivalent aggregation technique is developed to approximate the transient production rate with high accuracy.

CHAPTER VIII

SIMULATION AND ANALYSIS OF BREWERY PRODUCTION SYSTEM

8.1 Introduction

Due to insufficient production capacity in a local brewy company to meet increasing market demand, we constructed a discrete event model and use ARENA, one of the leading simulation software in the market, to simulate the brewy production line to assess capacity utilization and seek potential improvements.

The system consists of filtration machines, storage tanks and packaging lines (see Figure 8.1). There are three filtration machines: Pall1, Pall2 and CF, three storage tanks with finite capacity: CFT1, CFT2 and CFT3, and three packaging lines: B11, B14 and C10. Sufficient beer will be provided to filtration machines, i.e., they will never be starved, which fill the storage tanks with beers; packaging lines then pull the beers from storage tanks and fill them into bottles or cans of beers which will be shipped out. It is assumed that the packaging lines will never be blocked.

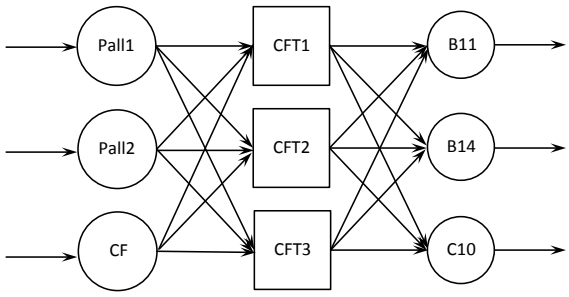


Figure 8.1: Brewy production line

The brewy line produces four brands of beers: G5, JK, JC and 17. The system currently operates six days a week (idle on Sunday) and 24 hours a day. The model replicates one month's production and will be verified by actual production of the brewy line.

8.2 Modeling

8.2.1 Model Discretization

The flow of beer in the simulation is modeled as discretized into a series of "entities", where each entity represents one barrel of beer. For a producing unit operating at a rate of 150 bbl/hr, one entity represents 24 seconds of production, which can be considered as sufficiently fine for weekly or monthly study. In this simulation model, we use discrete entities to approximate continuous flow of liquid. This technique is widely used in simulation studies. As long as the volume represented by each entity is significantly smaller than the production duration, the simulation model is valid. In the current setting, we assume each entity represent one barrel of beer, which is approximately 10.6 seconds of production at Pall1. In this case, we referred to the time interval between two consecutive entities as cycle time, or cycle. For instance Pall1's cycle time is 3600/340 bph = 10.59 seconds/entity.

8.2.2 Filtration Machines

Each filter is characterized by its maximum speed (Pall1speed = 340 bbls/hr, Pall2speed = 340 bbls/hr, CFspeed = 385 bbls/hr), maintenance schedule, initialization time, and brand change-over matrix. When the simulation runs, the filter operates reliably and continuously at its maximum speed unless the current batch finishes, the downstream storage tanks is full, or the scheduled maintenance occurs. Each filter is subject to four types of maintenance activities at every 6,000, 30,000, 100,000 and 250,000 bbls of production with durations of 3 hours, 18 hours, 6 hours, and 6 hours,

respectively. In practice, to allow flexibility, a pair of upper bound and lower bound is provided for each maintenance type so that an ending task will not be interrupted by maintenance (LB6K = 4,000 bbls, UB6K = 6,500 bbls, LB30K = 25,000 bbls, UB30K = 33,000 bbls, LB100K = 95,000 bbls, UB100K = 105,000 bbls, LB250K = 237,500 bbls, UB250K = 262,500 bbls). Moreover, before each batch of production, an initialization is required for the filters (3 hours for the Pall1, Pall2, and 3.5 hours for CF). A 4-by-4 matrix defines the change-over times among different brands for each filter with range from 10 minutes to 40 minutes. Finally, the CF must be shutdown for 6 hours after every 60 hours of continuous production.

The simulation logic for the Pall1 filter is shown in Figure 8.2. In this model, an entity is generated after 3 hours of initial startup and sterilization. Then, the model checks if the current order of beer has been completed. If the order is complete, then change over will be conducted on Pall1; otherwise, we hold the entity until its release is authorized. The entity is then sent through Pall1 and proceeds to CFT1 or other tanks depending on schedule. The production and system statistics are updated and the next entity will wait for one cycle to enter the filtration system. Other filtration machines maintain identical logic.

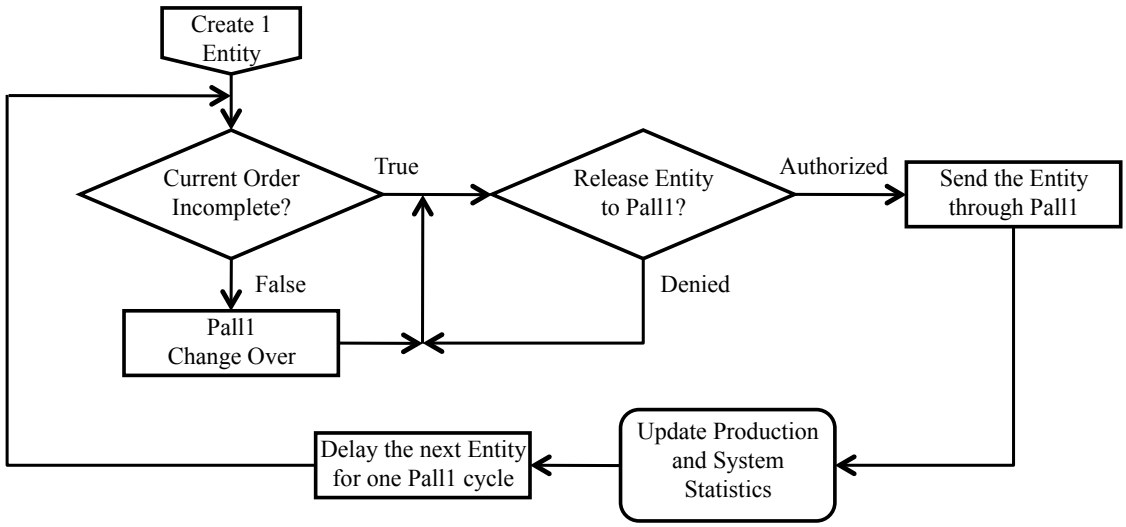


Figure 8.2: Simulation logic for Pall1 filter

8.2.3 Storage Tanks

The current simulation model has three tanks, with capacities of 1,350 bbls, 1,350 bbls and 1,850 bbls respectively. During production, the tank must maintain a minimum level of beer which is 85 bbls. On the other hand, when a tank is full, the upstream filter stops supplying beer to the tank until an empty space of at least 150 bbls is available (by the downstream draw) in this tank. In addition, at the beginning of each task, the tank is first "primed" with beer to 675 bbls before the downstream fillers start to pull. Finally, the residual left in storage tanks at the end of a task is "dumped" at the speed of 100 bbls/hr, or 25 minutes, whichever takes longer. Similar to the filters, a 4-by-4 matrix defines the change-over times among different brands for each tank.

The simulation module for storage tanks is constructed by continuously monitoring the occupancy and determine if they need change over or weekly cleaning/sterilization. Since no production from the filtration machines or the packaging lines can be performed before the tanks finish sterilization, this module will be responsible of sending central signals that will trigger the production of the entire system. In this module, however, the entity that flows around is not a unit of beer but a virtual monitor. When a sterilization process (either weekly or change over) starts, we set the status, for instance, CFT1 by using CFT1Sterilizing = 1. After the sterilization completes, we set the variable back to 0. If the sterilization is for the purpose of change over, then the module will send signal to Pall1 and all three packaging lines when its change over is finished. The logic is shown in Figure 8.3.

8.2.4 Packaging Lines

During production, the packaging lines operate at their maximum speeds (155 bbls/hr for B11, 191 bbls/hr for B14 and 308 bbls/hr for C10). Similar to the filtration machines module, each entity in packagine lines module also represents one barrel

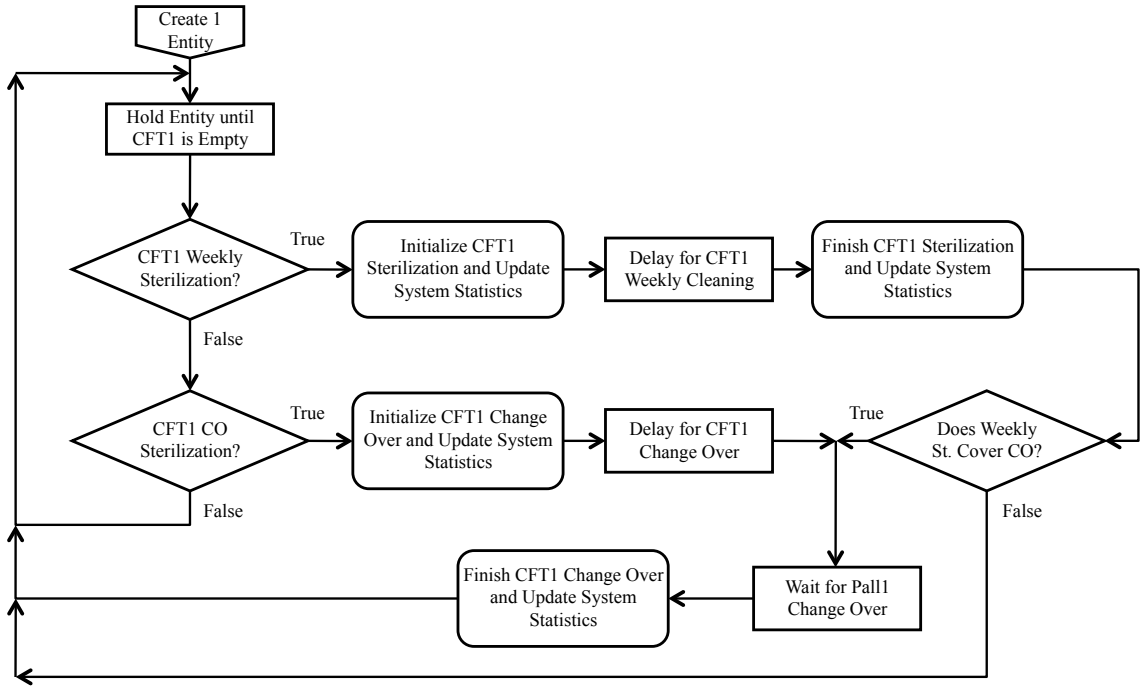


Figure 8.3: Simulation logic for CFT1

of beer. However, since packaging lines produce at a lower speed , the cycle time is different. In this case, assuming that each entity is one barrel, for instance, the cycle time of B11 is 3600/155 bph = 23.23 seconds/entity.

Each packaging line undergoes weekly preventive maintenance with duration 8 hours, which occurs on every Tuesday morning for B11, every Monday morning for B14, and every Wednesday morning for C10. Due to possible stops in the packaging lines, such as wrong labeling, stuck and other mechenical issues, each of them is considered as "unreliable", i.e., prone to failures of stoppages. The reliability is quantified by the efficiencies (88.94%, 79.16% and 81.75%, respectively) of the lines and average downtimes per occurrence. In addition, we assume the up- and downtime of the lines follow exponential distribution (average uptime is 24.12 minutes, 11.4 minutes and 13.44 minutes, respectively, and average downtime is 3 minutes). Finally, a 4-by-4 matrix defines the change-over times among different brands for each packaging line. The logic is shown in Figure 8.4.

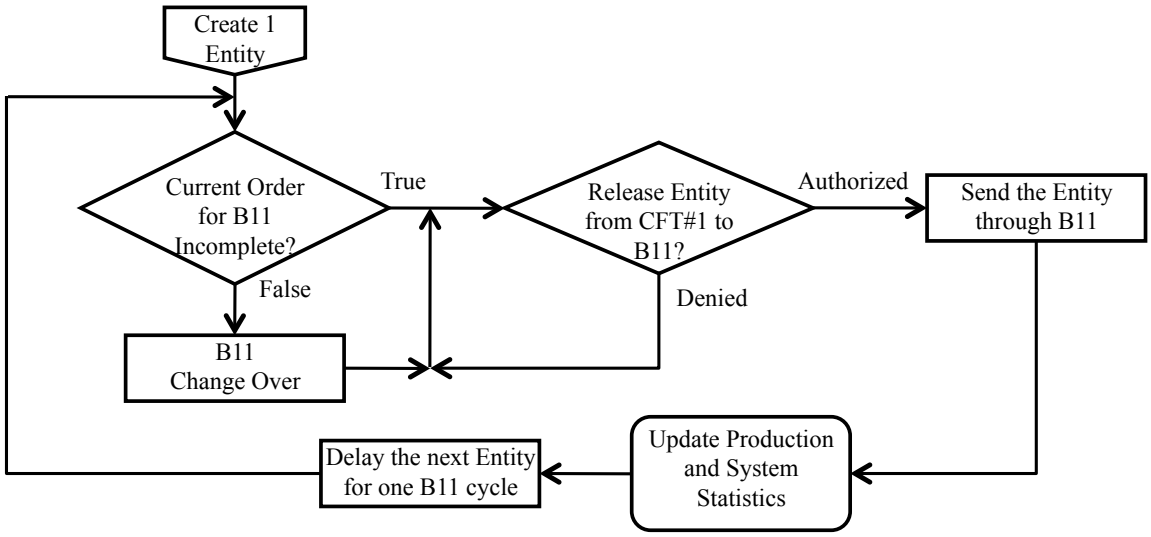


Figure 8.4: Simulation logic for B11

8.2.5 Interaction Among Components

According to operation practice, a filter can only supply beer to one tank at any time, while a tank can supply arbitrary number of packaging lines simultaneously. Some packaging lines can begin earlier than others. As mentioned above, during production, the downstream packaging lines stop drawing beer from the tank, if its occupancy hits the minimum level of beer. Also, when a tank is full, the upstream filter stops supplying beer to the tank until an empty space of 150 bbls is created in this tank.

The animation of this ARENA simulation model is shown in Figure 8.5 (the extra filter Pall3 and extra packaging line C11 are for further what-if analysis).

8.2.6 Schedule

One month's production will be scheduled ahead specifying the beginning of each task, beer brand, volume of beer, filter, tank, and packaging line/lines. Figure 8.6 summarizes the prodcution schedule in August, 2011. Note that CF is not in production in this month (it produced around 20,000 bbls in June, 2011). In the figure, the four draft brands are illustrated in different colors, while the production volumes

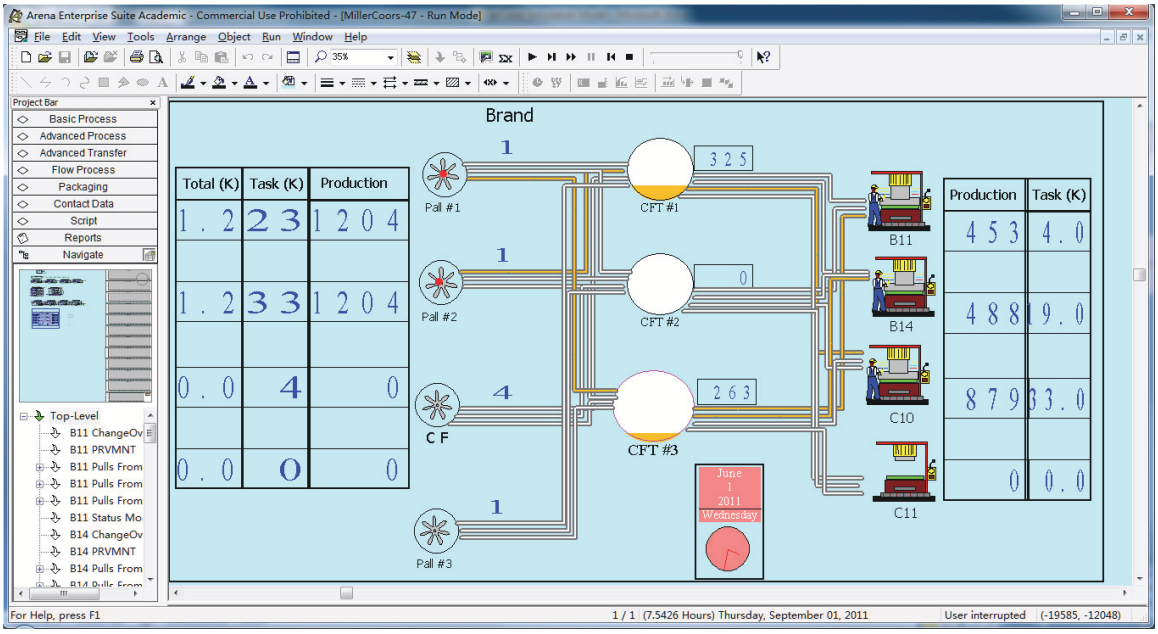


Figure 8.5: Simulation animation

Day		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	9/1
AM	Pall1	15199				9067				14088			10609				15238			9843			14297			13178				12194			
PM																																	
AM	Pall2	180	15697							3937			7615			439	16094						4055			13088				400		5733	
PM																																	
AM	B11										929		4351													3693							
PM																																	
AM	B14	180	15472							3008			3264			439	16094						4055			9395				400		5733	
PM																																	
AM	C10	15199				9067				14088			10609				15238			9843			14297			13178				12194			
PM																																	

Brands: G5 JK JC 17 Single underline: CFT1 Double underline: CFT2

Figure 8.6: August production schedule

scheduled for each equipment are illustrated by the numbers in the colored blocks. The filter-packing assignment for each task can be easily determined by looking at the production volumes, while the assignment of the tanks is illustrated by the underlines of the production volume (one line CFT1, double lines CFT2). During the investigation, a few interesting rules are used in the BrewAS schedule: Pall1 always releases beer to CFT1, and only supplies C10, while Pall2 always releases to CFT2, and only supplies B11 and B14. CF and CFT3 are not used in the schedule. In addition, Pall2 supplies beer to B11 and B14 simultaneous only when the current task at one of two packaging lines is almost finished so that the intermediate tank has sufficient work-in-process to avoid starvation.

To formulate the production schedule into a form that can be read by the simulation model, we introduce the notions:

Batch: Certain volume of beer of a single brand that is scheduled to be continuously produced through the same filter and same tank but can be packaged in one or more packaging lines;

Task: Part of a batch that belongs to the same packaging line.

Therefore, a batch consists of one (if a single packaging is involved) or more (if more than one packaging is involved) tasks. We introduce the notion of task so that beer of the same brand but different packaging can be produced in a serial matter to avoid potential starvation. Each batch/task is defined by its earliest starting time (day, hour, minute), brand, total volume required for each packaging line, the filter to be used, and the tank to be used. Note that only one filter and one tank are used in a single batch/task.

In addition, it is estimated that 5% of beer is lost in packaging and 2% extra beers will be supplied to the filters to avoid possible shortage of beer. Thus, the total volume of beer pumped to the filter is 7%.

8.2.7 Data Collection

During a simulation run, the program tracks the states of every component of the system (filters, tanks, and packaging lines). The summarized results in an external MS Excel file include the beginning time and finish time, the total time duration that the components reside in each state: Idle (during weekdays), Saturday, Sunday, Starvation, Downtime, Running, Changeover, Maintenance, and Blockage. Specifically, we define:

- A packaging line is *starved* at a time moment if a task is scheduled at the packaging line and the tank that supplies the line is below the minimum level.
- A filter is *blocked* if the filter still has beer to produce for the current batch/task but the tank assigned to the batch/task is full.

Based on this information, we can summarize the utilization and the performance for each component in this month. A chart based on the Excel output is shown in Figure 8.7.

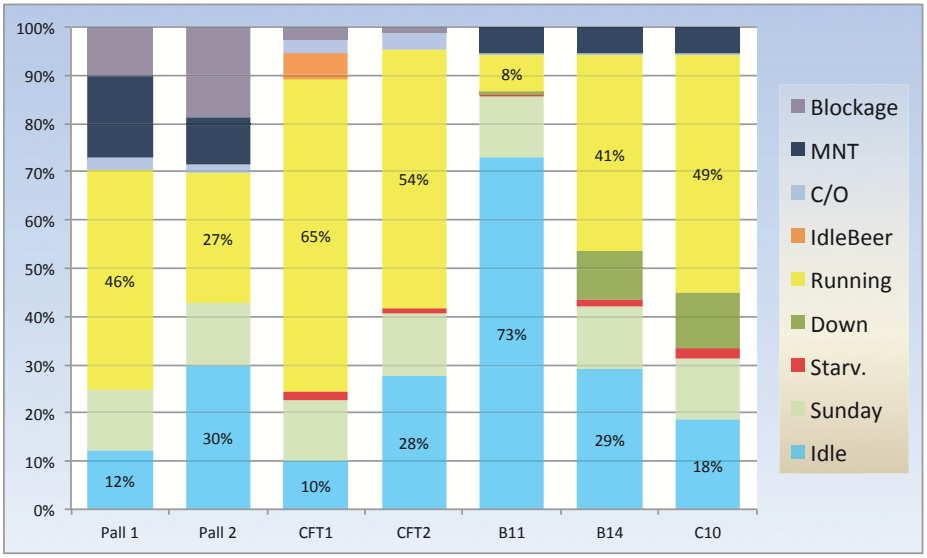
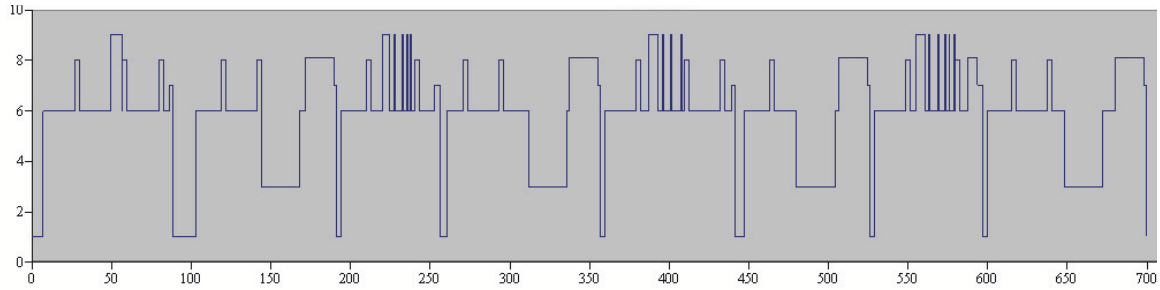
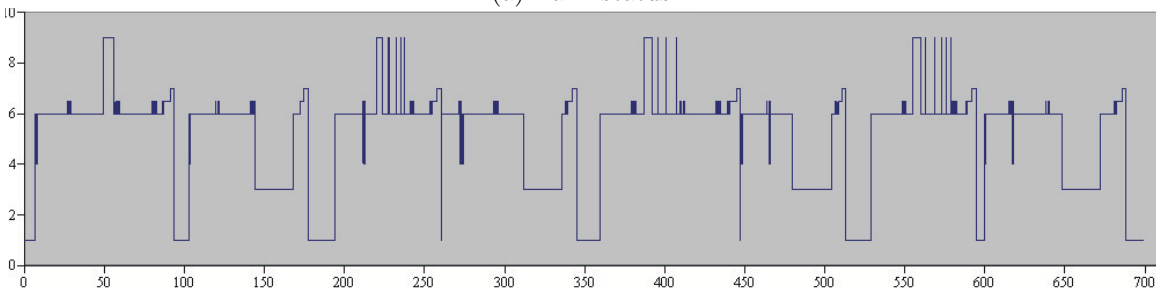


Figure 8.7: Components status chart

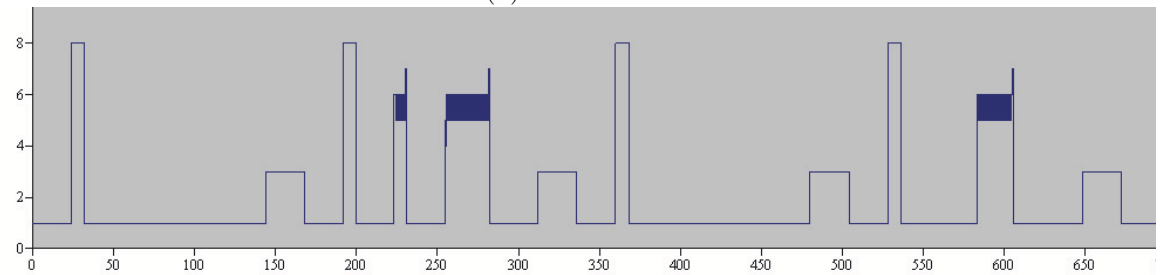
Besides the Excel file, the simulation model can also display the status of each component durign the one month produciton on the main screen of ARENA, which is excerpted in Figure 8.8.



(a) Pall1 status



(b) CFT1 status



(c) B11 status

Figure 8.8: Selected components status on ARENA screen

8.2.8 Customizability

All parameters mentioned above can be modified based on actual production scenario in the Excel file. The schedule of the batches/tasks can also be modified in the same file. This file is used as the input to the simulation program. In addition, the simulation can be expanded to include other parts of the brewery, for instance, other (non-draft) brands.

8.3 Model Validation

The model is run accroding to August schedule. The results are shown in Figure 8.7 and compared with the actual data collected from the packaging lines which is shown in Table 8.1.

Table 8.1: Packaging lines comparison (in hours)

Line		Qty(bbls)	Obs	Prod	Non_Prod	Downtime	Eff(%)
B11	Real	8,973	744	58	679	7.20	88.94
	Sim	8,973	748	58	683	6.25	90.26
B14	Real	58,042	744	304	360	80.01	79.16
	Sim	58,042	748	304	368	74.87	80.25
C10	Real	113,712	744	369	292	82.41	81.75
	Sim	113,712	748	369	291	86.61	81.00

Table 8.1 shows acceptable accuracy of the model compared with reality. However, as shown in Figure 8.7, there is significant starvation to B14 and C10, which should not be seen in practice. Possible reasons include the errors in the equipment parameters, scheduling pattern and volumes, and definition of starvations. Therefore, we modified the schedule a little as follows:

Separate the original task 5 into two new tasks: tasks 5 and 6. In original task 5, C10 pulls 14088 bbls from tank 1 which is fed by Pall2. In the modified tasks 5 and 6, C10 pulls 5000 and 9088 bbls from tank 1, which are still fed by Pall 2, respectively. The purpose of the modification is to avoid the 18-hour maintenance of Pall2 due to 30,000 bbls production happening during the task. Similar to the reasoning above, we also separate the original task 11 and task 19.

The simulation result with the modified schedul is shown in Figure 8.9. There is no significant starvation after modifying the schedule. It is reasonable and practical that a major maintenance should not occur during production, which will most likely lead to significant starvation and therefore production loss.

As seen from the above results, the ARENA simulation model produces similar

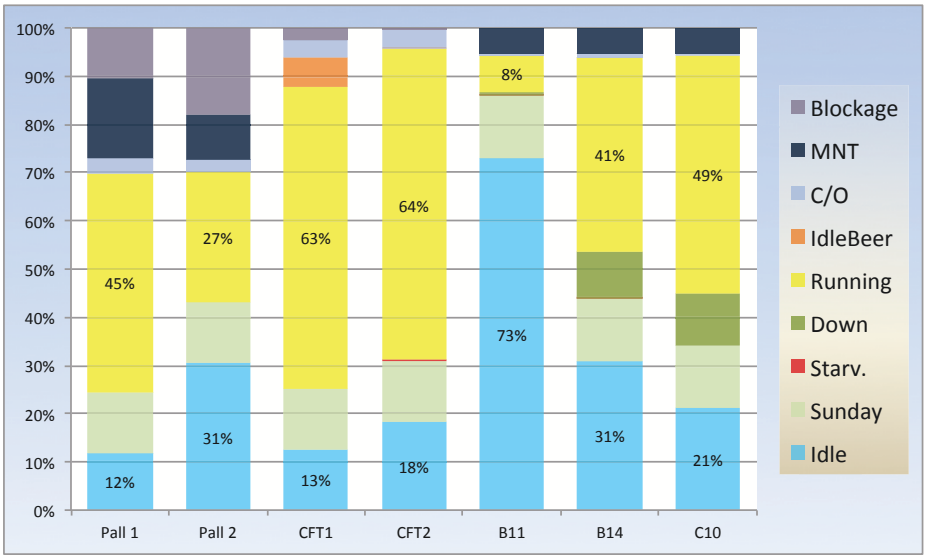


Figure 8.9: Components status chart separating tasks

results to those collected in reality and is valid.

8.4 Production Capacity Analysis

An experiment is carried out to modify the schedule so that all tasks are pushed to start as early as possible. Specifically, we combine the tasks appropriately and change the beginning time of all tasks to 0:00am on August 1st, 2011: If there are two or more consecutive tasks with the same assigned packaging line and the same brand, then we combine them as one single task. The purpose of the modification is to determine the earliest time for the system to finish all the schedule production of the month.

Figure 8.10 and Table 8.2 show the simulation results based on the modified schedules.

According to the simulation result, the total starvation time of B14 is 0.46 hours and that of C10 is 8.46 hours, almost all of which are due to in-task maintenance of the filters. As a result, the total running time under this schedule is 672 hours, or 28 days. In other words, the system needs about 28 days to finish all the tasks

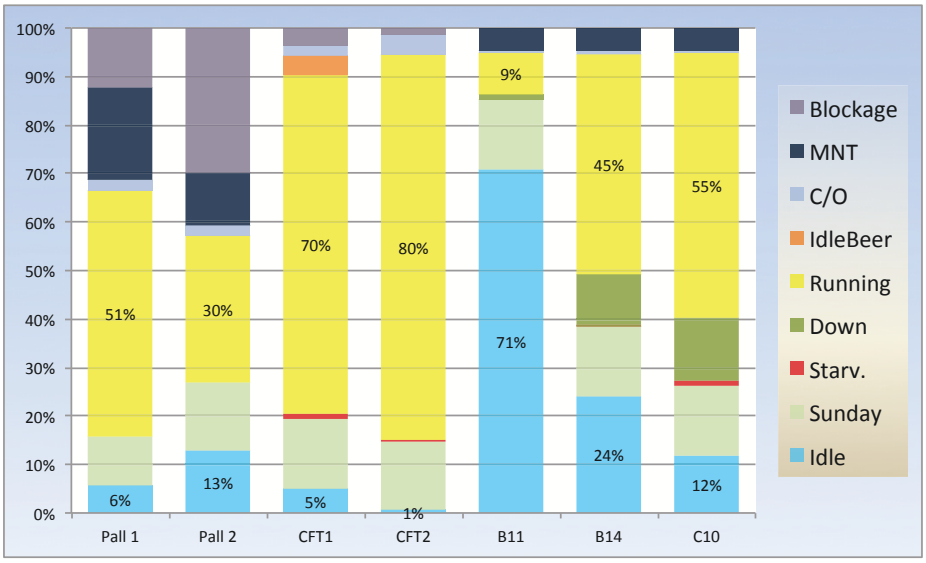


Figure 8.10: Components status chart in capacity analysis

Table 8.2: Packaging lines comparison in capacity analysis (in hours)

Line		Qty(bbls)	Obs	Prod	Non_Prod	Downtime	Eff(%)
B11	Real	8,973	744	58	679	7.20	88.94
	Sim	8,973	672	58	606	7.78	88.15
B14	Real	58,042	744	304	360	80.01	79.16
	Sim	58,042	672	304	295	73.08	80.61
C10	Real	113,712	744	369	292	82.41	81.75
	Sim	113,712	672	369	271	85.63	81.18

compared with 31 days in reality and therefore there is extra production capacity available. We can also see that the in-task maintenance of the filters will result to significant starvation time of packaging line.

8.5 Summary

The summaries for the ARENA simulation are:

- The model is valid and can mimic the production line with real time information;
- The simulation results show that there is at least 12% more productivity capacity in this system under appropriate production schedule. Therefore energy consumption will be reduced due to less production running time;
- Increasing the speeds of or purchase additional filters and packaging lines will increase productivity further;
- Purchase additional storage tanks will not increase productivity significantly.

CHAPTER IX

CONCLUSIONS AND FUTURE WORK

9.1 Conclusions

Our research construct mathematical modeling of serial production systems with multiple unreliable machines and finite capacity buffers in between, where machines have Bernoulli or geometric reliability. Then exact formulas and/or approximation approaches are developed to analyze system performance measures during transient period. Based on these tools, real-time operation control and optimization are investigated and practical examples are provided. Results show that the research findings lead to significant energy savings.

In Chapter II, Bernoulli and geometric serial lines are modeled and performance measures are provided.

Chapter III studies energy reduction problem in Bernoulli serial lines with stripping operations through optimal control of machine startup schedule. Specifically, using transient analysis, analytical mathematical model is developed, which describes the dynamics of the system. In addition, closed-form expressions are derived to evaluate the productivity and energy performances. Based on these expressions, the effects of startup schedule on system performances are discussed and procedures for developing optimal machine startup schedules are formulated. Numerical results show that the optimal schedule can lead to significant improvements in energy utilization efficiency.

Chapter IV studies productivity and energy performance in Bernoulli serial lines with machine startup and shutdown schedule. Since the system operates in transient regimes, steady state analysis is not applicable. Using Markovian analysis, closed-form expressions are provided to calculate the performance measures of Bernoulli serial lines with time-dependent machine efficiencies, and a recursive procedure based on aggregation is developed. Based on this technique, the effects of machine startup and shutdown schedule on system performances are discussed and a greedy algorithm-based procedure for obtaining sub-optimal operations schedules is formulated. Numerical results show that the operations schedule can lead to significant improvements in energy efficiency.

Chapter V investigates the structural properties of order completion time in Bernoulli serial lines and take advantage of recursive aggregation technique to estimate order completion time with high level of computation efficiency and accuracy. Experiment shows that the order completion time is approximately linear to batch size and is reversible.

In Chapter VI, two methods are provided and invested to transform geometric lines to Bernoulli lines to approximate PR in geometric lines. Results show that they only obtain acceptable level of accuracy in certain production lines. Further modifications of these two methods and explorations of other technique including closed-form formulas will be future work.

Chapter VII develops mathematical models for transient analysis and derives closed-form expressions for evaluating the production rate, consumption rate, work-in-process, and probabilities of machine starvation and blockage, during transients, in the framework of serial production lines with geometric machines and finite buffers. In addition, a computationally efficient algorithm based on recursive aggregation is developed to approximate the transient performance measures with high accuracy. Numerical experiments show that the methods developed can be applied to systems

with time-varying machine parameters as well.

In Chapter VIII, a simulation model is constructed using ARENA to mimic production with three processes and perform productivity capacity analysis. Simulation results show that the system has at least 12% extra productivity capacity.

In summary, this research develops formulas and algorithms to analyze serial production lines with Bernoulli or geometric reliabilities during transient period. Based on these mathematical tools, we perform real time analysis and investigate startup and shutdown schedules to reduce energy consumption.

9.2 Future Work

Future work can be summarized in the following directions.

9.2.1 Performance Measures

We pay most attention to PR evaluations since this is the major concerns in manufacturing. We also investigate the evaluations of CR, WIP, ST and BL. Based on these system performance metrics, we will be able to evaluation derivative performance measures, such as order completion time, total energy consumption, energy consumption per parts, etc. However, some other performance measures can also contribute to system performance analysis, such as bottleneck identification during transient, the variance of the performance measures we investigates and parts' residence time, which deserve our future endeavor.

Specially, chapter VI introduces two methods to transform geometric serial lines to Bernoulli serial lines, which does not obtain acceptable accuracy in most cases. We believe that some modifications of these methods will result to higher accuracy of estimating system performance. Besides, estimating other performance measures is also our future work direction.

On the other hand, in our research, most of the time we assume that buffers are

empty and, for geometric lines, machines are all up or down at the beginning. These are not necessarily the case in practice. Further investigations of systems with parts in buffers and/or machines are not all up or down at the beginning are needed.

9.2.2 System Theoretical Properties

We have focused on performance measures evaluation; however, system theoretical properties have not received much attention. For example, monotonicity, reversibility and effects of up- and downtime of geometric serial lines during transient are all important properties to help analyze production lines.

9.2.3 Schedule and Control

We only investigate control policies and mechanism for Bernoulli lines with threshold and steepest ascent algorithms. However, similar analysis should also be performed on geometric lines. Besides, other effective and efficient feedback controllers and robust optimization algorithms for machine startup and shutdown should also be investigated. Those algorithms should not only provide acceptable accuracy but also give results in an acceptable time range for the sake of "real time analysis".

9.2.4 Other Reliabilities or Complex Structures

In terms of reliabilities, we only investigate the most common (and also simple) reliabilities: Bernoulli and geometric. However, production lines with other practical reliabilities should also be investigated, such as Rayleigh, Weibull, Gamma and Log-normal reliabilities.

In terms of structures, we only investigate serial production lines. However, for example, parallel production lines, assembly lines, closed lines and lines with rework mechanism are also common in industry and we should evaluate performance of those production lines during transient.

9.2.5 Implementation

Our major effort is to develop formulas and methods to (exactly or approximately) analyze production systems. However, significant effort should also be placed on industrial implementations, which will validate our models and methodologies. Note that our research can not only apply to industrial production systems, but also apply to many other areas, such as service, logistics, etc., since the machines in our research can also represent doctors, trucks, etc., while buffers can also represent seats, warehouses, etc.

9.2.6 Other Directions

Other directions include: extending the results to machines with warm-up times, systems with other energy consumption models (such as machines having multiple power states) and continuous improvement of systems having geometric reliability.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] H. T. Papadopoulos, C. Heavy, and J. Browne, *Queueing Theory in Manufacturing Systems Analysis and Design*. Chapman & Hill, London, UK, 1993.

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

[3] S. B. Gershwin, *Manufacturing Systems Engineering*. Prentice Hall, Englewood Cliff, NJ, 1994.

[4] J. Li and S. M. Meerkov, *Production Systems Engineering*. Springer, 2009.

[5] C. Galitsky and E. Worrell, "Energy efficiency improvement and cost saving opportunities for the vehicle assembly industry," *Ernest Orlando Lawrence Berkeley National Laboratory (LBNL-50939-Revision)*, 2008.

[6] "U.S. Department of Commerce: Annual survey of manufacturers," http://www.census.gov/manufacturing/asm/index.html.

[7] K. Kissock, K. Hallinan, and W. Bader, "Energy and waste reduction opportunities in industrial processes," *Strategic Planning for Energy and the Environment*, vol. 21, no. 1, pp. 40–53, 2001.

[8] G. Boyd, "Development of a performance-based industrial energy efficiency indicator for automobile assembly plants," *Argonne National Laboratory (ANL/DIS-05-3)*, 2005.

[9] G. Boyd, E. Dutrow, and W. Tunnessen, "The evolution of the energy star energy performance indicator for benchmarking industrial plant manufacturing energy use," *Journal of Cleaner Production*, vol. 16, no. 6, pp. 709–715, 2008.

[10] T. Kolta, "Selecting equipment to control air pollution from automotive painting operations," *Society of Automotive Engineers (SAE) International Congress and Exposition*, 1992.

[11] G. J. Roelant, A. J. Kemppainen, and D. R. Shonnard, "Assessment of the automobile assembly paint process for energy, environmental, and economic improvement," *Journal of Industrial Ecology*, vol. 8, pp. 173–191, 2004.

[12] S. Papasavva, S. Kia, J. Claya, and R. Gunther, "Characterization of automotive paints: an environmental impact analysis," *Progress in Organic Coatings*, vol. 43, no. 1-3, pp. 193–206, 2001.

[13] S. M. Meerkov and L. Zhang, "Transient behavior of serial production lines with bernoulli machines," *IIE Transactions*, vol. 40, no. 3, pp. 297–312, 2008.

[14] W. K. Grassmann, "Transient solutions in Markovian queueing systems," *Computers & Operations Research*, vol. 4, no. 1, pp. 47–53, 1977.

[15] M. F. Neuts, *Structured stochastic matrices of $M/G/1$ type and their applications*. CRC Press, 1989, vol. 5.

[16] D. J. Bertsimas and D. Nakazato, "Transient and busy period analysis of the $GI/G/1$ queue: The method of stages," *Queueing Systems*, vol. 10, no. 3, pp. 153–184, 1992.

[17] W. Böhm and S. G. Mohanty, "The transient solution of $M/M/1$ queues under (M, N)-policy: A combinatorial approach," *Journal of Statistical Planning and Inference*, vol. 34, no. 1, pp. 23–33, 1993.

[18] I.-J. Lee and E. Roth, "A heuristic for the transient expected queue length of markovian queueing systems," *Operations Research Letters*, vol. 14, no. 1, pp. 25–27, 1993.

[19] Y. Narahari and N. Viswanadham, "Transient analysis of manufacturing systems performance," *IEEE Transactions on Robotics and Automation*, vol. 10, no. 2, pp. 230–244, 1994.

[20] J.-M. Garcia, O. Brun, and D. Gauchard, "Transient analytical solution of $M/D/1/N$ queues," *Journal of Applied Probability*, pp. 853–864, 2002.

[21] M. A. Gökçe, M. C. Dinçer, and M. A. Örnek, "Analysis of transient throughput rates of transfer lines with pull systems," in *Just-in-Time Systems*. Springer, 2012, pp. 287–304.

[22] R. Stolletz and S. Lagershausen, "Time-dependent performance evaluation for loss-waiting queues with arbitrary distributions," *International Journal of Production Research*, vol. 51, no. 5, pp. 1366–1378, 2013.

[23] P. G. Harrison, "Transient behaviour of queueing networks," *Journal of Applied Probability*, vol. 18, pp. 482–490, 1981.

[24] A. R. Odoni and E. Roth, "An empirical investigation of the transient behavior of stationary queueing systems," *Operations Research*, vol. 31, no. 3, pp. 432–455, 1983.

[25] W. D. Kelton, "Transient exponential-Erlang queues and steady-state simulation," *Commun. ACM*, vol. 28, pp. 741–749, 1985.

[26] A. Joseph and W. Ward, "Transient behavior of the M/M/1 queue via Laplace transforms," *Advances in Applied Probability*, vol. 20, pp. 145–178, 1988.

[27] J. Abate and W. Whitt, "Transient behavior of the M/G/1 workload process," *Operations Research*, vol. 42, no. 4, pp. 750–764, 1994.

[28] Y. Seki and N. Hoshino, "Transient behavior of a single-stage kanban system based on the queueing model," *International Journal of Production Economics*, vol. 60-61, pp. 369–374, 1999.

[29] G. Riaño, "Transient behavior of stochastic networks: Application to production planning with load-dependent lead times," Ph.D. dissertation, Georgia Institute of Technology, 2002.

[30] J. Griffiths, G. Leonenko, and J. Williams, "The transient solution to $M/E_k/1$ queue," *Operations Research Letters*, vol. 34, no. 3, pp. 349–354, 2006.

[31] A. Ingolfsson, E. Akhmetshina, S. Budge, Y. Li, and X. Wu, "A survey and experimental comparison of service-level-approximation methods for nonstationary M(t)/M/S(t) queueing systems with exhaustive discipline," *INFORMS Journal on Computing*, vol. 19, pp. 201–214, 2007.

[32] G.-A. Klutke and L. M. Seiford, "Transient behaviour of finite capacity tandem queues with blocking," *International Journal of Systems Science*, vol. 22, no. 11, pp. 2205–2215, 1991.

[33] M. N. Gopalan and U. D. Kumar, "On the transient behaviour of a merge production system with an end buffer," *International Journal of Production Economics*, vol. 34, no. 2, pp. 157–165, 1994.

[34] S. Mocanu, "Numerical algorithms for transient analysis of fluid queues," in *Proceedings of 5th International Conference on the Analysis of Manufacturing Systems*, Zakymthos, Greece, 2005, pp. 15–20.

[35] L. Lin and J. K. Cochran, "Metamodels of production line transient behaviour for sudden machine breakdowns," *International Journal of Production Research*, vol. 28, no. 10, pp. 1791–1806, 1990.

[36] E. J. Stahlman and J. K. Cochran, "Dynamic metamodelling in capacity planning," *International journal of production research*, vol. 36, no. 1, pp. 197–210, 1998.

[37] H. Missbauer, "Models of the transient behaviour of production units to optimize the aggregate material flow," *International Journal of Production Economics*, vol. 118, no. 2, pp. 387–397, 2009.

[38] J. Liu and F. Yang, "Transient analysis of general queueing systems via simulation-based transfer function modeling," submitted to *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 2011. [Online]. Available: http://www2.cemr.wvu.edu/∼yang/TOMACS%202011.pdf

[39] S. Shaaban and S. Hudson, "Transient behaviour of unbalanced lines," *Flexible Services and Manufacturing Journal*, vol. 24, no. 4, pp. 575–602, 2012.

[40] F. K. Pil and S. Rothenberg, "Environmental performance as a driver of superior quality," *Production and Operations Management*, vol. 12, no. 2, pp. 404–415, 2003.

[41] C. J. Corbett and R. D. Klassen, "Extending the Horizons: Environmental Excellence as Key to Improving Operations," *Manufacturing Service Operations Management*, vol. 8, no. 1, pp. 5–22, 2006.

[42] J.-T. Lim, S. M. Meerkov, and F. Top, "Homogeneous, asymtotically reliable serial production lines: theory and a case study," *IEEE Transactions on Automatic Control*, vol. 35, no. 5, pp. 524–534, 1990.

[43] J. Li, "Throughput analysis in automotive paint shops: a case study," *IEEE Transactions on Automation Science and Engineering*, vol. 1, no. 1, pp. 90–98, 2004.

[44] J. Li, D. E. Blumenfeld, and S. P. Marin, "Manufacturing system design to improve quality buy rate: An automotive paint shop application study," *IEEE Transactions on Automation Science and Engineering*, vol. 4, no. 1, pp. 75–79, 2007.

[45] J. Arinez, S. Biller, S. M. Meerkov, and L. Zhang, "Qualityquantity improvement in an automotive paint shop: A case study," *IEEE Transactions on Automation Science and Engineering*, vol. 7, no. 4, pp. 755–761, 2010.

[46] E. Sadeghipour, E. Westervelt, and S. Bhattacharya, "Painting green: Design and analysis of an environmentally and energetically conscious paint booth hvac control system," in *American Control Conference, 2008*, june 2008, pp. 3325 –3330.

[47] E. Müller and T. Löffler, "Improving energy efficiency in manufacturing plants - case studies and guidelines," Chemnitz University of Technology, Saxony, Germany, Tech. Rep., 2009.

[48] J. Wang, J. Li, and N. Huang, "Optimal scheduling to achieve energy reduction in automotive paint shops," in *Proceedings of ASME Manufacturing Science and Engineering Conference*, West Lafayette, IN, 2009, pp. 161–167.

[49] C. P. A. Guerrero, J. Wang, J. Li, J. Arinez, S. Biller, N. Huang, and G. Xiao, "Production system design to reduce energy consumption: A case study in automotive paint shop," in *Proceedings of International Symposium on Flexible Automation*, Tokyo, Japan, 2010.

[50] C. Wang and L. Zhang, "Transient analysis of bernoulli serial lines: Performance evaluation and system-theoretic properties," submitted to *18th IFAC World Congress*, 2010.

[51] S. M. Meerkov, N. Shimkin, and L. Zhang, "Transient behavior of two-machine geometric production lines," *IEEE Transactions on Automatic Control*, vol. 55, no. 2, pp. 453–458, 2010.

[52] S. M. Meerkov and L. Zhang, "Unbalanced production systems with floats: Analysis and lean design," to appear in *International Journal of Manufacturing Technology and Management*, 2010.

[53] J. Wang, Y. Hu, and J. Li, "Transient analysis to design buffer capacity in dairy filling and packing production lines," *Journal of Food Engineering*, vol. 98, pp. 1–12, 2010.

[54] P. Brucker, Y. N. Sotskov, and F. Werner, "Complexity of shop-scheduling problems with fixed number of jobs: A survey," *Mathematical Methods of Operations Research*, vol. 65, no. 3, pp. 461–481, 2007.

[55] G. Chen, L. Zhang, J. Arinez, and S. Biller, "Energy-efficient production systems through schedule-based operations," *IEEE Transactions on Automation Science and Engineering*, vol. 10, no. 1, pp. 27–37, 2013.

[56] W. Wiede Jr and G. V. Reklaitis, "Determination of completion times for serial multi product processes-3 mixed intermediate storage systems," *Computers and Chemical Engineering*, vol. 11, no. 4, pp. 357–368, 1987.

[57] K. Fang, N. Uhan, F. Zhao, and J. W. Sutherland, "A new approach to scheduling in manufacturing for power consumption and carbon footprint reduction," *Journal of Manufacturing Systems*, vol. 30, pp. 234–240, 2011.

[58] A. H. Kashan and B. Karimi, "An improved mixed integer linear formulation and lower bounds for minimizing makespan on a flow shop with batch processing machines," *International Journal of Advanced Manufacturing Technology*, vol. 40, pp. 582–594, 2009.

[59] P. Baptiste and C. Jacquemard, "Exact distribution of the makespan in a two machines flow shop with two kinds of jobs," *International Journal of Production Economics*, vol. 74, pp. 77–83, 2001.

[60] A. Azaron and S. F. Ghomi, "Lower bound for the mean project completion time in dynamic PERT networks," *European Journal of Operational Research*, vol. 186, pp. 120–127, 2008.

[61] T. Perera and K. Liyanage, "Methodology for rapid identification and collection of input data in the simulation of manufacturing systems," *Simulation Practice and Theory*, vol. 7, no. 7, pp. 645–656, 2000.

[62] B. Oosterman, M. Land, and G. Gaalman, "The influence of shop characteristics on workload control," *International Journal of Production Economics*, vol. 68, no. 1, pp. 107–119, 2000.

[63] G. J. Roelant, A. J. Kemppainen, and D. R. Shonnard, "A system-theoretic property of serial production lines - improvability," *International Journal of Systems Science*, vol. 26, no. 5, pp. 755–785, 1995.

[64] T. Altiok, *Performance Analysis of Manufacturing Systems*. Springer-Verlag, New York, NY, 1997.

[65] L. Zhang, C. Wang, J. Arinez, and S. Biller, "Transient analysis of bernoulli serial lines: Performance evaluation and system-theoretic properties," *IIE Transactions*, 2012.

[66] J. Geldermann, M. Treitz, and O. Rentz, "Integrated technique assessment based on the pinch analysis approach for the design of production networks," *European Journal of Operational Research*, vol. 171, no. 3, pp. 1020–1032, 2006.

[67] G. Chen, L. Zhang, J. Arinez, and S. Biller, "Real-time performance analysis of production lines: A system-theoretic approach," in *IEEE International Conference on Automation Science and Engineering*, Seoul, Korea, 2012.

[68] L. Zhang, J. Arinez, and S. Biller, "Energy consumption reduction in serial production lines via optimal startup schedule," submitted to *ASME Manufacturing Science & Engineering Conferences*, 2011.

[69] G. Chen, L. Zhang, J. Arinez, and S. Biller, "Energy-efficient production systems through schedule-based operations," *IEEE Transactions on Automation Science and Engineering*, vol. 10, no. 1, pp. 27–37, 2013.

[70] L. Zhang, C. Wang, J. Arinez, and S. Biller, "Transient analysis of Bernoulli serial lines: Performance evaluation and system-theoretic properties," *IIE Transactions*, vol. 45, no. 5, pp. 528–543, 2013.

CURRICULUM VITAE

Guorong Chen

Personal Data

PLACE OF BIRTH: China
EMAIL: gchen@uwm.edu

Education

MAY 2014	PhD, **University of Wisconsin-Milwaukee** Major: Industrial Engineering Minor: Finance
MAY 2009	M.S., **University of Nebraska-Lincoln** Major: Actuarial Science Minor: Finance
JUNE 2006	B.S., **China University of Geosciences (Beijing)** Major: Mathematics and Applied Mathematics

Work Experience

JAN 2011-MAY 2014	Research Assistant at UNIVERSITY OF WISCONSIN-MILWAUKEE, Wisconsin. Advisor: Prof. Liang Zhang
OCT 2013-DEC 2013	Intern at EATON CORPORATION, Wisconsin. Supervisor: Dr. Yang Yi
MAY 2012-FEB 2013	Energy Engineer at INDUSTRIAL ASSESSMENT CENTER IN UW-MILWAUKEE, Wisconsin. Director: Dr. Yingchun Yuan
MAY 2011-DEC 2011	Project Assistant at MILLERCOORS LLC AND UW-MILWAUKEE, Wisconsin. Advisor: Prof. Liang Zhang
APR 2010-JUL 2010	Financial Engineering Intern at CHANGSHENG FUND MANAGEMENT, Beijing, China. Supervisor: Dr. Zhongguang Bai

Honors and Awards

2011-2014	Reviewer of annual IEEE Conference on Automation Science and Engineering
2012-2014	Recipient of Chancellor's Graduate Student Awards

Publications and Presentations

[1] G. Chen, C. Wang, L. Zhang, J. Arinez and G. Xiao, "Transient analysis of serial production lines with geometric machines and finite buffers", submitted to *IEEE Transactions on Automatic Control*, 2014.

[2] G. Chen, L. Zhang, J. Arinez and S. Biller, "Energy-efficient production systems through schedule-based operations", *IEEE Transactions on Automation Science and Engineering*, vol. 10, no. 1, pp. 27-37, 2013.

[3] G. Chen, L. Zhang, J. Arinez and G. Xiao, "Analysis of production completion time in Bernoulli serial lines with finite buffers", presented in *IEEE International Conference on Automation Science and Engineering*, Madison, WI, 2013.

[4] G. Chen, L. Zhang and J. Arinez, "Analysis of production completion time in Bernoulli serial lines with finite buffers", *Industrial and Systems Engineering Research Conference*, San Juan, Puerto Rico, 2013.

[5] G. Chen, C. Wang, L. Zhang, J. Arinez and S. Biller, "Transient performance analysis of production lines: Theory and applications", *International Conference on Advanced Vehicle Technologies and Integration*, pp. 707-713, Changchun, China, 2012 (invited paper).

[6] G. Chen, L. Zhang, J. Arinez and S. Biller, "Real-time performance analysis of production lines: A system-theoretic approach", *IEEE International Conference on Automation Science and Engineering*, pp. 656-661, Seoul, Korea, 2012.

[7] G. Chen, L. Zhang, J. Arinez and S. Biller, "Feedback control of machine startup for energy-efficient manufacturing in Bernoulli production lines", presented in *Annual ASME International Manufacturing Science and Engineering Conference*, Notre Dame, IN, 2012.

[8] G. Chen, C. Wang, L. Zhang, J. Arinez and S. Biller, "Transient analysis of serial production lines with geometric machines", *IIE Annual Conference*, Orlando, FL, 2012.

[9] G. Chen, L. Zhang, J. Arinez and S. Biller, "Energy-efficient production systems through schedule-based operations", presented in *INFORMS Annual Conference*, Phoenix, AR, 2012.

[10] G. Chen, C. Wang, L. Zhang, J. Arinez and S. Biller, "Transient analysis of serial production lines with geometric machines", *IIE Annual Conference*, Orlando, FL, 2012.

[11] G. Chen, L. Zhang, J. Arinez and S. Biller, "Feedback control of machine startup for energy efficient manufacturing in Bernoulli serial lines", *IEEE Conference on Automation Science and Engineering*, pp. 666-671, Trieste, Italy, 2011.

[12] G. Chen, L. Zhang, J. Arinez and S. Biller, "Energy consumption reduction in serial production lines via optimal startup schedule", presented in *6th Annual ASME International Manufacturing Science and Engineering Conference*, pp. 155-161, Corvallis, OR, 2011.

[13] D. Ma and G. Chen, "The relationship between CG (Corporate Governance) structure and CG mechanism: a system approach", *Contemporary Economic Management*, vol. 27, no. 3, pp. 66-69, 2005 (in Chinese).