ANALYTICAL METHODS FOR PERFORMANCE ENHANCEMENT IN UNRELIABLE MULTISTAGE MANUFACTURING SYSTEMS WITH IMPERFECT PRODUCTION

CHEN RUIFENG

(B.Eng.)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2010

Acknowledgements

A simple thank you seems inadequate to represent my deep appreciation and gratitude for all the people who have shared part of my life during my PhD experience. I have learned as much from each of you as I have from all the courses I have taken and the books I have read.

First and foremost, I am profoundly grateful for the support, encouragement, and guidance of my advisor, Professor V. Subramaniam. From concept to production of this work, Professor V. Subramaniam has guided, challenged, and encouraged me in innumerable ways. He gave me the freedom to explore the research problems using various methods. I have learned immensely from the weekly (and sometimes even daily) discussions with Professor V. Subramaniam and this is helpful not only in the research but also for my future career.

I would also like to thank the National University of Singapore for offering me the research scholarship, research facilities, and valuable courses. Without this support, my graduate study will not have been as fruitful as it has been in the past five years.

I want to thank European Aeronautic Defence and Space Company (EADS) Singapore, for providing me with a valuable opportunity to work with their research group. This experience has broadened my horizon, and enriched my knowledge, especially on data processing. Special thanks go to Dr. David Woon, who offered helpful suggestions and kindly support in my research.

Expressed thanks are also due to my colleagues in the research group, Yang Rongling, Lin Yuheng, Cao Yongxin, Chanaka Dilhan Senanayake, and S.P. Singh. My gratitude is also extended to the friends in National University of Singapore, Huang Weiwei, Zhu Kunpeng, Weng Yulin, Zhou Longjiang, Feng Xiaobing, Chao Shuzhe, Yin Jun, Han Dongling, Wei Wei, Wan Jie, Zhao Guoyong, Kommisetti V. R. S. Manyam, and many others, for their enlightening discussion and suggestions.

I owe my deepest thanks to my family for the unconditional and selfless support. My parents have provided me with a lifetime of example for everything that I have done. My attempt to be as good and wise as them is all I can do in return. To my parents in laws, their encouragement in so many ways has made this journey easier. Last but not least, I would like to thank my wife, Li Lin, who has been patient, understanding, helpful, and insightful. She has reminded me to follow my heart and chase dreams, and inspired me to achieve more.

Table of Contents

Summary

To confront the fierce international and domestic competition, manufacturing companies are endeavoring to increase production rate, improve manufacturing quality, reduce inventory, cut down operational costs, and hence maintain competitive standing in the market. Performance enhancement is challenging in a multistage manufacturing system, because of the complex configuration and various uncertainties in the system. This thesis details a modeling framework for performance analysis of multistage manufacturing systems. This modeling framework characterizes the uncertain properties of manufacturing systems that undermine system performance, in particular: 1) machines are unreliable and may experience deterioration; 2) production is imperfect and defective parts are generated randomly.

The modeling framework can be used to estimate a variety of quantitative and qualitative performance measures. These estimates may enable one to assess and improve the management of a multistage manufacturing system. A managerial issue investigated in this research is preventive maintenance, which is widely implemented in manufacturing systems for improving machine reliability. Although analytical models of single or two-machine systems with preventive maintenance have been proposed in the literature, similar study on multistage systems remains limited. Based on the modeling framework, the author presents an algorithm to determine the frequency of preventive maintenance on each machine of a multistage manufacturing system. Performing preventive maintenance at the frequency prescribed by

the algorithm may avoid excessive or insufficient maintenance, resulting in improved production rate.

In addition to machine unreliability, imperfect production may also substantially increase the cost of a manufacturing system. In order to mitigate the corrupting effects of defective parts generated due to imperfect production, the quality inspection of the multistage manufacturing system is also investigated in this thesis. An algorithm is formulated for determining the placement of inspection machines in such a system. With the inspection allocation scheme indicated by this algorithm, the quality of material flow in the multistage manufacturing system is improved. This may reduce the waste on processing defective parts and penalty resulting from defective parts shipped to customers.

Based on the modeling framework, the author further explores the extension for multistage manufacturing systems with batch operations and generally distributed processing times. This extension makes it possible to model a wider range of real manufacturing systems.

Keywords: Multistage Manufacturing Systems; Quantity and Quality Performance; Preventive Maintenance; Inspection Allocation; Batch Operations; Decomposition

List of Tables

List of Figures

Chapter 1.

Introduction

1.1. Research Background

Uncertainty associated with production activities has long been considered to be the "enemy of manufacturing management" (Gershwin, 2009). A manufacturing system may experience various uncertain events (Liberopoulos et al., 2006): machines may deteriorate and break down; defective parts may be generated; inspection errors may occur; machine processing times may vary; demand may fluctuate; raw material supply may be delayed; etc (some commonly observed uncertain characteristics of manufacturing systems are summarized in Figure 1.1). Due to the uncertainty, manufacturing systems rarely perform exactly as expected, and this substantially complicates the decision-making in the control and configuration of such systems.

Manufacturing systems may be roughly divided into two groups: single stage systems and multistage systems. Single stage systems are usually used in the manufacturing of relatively simple products. Multistage systems, on the other hand, integrate a number of manufacturing stages (i.e. machines) to fabricate products with high complexity. The automotive assembly system illustrated in Figure 1.2 is one typical example of the multistage manufacturing system, which consists of hundreds of machines with various functionalities (Sakai and Amasaka, 2007). Compared with single stage systems, the impact of uncertainty in multistage manufacturing systems is much more complex and unpredictable, because machines are influenced by each other. For instance, the failure of a machine may induce material starvation of its downstream machines, and hence interrupt their production. To mitigate the corrupting effects of uncertainty on a system, an analytical model for performance evaluation is beneficial. Such a model may provide useful insights for improving the management of manufacturing. As illustrated in Figure 1.3, an analytical model of the multistage manufacturing system may allow a line manager to evaluate various alternate options to configure a system (for example, one such configuration problem is to determine the size of each buffer in the system (Li and Meerkov, 2009)). Based on the performance measures provided by the model, the manager may identify the best option, and subsequently implement it in the real system. This practice may result in improved system performance.

In this thesis, the author investigates the multistage manufacturing system with unreliable machines (machines may deteriorate and break down) and imperfect production (defective parts are generated). This research provides the analysis for investigating the influence of production reliability and quality on system performance. Based on the proposed models, methods for enhancing the quantitative and qualitative performance of the multistage manufacturing system are also explored. Preventive maintenance (a widely implemented strategy for improving production rate) and quality inspection (a common practice for improving the quality of material flow in manufacturing systems) are two focuses of this thesis. The motivation of this research will be further elaborated in the following subsections.

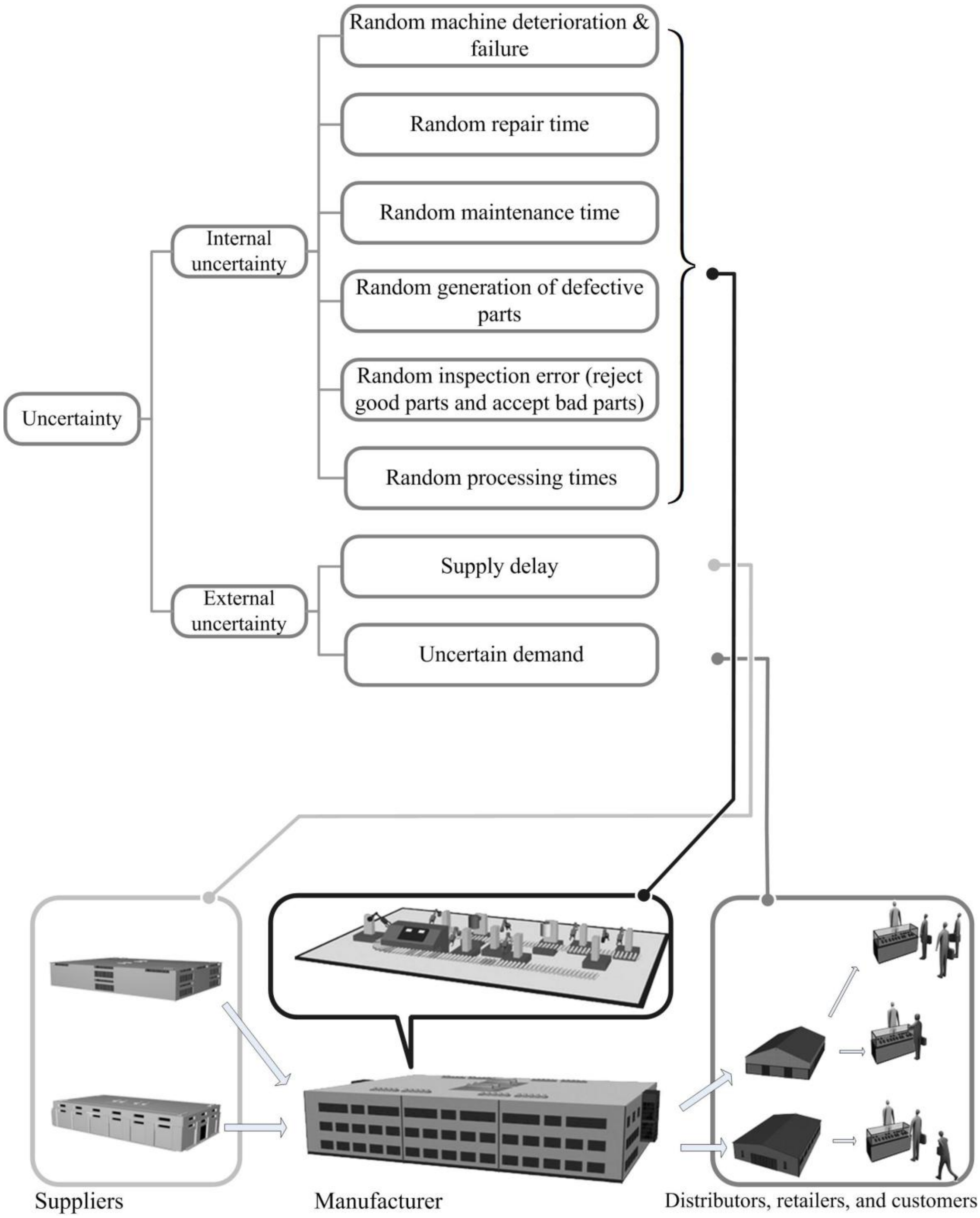


Figure 1.1. Uncertainty in a manufacturing system. Internal uncertainty is associated with the operation inside a manufacturing system (for instance, machines may deteriorate and break down, repair time may fluctuate, defective parts may be generated, inspection error may occur, processing time may be random, etc). The external uncertainty mainly originates from supply delay and demand fluctuation. Both internal and external uncertainty may influence the performance of a system.

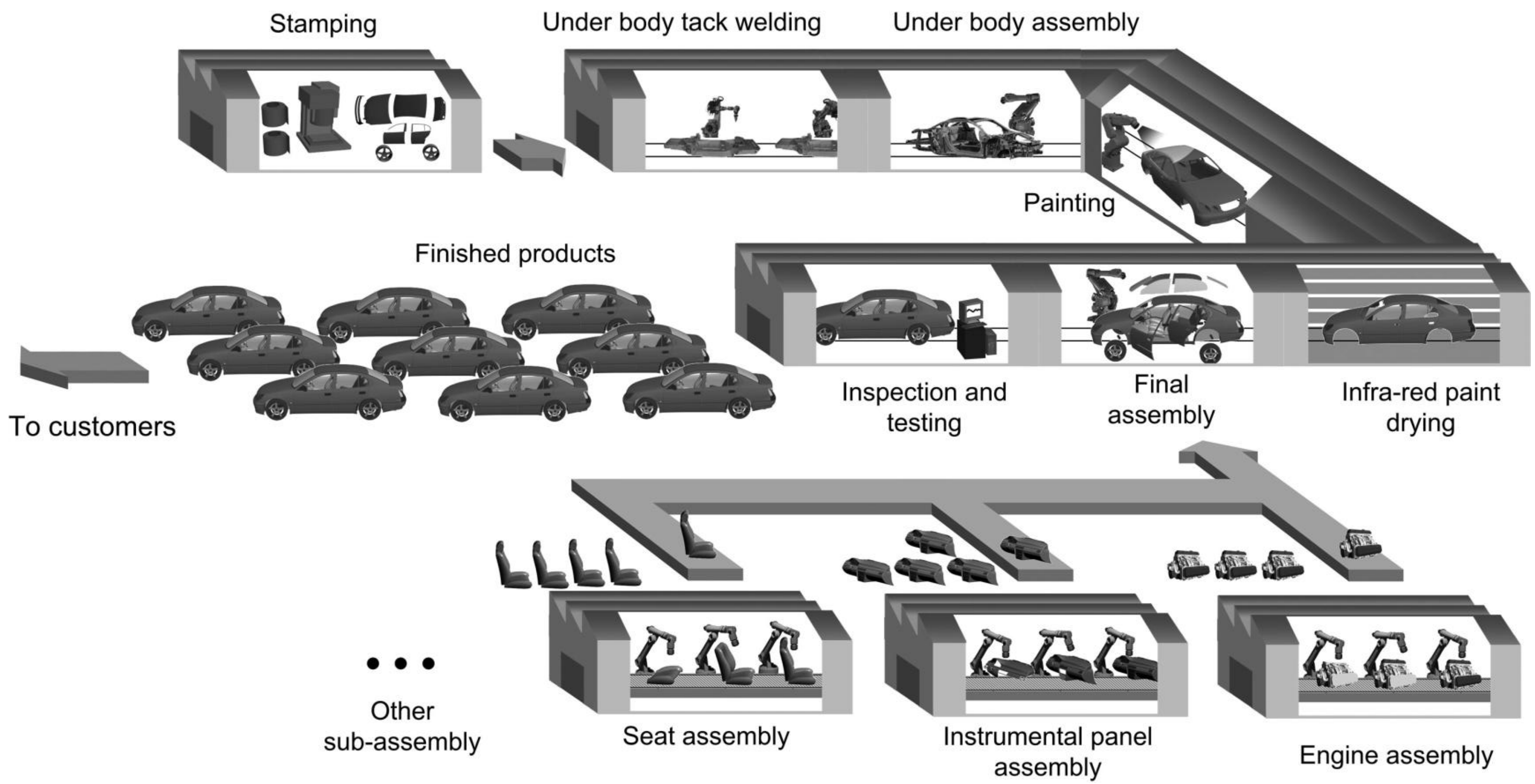


Figure 1.2. A typical example of the multistage manufacturing system: the automotive assembly system. This system consists of a large number of machines dedicated to different processes, such as stamping, under body tack assembly, painting, paint drying, final assembly, inspection, etc. Additionally, such a system may also include the sub-assembly lines for some major components, such as seats, instrumental panels, engines, etc.

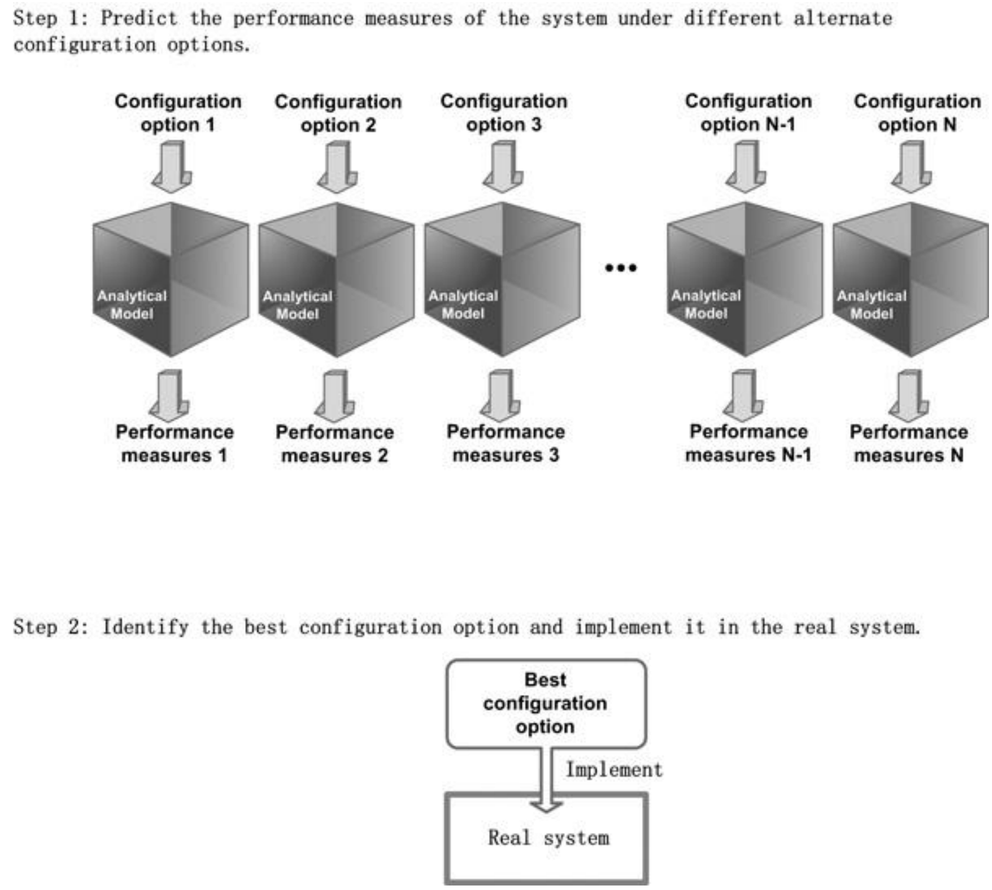
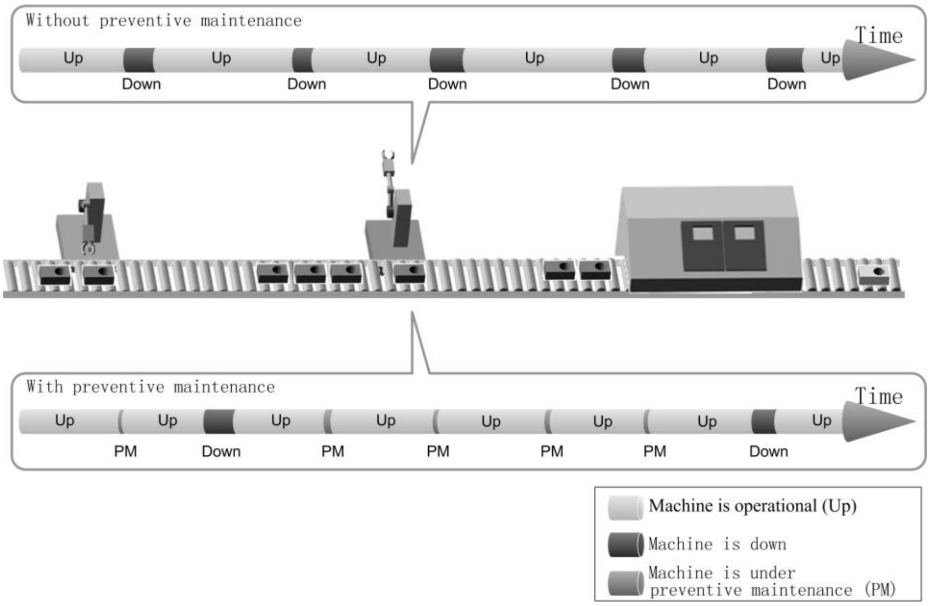


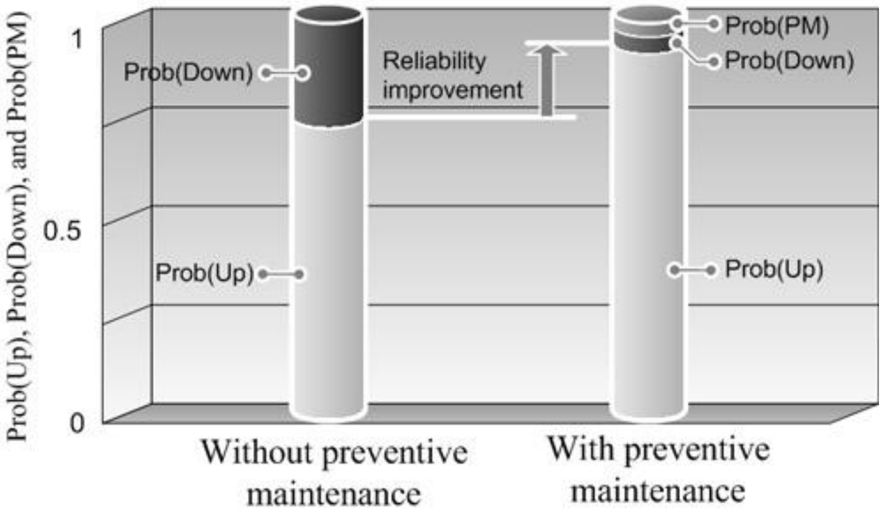
Figure 1.3. A typical application of the model in the management of manufacturing systems. The model is used to predict performance measures of a manufacturing system under different feasible configuration alternatives. Based on the performance measures, the best option can be identified and then implemented in the real system.

1.1.1. Machine Deterioration and Strategy for Improving System Reliability

Production rate of the manufacturing system is viewed as a key performance indicator of competitiveness in the global marketplace (Gerold, 2004). A major impediment to high production rate, as pointed out by many practitioners and scholars, is machine deterioration and failure (Montoro-Cazorla and Perez-Ocon, 2006). Unpredictable failures may delay production and also induce repair costs, resulting in a significant loss of profit. For example, a case study on a paper production company by Alsyouf (2006) indicates that machine failures had reduced profit by approximately 9%. Fortunately, the incidence of machine failures may be reduced by preventive maintenance, a mainstream strategy for improving the reliability of manufacturing systems (Garg and Deshmukh, 2006, Bao and Jaishankar, 2008). For instance, by regularly replacing worn gears of robot arms in car body assembly lines, uptimes of these machines are substantially extended (Sakai and Amasaka, 2007). As depicted in Figure 1.4, preventive maintenance may eliminate accumulated deterioration of a machine before it results in machine failure. However, frequent preventive maintenance may also interrupt the processing of machines and thus undermine production rate (Ambani, et al., 2009). Therefore, to increase production rate, manufacturers need to find a reasonable tradeoff between the interruptions caused by machine failures and preventive maintenance. Striking this tradeoff may require an analytical model that reflects the influence of machine failures and preventive maintenance on the performance of the system. Analytical models that have been proposed in the literature for this purpose predominantly focus on single-machine systems (Kenne and Gharbi, 1999; Bloch-Mercier, 2002; Gurler and Kaya, 2002; Moustafa et al., 2004; Zequeira et al., 2004; Chen and Trivedi, 2005; Chen and Wu, 2007; Wu and Makis, 2008). Recently, several studies (Kyriakidis and Dimitrakos, 2006; Pavitsos and Kyriakidis, 2009; Ambani et al., 2009) have explored other systems consisting of two or three machines.



(a) If preventive maintenance is not performed, deterioration accumulates in a machine and this may induce frequent machine failures. If on the other hand, preventive maintenance is performed, this practice may eliminate the accumulated deterioration. Therefore, the average time between two consecutive machine failures may be substantially extended.



(b) The probability that a machine is up (operational) is improved when preventive maintenance is performed. Repairing a machine from complete failures usually requires much more time than preventive maintenance. Therefore, although preventive maintenance may also interrupt machine processing, it reduces the overall interruption to production, resulting in improved machine reliability.

Figure 1.4. The effect of preventive maintenance. Preventive maintenance may reduce the probability of machine failure and enhance the reliability of machines.

1.1.2. Imperfect Production and Solution for Quality Improvement

In addition to machine deterioration and failure, imperfect production is another uncertain factor that may substantially undermine the system performance. In a multistage manufacturing system, machines with various functionalities are connected as a network. Each of the machines may generate defective parts randomly (Heredia-Langner et al., 2002). For example, in some PCB assembly lines, defects account for up to 10% of production (Shina, 2002). If these defective parts are left undetected, they will progress downstream of the manufacturing process and consume valuable machine capacity. Hence, it is common practice to place inspection machines at different locations in the manufacturing system to detect and remove defective parts, as demonstrated in Figure 1.5. Determining the exact placement of inspection machines in a multistage manufacturing system is a complex problem as it affects not only the quality of parts, but also the quantitative performance of the system, such as production rate and *WIP*. Therefore, solving this problem requires an analytical model that reflects the influence of inspection machines on both quantitative and qualitative performance measures of the system. In the literature, a number of analytical models have been proposed for performance analysis of multistage manufacturing systems, which may be roughly categorized as quantitative and qualitative models. Quantitative models are usually dedicated to estimating production rate and *WIP* by considering random processing times and unreliable machines. In comparison, qualitative models focus on evaluating the quality of parts in manufacturing systems.

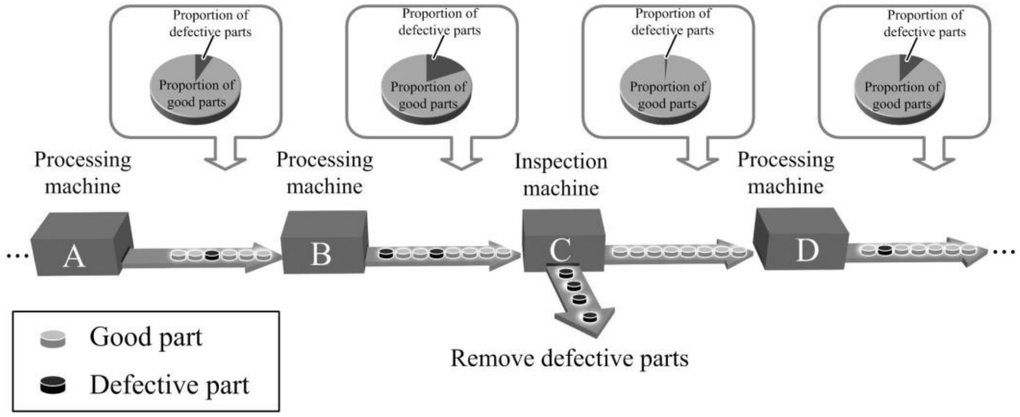


Figure 1.5. The effect of inspection. Each processing machine (such as machine A, B, and D) may produce defective parts randomly. Therefore, after each processing machine, the proportion of defective parts in the material flow may increase. To improve the quality of material flow, the inspection machine (machine C) is placed to remove defective parts. This may prevent wasting the capacity of machine D by eliminating the processing of defective parts generated by machines A and B. Therefore, the cost due to imperfect production may be reduced via inspection.

1.2. Motivation

As discussed in Section 1.1.1, the investigation of machine deterioration and preventive maintenance on multistage manufacturing systems remains limited, especially for non-serial systems with intermediate buffers between machines. In multistage systems, manufacturers usually maintain a relatively small number of parts in each buffer to reduce the inventory holding cost. This makes the systems more vulnerable to machine failures and excessive preventive maintenance (Rezg et al., 2004; Alsyouf, 2009). Therefore, the research on preventive maintenance is of practical value for the management of multistage manufacturing systems. Although analytical models of such systems with unreliable machines have been proposed (Kuo et al., 1997; Gershwin and Burman, 2000; Chiang et al., 2000; Baynat et al., 2001; Li, 2005), these studies generally assume that machine failures are unpreventable and have not accounted for preventive maintenance.

In Chapter 3 of the thesis, the author formulates an approximate model for analyzing machine deterioration and preventive maintenance in the multistage manufacturing systems. This model is based on the decomposition method, which was first proposed in the 1960s (Sevastyanov, 1962) and has been extensively applied to the analysis of multistage manufacturing systems. In this model, a multistage manufacturing system is decomposed into mathematically tractable primitive line segments. This feature facilitates the modeling of multistage manufacturing systems with different numbers of machines and various configurations. The proposed model provides estimates of various commonly used performance measures, such as production rate, work-in-process (*WIP*), availability of each machine (i.e. the fraction of time that a machine is operational), probability of machine failures, probability of a machine being maintained, etc. The numerical experiments of Section 3.5 (which compare the analytical results obtained from the decomposition model with simulation results) demonstrate that these estimates are of satisfactory accuracy. Based on this model, the author also formulates an optimization problem to determine the frequency of preventive maintenance for each machine. An algorithm is provided for solving this problem in Section 3.4.

In addition, as mentioned in Section 1.1.2, quantitative and qualitative models of multistage manufacturing systems have been previously approached as two separate areas. On the one hand, quantitative models were proposed for multistage manufacturing systems with perfect production (i.e. no defects). This condition may not be encountered frequently in many real systems, since imperfect production is widely observed in practice (Mandroli et al., 2006). On the other hand, qualitative models rarely explore the influence of quality control on the quantitative performance of multistage manufacturing systems.

This may make it difficult to evaluate the configuration of inspection machines comprehensively.

In Chapter 4, the author analyzes the inspection allocation problem in multistage manufacturing systems by simultaneously considering both quantitative and qualitative issues. To evaluate the configuration of inspection machines, an integrated quantitative and qualitative model is formulated. As pointed out in the recent literature (Kim and Gershwin, 2005; Lee et al., 2007; Li et al., 2009), an integrated description of quantity and quality is necessary because these two issues are usually tightly coupled in real manufacturing systems. The model may be used to estimate various quantitative and qualitative performance measures, with which the author develops a profit function consisting of the following factors: revenue, inventory holding cost, processing cost, inspection cost, and penalty cost due to shipping defective parts. The placement of inspection machines is then formulated as a maximization problem of the profit function. A heuristic approach is developed for providing a good feasible solution to this problem and this is discussed in Section 4.4.

The modeling framework in this thesis is motivated by the decomposition model proposed by Gershwin (1994, 2000). However, this research is not just a simple variation of Gershwin's study, and it is also not a creative application of decomposition. We consider the multistage manufacturing system subjected to machine deterioration and preventive maintenance. These two factors may substantially influence the performance of a manufacturing system. In order to characterize this influence, the proposed model introduces multiple upstates for a machine to represent different levels of deterioration. Furthermore, an additional state is also included to model preventive maintenance. By contrast, in Gershwin's model, each machine has only two states, viz. up and down. The author formulated a new set of equations to characterize the state transitions due to machine deterioration and preventive maintenance, as presented in the following chapters. In addition, the author also considers various common characteristics that have not been incorporated in Gershwin's model. For instance, the following issues have been included in the model presented in this thesis:

- Defective parts are removed from the manufacturing process. This is commonly practiced to improve the quality of material flow in a manufacturing system and to reduce wastage of machine capacity.
- Machines are operated in batches (i.e. machines are capable of processing several parts simultaneously). The implementation of batch operations improves the utilization of machines and production rate. Therefore, batch machines are employed in many industries, such as electrical appliance manufacture (e.g. chemical coating processes), wafer fabrication (e.g. diffusion and oxidation processes), etc (Chen et al., 2010).
- The processing times of machines are generally distributed. In the decomposition models proposed in the literature previously, processing times are assumed to be either deterministic or exponentially distributed (exponential distribution can be used to characterize the processing times of a machine only when their standard deviation is equal to the mean (Bolch et al., 2006)). This was assumed to make the models mathematically tractable. However, this assumption may be inadequate to

model the non-deterministic nature of many industrial processes, such as random disturbances, operator inconsistencies, etc.

The model presented in this thesis is a substantial expansion of the previous decomposition models that have been proposed in the literature. It can be applied to a wide range of manufacturing systems, which were impossible with the models proposed previously.

1.3. Thesis Outline

The remainder of this thesis is organized as follows: a literature review pertaining to performance evaluation of multistage manufacturing systems is presented in Chapter 2. In Chapter 3, multistage manufacturing systems with machine deterioration and preventive maintenance are investigated. An analytical model is formulated for performance evaluation of such systems and subsequently used to improve machine reliability. In Chapter 4, the author develops an integrated quantitative and qualitative model for multistage manufacturing systems with imperfect production. An algorithm is also provided for determining the placement of inspection machines. In Chapter 5, the author analyzes the extension of the models presented in Chapters 3 and 4 for multistage manufacturing systems with batch operations (i.e. machines can process more than one part each time) and generally distributed processing times. This extension may facilitate the models in the thesis to adapt to more complex conditions. A discussion on future research opportunities is provided in Chapter 6. Finally, this thesis concludes with a summary of the key findings.

Chapter 2.

Performance Evaluation and Enhancement of Multistage Manufacturing Systems: a State of the Art

2.1. Overview

From car body assembly to wafer fabrication, from food processing to garment production, multistage manufacturing systems play an important role in modern industry. The prevalence of multistage manufacturing systems has attracted substantial research attention and resulted in the development of several analytical models for performance evaluation of such systems. One of the major objectives to develop these models is to predict system performance (e.g. production rate, inventory, production lead time, etc). Since these performance measures may be used to assess the impact of uncertainty, they are vital factors in the control and configuration of manufacturing systems. In the following section, the commonly used performance measures of manufacturing systems are discussed. Subsequently, in Section 2.3, analytical models for performance evaluation of multistage manufacturing systems are reviewed. In Section 2.4 and 2.5, we shall discuss analytical studies pertaining to preventive maintenance and inspection, which are two important strategies for improving performance of manufacturing systems.

2.2. Performance Measures of Manufacturing Systems

The increasing competitive pressure, resulting from the globalization of manufacturing activities and markets, stimulates manufacturing companies to continuously reorient their strategies, improve production efficiency, and reduce cost. To achieve the competitive standing, manufacturing companies must be able to measure different facets of performance of their systems, as reflected in Figure 2.1. Without the ability to measure performance, benchmarking efforts aimed at deploying the best manufacturing practices will not bear fruit. A variety of performance measures are used in practice, which may be roughly divided into two groups (Yang, 2007): 1) cost measures (the lower the better), such as inventory, production lead time, backorder, etc; 2) benefit measures (the higher the better), such as production rate, system yield, utilization, etc. Some commonly used performance measures in practice are highlighted as follows.

Figure 2.1. An important task in managing manufacturing systems is to predict system performance. The knowledge of performance measures may enable line managers to assess the system, and develop strategies for improving system performance.

- **Production rate**

Production rate is defined as the average number of parts a manufacturing system produces per unit time (Altiok, 1996). In some literature, it is also referred to as throughput (Bonvik et al., 2000). Production rate is a key performance measure of manufacturing systems, and it may be used to estimate the revenue of the systems.

- **System yield**

  System yield is a metric to evaluate the quality of production. It is defined as the fraction of input to a system that is transformed into output of products without defects (Kim, 2005). Another commonly used qualitative performance measure is effective production rate, i.e. the number of good parts a system produces per unit time.

- **Utilization**

  Utilization is defined as the fraction of time a machine is working (Gershwin, 1994). To improve production rate, machines in a manufacturing system should maintain relatively high utilization. One impediment for achieving this is random machine failure. By performing preventive maintenance, the probability of machine failure may be reduced, and hence the utilization is increased.

- **Inventory**

  Studies have demonstrated that inventory may comprise of up to 30 percent of a company's assets and perhaps as much as 90 percent of its working capital (Stevenson, 1992). Therefore, inventory has long been considered as a key performance indicator of manufacturing systems. The inventory in a manufacturing system is usually divided into three categories: raw materials, work-in-process (*WIP*), and finished goods. Raw materials are kept for two major reasons: to avoid frequent material transportations; and to reduce the impact of supply uncertainty on the production (Silver et al., 1998). *WIP* is defined as the total number of parts in the manufacturing system (in machines and intermediate buffers) (Meow 2001). Finished goods are held mainly to cope with the variability of demand and to shorten delivery time. The concentration of inventory investment varies in different industries. For example, in the primary steel industry of Canada, raw materials, *WIP*, and finished goods cost 46%, 25%, and 29% of the total inventory investment respectively, as illustrated in Figure 2.2. For railroad rolling stock manufacturers, the corresponding investments are 35%, 61%, and 4% respectively; and in the rubber industry, the numbers are 27%, 12%, and 61% respectively.

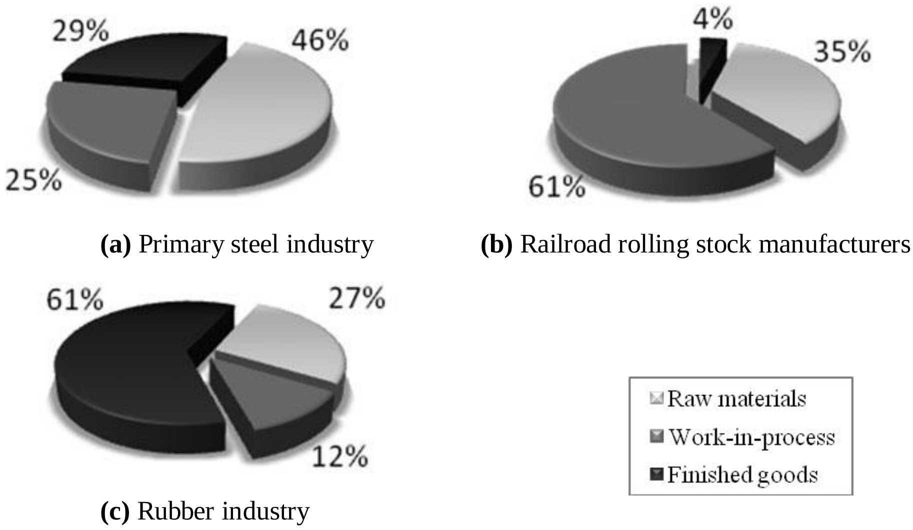


Figure 2.2. The relative concentration of inventory investment in three Canadian industries (in percent of total inventory investment) (Silver, et al., 1998).

- **Production lead time**

Production lead time is defined as the duration of time from the moment a part is released into a system until it finishes all the processes (Gershwin, 1994). According to Little’s law (Little, 1961), production lead time can be calculated as:

$$\text{Production lead time} = \frac{WIP}{\text{Production rate}} \tag{2.1}$$

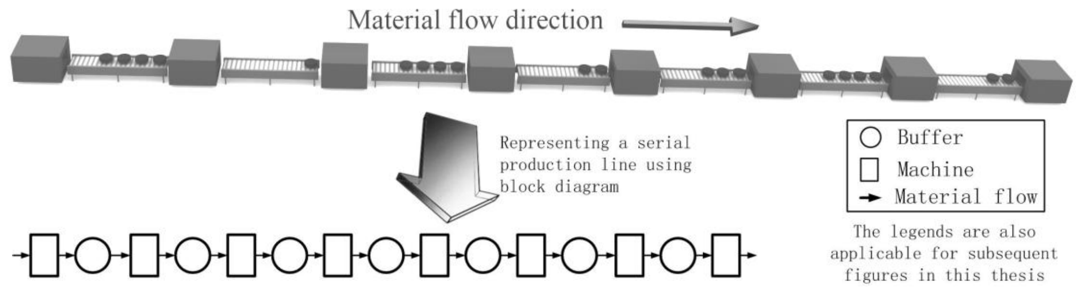
Hence, there are two ways to shorten the production lead time: reducing *WIP* and increasing production rate.

- **Backorder**

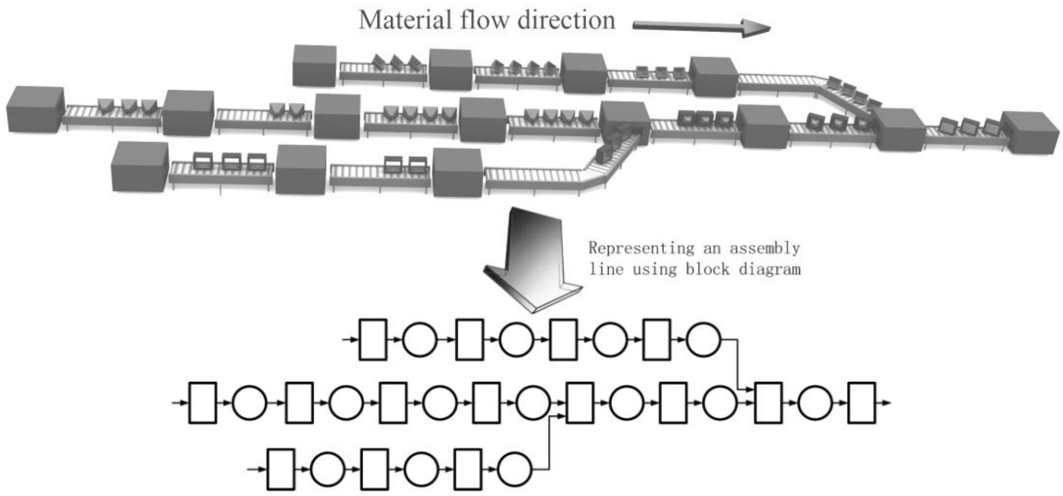
  Backorder is defined as the average amount of orders waiting to be served (Bonvik et al., 2000), and it is a performance indicator of customer service. Generally, low backorder usually implies good on-time delivery. An alternate measure to evaluate on-time delivery is the service level, which is the percentage of orders served before due times (Yang, 2007).

2.3. Analytical Models for Performance Evaluation of Multistage Manufacturing Systems

Reliable performance evaluation is desirable in the management of multistage manufacturing systems (Matta et al., 2005). Unfortunately, for multistage manufacturing systems (such as the serial production line and assembly line illustrated in Figure 2.3), providing reliable estimates of performance measures is a challenging task due to the large number of machines, complex configurations, and the uncertain characteristics of the systems. Computer simulation is widely used in practice for predicting performance measures of manufacturing systems (Takahashi et al., 2005; Yang et al., 2006; Carlson and Yao, 2008; Sandanayake et al., 2008, 2009; Hao and Shen, 2008; Betterton et al., 2009; Subramaniam et al., 2009). However, a relatively long computational time is usually required for obtaining performance measures with high confidence via simulation (Li and Meerkov, 2009). In some instances, especially when numerous alternate configurations must be analyzed, simulation may become prohibitively time consuming.



(a) A serial production line.



(b) An assembly line. In an assembly line, parts from different branches are merged to form a new one and hence the system has a non-serial configuration.

Figure 2.3. Two representative multistage manufacturing systems. In this figure, a rectangle represents a machine and a circle represents a buffer.

Analytical models of manufacturing systems have been developed as alternatives to simulation for providing performance measures with less computational time. As building exact models for multistage manufacturing systems is usually not tractable or too limited to be of interest (Dallery et al., 1992), many approximate models have been proposed in the literature, and these can be roughly categorized into aggregation (Ancelin, et al., 1987) and decomposition (Zimmern, 1956) models.

The fundamental idea of aggregation is to replace a two-machine-one-buffer section of the line with an equivalent machine, and this process is repeated until only one machine remains. This approach was initially proposed for serial production lines (Ancelin, et al., 1987). Kuo et al. (1997), Chiang et al. (2000), Li and Huang (2005) apply the aggregation approach to model assembly lines with exponential processing times. In these studies, assembly lines are divided into several serial production lines, each of which is then aggregated into an equivalent machine for calculating production rate. The aggregation approach was extended to include machine unreliability by Li and Meerkov (2005).

The decomposition approach, on the other hand, divides a multistage manufacturing system into a series of primitive line segments. The development of a decomposition model generally includes the following three steps (Dallery et al., 1992): (1) characterizing the primitive line segment; (2) deriving the equations to determine the parameters of each line segment; (3) developing an algorithm to solve these equations. The first step is critical, as it determines how the production line should be decomposed. One way to characterize the primitive line segment is using existing queuing models (Atiok et al., 1985; Dallery et al., 1989; Tempelmeier et al., 2001; Manitz et al., 2008). However, this may limit the extensibility of the decomposition approach for including various uncertainties in a manufacturing system. For example, if machines are subjected to some commonly observed random events, such as machine deterioration or quality failures, the existing queuing models may be insufficient to model such phenomena. For this reason, most of these researches focus only on production lines consisting of reliable machines and without quality issues (Dallery et al., 1992).

An alternative mathematical tool to characterize the primitive line segment is Markov theory, and this is used in the decomposition method proposed by Gershwin (1987). Based on this approach, a multistage manufacturing system is divided into a series of two-machine-one-buffer (2M1B) line segments. The state of each line segment is defined as $\left(x, \alpha^u, \alpha^d\right)$, where x represents the *WIP* in the line segment, α^u (or α^d) indicates whether the upstream (or downstream) machine is "up" or "down". A Markov model is formulated for each 2M1B line segment and provides the limiting probabilities of the states. These limiting probabilities are then used to calculate the performance measures, such as production rate and *WIP* of the system. The use of Markov theory in a decomposition model makes it possible to characterize various uncertainties in multistage manufacturing systems. Tolio et al. (2002) and Levantesi et al. (2003) explored production lines where machines have multiple failures, i.e. a machine may have different types of failures with distinct repair times. Kim and Gershwin (2005, 2008) and Colledani and Tolio (2006, 2009) extended the decomposition model to serial production lines where machines may experience quality failures. In addition to serial production lines, the decomposition method based on Markov theory has also been applied in multistage manufacturing systems with various configurations, including assembly/disassembly lines (Gershwin, 1991; Gershwin and Burman 2000) and multiple-part systems (Colledani et al., 2005, 2008; Gurgur and Altiok 2007, 2008).

In the literature, several case studies on the application of decomposition models in real manufacturing systems have been published. Burman et al.

(1998) investigated an ink-jet printer production line at Hewlett-Packard Corporation, and developed a model for performance evaluation of this system. Liberopoulos and Tsarouhas (2002) formulated a model for the croissant production line of Chipita International Inc., one of the largest Greek manufacturers of bakery products and snacks. Their model was used to determine the size of each buffer in the production line. Patchong et al. (2003) presented a case study on the car body assembly line at PSA Peugeot Citroen. An analytical model was formulated and subsequently used to examine the impact of machine failures on production rate. Alden et al. (2006) analyzed the performance of a car assembly line of General Motors Corporation. Their model was used to identify the bottleneck machines and improve buffer allocation. Colledani et al. (2010) studied a production line of Scania, a manufacturer of heavy trucks and buses, as well as industrial and marine diesel engines, and proposed a model for the purpose of performance evaluation. In all these case studies, machine unreliability is considered as an important factor that undermines production rate. For mathematical tractability, these case studies generally assume that machines have only two states (i.e. machines are either up or down). However, this assumption is inadequate for modeling systems with machine deterioration and preventive maintenance. Additionally, production is assumed to be perfect in these case studies, and inspection is not considered. Due to the inadequacy of the previous decomposition models in the literature, the study on real manufacturing systems with preventive maintenance and inspection was not attempted to the best knowledge of the author. However, with the model proposed in this thesis, we are able to simultaneously analyze the quantitative and qualitative performance of an unreliable multistage manufacturing system with imperfect production. The proposed model can also be used to plan preventive maintenance and determine the allocation of inspection machines. This may further improve the performance of a manufacturing system.

2.4. Analytical Studies of Manufacturing Systems with Unreliable Machines and Preventive Maintenance

In previous analytical models of multistage manufacturing systems that consider unreliability, machines are usually assumed to have two states: "up" and "down". If a machine is "up", it has a constant transition rate to break down. However, as Yao et al. (2005) pointed out, this two-state description of machine reliability may not be accurate if machines are subjected to continuous deterioration, a phenomenon widely observed in practice (Gurler and Kaya, 2002; Moustafa et al., 2004; Chen and Trivedi, 2005). In many real systems, machines continuously degrade due to various reasons, such as gear wear, corrosion, fatigue, ageing, etc (Montoro-Cazorla and Perez-Ocon, 2006). As deterioration accumulates, machines become more and more failure prone, and eventually break down. Through preventive maintenance, manufacturers may effectively reduce the accumulated deterioration and hence prevent the occurrence of machine failures. Therefore, preventive maintenance may substantially improve production rate of a manufacturing system. In order to provide reliable performance evaluation of a manufacturing system, it may be necessary to incorporate preventive maintenance in the analytical model (Li et al., 2009; Chen and Subramaniam, 2010).

The majority of analytical studies on preventive maintenance focus on single-machine manufacturing systems (Bloch-Mercier, 2002; Gurler and Kaya, 2002; Moustafa et al., 2004; Chen and Trivedi, 2005; Montoro-Cazorla and Perez-Ocon, 2006; Bao and Jaishankar, 2008; Wu and Makis, 2008). In comparison with the two-state ("up" and "down") description of machine reliability, these studies generally incorporate a number of additional states between the best and worst states of a machine to represent different levels of deterioration. In addition, a state of preventive maintenance is also introduced. Markov models have been formulated to describe the transitions between all these states, and based on these models, the limiting probabilities of the states are calculated. These probabilities are subsequently used to estimate performance measures of a manufacturing system, including machine availability (i.e. the fraction that a machine is neither down nor under maintenance), average repair and maintenance costs, etc. Based on these performance measures, maintenance managers may be able to evaluate the reliability and cost of a system and hence determine an appropriate frequency to perform preventive maintenance.

Some recent researches explored preventive maintenance in more complex manufacturing systems rather than single-machine systems. Ambani et al. (2009) analyzed a three-machine serial production line and formulated a Markov model for calculating the availability of each machine and the whole line. However, some key performance measures of the manufacturing system, such as the inventory, were not provided in this study. In addition, this study assumes that the system is without intermediate buffers. However, in production lines with unreliable machines, buffers are usually placed to reduce the impact of machine failures. This has been considered by Kyriakidis and Dimitrakos (2006), who explored a two-machine system with an intermediate finite buffer. The upstream machine is assumed to be subjected to deterioration and the downstream machine is reliable. A model was formulated and then used to plan preventive maintenance for the upstream machine. Pavitsos and Kyriakidis (2009) analyzed a similar system where upstream and downstream machines are swapped (i.e. the upstream machine is reliable while the downstream machine is unreliable).

Previous analytical models pertaining to preventive maintenance generally focus on small manufacturing systems. Their extensions to large-scale manufacturing systems have been studied limitedly. However, these extensions are necessary as many real manufacturing systems, such as auto assembly lines, usually consist of hundreds of machines (Sakai and Amasaka, 2007). In a multistage manufacturing system, the relationship between preventive maintenance and production rate is more complex than that in small systems, due to the large number of machines. To describe this relationship, the influence of machines on each other should be incorporated in the model. Therefore, the author formulates an analytical model for performance evaluation of multistage manufacturing systems with preventive maintenance, and this will be discussed in Chapter 3.

2.5. Analytical Studies of Manufacturing Systems with Imperfect Production and Quality Inspection

Previous analytical models of multistage manufacturing systems focus on predicting quantitative performance measures, particularly production rate. This emphasis on production rate is necessary for achieving high revenue. However, it is equally important to maintain high-quality production, since products with inferior quality may incur expensive penalty cost and a loss of market share (Montgomery, 2001; Mandroli, et al., 2006). Therefore, to comprehensively assess the performance of a multistage manufacturing system, it is necessary to develop an integrated model that provides both quantitative and qualitative performance measures (Cao et al., 2010). In such an integrated model, the conservation of part flow, which was usually assumed in previous quantitative models of multistage manufacturing systems, is no longer satisfied. The flow rate at each machine is altered (Penn and Raviv, 2007, 2008) as defective parts are removed by inspection machines. This phenomenon needs to be considered for reliable performance evaluation of the multistage manufacturing system.

A feature of the manufacturing system rarely reflected in previous quantitative models in the literature is the quality of material flow (which may be alternatively interpreted as the fraction of parts without defect after each machine). This is an important performance indicator of a multistage manufacturing system. High fraction of defective parts in the material flow usually implies that a substantial portion of processing capacity is lost on these defective parts. To provide a reliable estimate for the quality of material flow in the manufacturing system, a number of qualitative models have been proposed, and the majority of these studies are based on serial production lines (Bai and Yun, 1996; Lee and Unnikrishnan, 1998; Heredia-Langner et al., 2002; Kakade et al., 2004; Rau and Chu, 2005; Freiesleben, 2006; Van

Volsem et al., 2002, 2007; Penn and Raviv, 2007, 2008). These models describe the impact of imperfect production and inspection on the quality of material flow. One major application of these models is to determine the allocation of inspection machines. By appropriately placing inspection machines in a manufacturing system, the quality of material flow may be improved, resulting in a reduction of cost. These studies generally assume that each processing machine randomly generates defective parts with a constant probability, and this type of quality failure is referred to as the Bernoulli-type quality failure (Montgomery, 2001). Bernoulli-type quality failure is inherent in the design of a machine and cannot be removed by maintenance or repair. Since this type of quality failure is common in practice, many analytical models of multistage manufacturing systems with imperfect production have assumed this type of quality failure. Another type of quality failure is persistent quality failure, which is attributed to physical causes, such as the breakdown of tools. Once a processing machine experiences a persistent quality failure, it will continue producing defective parts, until the quality failure is detected and repaired. A recent study by Kim and Gershwin (2005) investigated a serial production line with persistent quality failure and provided an analytical model for predicting the yield of such a system.

In assembly lines, parts delivered from different branches are merged to form a new one. Predicting the quality of material flow in assembly lines is more complex than that in tandem production lines due to the non-serial configuration of the systems. An early analytical study of inspection and qualitative issues in assembly lines was analyzed by Britney (1972) based on a case consisting of six processing machines. This model was later extended by

Taneja and Viswanadham (1994) and Viswanadham, et al. (1996) to assembly lines with arbitrary number of machines. These studies incorporate the type I and type II inspection errors (viz. rejecting good parts and accepting non-conforming parts respectively). In these models, the quality of material flow and the proportion of parts rejected by each inspection machine are calculated through an iterative procedure. These values are then used to estimate different types of costs, such as processing cost, inspection cost, penalty cost due to shipping nonconforming parts to customers, etc. Based on these estimates, the placement of inspection machines is formulated as a problem to minimize the total cost required for producing one part. In recent researches, qualitative models have been formulated for some particular assembly lines by incorporating their specific features. For instance, Vivek et al. (2004) explored a printed circuit board (PCB) assembly line, in which the inspection of all components on the PCB is costly. Therefore, only the most defect-prone components on the PCB are tested in each inspection machine. Trichy et al. (2001) and Shi and Sandborn (2003, 2006) investigated assembly lines where parts are repaired immediately after they are classified as defective. By repairing defective parts, the value of these parts may be salvaged, and hence the total cost of the assembly line is further reduced.

In previous qualitative models of assembly lines, the influence of inspection machines on the quantitative performance is investigated limitedly. As pointed out by Drezner et al. (1996), inspection may substantially affect production rate as well as *WIP*. Avoiding this consideration may undermine the effectiveness of inspection allocation. This was also reiterated by Penn and Raviv (2007, 2008), who explored the inspection allocation problem in a serial production line with exponential processing times and infinite buffers. However, the extension of this model for non-serial production systems, such as assembly lines, was not provided. In many systems, buffers are intentionally chosen to be finite in order to prevent excessive WIP. Therefore, the study reported by Penn and Raviv (2007), in which infinite buffers were considered, may not be applicable for such systems. Furthermore, the study restricted the focus on production lines where machines never break down and inspection is error-free. This condition may not be satisfied in many real production lines, as machine failures and inspection errors are widely observed in practice, and they substantially affect the performance of manufacturing systems.

Chapter 3.

Performance Enhancement of Multistage Manufacturing Systems with Unreliable Machines

—An Investigation of Machine Deterioration and Preventive Maintenance

3.1. Overview

In real manufacturing systems, machines may deteriorate due to corrosion, gear wear, fatigue, ageing, and many other reasons (Montoro-Cazorla and Perez-Ocon, 2006). The deterioration may accumulate in a machine and eventually cause breakdown. To eliminate machine deterioration and reduce subsequent failures, preventive maintenance is widely implemented in manufacturing industries. However, as reflected in Figure 3.1, excessive preventive maintenance also induces frequent interruption to production. Therefore, preventive maintenance ought to be performed at suitable frequencies such that the interruption to production is minimized. In this chapter, the author investigates preventive maintenance in multistage manufacturing systems, and develops an analytical model for such systems. In Section 3.3, this model is presented for assembly line systems, the multistage systems of choice in the manufacturing of various products, such as automobile, LCD television, personal computer, etc. In addition to assembly lines, the proposed model may also be applicable for serial production lines. In Section 3.4, we discuss the application of the model in determining the frequency of preventive maintenance of each machine in an assembly line, and this is formulated as an optimization problem for maximizing production rate. Numerical illustrations are provided in Sections 3.5 to 3.10.

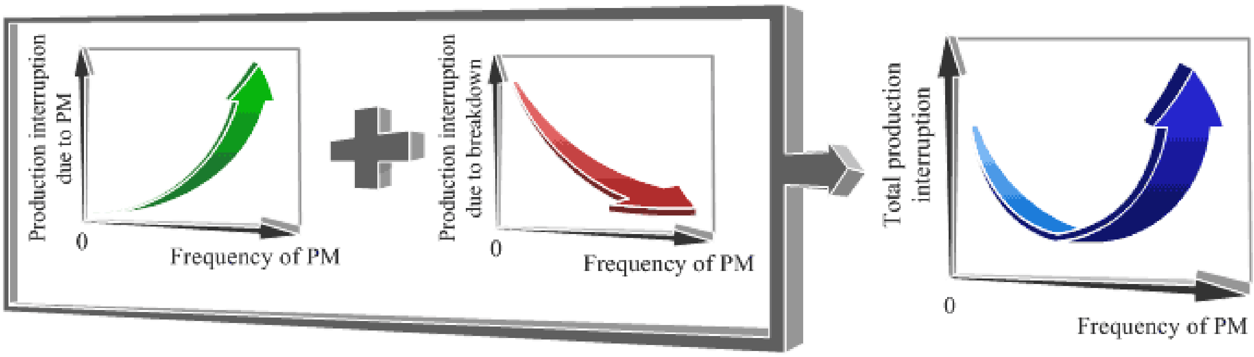


Figure 3.1. An appropriate preventive maintenance frequency may minimize production interruptions due to maintenance or machine failures, resulting in high machine reliability.

3.2. Definition of Notations

The notations used in this chapter to describe production lines (such as the systems illustrated in Figure 3.2) are listed below. The notations that denote "rate" (e.g. preventive maintenance rate and deterioration rate) represent the transition rate of the occurrence of an event (e.g. preventive maintenance or machine deterioration). These rates are used to characterize transitions of states in Markov models and are defined as (Gross and Harris, 1998):

$$\lim_{\delta t \to 0} \frac{Prob\left(\text{Event } E \text{ has occured at } t+\delta t \mid \text{Event } E \text{ has not occured at } t\right)}{\delta t} \tag{3.1}$$

K: The number of machines in a system. $K = 2$ for the 2M1B system.

k: The index of machine in a system, i.e. $M_k \quad k \in \{1, 2, ..., K\}$. For convenience, we also use the index of machine, if it is not the last machine, to denote its immediately downstream buffer (B_k). It should be noted that there are only $K-1$ intermediate buffers in the system.

X_k: The maximum capacity of buffer B_k, where $k \in \{0,1,...,K-1\}$.

x_k: The number of parts in buffer B_k, $x_k \in \{0,1,...,X_k\}$ and $k \in \{0,1,...,K-1\}$.

μ_k: Processing rate of M_k.

N_k: The number of upstates of M_k. As illustrated in Figure 3.3, a machine has N_k upstates representing different levels of deterioration (Gurler and Kaya, 2002). A machine may degrade from one upstate to its subsequent upstate until it finally breaks down.

n_k: The index of upstates of machine M_k, $n_k \in \{1,2,...,N_k\}$.

γ_k: The state of M_k. $\gamma_k \in \{1,2,\cdots,N_k,N_k+1,N_k+2\}$, where $\{1,2,\cdots,N_k\}$ are the N_k upstates of increasing deterioration. $\gamma_k = 1$ represents the "best" condition state of M_k. As illustrated in Figure 3.3, it is assumed that $\gamma_k = N_k$ is the worst possible up condition, and further deterioration will lead to the failure of the machine and hence transit to the down state, $\gamma_k = N_k + 1$. In addition, $\gamma_k = N_k + 2$ represents the preventive maintenance state.

p_{k,n_k}: The deterioration rate of M_k from the n_k^{th} upstate to the subsequent ($(n_k+1)^{\text{th}}$) state, as illustrated in Figure 3.3.

r_k: The repair rate of M_k.

π_{k,n_k}: Preventive maintenance rate of machine M_k in the n_kth upstate. As illustrated in the transition diagram of Figure 3.3, M_k is subjected to preventive maintenance at any of its upstate ($\gamma_k \leq N_k$). In the literature, π_{k,n_k} is usually assumed to be identical for each upstate (Bao and Jaishankar, 2008; Ambani et al., 2009). In this study, the analytical model is formulated without this restriction. The proposed analytical model does not require the maintenance rates for the various upstates to be identical. However, for reasons of simplicity, in the numerical experiments, the maintenance rates are chosen to be homogeneous, and we have:

$$\pi_{k,1} = \pi_{k,2} = \cdots = \pi_{k,N_k} = \frac{1}{Mean\left(\text{inter-maintenance time of } M_k\right)} \quad (3.2)$$

ρ_k: The transition rate to complete preventive maintenance of M_k.

$$\rho_k = \frac{1}{Mean\left(\text{maintenance time of } M_k\right)} \quad (3.3)$$

Fq_k: The frequency of preventive maintenance on machine M_k. Based on Eqns (3.2) and (3.3), Fq_k may be calculated as:

$$\begin{aligned} Fq_k &= \frac{1}{Mean\left(\text{inter-maintenance time of } M_k\right) + Mean\left(\text{maintenance time of } M_k\right)} \\ &= \frac{1}{\dfrac{1}{\pi_{k,n_k}} + \dfrac{1}{\rho_k}} \\ &= \frac{\pi_{k,n_k} \cdot \rho_k}{\pi_{k,n_k} + \rho_k} \end{aligned} \quad (3.4)$$

3.3. Model Development

In this section, the author will first present a continuous-time-discrete-state Markov model of a two-machine-one-buffer (2M1B) line with machine deterioration and preventive maintenance. This 2M1B line is then used as "building blocks" for modeling assembly lines in Sections 3.3.2 and 3.3.3.

3.3.1. A 2M1B Line with Machine Deterioration and Preventive Maintenance

To formalize the model, the following assumptions are used in this section.

- The machine deterioration is assumed to be operation dependent (i.e. a machine may deteriorate only when it is processing and it will not deteriorate when it is idle) (Gershwin, 1994).

  This is a common assumption in many analytical studies of manufacturing systems with unreliable machines (Bonvik, et al., 2000; Colledani, et al., 2005, 2006, 2008, 2009; Gershwin, 1994, 2000; Gurgur and Altiok, 2007, 2008; Kim and Gershwin, 2005, 2009; Matta, et al., 2005). In real systems, machines may deteriorate due to various reasons, and tool wear is a widely reported reason of machine deterioration (Montoro-Cazorla and Perez-Ocon, 2006). As mentioned by Li and Meerkov (2009), tool wear does not occur when a machine is idle, and the operation dependent assumption reflects this feature. However, this assumption is not restrictive in this study. The state transitions of the Markov model in this thesis may be extended to account for time dependent deterioration, where a machine may deteriorate even if it is not processing

(Buzacott and Shanthikumar, 1993; Mourani et al., 2007).

- As illustrated in Figure 3.3, we shall assume that both repair and preventive maintenance will revert the state of a machine to the best condition (i.e. $\gamma_k = 1$).

  This assumption is widely used in many researches on preventive maintenance (Smith and Dekker, 1997; Tomasevicz and Asgarpoor, 2006; Bao and Jaishankar, 2008; Ambani et al., 2009). The validity of this assumption is supported by a case study on the Toyota production system (Sakai and Amasaka, 2007). Sakai and Amasaka analyzed the reliability of robots in such a system, and pointed out that gear wear is one major cause of robot deterioration and failure. In the preventive maintenance implemented in this system, gears of robots are replaced with new ones, which can effectively remove accumulated deterioration. Additionally, when a robot breaks down due to the gear wearing, an overhaul is performed, and then the robot is restored to the best operating condition.

- The deterioration of machines is assumed to be unobservable (i.e. maintenance operators are incapable of determining the current upstate of a machine).

  Unobservable machine deterioration is assumed in many analytical studies pertaining to preventive maintenance (Yeh, 2003; Kuo, 2006; Bao and Jaishankar, 2008; Ghasemi et al., 2010). Unobservable deterioration is common in practice. For instance, the deterioration of the drilling tool in a CNC machine may result in an increment of vibration, which is difficult to detect to the naked eye (Naveen Prakash and Ravindra, 2008). Although it is possible to use sophisticated sensors to monitor the state of a machine, these sensors may be too expensive to implement. In Section 3.10, we shall discuss the extension of the model for systems where the deterioration of machines can be detected via inspecting the machines (Ambani et al., 2009).

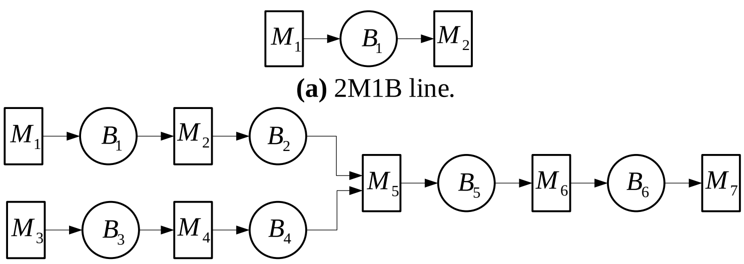
- All the machines are assumed to be single-item machines (i.e. each machine may process one part each time) in this chapter.

  In the literature, the majority of analytical models of multistage manufacturing systems also focus on single-item machines (Chiang et al., 2000; Penn and Raviv, 2007, 2008; Li and Meerkov, 2009; Gershwin, 1994, 2000). This is because single-item machines are commonly observed in real manufacturing systems. However, in some systems, such as wafer fabrication lines, batch machines (i.e. a machine may process several parts simultaneously) are also used to improve the production rate. For this reason, the author will discuss the incorporation of batch machines in the decomposition model in Chapter 5.

- We also assume that the first machine in the system is never starved of raw material and the last machine is never blocked as finished parts are removed immediately from the system.

  Stock out of raw material may induce an expensive loss of machine capacity and hence a reduction of profit (Huang and Wu, 2010). In practice, manufacturers usually implement sophisticated replenishment policies, which prevent the first machine from running out of raw material. Therefore, it is reasonable to assume that the system is never starved of raw material. On the other hand, the assumption that the last machine never gets blocked is made for reasons of simplicity in many studies (Gershwin, 1994, 2000; Chiang et al., 2000; Levantesi et al., 2003; Kim, 2005). This condition may not be true if a system is running at a demand rate, which is lower than the system capacity. However, as discussed by Li and Meerkov (2009), the customer demand can be approximated as an additional processing machine in this case. With this approximation, the proposed model in this chapter may also be extended to the systems with finite demand.

Figure 3.3 depicts state transitions of machine M_k, $k \in \{1, 2\}$, in a 2M1B line due to deterioration and preventive maintenance. If $\gamma_k \le N_k$, M_k may deteriorate to a worse condition with a transition rate of p_{k,γ_k}. When M_k is at the worst upstate ($\gamma_k = N_k$), further deterioration may result in machine failure ($\gamma_k = N_k + 1$) with a transition rate of p_{k,N_k}. Once M_k breaks down, repair is triggered, and the transition rate to complete repair of M_k is r_k. If M_k is operational ($\gamma_k \le N_k$), preventive maintenance is performed with a transition rate of π_{k,γ_k}, and this operation is completed at a transition rate of ρ_k.



(b) Assembly line.

Figure 3.2. Multistage manufacturing systems.

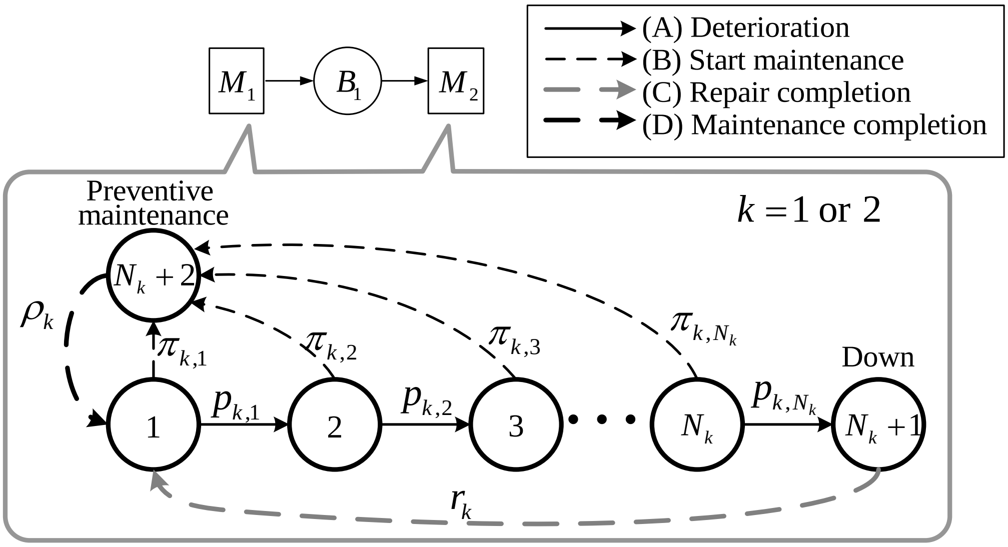


Figure 3.3. Transitions of machine states (γ_k) due to deterioration and preventive maintenance. Note: the transitions illustrated in this figure are for machine M_k only. While the transitions of all machines may be similar, the parameters associated with transitions may differ for different machines.

The 2M1B line is modeled as a continuous-time-discrete-state Markov process, with the state defined as:

$$S = (x_1, \gamma_1, \gamma_2) \tag{3.5}$$

The balance equations (Gershwin 1994), which equate the rate of leaving a particular state and the rate of entering it, are used to describe state transitions of the 2M1B line. γ_k , $k \in \{1, 2\}$, may be roughly categorized into the following four conditions:

- M_k is at the best state ($\gamma_k = 1$).
- M_k is at a deteriorated state ($2 \le \gamma_k \le N_k$).
- M_k is down ($\gamma_k = N_k + 1$).
- M_k is under preventive maintenance ($\gamma_k = N_k + 2$).

Based on the combination of these four conditions of both machines, we have 16 groups of balance equations, as listed in Table 3.1. For reasons of brevity, we do not provide the complete balance equations in this subsection.

Instead, Group 6 in Table 3.1 is selected as an example to illustrate the development of balance equations. The complete balance equations are provided in Appendix A.

Table 3.1. Balance equation groups based on γ_1 and γ_2.

Group 1	$\gamma_1 = 1, \gamma_2 = 1$	**Group 2**	$2 \le \gamma_1 \le N_1, \gamma_2 = 1$
Group 3	$\gamma_1 = N_1 + 1, \gamma_2 = 1$	**Group 4**	$\gamma_1 = N_1 + 2, \gamma_2 = 1$
Group 5	$\gamma_1 = 1, 2 \le \gamma_2 \le N_2$	**Group 6**	$2 \le \gamma_1 \le N_1, 2 \le \gamma_2 \le N_2$
Group 7	$\gamma_1 = N_1 + 1, 2 \le \gamma_2 \le N_2$	**Group 8**	$\gamma_1 = N_1 + 2, 2 \le \gamma_2 \le N_2$
Group 9	$\gamma_1 = 1, \gamma_2 = N_2 + 1$	**Group 10**	$2 \le \gamma_1 \le N_1, \gamma_2 = N_2 + 1$
Group 11	$\gamma_1 = N_1 + 1, \gamma_2 = N_2 + 1$	**Group 12**	$\gamma_1 = N_1 + 2, \gamma_2 = N_2 + 1$
Group 13	$\gamma_1 = 1, \gamma_2 = N_2 + 2$	**Group 14**	$2 \le \gamma_1 \le N_1, \gamma_2 = N_2 + 2$
Group 15	$\gamma_1 = N_1 + 1, \gamma_2 = N_2 + 2$	**Group 16**	$\gamma_1 = N_1 + 2, \gamma_2 = N_2 + 2$

In Group 6, machines M_1 and M_2 are operational. The balance equations of this group may be further divided into three sub-categories:

1) Internal state (where M_1 is not blocked and M_2 is not starved).

Based on the transition diagram in Figure 3.4(a) for the internal state, the balance equation may be formulated as:

$$P(x_1, \gamma_1, \gamma_2) \cdot \left(\mu_1 + \mu_2 + p_{1,\gamma_1} + p_{2,\gamma_2} + \pi_{1,\gamma_1} + \pi_{2,\gamma_2}\right) = P(x_1 - 1, \gamma_1, \gamma_2)\mu_1 +$$
$$P(x_1 + 1, \gamma_1, \gamma_2)\mu_2 + P(x_1, \gamma_1 - 1, \gamma_2)\, p_{1,\gamma_1 - 1} + P(x_1, \gamma_1, \gamma_2 - 1)\, p_{2,\gamma_2 - 1} \qquad (3.6)$$

The left side of Eqn (3.6) represents the total transition rate from state $(x_1, \gamma_1, \gamma_2)$ to other states. The transitions out of this state include:

- Part arrival (M_1 delivers parts to the buffer at the rate of μ_1).
- Part departure (M_2 sends out parts at the rate of μ_2).

- Deterioration of M_1 and M_2 (with the rate p_{1,γ_1} and p_{2,γ_2} respectively).
- Preventive maintenance of M_1 and M_2 (with the rate π_{1,γ_1} and π_{2,γ_2}).

The right side of Eqn (3.6) indicates the total transition rate from other states to state (x_1,γ_1,γ_2). These transitions include part arrival, part departure, and deterioration of both machines.

2) Boundary state with $x_1 = 0$ (i.e. M_2 is starved).

When the buffer is empty, M_2 is starved and remains idle. In this case, M_2 will not deteriorate due to the operation dependent assumption. As illustrated in the transition diagram of Figure 3.4(b), we have the balance equation as:

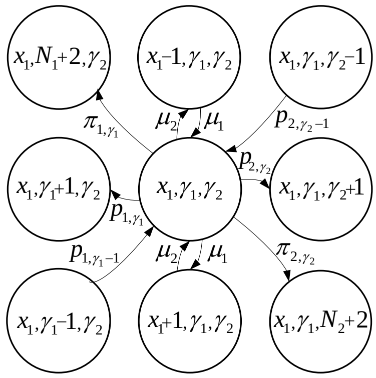
$$P(0,\gamma_1,\gamma_2)\cdot(\mu_1 + p_{1,\gamma_1} + \pi_{1,\gamma_1} + \pi_{2,\gamma_2}) = P(1,\gamma_1,\gamma_2)\mu_2 + P(0,\gamma_1 - 1,\gamma_2)\, p_{1,\gamma_1 - 1} \tag{3.7}$$

3) Boundary state with $x_1 = X_1$ (i.e. M_1 is blocked).

When the buffer is full, i.e. $x_1 = X_1$, M_1 is blocked, and remains idle. Hence, M_1 will not deteriorate. The balance equation for this state can be derived based on Figure 3.4(c) as:

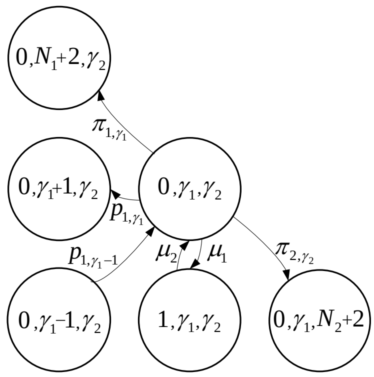
$$P(X_1,\gamma_1,\gamma_2)\cdot(\mu_2 + p_{2,\gamma_2} + \pi_{1,\gamma_1} + \pi_{2,\gamma_2}) = P(X_1 - 1,\gamma_1,\gamma_2)\mu_1 + P(X_1,\gamma_1,\gamma_2 - 1)\, p_{2,\gamma_2 - 1} \tag{3.8}$$

Similarly, balance equations for the other groups can be derived. By solving the balance equations and the normalization equation (i.e. the sum of all probabilities equals one), the limiting probabilities of all states, $P(x_1,\gamma_1,\gamma_2)$ are obtained.



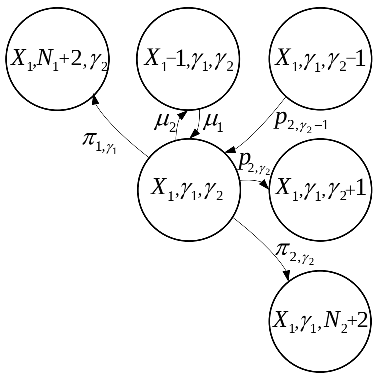
$0 < x_1 < X_1$

(a) Internal transition.



$x_1 = 0$

(b) Boundary transition when buffer is empty (M_2 is starved).



$x_1 = X_1$

(c) Boundary transition when buffer is full (M_1 is blocked).

Figure 3.4. Transition diagrams for state (x_1,γ_1,γ_2) (the group with $2 \le \gamma_1 \le N_1$ and $2 \le \gamma_2 \le N_2$).

3.3.2. Assembly Lines with Preventive Maintenance

An assembly line with K machines is approximately decomposed into $K-1$

primitive line segments, as illustrated in Figure 3.5. Each line segment is denoted with the index, k ($k \in \{1, 2, \cdots, K-1\}$). Each line segment consists of an upstream machine (M_k^u), a downstream machine (M_k^d), and an intermediate buffer (B_k). The superscripts "u" and "d" are used to differentiate the parameters of upstream and downstream machines.

To capture the influence of primitive line segments on each other, the author introduces a new state, referred to as "pseudo down" for both M_k^u and M_k^d. A machine in a primitive line segment is said to be "pseudo down" if it is starved or blocked by the upstream or downstream segments, and hence remains idle. If M_k^u (or M_k^d) is "pseudo down", it cannot process, which is similar to the condition that a machine is physically down. However, "pseudo down" is distinct from machine failure because preventive maintenance may still be performed on M_k^u (or M_k^d) in this state; while only repair may be performed if M_k^u (or M_k^d) is physically down. Two variables α_k and β_k are used to denote whether M_k^u and M_k^d are "pseudo down" respectively, and these two variables are defined as follows:

- α_k=0: The upstream machine M_k^u is "pseudo down", and this represents the condition that M_k^u's corresponding machine in the assembly line (i.e. M_k) is starved. For instance, in Figure 3.6(a), M_6^u represents a non-assembly machine M_6, and M_6^u is "pseudo down" if the upstream buffer of M_6 (B_5) is empty. While in the primitive line segment of Figure 3.6(b), M_5^u represents an assembly machine, M_5. M_5^u is "pseudo down" if either

B_2 or B_4 is empty.

- α_k =1: M_k^u is not "pseudo down", which implies that the corresponding machine in the assembly line, M_k, is not starved. If α_k =1, M_k^u continues processing as long as M_k^u is operational and B_k is not full.
- β_k =0: The downstream machine M_k^d is "pseudo down". In Figure 3.6(c), the downstream machine in the primitive line segment, M_5^d, represents a non-assembly machine, M_6. In this case, M_5^d is "pseudo down" if buffer B_6 is full ($x_6 = X_6$). A more complex case is as illustrated in Figure 3.6(d), where M_2^d represents the assembly machine M_5. M_2^d is "pseudo down" under one of the following conditions:

  1) $x_4 = 0$ (i.e. parts required for assembly are lacking).

  2) $x_5 = X_5$ (i.e. the downstream buffer of M_5 is full).
- β_k =1: M_k^d is not "pseudo down", and thus it continues processing as long as M_k^d is operational and B_k is not empty.

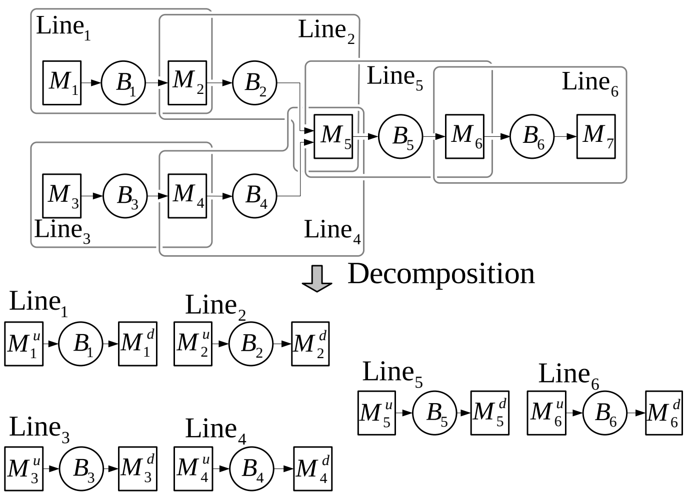
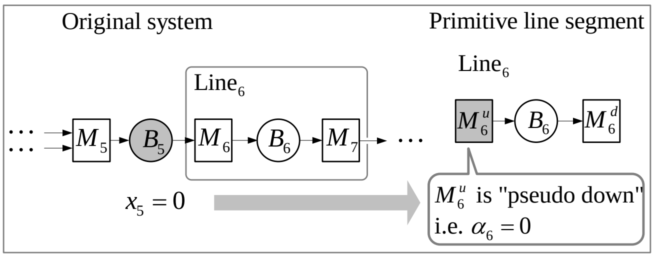
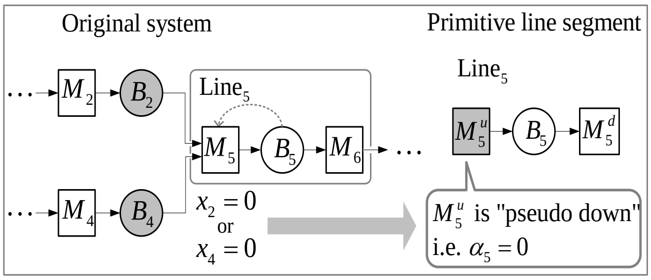


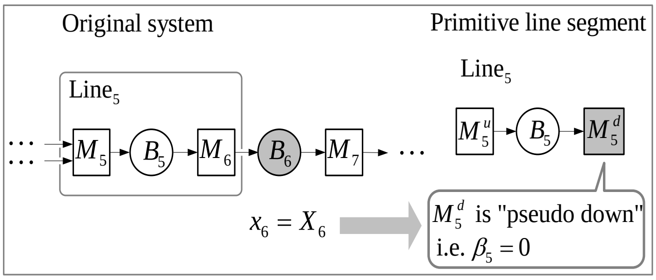
Figure 3.5. Decomposing an assembly line into primitive line segments.



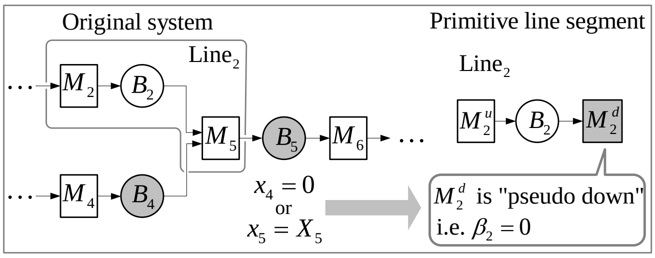
(a) M_k^u represents a non-assembly machine.



(b) M_k^u represents an assembly machine.



(c) M_k^d represents a non-assembly machine.



(d) M_k^d represents an assembly machine.

Figure 3.6. The interpretation of the "pseudo down" state. This example is based on the assembly line depicted in Figure 3.5.

With α_k and β_k, the state of the k^{th} primitive line segment may be extended based on Eqn (3.5) as:

$$S_k = (x_k, \gamma_k^u, \gamma_k^d, \alpha_k, \beta_k) \tag{3.9}$$

The balance equations of the primitive line segment can be easily derived based on the discussion in Section 3.3.1 and by considering the additional transitions of states α_k and β_k, as elaborated below:

1) Transition from $\alpha_k = 1$ to $\alpha_k = 0$.

M_k^u may be switched to "pseudo down" state only when it completes processing a part. This is because the starvation of a machine (e.g. M_k) in the assembly line, if it occurs, commences at the moment when M_k completes processing a part and cannot obtain the parts for its next processing. Hence, when M_k^u delivers a part to buffer B_k, it has a probability of becoming "pseudo down". This probability is denoted as p_k^α, and used as the parameter to describe the transition from $\alpha_k = 1$ to $\alpha_k = 0$:

$$Prob\left(\alpha_k\left(t+\delta t\right)=0 \mid \alpha_k\left(t\right)=1, \gamma_k^u\left(t\right)\le N_k^u, x_k\left(t\right)< X_k\right)= p_k^\alpha \cdot \mu_k^u \cdot \delta t \qquad (3.10)$$

where $\alpha_k\left(t\right)=1$, $\gamma_k^u\left(t\right)\le N_k^u$, and $x_k\left(t\right)< X_k$ indicate M_k^u is processing at time *t*. μ_k^u is the transition rate that M_k^u completes processing a part. $p_k^\alpha \cdot \mu_k^u$ implies that M_k^u becomes "pseudo down" with the probability p_k^α when it completes processing a part.

2) Transition from $\beta_k = 1$ to $\beta_k = 0$.

Similarly, when M_k^d delivers a part out of the primitive line segment, it has a probability of becoming "pseudo down". This probability, denoted as p_k^β, is used to describe the transition from $\beta_k = 1$ to $\beta_k = 0$.

$$Prob\left(\beta_k\left(t+\delta t\right)=0 \mid \beta_k\left(t\right)=1, \gamma_k^d\left(t\right)\le N_k^d, x_k\left(t\right)> 0\right)= p_k^\beta \cdot \mu_k^d \cdot \delta t \qquad (3.11)$$

3) Transition from $\alpha_k = 0$ to $\alpha_k = 1$.

M_k^u recovers from "pseudo down" with the transition rate of r_k^α, i.e.

$$Prob\left(\alpha_k\left(t+\delta t\right)=1 \mid \alpha_k\left(t\right)=0\right)= r_k^\alpha \cdot \delta t \qquad (3.12)$$

4) Transition from $\beta_k = 0$ to $\beta_k = 1$.

Similarly, if M_k^d is "pseudo down", it recovers with the transition rate of r_k^β :

$$Prob\left(\beta_k\left(t+\delta t\right)=1 \mid \beta_k\left(t\right)=0\right)=r_k^\beta \cdot \delta t \tag{3.13}$$

With these considerations, we may extend the balance equations for 2M1B line (discussed in Section 3.3.1) to a primitive line segment. For instance, for the group of states $(x_k, \gamma_k^u, \gamma_k^d, \alpha_k, \beta_k)$ with $2 \le \gamma_k^u \le N_k^u$ (M_k^u is in the internal deterioration state), $2 \le \gamma_k^d \le N_k^d$ (M_k^d is in the internal deterioration state), $\alpha_k = 1$ (M_k^u is not pseudo down), and $\beta_k = 1$ (M_k^d is not pseudo down), we may have balance equations as follows.

1) Internal state ($0 < x_k < X_k$).

For the internal state, M_k^u is not blocked and M_k^d is not starved. The balance equation can be formulated as:

$$\begin{aligned}
&P\left(x_k,\gamma_k^u,\gamma_k^d,1,1\right)\cdot\left(\mu_k^u p_k^\alpha+\mu_k^u\left(1-p_k^\alpha\right)+\mu_k^d p_k^\beta+\mu_k^d\left(1-p_k^\beta\right)+p^u_{k,\gamma_k^u}+p^d_{k,\gamma_k^d}+\pi^u_{k,\gamma_k^u}+\pi^d_{k,\gamma_k^d}\right)= \\
&\quad P\left(x_k-1,\gamma_k^u,\gamma_k^d,1,1\right)\mu_k^u\left(1-p_k^\alpha\right)+P\left(x_k+1,\gamma_k^u,\gamma_k^d,1,1\right)\mu_k^d\left(1-p_k^\beta\right)+ \\
&\quad P\left(x_k,\gamma_k^u-1,\gamma_k^d,1,1\right)p^u_{k,\gamma_k^u-1}+P\left(x_k,\gamma_k^u,\gamma_k^d-1,1,1\right)p^d_{k,\gamma_k^d-1}+P\left(x_k,\gamma_k^u,\gamma_k^d,0,1\right)r_k^\alpha+ \\
&\quad P\left(x_k,\gamma_k^u,\gamma_k^d,1,0\right)r_k^\beta
\end{aligned} \tag{3.14}$$

The left side of Eqn (3.14) represents the transitions out of state $\left(x_k,\gamma_k^u,\gamma_k^d,1,1\right)$, including:

- M_k^u completes processing a part and becomes "pseudo down" with the transition rate $\mu_k^u p_k^\alpha$; M_k^u completes processing a part and remains not "pseudo down" with the transition rate $\mu_k^u\left(1-p_k^\alpha\right)$.

- M_k^u completes processing a part and becomes "pseudo down" with the transition rate $\mu_k^d p_k^\beta$; it may also remain not "pseudo down" with the transition rate $\mu_k^d\left(1-p_k^\beta\right)$.
- M_k^u and M_k^d deteriorates with the transition rates $p_{k,\gamma_k^u}^u$ and $p_{k,\gamma_k^d}^d$ respectively.
- Preventive maintenance is triggered on M_k^u and M_k^d with the transition rates $\pi_{k,\gamma_k^u}^u$ and $\pi_{k,\gamma_k^d}^d$ respectively.

The right side of Eqn (3.14) represents the transitions into state $\left(x_k,\gamma_k^u,\gamma_k^d,1,1\right)$, and these involve: part arrival, part departure, deterioration of M_k^u and M_k^d, and recovery of M_k^u and M_k^d from "pseudo down".

2) Lower boundary state (where $x_k = 0$).

For the lower boundary state, M_k^d is starved and idle. Therefore, it will not deteriorate. The balance equation for the lower boundary state is:

$$P\left(0,\gamma_k^u,\gamma_k^d,1,1\right)\cdot\left(\mu_k^u p_k^\alpha+\mu_k^u\left(1-p_k^\alpha\right)+p_{k,\gamma_k^u}^u+\pi_{k,\gamma_k^u}^u+\pi_{k,\gamma_k^d}^d\right)=P\left(1,\gamma_k^u,\gamma_k^d,1,1\right)\mu_k^d\left(1-p_k^\beta\right)+$$
$$P\left(0,\gamma_k^u-1,\gamma_k^d,1,1\right)p_{k,\gamma_k^u-1}^u+P\left(0,\gamma_k^u,\gamma_k^d,0,1\right)r_k^\alpha+P\left(0,\gamma_k^u,\gamma_k^d,1,0\right)r_k^\beta \qquad (3.15)$$

3) Upper boundary state (where $x_j = X_j$).

For the upper boundary state, M_k^u is blocked and idle, and hence it will not deteriorate. The balance equation of the upper boundary state can be formulated as:

$$P\left(X_k,\gamma_k^u,\gamma_k^d,1,1\right)\cdot\left(\mu_k^d p_k^\beta+\mu_k^d\left(1-p_k^\beta\right)+p_{k,\gamma_k^d}^d+\pi_{k,\gamma_k^u}^u+\pi_{k,\gamma_k^d}^d\right)=$$
$$P\left(X_k-1,\gamma_k^u,\gamma_k^d,1,1\right)\mu_k^u\left(1-p_k^\alpha\right)+P\left(X_k,\gamma_k^u,\gamma_k^d-1,1,1\right)p_{k,\gamma_k^d-1}^d+P\left(X_k,\gamma_k^u,\gamma_k^d,0,1\right)r_k^\alpha+$$
$$P\left(X_k,\gamma_k^u,\gamma_k^d,1,0\right)r_k^\beta \qquad (3.16)$$

p_k^α , r_k^α , p_k^β , and r_k^β are parameters of a primitive line segment. An algorithm to calculate these parameters is provided in Appendix B. With these parameters, we may derive the balance equations for solving the limiting probabilities of states, $P\left(x_k, \gamma_k^u, \gamma_k^d, \alpha_k, \beta_k\right)$ for each primitive line segment.

3.3.3. Performance Measures

Predicting the performance measures is one of the major objectives of this research. With the limiting probabilities of states $P\left(x_k, \gamma_k^u, \gamma_k^d, \alpha_k, \beta_k\right)$ of each primitive line segment, we are able to compute the following performance measures.

- Production rate (*PR*).

  Production rate of an assembly line is calculated based on the limiting probabilities of the $(K-1)^{\text{th}}$ primitive line segment. The probability that M_{K-1}^d is operational, not 'pseudo down', and not starved is: $\sum_{\alpha_{K-1}} \sum_{\gamma_{K-1}^d \le N_{K-1}^d} \sum_{\gamma_{K-1}^u} \sum_{x_{K-1} \ge 1} P\left(x_{K-1}, \gamma_{K-1}^u, \gamma_{K-1}^d, \alpha_{K-1}, 1\right)$. Under these conditions, M_{K-1}^d processes at the speed of μ_{K-1}^d. Hence, the production rate is:

$$PR = \sum_{\alpha_{K-1}} \sum_{\gamma_{K-1}^d \le N_{K-1}^d} \sum_{\gamma_{K-1}^u} \sum_{x_{K-1} \ge 1} P\left(x_{K-1}, \gamma_{K-1}^u, \gamma_{K-1}^d, \alpha_{K-1}, 1\right) \mu_{K-1}^d \tag{3.17}$$

- Work-In-Process (*WIP*).

  We may obtain the *WIP* in a buffer as:

$$WIP_k^B = \sum_{\beta_k} \sum_{\alpha_k} \sum_{\gamma_k^d} \sum_{\gamma_k^u} \sum_{x_k} \left(x_k \cdot P\left(x_k, \gamma_k^u, \gamma_k^d, \alpha_k, \beta_k\right)\right),\ 1 \le k \le K-1 \tag{3.18}$$

  Since part rejection is not considered in this chapter, machines in the multistage systems have the same flow rates, which are equal to production rate (*PR*) of the system. The *WIP* in a machine is equal to the probability that a machine is busy, which may be estimated by:

$$WIP_k^M = Prob\left(M_k \text{ is busy}\right)$$

$$= \frac{PR}{\mu_k} \tag{3.19}$$

Therefore, the *WIP* of the system is:

$$WIP = \sum_{k=1}^{K-1} WIP_k^B + \sum_{k=1}^{K} WIP_k^M \tag{3.20}$$

- Probability of machine failure (MF_k).

  The probability of machine failure for M_k is $MF_k = Prob\left(\gamma_k = N_k + 1\right)$, and this can be estimated as:

$$MF_k = \sum_{\beta_k}\sum_{\alpha_k}\sum_{\gamma_k^d}\sum_{\gamma_k^u = N_k^u + 1}\sum_{x_k} P\left(x_k, \gamma_k^u, \gamma_k^d, \alpha_k, \beta_k\right), \text{ for } k \le K-1 \tag{3.21}$$

$$MF_K = \sum_{\beta_{K-1}}\sum_{\alpha_{K-1}}\sum_{\gamma_{K-1}^d = N_{K-1}^d + 1}\sum_{\gamma_{K-1}^u}\sum_{x_{K-1}} P\left(x_{K-1}, \gamma_{K-1}^u, \gamma_{K-1}^d, \alpha_{K-1}, \beta_{K-1}\right), \text{ for } M_K \tag{3.22}$$

- Probability of preventive maintenance (PM_k).

  Similarly, PM_k (i.e. $Prob\left(\gamma_k = N_k + 2\right)$) can be calculated as:

$$PM_k = \sum_{\beta_k}\sum_{\alpha_k}\sum_{\gamma_k^d}\sum_{\gamma_k^u = N_k^u + 2}\sum_{x_k} P\left(x_k, \gamma_k^u, \gamma_k^d, \alpha_k, \beta_k\right), \quad \text{for } k \le K-1 \tag{3.23}$$

$$PM_K = \sum_{\beta_{K-1}}\sum_{\alpha_{K-1}}\sum_{\gamma_{K-1}^d = N_{K-1}^d + 2}\sum_{\gamma_{K-1}^u}\sum_{x_{K-1}} P\left(x_{K-1}, \gamma_{K-1}^u, \gamma_{K-1}^d, \alpha_{K-1}, \beta_{K-1}\right), \text{ for } M_K \tag{3.24}$$

- Availability of machine (A_k).

  The availability of a machine ($A_k = Prob\left(\gamma_k \le N_k\right)$) can be estimated by:

$$A_k = 1 - MF_k - PM_k, \quad \text{for } k \in \{1, 2, \cdots, K\} \tag{3.25}$$

3.4. An Application of the Model: Determining the Frequency of Preventive Maintenance for Improving Production Rate

In this section, the author formulates the determination of maintenance rate as an optimization problem to maximize production rate. With an appropriate maintenance rate for each machine obtained by solving this problem, we may subsequently calculate the frequency of preventive maintenance using Eqn (3.4). For simplicity, we shall assume the maintenance rate of each upstate is identical, i.e. $\pi_{k,1} = \pi_{k,2} = \cdots = \pi_{k,N_k} = \pi_k$.

Performance Enhancement Problem 3-1:
Preventive Maintenance for Production Rate Improvement.

Maximize: $$PR = \sum_{\alpha_{K-1}} \sum_{\gamma_{K-1}^{d} \le N_{K-1}^{d}} \sum_{\gamma_{K-1}^{u}} \sum_{x_{K-1} \ge 1} P\left(x_{K-1}, \gamma_{K-1}^{u}, \gamma_{K-1}^{d}, \alpha_{K-1}, 1\right) \mu_{K-1}^{d} \qquad (3.26)$$

Subject to: $0 \le \pi_k < \mu_k$, $k \in \{1, 2, \cdots, K\}$

Where $P\left(x_{K-1}, \gamma_{K-1}^{u}, \gamma_{K-1}^{d}, \alpha_{K-1}^{u}, 1\right)$ is calculated based on the decomposition model. Generally, the preventive maintenance rate π_k is smaller than the processing rate μ_k (the inter-maintenance times are usually many orders larger than the processing times of machines in practice). Hence, we use μ_k as an upper bound for π_k.

There are many possible methods for solving the problem above. In this chapter, we use the steepest ascent method as an example to illustrate the determination of π_k. This method identifies the search direction via the sensitivity analysis of the objective function (i.e. the production rate) with respect to each decision variable (i.e. π_k). Let $\Pi = \left[\pi_1, \pi_2, \cdots, \pi_K\right]$. Define

$\Delta\pi_k$ as a small increment of π_k. In addition, we also define a vector $\Delta\Pi_k = [0,\cdots 0, \Delta\pi_k, 0,\cdots,0]$, where all the elements are 0 except the k^{th} element is equal to $\Delta\pi_k$. Then, the derivative of production rate with respect to π_k may be approximately calculated as:

$$\frac{dPR(\Pi)}{d\pi_k} = \frac{PR(\Pi+\Delta\Pi_k)-PR(\Pi)}{\Delta\pi_k},\ k\in\{1,2,\cdots,K\} \tag{3.27}$$

where $PR(\Pi)$ is the production rate corresponding to Π. At Π, the search direction with the maximum gradient may be calculated as:

$$D=\left[\frac{PR(\Pi+\Delta\Pi_1)-PR(\Pi)}{\Delta\pi_1},\frac{PR(\Pi+\Delta\Pi_2)-PR(\Pi)}{\Delta\pi_2},\cdots,\frac{PR(\Pi+\Delta\Pi_K)-PR(\Pi)}{\Delta\pi_K}\right] \tag{3.28}$$

where D is a $1\times K$ vector representing the search direction. In the direction of D, a single-dimensional optimization approach, such as the golden section search (Kiefer, 1953), is applied to identify the next Π, which induces a maximum increment of $PR(\Pi)$. This procedure is repeated until no better Π can be found. The algorithm can be summarized as:

Algorithm for Calculating Frequency of Preventive Maintenance (and Maintenance Rate)

- Initialize the iteration number, i =0. Let Π^i denote the vector of maintenance rates obtained in the i^{th} iteration of the algorithm. Choose an initial value for this vector (e.g. $\Pi^0 = [0,0,\cdots,0]$). Choose a small value ε as the tolerance limit of the algorithm.

- **Loop**

  Calculate D^i from Eqn (3.28). Perform a golden section search (Kiefer, 1953) in the direction of D^i to maximize production rate $PR\left(\Pi^i + J \cdot D^i\right)$, where J is a nonnegative step size that maximizes production rate.

- If $PR\left(\Pi^i + J \cdot D^i\right) - PR\left(\Pi^i\right) > \varepsilon$, $\Pi^{i+1} = \Pi^i + J \cdot D^i$, $i = i+1$, go to Loop. Otherwise, calculate the frequency of preventive maintenance, Fq_k based on Eqn (3.4), and terminate the algorithm.

Optimization Problem for Maximizing Profit

The above optimization problem uses production rate as the objective function. Improving production rate may not necessarily be the only managerial goal in many manufacturing systems. Reducing operational costs associated with machine repair, preventive maintenance, and holding WIP may be other additional considerations. The decomposition model also makes it possible to assess the cost structure of a multistage manufacturing system with unreliable machines. This may further improve the planning of preventive maintenance. In a multistage manufacturing system, preventive maintenance may give rise to the following aspects:

1) Revenue. Revenue of a manufacturing system can be estimated by multiplying production rate and unit price (Gershwin and Schor, 2000). Since the unit price of a product is usually a constant value in many manufacturing systems, revenue is proportional to the production rate. Therefore, the influence of preventive maintenance on production rate is also reflected in revenue.

2) Preventive maintenance cost. Performing preventive maintenance on a machine requires spending man labor as well as other resources, and this induces additional operational costs. Since excessive preventive maintenance cost may also undermine the profitability of a system, this cost is considered as an important factor in the decision making of preventive maintenance (Ambani, et al., 2009).
3) Machine repair cost. Similar to preventive maintenance, repairing a machine from complete failure also increases the operational costs. For this reason, many studies incorporate machine repair cost in the evaluation of the preventive maintenance scheme (Meller and Kim, 1996; Ambani, et al., 2009).
4) WIP holding cost. The interruption to production of a machine, either due to preventive maintenance or machine failure, may affect the WIP level in the buffers (Meller and Kim, 1996). Therefore, to examine the impact of preventive maintenance comprehensively, the WIP holding cost should also be accounted for. This cost is usually estimated by multiplying the average WIP level and the unit holding cost (Gershwin and Schor, 2000).

Revenue, preventive maintenance cost, machine repair cost, and WIP holding cost of a system can be estimated based on the decomposition model. With these estimates, we may then estimate the expected profit (*EP*) of the system, and use it as the objective function to determine the maintenance rate. This requires the following information:

$Price$: The unit price of a finished part.

$Cost_k^{PM}$: The cost of preventive maintenance on machine M_k.

$Cost_k^{Repair}$: The cost for repairing machine M_k.

$Cost^{WIP}$: *WIP* holding cost per part.

With the above information, we have the optimization problem as follows. This problem may also be solved using the above algorithm by replacing $PR(\Pi)$ with the new objective function, $EP(\Pi)$.

Performance Enhancement Problem 3-2:

Preventive Maintenance for Profit Improvement.

Maximize:

$$EP = revenue - preventive\ maintenance\ cost - machine\ repair\ cost - WIP\ holding\ cost$$

$$= PR \cdot Price - \sum_{k=1}^{K} \left(Cost_k^{PM} \cdot PM_k \right) - \sum_{k=1}^{K} \left(Cost_k^{Repair} \cdot MF_k \right) - Cost^{WIP} \cdot WIP \tag{3.29}$$

Subject to: $0 \le \pi_k < \mu_k$, $k \in \{1, 2, \cdots, K\}$

Where the objective function Eqn (3.29) includes the following four factors and their relationship is as reflected in Figure 3.7.

- $PR \cdot Price$ represents the expected revenue per unit time.
- $\sum_{k=1}^{K} \left(Cost_k^{PM} \cdot PM_k \right)$ is the preventive maintenance cost of the system. PM_k is the probability of preventive maintenance and is calculated using Eqn (3.23) or (3.24).
- $\sum_{k=1}^{K} \left(Cost_k^{Repair} \cdot MF_k \right)$ is the repair cost of the system due to machine failures. MF_k is the probability of machine failure and it is calculated using Eqn (3.21) or (3.22).
- $Cost^{WIP} \cdot WIP$ is the *WIP* holding cost.

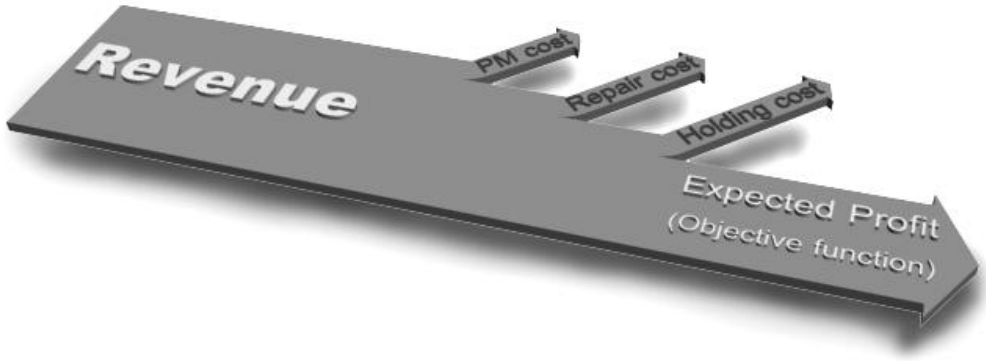


Figure 3.7. The relationship between expected profit, revenue, and cost factors.

3.5. Model Validation

In this section, we shall first discuss the convergence of the decomposition algorithm. Subsequently, the decomposition model is evaluated based on systems with homogeneous machines. In Section 3.5.3, the accuracy of the decomposition model in systems with non-homogeneous machines is examined.

3.5.1. Convergence of the Decomposition Algorithm

Due to the complexity of the model, a mathematical proof of convergence of the decomposition algorithm presented in Appendix B is not possible. However, the author has used the algorithm for a variety of problems (in excess of 5000 cases), in which the parameters and numbers of machines are randomly generated. In all of these experiments, the algorithm has always converged.

3.5.2. Model Validation in Systems with Homogeneous Machines

To validate the decomposition model, the author conducted a large number of experiments with different numbers of machines and parameter values, as well as different serial and assembly line configurations. Comparisons between analytical and simulation results demonstrate that the model can estimate performance measures of production lines with good accuracy.

In this subsection, three cases of experiments will be discussed. The configurations of these production lines are illustrated in Figure 3.8. Although the model was developed for non-homogeneous parameters, for reasons of simplicity, we shall assume that the machine parameters in this set of experiments are homogeneous. In these experiments, the unit of time is an hour. The parameters for the validation experiments are chosen as follows:

- $X_k = 5$, μ_k =1, N_k =4, r_k =0.1, ρ_k =1, for $k \in \{1, 2, \cdots, K\}$.
- p_{k,n_k} =0.05, for $n_k \in \{1, 2, \cdots, N_k\}$, $k \in \{1, 2, \cdots, K\}$.

For each of the three production line configurations, three conditions with different maintenance rates will be tested:

1) No maintenance (NM): π_{k,n_k} =0, for $n_k \in \{1, 2, \cdots, N_k\}$, $k \in \{1, 2, \cdots, K\}$.

2) Medium maintenance (MM): π_{k,n_k} =0.05, for $n_k \in \{1, 2, \cdots, N_k\}$, $k \in \{1, 2, \cdots, K\}$.

3) High maintenance (HM): π_{k,n_k} =0.25, for $n_k \in \{1, 2, \cdots, N_k\}$, $k \in \{1, 2, \cdots, K\}$.

Both analytical and simulation results are obtained on a Personal Computer with Intel Core 2 Duo CPU (2.33GHz) and 4Gb RAM (this computer is also used to perform other numerical experiments in this thesis). Each simulation was run for 1 million time units with a warm up period of 0.1 million time units (the same settings of simulation will also be used in the experiments of model validation in the following chapters). For each experiment, 10 simulation runs are performed, which assure that the 95% confidence intervals of production rate and *WIP* are less than 0.2% for each estimate, as indicated in Table 3.2. The numerical results include production rate, *WIP*, and CPU time, which are provided in Table 3.2. Based on the CPU times of Table 3.2, we find that the decomposition model requires only a fraction of the time required by simulation. For all cases, the decomposition model provides the results in less than one second. On the other hand, it takes more than two minutes to complete one simulation run. To obtain the performance measures with significant confidence, ten such runs are required. Based on the results of production rate and *WIP* in Table 3.2, the decomposition model yields results that are very similar to the results obtained with simulation. The relative difference of production rate and *WIP* between decomposition and simulation is generally smaller than 4%.

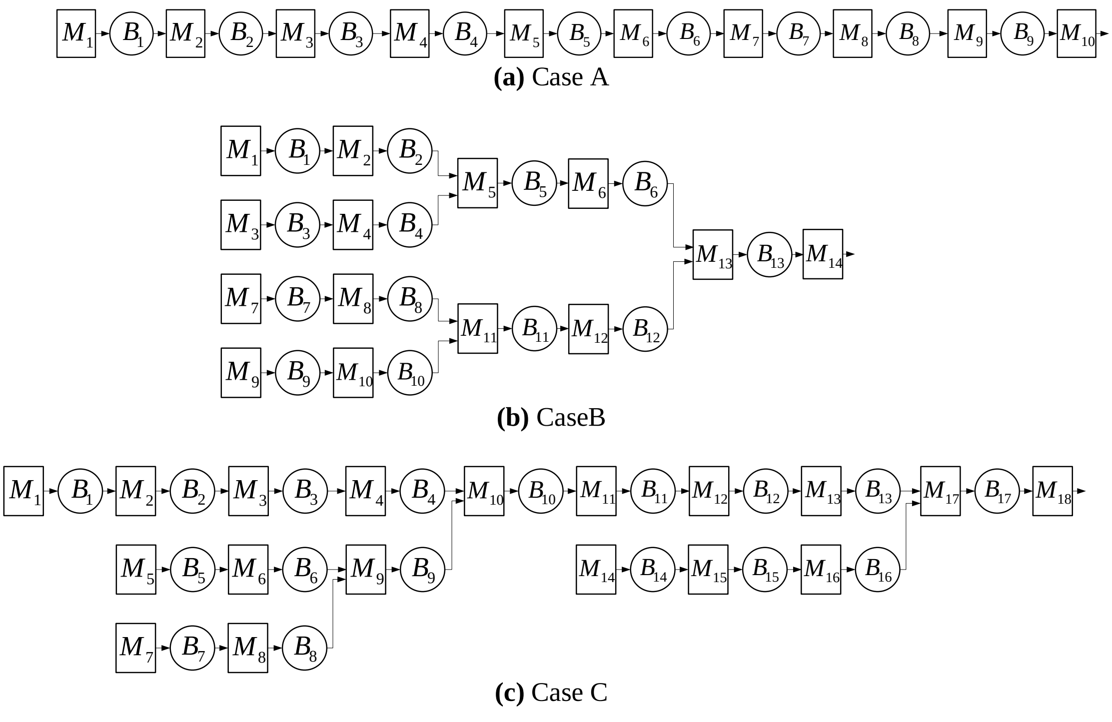


Figure 3.8. The production lines in the experiments.

Table 3.2. Comparison of performance measures and CPU times between decomposition model (Dec) and simulation (Sim) for homogeneous systems. Diff=100%·|Dec-Sim|/Sim. The CPU time of simulation is the average time required for finishing one simulation run for each experiment.

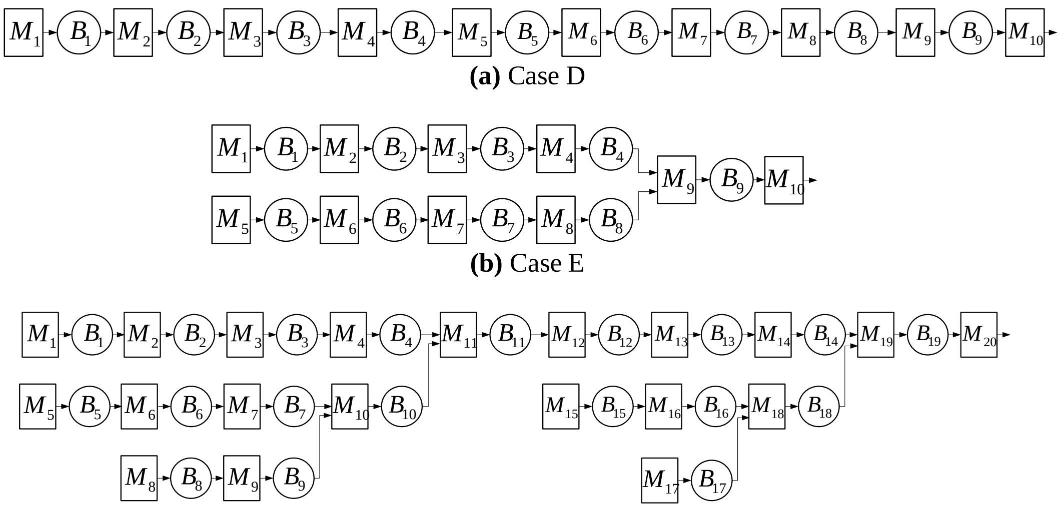
Case	Production rate (parts/hour)			*WIP* (parts)			CPU time (seconds)	
	Dec	Sim	Diff (%)	Dec	Sim	Diff (%)	Dec	Sim (per run)
A (NM)	0.51922	0.50668 ± 0.00047	2.4749	25.594	25.634 ± 0.022	0.1560	0.3527	124.5
A (MM)	0.64695	0.64305 ± 0.00056	0.6065	25.974	26.309 ± 0.025	1.2733	0.3643	131.4
A (HM)	0.52993	0.53863 ± 0.00028	1.6152	25.419	26.273 ± 0.038	2.4892	0.3674	137.5
B (NM)	0.48173	0.46802 ± 0.00069	2.9294	44.6974	46.075 ± 0.022	2.9899	0.4647	163.7
B (MM)	0.62492	0.62122 ± 0.00029	0.5956	44.1798	45.023 ± 0.013	1.8728	0.4432	164.6
B (HM)	0.51243	0.52251 ± 0.00012	1.9291	43.572	44.508 ± 0.024	2.1029	0.4753	165.5
C (NM)	0.47612	0.45172 ± 0.00057	2.5237	56.312	58.134 ± 0.035	2.7901	0.5542	192.7
C (MM)	0.61943	0.61169 ± 0.00024	1.2653	55.898	57.347 ± 0.034	2.5267	0.5753	194.3
C (HM)	0.50311	0.51732 ± 0.00019	2.7468	54.952	56.767 ± 0.037	3.1972	0.5822	201.5

For each production line configuration, based on the comparison between the production rates of the system under no, medium, and high maintenance conditions, we observe that the production line exhibits relatively low production rate when no preventive maintenance is performed. In addition, excessive maintenance also leads to lower production rate. This observation reinforces our hypothesis that determining a suitable maintenance rate (or equivalently the frequency of preventive maintenance) is important for a multistage manufacturing system to improve production rate. This will be further elaborated in the experiments of the next subsection.

3.5.3. Model Validation in Systems with Non-homogeneous Machines

In this subsection, non-homogeneous manufacturing systems, where machines have distinct processing rates and deterioration rates, are investigated. The configurations of the manufacturing systems studied in this subsection are as illustrated in Figure 3.9. The parameters of these three systems are chosen as follows:

- $X_k = 5$, N_k =4, r_k =0.1, ρ_k =1, for $k \in \{1, 2, \cdots, K\}$.
- The processing rates and deterioration rates of machines in the three experiments are summarized in Tables 3.3, 3.4, and 3.5 respectively.



(a) Case D

(b) Case E

(c) Case F

Figure 3.9. The production lines in the experiments.

Table 3.3. Processing rate and deteriorate rate of each machine in Case D.

Machine Number	1	2	3	4	5	6	7	8	9	10
Processing rate	1	1.1	0.9	0.9	1.2	1.1	1	0.9	0.9	1
Deterioration rate	0.05	0.02	0.01	0.03	0.05	0.01	0.04	0.02	0.05	0.04

Table 3.4. Processing rate and deteriorate rate of each machine in Case E.

Machine Number	1	2	3	4	5	6	7	8	9	10
Processing rate	0.9	1	1.1	0.9	1.2	0.8	1.1	1.1	1.2	0.9
Deterioration rate	0.04	0.01	0.03	0.02	0.01	0.05	0.02	0.01	0.02	0.01

Table 3.5. Processing rate and deteriorate rate of each machine in Case F.

Machine Number	1	2	3	4	5	6	7	8	9	10
Processing rate	1	0.9	1	0.9	1.1	0.9	1.2	1	1.1	1
Deterioration rate	0.02	0.05	0.02	0.05	0.03	0.02	0.03	0.04	0.01	0.02
Machine Number	11	12	13	14	15	16	17	18	19	20
Processing rate	1	1.2	0.9	1.1	1.1	1	0.9	1.1	1.2	1
Deterioration rate	0.02	0.01	0.02	0.01	0.02	0.03	0.01	0.02	0.05	0.04

As in Section 3.5.2, the analytical results obtained from the decomposition model are compared with simulation results. Ten simulation runs are performed for each experiment, and each simulation was run for 1 million time units with a warm up period of 0.1 million time units. As indicated in Table 3.6, the relative difference between the analytical and simulation results is generally lower than 4%. This observation demonstrates that the decomposition model can provide reliable estimates of performance measures for multistage manufacturing systems with non-homogeneous machines.

Table 3.6. Comparison of performance measures and CPU times between decomposition model (Dec) and simulation (Sim) for non-homogeneous systems. Diff=100%·|Dec-Sim|/Sim. The CPU time of simulation is the average time required for finishing one simulation run for each experiment.

Case	Production rate (parts/hour)			*WIP* (parts)			CPU time (seconds)	
	Dec	Sim	Diff (%)	Dec	Sim	Diff (%)	Dec	Sim (per run)
D (NM)	0.57745	0.56351 ± 0.00032	2.474	26.683	26.605 ± 0.018	0.293	0.343	130.7
D (MM)	0.64654	0.66416 ± 0.00053	2.652	27.121	26.549 ± 0.021	2.155	0.356	135.7
D (HM)	0.51907	0.53696 ± 0.00034	3.331	26.333	26.300 ± 0.018	0.123	0.372	137.9
E (NM)	0.61356	0.5990 ± 0.00045	2.430	28.783	29.369 ± 0.019	1.992	0.367	135.2
E (MM)	0.66120	0.66905 ± 0.00034	1.173	28.829	29.574 ± 0.018	2.516	0.378	138.4
E (HM)	0.52957	0.535629 ± 0.00025	1.131	28.972	29.674 ± 0.013	2.365	0.399	140.2
F (NM)	0.52990	0.5245 ± 0.00040	1.029	64.938	66.174 ± 0.032	1.868	0.623	201.3
F (MM)	0.63483	0.64077 ± 0.00039	0.927	62.983	65.021 ± 0.037	3.135	0.644	204.5
F (HM)	0.50624	0.51371 ± 0.00033	1.454	62.894	64.711 ± 0.039	2.808	0.673	207.8

3.6. A Case Study for Determining the Maintenance Rate of Each Machine in an Assembly Line

In this set of experiments, the decomposition model is used to determine the frequency of preventive maintenance for each machine in a multistage manufacturing system. An assembly line with nine machines (Figure 3.10) will be studied in this set of experiments. We shall consider the case where machines have different deterioration rates. M_3 and M_7 are intentionally assigned with higher deterioration rates. Machines with higher deterioration rates may require more frequent preventive maintenance, and this will be examined later based on the numerical results.

For simplicity, parameters of the machines are chosen to be homogeneous (except for the deterioration rates). The decision variables of the experiments are the individual maintenance rates. Similar experiments may also be conducted when the parameters are non-homogenous. In this experiment, the unit of time is an hour. Parameters used in this set of experiments are as follows:

- $X_k = 5$, μ_k =1, N_k =4, r_k =0.05, ρ_k =1, for $k \in \{1, 2, \cdots, K\}$.
- p_{1,n_1} =0.005, for $n_1 \in \{1, 2, \cdots, N_1\}$; p_{2,n_2} =0.01, for $n_2 \in \{1, 2, \cdots, N_2\}$;
  p_{3,n_3} =0.02, for $n_3 \in \{1, 2, \cdots, N_3\}$; p_{4,n_4} =0.01, for $n_4 \in \{1, 2, \cdots, N_4\}$;
  p_{5,n_5} =0.005, for $n_5 \in \{1, 2, \cdots, N_5\}$; p_{6,n_6} =0.01, for $n_6 \in \{1, 2, \cdots, N_6\}$;
  p_{7,n_7} =0.02, for $n_7 \in \{1, 2, \cdots, N_7\}$; p_{8,n_8} =0.005, for $n_8 \in \{1, 2, \cdots, N_8\}$;
  p_{9,n_9} =0.01, for $n_9 \in \{1, 2, \cdots, N_9\}$.

In addition, we shall assume that the maintenance rate at each upstate of a machine is identical, i.e. $\pi_{k,1} = \pi_{k,2} = \cdots = \pi_{k,N_k} = \pi_k$ for $k \in \{1, 2, \cdots, K\}$.

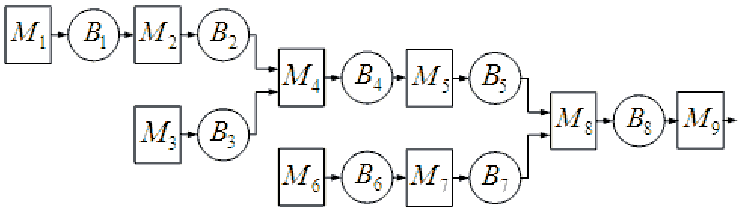


Figure 3.10. The assembly line studied in the experiment.

Using the algorithm presented in Section 3.4, we obtain the maintenance rate (π_k) for each individual machine in the assembly line, as presented in Table 3.7. With the maintenance rate, the frequency of preventive maintenance (Fq_k) for each machine is then calculated using Eqn (3.4). Fq_k represents the number of preventive maintenance performed on machine M_k in an hour. For example, $Fq_1 = 0.00891$ in Case H implies that preventive maintenance is performed 0.00891 times per hour (or once every 112.2 hours). As M_3 and M_7 have higher deterioration rates, they may require more preventive maintenance than other machines. This is also reflected in the results of Table 3.7, where Fq_3 and Fq_7 are higher than that of the other machines.

The production rate of the system under the maintenance rate predicted by the algorithm in Section 3.4 is provided in Table 3.7. For comparison, we also provide the production rate of the system where no preventive maintenance is performed. In this subsection, we shall refer to the experiment without preventive maintenance as Case G, and the experiment with the predicted maintenance rate as Case H. Based on results presented in Table 3.7, the production rate of Case H (0.64392) is 15.96% higher than that of Case G (0.55531). This difference indicates that preventive maintenance, if performed appropriately, may result in a substantial improvement of production rate.

One of the reasons for the production rate enhancement in Case H is that preventive maintenance may improve the availability of each machine. As illustrated in Figure 3.11, each machine has a relatively high probability of failure in Case G. While in Case H, the total probability of interruption (including machine failure and preventive maintenance) of each machine is reduced by approximately 40% when compared to Case G.

A second reason of the production rate enhancement is that in Case H, the probability of a machine being down ($Prob(Down)$, which can be estimated using Eqns (3.21) and (3.22)) is much lower than the probability of being under preventive maintenance ($Prob(PM)$, which is calculated using Eqns (3.23) and (3.24)). In practice, repairing a machine from complete failure usually requires much more time than preventive maintenance (Bao and Jaishankar, 2008). Therefore, preventive maintenance may cause short interruptions to processing while machine failures result in longer interruptions. In an assembly line, a small number of parts are held in each buffer, which may sustain the production of the downstream system of a machine for a short period of time when the processing of this machine is interrupted. This feature effectively mitigates the impact of short interruptions on production rate of the system. Machine failures, on the other hand, usually require much more time before a machine can resume processing, and this may significantly undermine the production rate. Therefore, reducing the proportion of long interruptions is also a factor for achieving high production rate.

As shown in Figure 3.11, M_3 and M_7 are the most unreliable machines in the system, as they have relatively higher probability of failure than other machines. Hence, preventive maintenance on these two machines is more important than that on other machines. To provide an intuitive description of the influence of preventive maintenance on the production rate, the maintenance rate of these two machines are varied from 0 to 0.1, and a surface of production rate vs. π_3 and π_7 is generated, as illustrated in Figure 3.12. The maintenance rate of the other machines (i.e. π_k where $k \notin \{3,7\}$) are chosen as for Case H in Table 3.7. From Figure 3.12, we may observe that the point representing Case H is close to the peak point of the surface. This agreement indicates that the algorithm in Section 3.4 may provide a good solution for the maintenance rate of machines in an assembly line that maximizes production rate.

Table 3.7. Numerical results in the experiment for determining the frequency of preventive maintenance. The maintenance rate, frequency of preventive maintenance of each machine, and production rate of the system in Cases H (preventive maintenance is performed according to the maintenance rate provided by the algorithm in Section 3.4) and G (no preventive maintenance).

Case	Maintenance rate									Production rate
	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	
H	0.01084	0.02045	0.04329	0.01906	0.01007	0.02156	0.04076	0.009845	0.02294	0.64392
G	0	0	0	0	0	0	0	0	0	0.55531
Case	**Frequency of preventive maintenance**									
	Fq_1	Fq_2	Fq_3	Fq_4	Fq_5	Fq_6	Fq_7	Fq_8	Fq_9	
H	0.00891	0.01451	0.02320	0.01379	0.00838	0.01506	0.02245	0.00822	0.01572	
G	0	0	0	0	0	0	0	0	0	

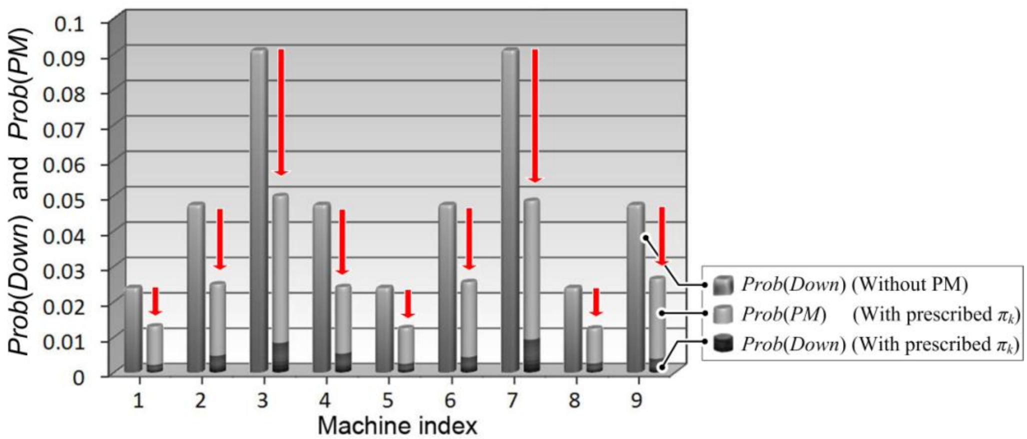


Figure 3.11. The probability of machine failure (*Prob*(*Down*)) of Case G, and the probability of machine failure (*Prob*(*Down*)) and preventive maintenance (*Prob*(*PM*)) of Case H.

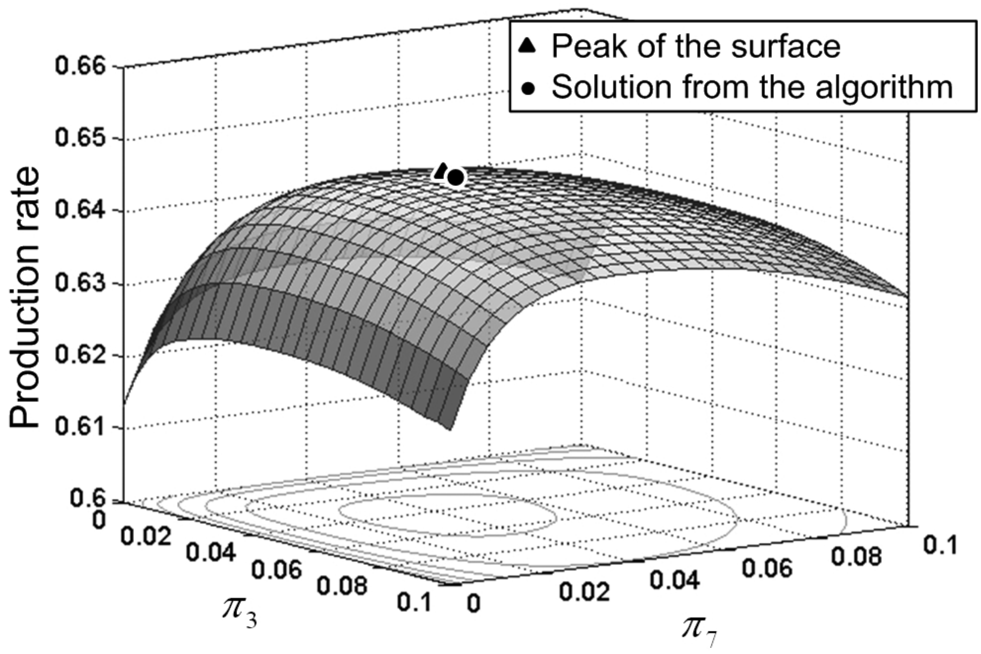


Figure 3.12. Production rate vs. π_3 and π_7.

3.7. Impact of Costs and Buffer Sizes on Preventive Maintenance: A Numerical Study

Improving production rate may not necessarily be the only managerial goal in many manufacturing systems. Other additional considerations may include reducing operational costs associated with machine repair, preventive maintenance, and holding WIP. In addition to performing preventive maintenance, other operational strategies such as configuring buffer sizes, may also influence the response of a system towards its goal of maximizing profit. In this regards, we use the decomposition model to formulate the following optimization problem that will simultaneously determine effective sizes of all buffers and maintenance rates of all machines:

Performance Enhancement Problem 3-3:

Preventive Maintenance and Buffer Allocation for Profit Improvement.

Maximize:

$$EP = revenue - preventive\ maintenance\ cost - machine\ repair\ cost - WIP\ holding\ cost$$

$$= PR \cdot Price - \sum_{k=1}^{K}\left(Cost_k^{PM} \cdot PM_k\right) - \sum_{k=1}^{K}\left(Cost_k^{Repair} \cdot MF_k\right) - Cost^{WIP} \cdot WIP \tag{3.30}$$

Subject to: $0 \le \pi_k < \mu_k$, $k \in \{1, 2, \cdots, K\}$

$X_k > 0$, $k \in \{1, ..., K-1\}$

The algorithm in Section 3.4 can be easily modified to solve this problem, and this requires the following extensions:

1) The original objective function (Eqn (3.26)) is replaced with profit (Eqn(3.30)).
2) Buffer sizes, X_k, $k \in \{1, ..., K-1\}$, are additional decision variables.

We use the assembly line discussed in Section 3.6 as an example to demonstrate the application of the decomposition model in simultaneously determining maintenance rates and buffer sizes. This experiment is referred to as Case I. The parameters are chosen as in Case H. Additionally, the following parameters are also used:

$Price$: 1000$/part

$Cost^{WIP}$: 5$/hour

$Cost_k^{Repair}$: Repair cost for each machine is 2000$/hour

$Cost_k^{PM}$: Preventive maintenance cost for each machine is 200$/hour

Using the algorithm in Section 3.4, we obtain the maintenance rates and buffer sizes for both Cases I and H. The results are summarized in Table 3.8.

The profit and various costs of the assembly line for both experiments are itemized in Figure 3.13. We may observe the following:

1) Profit of Case I is higher than that of Case H.
2) Revenue of Case I is significantly higher than that of Case H. This is because buffers are configured better in Case I. Results in Table 3.8 demonstrate that more space is provided for buffers after machines with low reliability (e.g. machines M_3 and M_7). This buffer configuration contributes to the improvement of production rate with a slight increase in holding cost.
3) The sum of repair and preventive maintenance costs in Case I is smaller than that of Case H. In cases where repairing a machine from complete failure is costly, line managers may perform frequent preventive maintenance to reduce the repair cost. Although performing maintenance more frequently may decrease the availability of machines for production, the reduction in repair cost may compensate for this loss in production.

Table 3.8. The maintenance rates of machines and buffer sizes obtained from Case I.

Maintenance rate	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9
	0.01612	0.03098	0.06693	0.02952	0.01558	0.03598	0.06826	0.01500	0.03253
Buffer size	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	
	3	4	6	5	4	4	6	6	

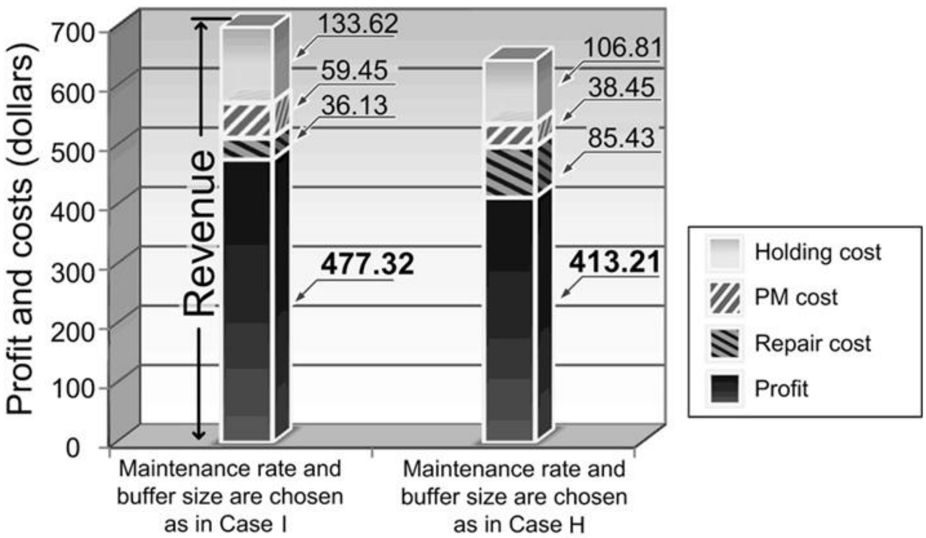


Figure 3.13. Profit and costs of the assembly line under two conditions: 1) maintenance rate and buffer size are chosen as in Case I; 2) maintenance rate and buffer size are chosen as in Case H.

3.8. Numerical Comparison of the Decomposition and Single-Machine Models

Much of the literature pertaining to preventive maintenance focuses on single-machine models. These models are usually proposed to minimize the sum of repair and preventive maintenance costs for a single machine (Montoro-Cazorla and Perez-Ocon, 2006). However, in a multi-machine system, merely minimizing these two costs is inadequate, because other economic measures, such as WIP holding cost and revenue also affect profit. WIP holding cost and revenue cannot be calculated using the single-machine model. By contrast, they can be easily estimated using the decomposition model. With these estimates, we can examine the impact of preventive maintenance more comprehensively and therefore determine maintenance rates more effectively.

In this subsection, we use the assembly line discussed in Section 3.7 to further elaborate the advantage of the decomposition model. In Section 3.7, we had identified maintenance rates for all machines in the assembly line based on the decomposition model. For comparison, we use the single-machine model proposed by Bao and Jaishankar (2008) to estimate the repair and preventive maintenance costs of the machines in Case I. We can then estimate the maintenance rate of each machine by minimizing the sum of repair and preventive maintenance costs. Maintenance rates obtained using the decomposition and single-machine models are provided in Table 3.9. These maintenance rates can then be used to estimate the various performance measures summarized in Table 3.9. In this analysis, buffer sizes are chosen as in Section 3.7.

The results in Table 3.9 indicate that the sum of repair and preventive maintenance costs of the assembly line is 94.32 for the experiment using the single-machine model. The corresponding value for the assembly line where maintenance rates are chosen using the decomposition model is 95.58. The latter is 1.34% higher than the former. However, if we compare the profit of these two cases, we notice that profit of the assembly line using the decomposition model (477.32) is 1.27% higher than the case using the single-machine model (471.34). This is because the decomposition model incorporates the WIP holding cost and revenue in determining maintenance rates, and this results in better overall performance.

Table 3.9. Maintenance rates and performance measures determined using the decomposition and single-machine models.

Maintenance rate	π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9
Decomposition	0.01612	0.03098	0.06693	0.02952	0.01558	0.03598	0.06826	0.01500	0.03253
Single-machine	0.01828	0.03672	0.07461	0.03661	0.01820	0.03675	0.07479	0.01818	0.03664

Performance measure	Repair cost	PM cost	Holding cost	Revenue	Profit
Decomposition	36.13	59.45	133.62	706.52	477.32
Single-machine	27.59	66.73	134.41	700.07	471.34

3.9. Analyzing CPU time and Accuracy of the Decomposition Model

In this subsection, we perform a set of experiments to investigate the impact of system size and number of upstates on CPU time and accuracy of the decomposition model. We consider a tandem production line, and the number of machines is varied from 2 to 20 in steps of 2. Additionally, four cases are studied, where the upstates of machines are 1, 2, 4, and 8 respectively. The other parameters are chosen as follows:

$$X_k = 5, \ \ \mu_k = 1, \ \ r_k = 0.1, \ \ \rho_k = 1, \ \text{ for } k \in \{1, ..., K\}$$

$$p_{k,n_k} = 0.01 \ \text{ for } \ \ n_k \in \{1, ..., N_k\} \ \ \text{ and } \ \ k \in \{1, ..., K\}$$

In each experiment, we develop a decomposition model for estimating the performance measures of the system. The CPU time of the decomposition model is summarized in Figure 3.14. Simulation is also performed for comparison. Each simulation was run for 1 million time units, including a warm up period of 0.1 million time units. For each experiment, ten simulation runs were performed to guarantee that the 95% confidence intervals of the simulation results (such as production rate, etc) are less than 0.2% of the point estimates. The CPU time of simulation is plotted in Figure 3.15. Additionally, we also compare the production rates obtained using the decomposition model and simulation. The absolute relative difference (i.e. 100%·|Dec-Sim|/Sim, where Dec and Sim denote the production rate obtained from the decomposition model and simulation respectively) is plotted in Figure 3.16. From Figures 3.14, 3.15, and 3.16, we make the following observations:

1) The comparison between the CPU time of the decomposition model (Figure 3.14) and simulation (Figure 3.15) demonstrates that the decomposition model is superior in terms of computational speed. For instance, in the experiment where the number of machines is 20 and number of upstates is 8, the CPU time of the decomposition model is 1.18 seconds. By contrast, it requires 279.12 seconds to finish one simulation run for the same configuration. Based on Figures 3.14 and 3.15, we observe that, in each experiment, the CPU time of the decomposition model is less than 0.5% of the CPU time required for performing one simulation run. This advantage in computational speed makes the decomposition model a time efficient alternative to simulation in the performance analysis of multistage manufacturing systems.

2) As illustrated in Figure 3.14, a large-scale system may require more CPU time than a small system. In the decomposition model, a system with K machines is decomposed into K-1 primitive line segments. Therefore, a decomposition model with more line segments requires more computational effort. For example, in the case with 8 upstates, the CPU times of the decomposition models of the systems with 10 and 20 machines are 0.56 and 1.18 seconds respectively. The latter is approximately twice of the former. Likewise, the CPU time of simulation also increases with the number of the machines, as illustrated in Figure 3.15.
3) Figure 3.14 demonstrates that the CPU time of the decomposition model is influenced by the number of upstates. This is because the number of balance equations for each primitive line segment is affected by the number of upstates. The system with a large number of upstates requires solving more balanced equations, and this may increase the CPU time. Comparing Figure 3.15 with Figure 3.14, we may observe that the influence of the number of upstates on the CPU time of simulation is less significant.
4) The absolute relative difference between the production rate obtained from the decomposition model and simulation is lower than 3%, as reflected in Figure 3.16. This indicates that the decomposition model is of reasonable accuracy.
5) Figure 3.16 indicates that the difference increases as the system size increases. Machines in a manufacturing system are subjected to various random events, such as deterioration, machine repair, etc. The more machines a system has, the more random events may occur. The decomposition model has slightly lower accuracy if the system is

structurally more complex and subjected to more random events. Additionally, as the number of upstates increases, the system becomes more complex, and this also results in a slight increase of the difference.

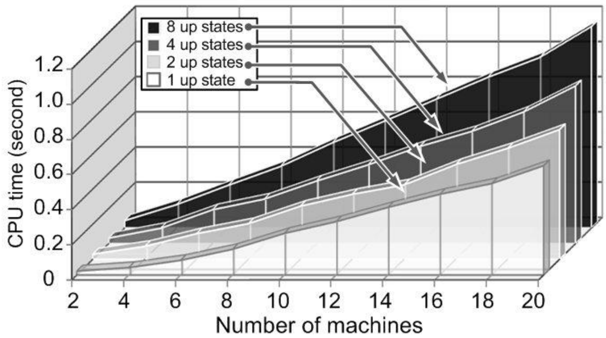


Figure 3.14. CPU time of the decomposition model vs. the number of machines and number of upstates.

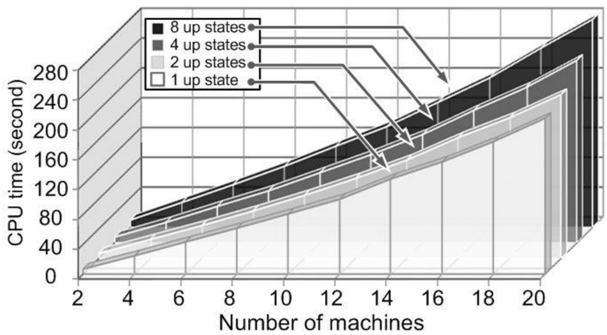


Figure 3.15. CPU time of the simulation per run vs. the number of machines and number of upstates. Ten runs are performed for each experiment.

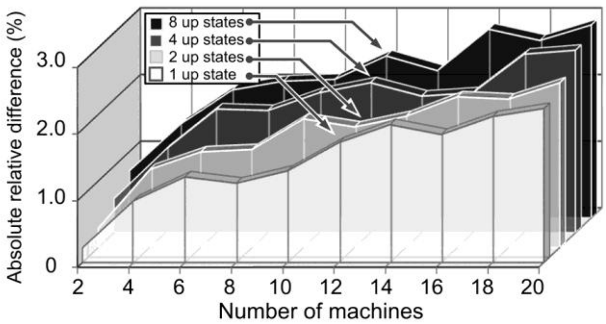


Figure 3.16. Absolute relative difference between the decomposition model and simulation vs. the number of machines and number of upstates.

3.10. Extension of the Model for Incorporating Machine State Inspection

The analytical model discussed in Section 3.4 is based on the assumption that upstates of the machines are unobservable (i.e. the maintenance operator is incapable of identifying the current upstate of a machine). In some manufacturing systems, machine state inspection is implemented for identifying the upstate of a machine (Moustafa et al., 2004). Once the inspection indicates a machine has deteriorated to a specific level, preventive maintenance is triggered immediately; otherwise, the state of a machine is not altered after the inspection. To incorporate the machine state inspection, two additional notations are defined:

λ_k: Inspection rate of M_k when it is operational (i.e. $\gamma_k \le N_k$).

TH_k: The threshold to perform preventive maintenance on M_k. Preventive maintenance may be triggered only when $\gamma_k \ge TH_k$.

The transition diagram of Figure 3.3 is extended to incorporate machine state inspection, as illustrated in Figure 3.17. Based on the discussion in Section 3.3.1, the balance equations can also be derived for the 2M1B line with machine state inspection and preventive maintenance. By solving these balance equations, the limiting probabilities of states of such a system are obtained.

The decomposition model presented in Section 3.4 can be easily extended to the multistage manufacturing system with machine state inspection. Similarly, we may estimate the performance measures of such a system, which can then be used to determine the inspection rate λ_k and the threshold of preventive maintenance TH_k for each machine. For example, the determination of λ_k and TH_k, $k \in \{1, 2, \cdots, K\}$ may be formulated as the optimization problem to maximize expected profit (*EP*) of the system. We may derive the optimization problem as below:

Performance Enhancement Problem 3-4:
Machine State Inspection and Preventive Maintenance for Profit Improvement.

Maximize:

$$
\begin{aligned}
EP &= revenue - preventive\ maintenance\ cost - machine\ repair\ cost - \\
&\quad WIP\ holding\ cost - machine\ state\ inspection\ cost \\
&= PR \cdot Price - \sum_{k=1}^{K}\left(Cost_k^{PM} \cdot PM_k\right) - \sum_{k=1}^{K}\left(Cost_k^{Repair} \cdot MF_k\right) \\
&\quad - Cost^{WIP} \cdot WIP - \sum_{k=1}^{K}\left(Cost_k^{Inspection} \cdot Prob\left(\gamma_k \le N_k\right) \cdot \lambda_k\right)
\end{aligned}
\tag{3.31}
$$

Subject to: $0 \le \lambda_k < \mu_k$, $k \in \{1, 2, \cdots, K\}$

$1 \le TH_k < N_k$, $k \in \{1, 2, \cdots, K\}$

where $Price$ denotes the unit price of a finished part. $Cost_k^{PM}$ is the cost of preventive maintenance on machine M_k. $Cost_k^{Repair}$ is the cost for repairing machine M_k. $Cost^{WIP}$ denotes the WIP holding cost per part. $Cost_k^{Inspection}$ is the operation cost per inspection on machine M_k, and $Prob\left(\gamma_k \le N_k\right) \cdot \lambda_k$ is the frequency that inspection is performed on machine M_k (if M_k is operational, i.e. $\gamma_k \le N_k$, inspection is triggered with the transition rate λ_k).

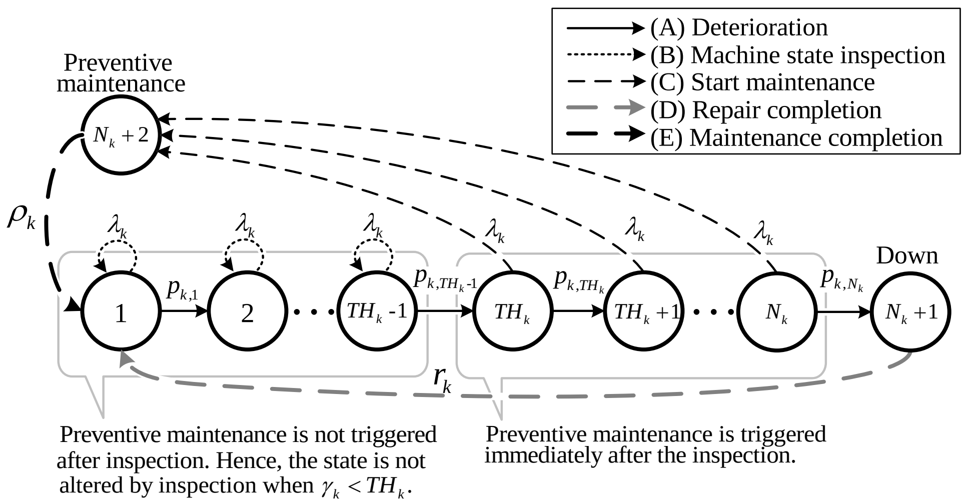


Figure 3.17. Machine state inspection and preventive maintenance.

Numerical Example

In this subsection, a numerical example is used to demonstrate the influence of TH_k and inspection rate λ_k on the performance of an assembly line. The system discussed in Section 3.6 is used in this experiment, and we assume that the upstates of machine can be detected via machine state inspection in this experiment. The parameters are chosen as in Section 3.6 with the following exception:

- Unit price: 1000 dollars/part
- Repair cost: 2000 dollars/hour (for each machine)
- Preventive maintenance cost: 1000 dollars/hour (for each machine)
- Machine state inspection cost: 100 dollars/inspection (for each machine)
- *WIP* holding cost per part: 1 dollar/(part·hour)

In this example, we shall choose TH_k identical for each machine to simplify the problem. The algorithm presented in Section 3.4 is used to determine the optimal inspection rate λ_k, $k \in \{1, 2, \cdots, K\}$ for each machine with different thresholds of preventive maintenance, TH_k. As the number of upstate of each machine, N_k is chosen to be 4 in this example, TH_k have four possible values, i.e. TH_k=1, 2, 3, or 4. The numerical results are provided in Table 3.10. Based on these results, we observe that when TH_k=3 and the inspection rates are chosen as in Table 3.10, profit of the system is the highest. In addition, the inspection rate obtained from the algorithm is also depicted in Figure 3.18. From this figure, we notice that for machines with higher deterioration rate (e.g. M_3 and M_7), inspection should also be performed more frequently.

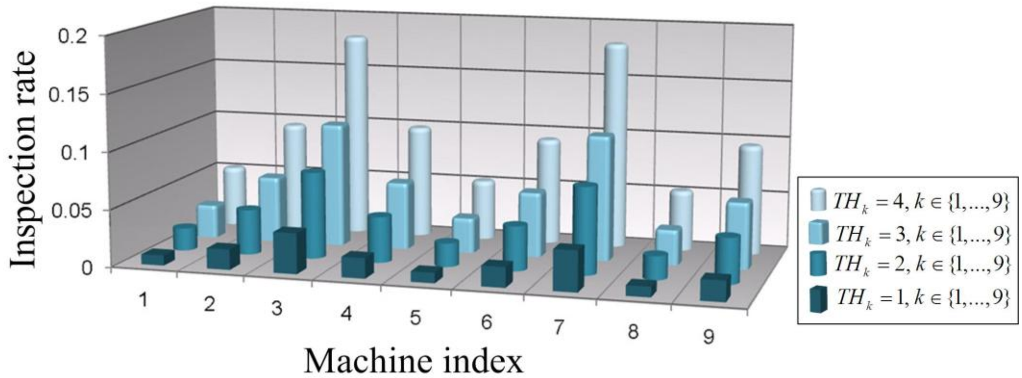


Figure 3.18. The inspection rate (λ_k) of each machine under four conditions, viz. TH_k =1, TH_k =2, TH_k =3, and TH_k =4, $k \in \{1, 2, \cdots, K\}$.

Table 3.10. The inspection rate of each machine and profit of the system under different TH_k. For simplicity, we choose $TH_1 = TH_2 = \ldots = TH_K$ in this example. The algorithm presented in Section 3.4 is used to determine the inspection rate for each machine with different maintenance threshold, TH_k.

TH_k	Inspection rate									Revenue ($/hour)	Repair cost ($/hour)	PM cost ($/hour)	Inspection cost ($/hour)	Holding cost ($/hour)	Expected profit ($/hour)
	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9						
1	0.0089	0.0177	0.0354	0.0177	0.0089	0.0177	0.0353	0.0089	0.0179	643.62	107.96	167.24	16.72	20.90	330.79
2	0.0198	0.0396	0.0760	0.0402	0.0207	0.0391	0.0765	0.0206	0.0398	664.93	22.39	75.11	38.78	21.55	507.10
3	0.0292	0.0579	0.1091	0.0593	0.0308	0.0568	0.1097	0.0308	0.0581	666.90	34.04	42.78	58.04	21.70	**510.34**
4	0.0533	0.0962	0.0962	0.0992	0.0529	0.0936	0.1822	0.0533	0.0964	659.50	102.20	27.91	102.68	21.55	405.17

Chapter 4.

Performance Enhancement of Multistage Manufacturing Systems with Imperfect Production

—A Study on Defective Material Flow and Inspection with Errors

4.1. Overview

Improving production quality is a key issue in many manufacturing companies to improve or at least maintain profitability, competitiveness, and market share. In a multistage manufacturing system, each processing machine may generate defective parts randomly. These defective parts often result in a waste of machine capacity, money, labor, etc, which may undermine the profit of a manufacturing system (Chen and Subramaniam, 2010), as reflected in Figure 4.1. Additionally, shipping defective parts to customers also induces costly penalty. In order to analyze the corrupting effects of imperfect production and provide insights for quality management of multistage manufacturing systems, the author formulates an integrated quantitative and qualitative model in Section 4.3. With this model, the author also investigates the placement of inspection machines in multistage manufacturing systems in Section 4.4. By appropriately allocating the inspection machines, profit of a multistage manufacturing system is improved. Numerical experiments of the proposed model are provided and discussed in Sections 4.5 to 4.8.

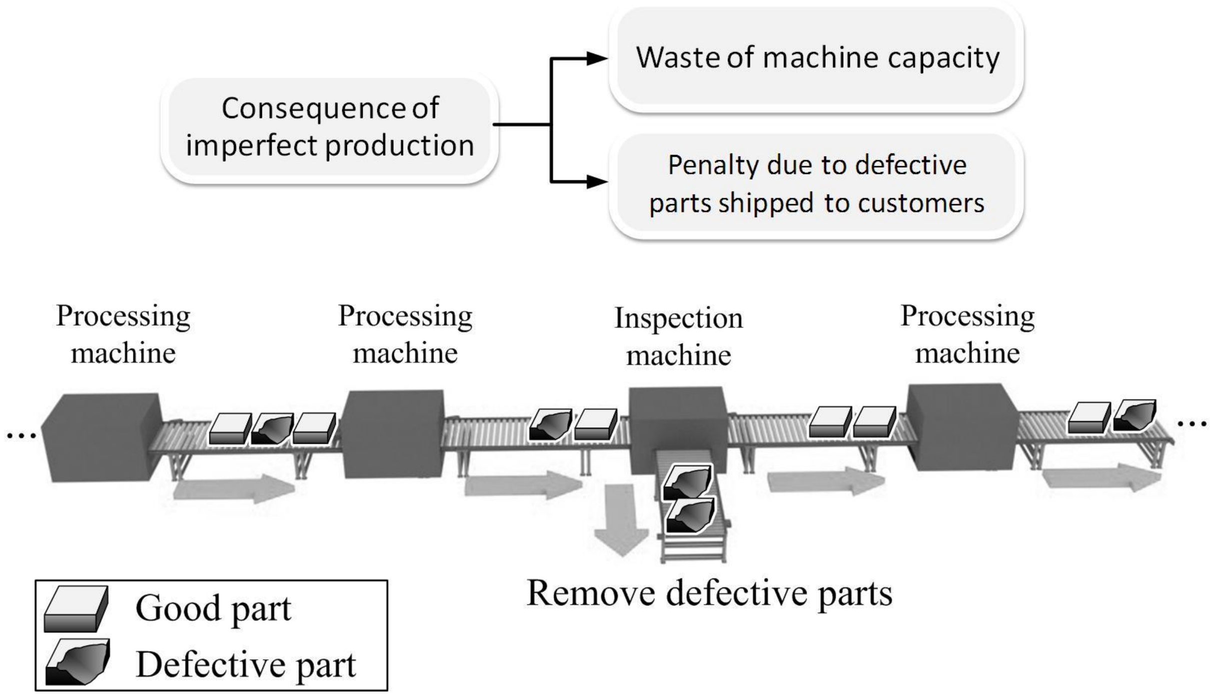


Figure 4.1. An important cause of profit loss: imperfect production. Each machine in a multistage manufacturing system may generate defective parts randomly. Imperfect production not only results in the waste of machine capacity and other resources, but also induces expensive penalty if defective parts are shipped to customers.

4.2. Definition of Notations

In addition to the notations defined in Section 3.2, the following notations will also be used in this chapter.

$U(k)$: A function that returns the immediately upstream machine of M_k if it is a non-assembly machine; and it returns the set of immediately upstream machines if M_k is an assembly machine. For example, $U(3)=2$ and $U(7)=\{3,6\}$ in Figure 4.2.

l_k: Defective rate (i.e. the probability that M_k generates defective parts). $l_k=0$ for inspection machines.

u_k: The probability that type I inspection error (classifying good parts as defective) occurs at M_k, if M_k is an inspection machine.

v_k: The probability that type II inspection error (classifying defective parts as good) occurs at M_k, if M_k is an inspection machine.

θ_k: Outgoing quality of parts of M_k, i.e. the fraction of good parts in the material flow out of M_k.

η_k: Reject rate of M_k. The fraction of the parts flowing into M_k that are rejected if M_k is an inspection machine. $\eta_k = 0$ for processing machines.

y_k: The state of machine M_k, $y_k \in \{0,1\}$. We assume that there are only two states of M_k, viz. up ($y_k = 1$) and down ($y_k = 0$).

p_k: The failure rate of M_k (i.e. the transition rate from state y_k=1 to state y_k=0).

r_k: The repair rate of M_k (i.e. the transition rate from state y_k=0 to state y_k=1).

$Price$: The unit price of a finished part.

$Cost_k^{Op}$: The operation cost per part of machine M_k. $Cost_k^{Op}$ is the processing cost per part if M_k is a processing machine; on the other hand, if M_k is an inspection machine, $Cost_k^{Op}$ represents the inspection cost per part.

$Cost^{WIP}$: *WIP* holding cost per part.

$Penalty$: The penalty cost per defective part shipped to customers.

4.3. Model Development

In this section, the quantitative and qualitative performance of assembly lines with imperfect production is investigated. This investigation may also be applicable for serial production lines. The quality of material flow in such systems will be analyzed first in Section 4.3.1. This analysis is later incorporated in the decomposition model of assembly lines presented in Section 4.3.2. This model decomposes an assembly line into a number of primitive line segments, each of which is characterized as a continuous-time-discrete-state Markov chain. The model is used to estimate a variety of performance measures, as discussed in Section 4.3.3.

The following assumptions are used in this chapter.

- Raw materials are without defects.

  Many previous analytical studies on multistage manufacturing systems with imperfect production also make this assumption for reasons of simplicity (Kim and Gershwin, 2005, 2009; Colledani and Tolio, 2006, 2009; Bai and Yun, 1996; Rau and Chu, 2005; Freiesleben, 2006). This assumption is appropriate in many real manufacturing systems, as most companies have very strict quality requirements for raw materials in order to assure the quality of final products. Raw materials are usually inspected before they are accepted by the manufacturers. Once the inspection results indicate that the raw materials are with a high level of defects, the manufacturers will not accept these raw materials. For this reason, raw materials used in the manufacturing system are usually of good quality, and hence it is reasonable to assume that they are without defects.

- Processing machines (including non-assembly and assembly machines) are subjected to Bernoulli-type quality failures (Montgomery, 2001), i.e. each processing machine generates defective parts randomly with a constant probability.

  Bernoulli-type quality failures are commonly observed in manufacturing systems (Bai and Yun, 1996; Heredia-Langner et al., 2002; Rau and Chu, 2005; Freiesleben, 2006; Van Volsem et al., 2007). This assumption reflects that in many systems, defective parts are generated due to various unpreventable factors associated with the product design and manufacturing environment. For instance, in the wafer fabrication industry, the dust may induce the random generation of defective silicon chips, and this is a typical example of Bernoulli-type quality failures.

- Ubiquitous inspection (i.e. inspecting parts after each processing machine) is costly, and hence a feasible inspection allocation scheme is desirable.

  In the literature, only a few papers pertaining to multistage manufacturing systems are based on the assumption of ubiquitous inspection (Kim and Gershwin, 2005, 2009). Ubiquitous inspection is a restrictive condition, which may not be encountered in practice. Therefore, the majority of researches consider un-ubiquitous inspection, a more general condition in reality (Colledani and Tolio, 2006, 2009; Penn and Raviv, 2007, 2008). This thesis will likewise study systems with un-ubiquitous inspection.

- Parts classified as defective by inspection machines are rejected.

  In manufacturing systems, defective parts are handled in one of the following three ways: reject, repair, or rework. Defective parts are usually rejected when they are impossible (or very expensive) to repair or rework. For instance, in the wafer fabrication industry, silicon chips contaminated by dust are usually abandoned. Since part rejection is one of the most commonly used defective parts handling mechanisms in practice, many researches adopt this assumption (Taneja and Viswanadham, 1994; Viswanadham et al., 1996; Penn and Raviv, 2007, 2008).

- The first machine is never starved of raw materials and the final machine is never blocked. This is a common assumption in many analytical studies of multistage manufacturing systems, and its validity has been discussed in Section 3.3.

- Both processing and inspection machines are subjected to operation-dependent failures (i.e. machines may break down only when it is processing or inspecting). Each machine has two states: “up” and “down”.

  The model presented in this section may be easily extended to incorporate machine deterioration (where a machine has multiple upstates with different levels of deterioration), as discussed in Section 3.3.

- All the machines are single-item machines, i.e. each machine can process or inspect one part each time (refer to Chapter 5 for the incorporation of batch machines).

4.3.1. Quality of Material Flow

Imperfect production, as well as imperfect inspection, has a substantial influence on the quality of material flow in an assembly line. In such a system, both the outgoing quality of parts (θ_k) and reject rate (η_k) of machine M_k can be calculated based on its upstream machines and this may fall into the following four conditions:

1) M_k is the first machine (and therefore a processing machine) in a branch of the assembly line (e.g. M_1 or M_4 in Figure 4.2). Since raw materials are assumed to be without defects, we may calculate θ_k and η_k as below, and this is also reflected in Figure 4.3(a):

   $$\theta_k = 1 - l_k \tag{4.1}$$

   $$\eta_k = 0 \qquad \text{(because } M_k \text{ is a processing machine)} \tag{4.2}$$

2) M_k is not the first machine of any branch, and it is a non-assembly processing machine (e.g. M_3 in Figure 4.2). As illustrated in Figure 4.3(b), θ_k and η_k can be calculated as:

   $$\theta_k = \theta_{U(k)} \cdot \left(1 - l_k\right) \tag{4.3}$$

   $$\eta_k = 0 \tag{4.4}$$

3) M_k is an assembly machine (e.g. M_7 in Figure 4.2). M_k may obtain good parts from its upstream machine M_i, $i \in U\left(k\right)$ with probability θ_i. Hence, the probability that all the parts for an assembly process are conforming is $\prod_{i \in U(k)} \theta_i$. As indicated in Figure 4.3(c),

   $$\theta_k = \left(\prod_{i \in U(k)} \theta_i \right) \cdot \left(1 - l_k\right) \tag{4.5}$$

   $$\eta_k = 0 \tag{4.6}$$

4) M_k is an inspection machine (e.g. M_2 in Figure 4.2). As illustrated in Figure 4.3(d), the parts out of an inspection machine may be divided into four categories:

- Good parts rejected due to type I inspection error (with probability $\theta_{U(k)} \cdot u_k$).
- Good parts sent to the next machine (with probability $\theta_{U(k)} \cdot (1-u_k)$).
- Bad parts accepted due to type II error (with probability $\left(1-\theta_{U(k)}\right) \cdot v_k$).
- Bad parts rejected correctly (with probability $\left(1-\theta_{U(k)}\right) \cdot (1-v_k)$).

Hence,

$$\theta_k = \frac{\theta_{U(k)} \cdot (1-u_k)}{\theta_{U(k)} \cdot (1-u_k) + \left(1-\theta_{U(k)}\right) \cdot v_k} \tag{4.7}$$

$$\eta_k = \theta_{U(k)} \cdot u_k + \left(1-\theta_{U(k)}\right) \cdot (1-v_k) \tag{4.8}$$

Based on Eqns (4.1) to (4.8), θ_k and η_k can be calculated for each machine. These values reflect the quality of material flow in the assembly line, as illustrated in Figure 4.4.

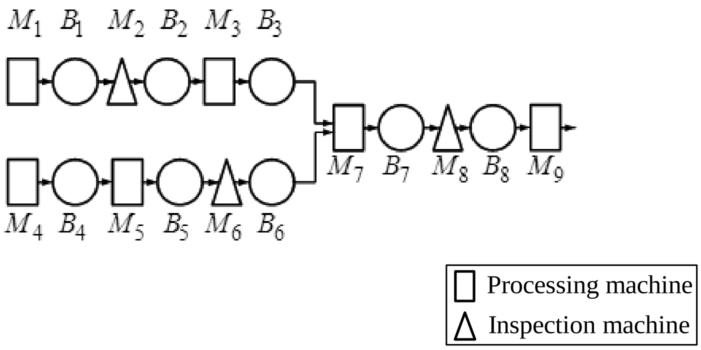


Figure 4.2. An assembly line with inspection machines.

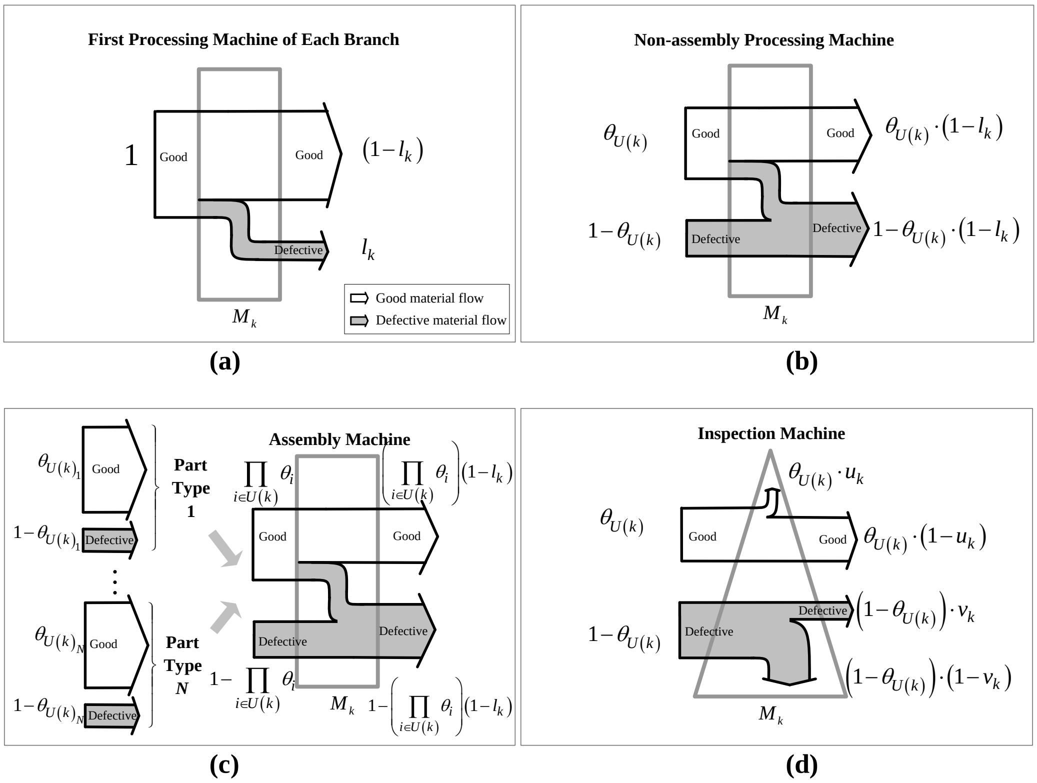


Figure 4.3. Quality of material flow before and after each machine.

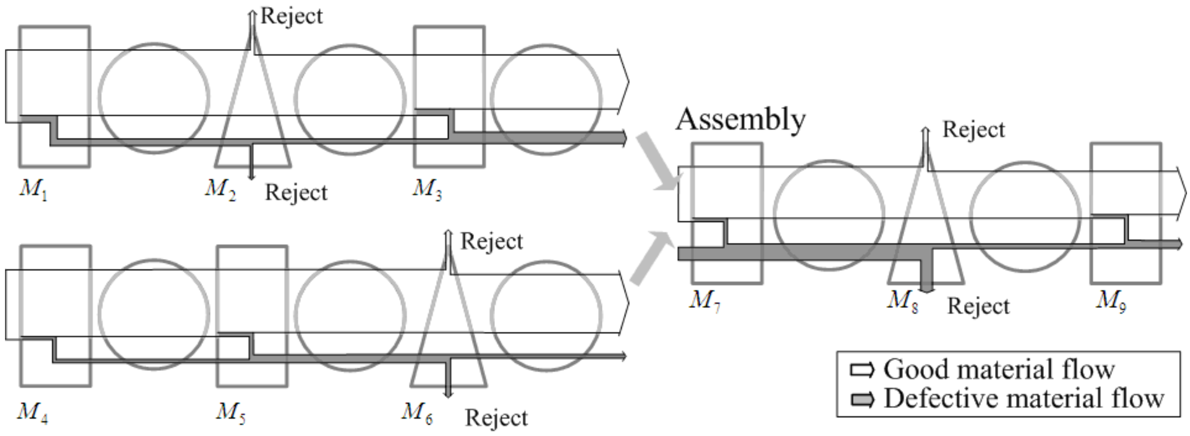


Figure 4.4. Quality of material flow in the assembly line.

4.3.2. Decomposition of Assembly Lines

As in Chapter 3, the author formulates a decomposition model for assembly lines with imperfect production. An important feature of the decomposition model in this chapter is that it incorporates the quality of material flow in the system (which is discussed in Section 4.3.1). This feature facilitates the prediction of system performance from the quantitative (such as total production rate and inventory) and qualitative (such as percentage of good finished parts, wasted machine capacity, and wasted processing cost due to defective parts) perspectives simultaneously. This model decomposes an assembly line with *K* machines into (*K*-1) two-machine-one-buffer (2M1B) primitive line segments, as illustrated in Figure 3.5.

A primitive line segment consists of two machines (the upstream machine M_k^u and downstream machine M_k^d) and a buffer B_k. The machines (M_k^u or M_k^d) in a primitive line segment may be up ($y_k^u = 1$ or $y_k^d = 1$) or down ($y_k^u = 0$ or $y_k^d = 0$). In a primitive line segment, if buffer B_k is full, M_k^u is blocked; and if B_k is empty, M_k^d is starved. The blockage and starvation of machines due to the buffer inside a primitive line segment can be easily captured by developing a Markov model for such a line segment. However, M_k^u and M_k^d may also be starved or blocked by buffers in the upstream or downstream line segments. Therefore, to account for the starvation or blockage due to the upstream or downstream line segments, an additional state, referred to as "pseudo down" is introduced for the machines in the line segment, and this is elaborated with details in Section 3.3.2. Two variables (α_k and β_k) are used to denote whether the upstream and downstream machines in a line segment are "pseudo down" respectively. As discussed in Section 3.3.2, the occurrence and disappearance of "pseudo down" state of machines in a line segment are characterized by four parameters (viz. p_k^α, r_k^α, p_k^β, and r_k^β). The calculation of these four parameters of a primitive line segment is presented in Appendix B. With these parameters, we may derive the balance equations for each primitive line segment, and this will be presented in the next subsection.

4.3.3. Deriving Balance Equations for the Primitive Line Segment

The decomposition model analyzed in this chapter considers part rejection due to inspection. The rejection of parts makes the primitive line segment behave differently from the primitive line segment presented in Chapter 3, where no parts are rejected. Therefore, a new set of balance equations should be derived to characterize state transitions of the primitive line segment with rejected parts, and this is presented in this section.

The state of the k^{th} primitive line segment may be defined as:

$$S_k = (x_k, y_k^u, y_k^d, \alpha_k, \beta_k) \tag{4.9}$$

State transitions of a primitive line segment are described using balance equations. Based on different combinations of y_k^u ($y_k^u = 1 \text{ or } 0$), y_k^d ($y_k^d = 1 \text{ or } 0$), α_j ($\alpha_k = 1 \text{ or } 0$), and β_k ($\beta_k = 1 \text{ or } 0$), the states of a primitive line segment may be roughly categorized into 2^4 or 16 groups. In the following paragraphs, we shall select the group of states with $y_k^u = 1$, $y_k^d = 1$, $\alpha_k = 1$, and $\beta_k = 1$ (i.e. M_k^u and M_k^d are up and not in the "pseudo down" state) as the example to illustrate the derivation of balance equations. The balance equations for the other groups may be obtained similarly. By solving the balance equations and the normalization equation (i.e. the sum of all probabilities equals one), the limiting probabilities of all states, $P\left(x_k, y_k^u, y_k^d, \alpha_k, \beta_k\right)$ are obtained. The balance equations of states in this group ($y_k^u = 1$, $y_k^d = 1$, $\alpha_k = 1$, and $\beta_k = 1$) may be further divided into three sub-categories:

1) Internal state ($0 < x_k < X_k$).

In this case, buffer B_k is neither empty nor full. Based on the transition diagram in Figure 4.5 (a), the balance equation may be formulated as:

$$
\begin{aligned}
P(x_k,1,1,1,1)\cdot\Big(p_k^u + p_k^d + \mu_k^u\eta_k^u p_k^\alpha + \mu_k^u\left(1-\eta_k^u\right)p_k^\alpha + \mu_k^u\left(1-\eta_k^u\right)\left(1-p_k^\alpha\right) + \\
\mu_k^d p_k^\beta + \mu_k^d\left(1-p_k^\beta\right)\Big) = P(x_k,0,1,1,1)r_k^u + P(x_k,1,0,1,1)r_k^d + \\
P(x_k,1,1,0,1)r_k^\alpha + P(x_k,1,1,1,0)r_k^\beta + P(x_k-1,1,1,1,1)\mu_k^u\left(1-\eta_k^u\right)\left(1-p_k^\alpha\right) + \\
P(x_k+1,1,1,1,1)\mu_k^d\left(1-p_k^\beta\right) \qquad (4.10)
\end{aligned}
$$

The left side of Eqn (4.10) represents the transitions out of state $(x_k,1,1,1,1)$, including:

- M_k^u and M_k^d may break down with transition rates of p_k^u and p_k^d respectively.
- M_k^u rejects a part and becomes "pseudo down" with the transition rate of $\mu_k^u\eta_k^u p_k^\alpha$, where μ_k^u is the processing rate of M_k^u, η_k^u is the probability that M_k^u rejects a part, and p_k^α is the probability that M_k^u becomes "pseudo down" after it finishes processing a part.
- M_k^u delivers a part to buffer B_k (i.e. the part is not rejected) and becomes "pseudo down" with the transition rate of $\mu_k^u\left(1-\eta_k^u\right)p_k^\alpha$.
- M_k^u delivers a part to buffer B_k and does not become "pseudo down". The corresponding transition rate is $\mu_k^u\left(1-\eta_k^u\right)\left(1-p_k^\alpha\right)$.

- M_k^d discharges a part and then becomes "pseudo down" with the transition rate of $\mu_k^d p_k^\beta$.
- M_k^d discharges a part and does not become "pseudo down". The rate of this transition is $\mu_k^d \left(1-p_k^\beta\right)$.

The right side of Eqn (4.10) indicates the transitions from other states to state $\left(x_k,1,1,1,1\right)$, including the repair of M_k^u and M_k^d from down states respectively, the recovery of M_k^u and M_k^d from "pseudo down" states respectively, part arrival, and part departure.

2) Lower boundary state (where $x_k = 0$).

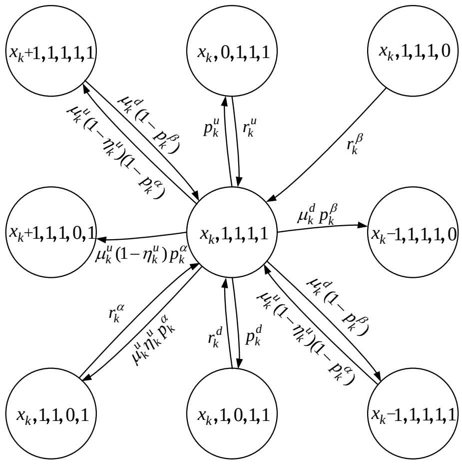
In this case, the buffer is empty, and as illustrated in Figure 4.5 (b), we have:

$$P\left(0,1,1,1,1\right)\cdot\left(p_k^u+\mu_k^u\eta_k^u p_k^\alpha+\mu_k^u\left(1-\eta_k^u\right)p_k^\alpha+\mu_k^u\left(1-\eta_k^u\right)\left(1-p_k^\alpha\right)\right)=$$
$$P\left(0,0,1,1,1\right)r_k^u+P\left(0,1,1,0,1\right)r_k^\alpha+P\left(0,1,1,1,0\right)r_k^\beta+P\left(1,1,1,1,1\right)\mu_k^d\left(1-p_k^\beta\right) \tag{4.11}$$

3) Upper boundary state (where $x_k = X_k$).

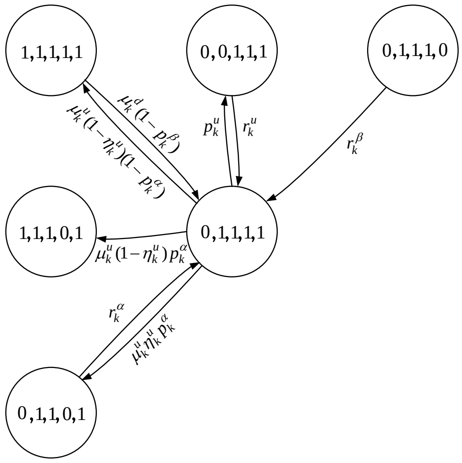
The buffer is full in this case, and the balance equation can be derived based on Figure 4.5 (c) as:

$$P\left(X_k,1,1,1,1\right)\cdot\left(p_k^d+\mu_k^d p_k^\beta+\mu_k^d\left(1-p_k^\beta\right)\right)=P\left(X_k,1,0,1,1\right)r_k^d+P\left(X_k,1,1,0,1\right)r_k^\alpha+$$
$$P\left(X_k,1,1,1,0\right)r_k^\beta+P\left(X_k-1,1,1,1,1\right)\mu_k^u\left(1-\eta_k^u\right)\left(1-p_k^\alpha\right) \tag{4.12}$$



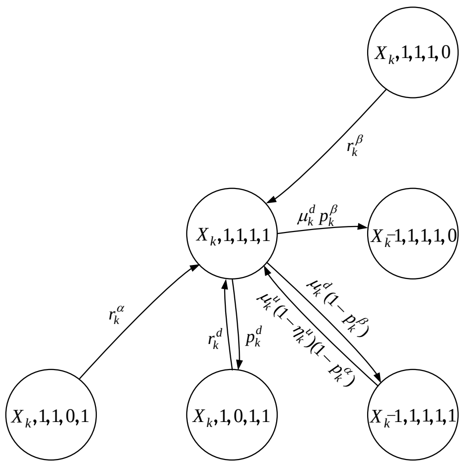
$0 < x_k < X_k$

(a) Internal transition.



$x_k = 0$

(b) Boundary transition when buffer is empty.



$x_k = X_k$

(c) Boundary transition when buffer is full.

Figure 4.5. Transition diagrams for state (x_k,1,1,1,1).

4.3.4. Performance Measures

With the limiting probabilities of the states of each primitive line segment, we may estimate various performance measures of the manufacturing system. For instance, Work-In-Process (*WIP*) may be calculated similarly as discussed in Section 3.3.3. In the remainder of this subsection, the author will highlight the calculation of performance measures relevant to qualitative issues that were not discussed in Chapter 3.

- Flow rate and production rate of the assembly line.

  The flow rate into machine M_k is the average number of parts delivered into M_k in a unit of time. This can be calculated as:

$$FR_k^{In} = \sum_{\beta_k} \sum_{y_k^d} \sum_{x_k < X_k} P\left(x_k, 1, y_k^d, 1, \beta_k\right) \mu_k^u, \quad 1 \le k \le K-1 \tag{4.13}$$

$$FR_K^{In} = \sum_{\alpha_{K-1}} \sum_{y_{K-1}^u} \sum_{x_{K-1} > 1} P\left(x_{K-1}, y_{K-1}^u, 1, \alpha_{K-1}, 1\right) \mu_{K-1}^d \text{, for machine } M_K \tag{4.14}$$

  The flow rate out of M_k and sent to the downstream machine can be calculated based on FR_k^{In} and the reject rate η_k as:

$$FR_k^{Out} = FR_k^{In} \cdot \left(1 - \eta_k\right) \tag{4.15}$$

  For the last machine M_K, FR_K^{Out} is also the total production rate (PR^{Total}) of the assembly line.

$$PR^{Total} = FR_K^{Out} \tag{4.16}$$

  Since the fraction of finished parts that are good is θ_K, the effective (good parts) and defective (defective parts not detected) production rates are:

$$PR^{Effect} = PR^{Total} \cdot \theta_K \tag{4.17}$$

$$PR^{Defect} = PR^{Total} \cdot \left(1 - \theta_K\right) \tag{4.18}$$

- Expected processing and inspection cost.

With FR_k^{In}, the expected processing and inspection cost of the assembly line may be estimated as:

$$EPIC = \sum_{k=1}^{K} \left(FR_k^{In} \cdot Cost_k^{Op} \right) \tag{4.19}$$

- Input quality of M_k.

The input quality of M_k is defined as the fraction of parts sent into M_k that are good.

$$\phi_k = \prod_{i \in U(k)} \theta_i \tag{4.20}$$

- Wasted capacity and wasted processing cost.

For each processing machine, we may calculate the fraction of capacity wasted on processing defective parts. We shall refer to this value as wasted capacity. It can be calculated as:

$$WC_k = \frac{FR_k^{In} \cdot (1 - \phi_k)}{\mu_k} \tag{4.21}$$

where $FR_k^{In} \cdot (1 - \phi_k)$ represents the defective flow rate into machine M_k.

To improve the quantitative performance of a manufacturing system, it is imperative that the machine capacity used to process defective parts is minimized. Hence, a processing machine with high wasted capacity usually implies that inspection is required before this machine.

Similarly, the wasted processing cost can be estimated as:

$$WPC_k = FR_k^{In} \cdot (1 - \phi_k) \cdot Cost_k^{Op} \tag{4.22}$$

Both wasted capacity and wasted processing cost will be used in the algorithm for determining inspection allocation in Section 4.4.

4.4. Inspection Allocation in Assembly Lines

The integrated quantitative and qualitative model in this chapter may be used to determine the placement of inspection machines in an assembly line for maximizing expected profit. In a real manufacturing system, finished parts are generally inspected before being shipped to customers. This practice may result in a reduction of the penalty due to undetected nonconforming parts (Penn and Raviv, 2007). Therefore, we shall assume that an inspection machine is always placed at the end of an assembly line. As illustrated in Figure 4.6, an assembly line with N processing machines may have $(N-1)$ possible locations for placing inspection machines, excluding the inspection machine at the end of the line. These candidate inspection machines may be placed between any two consecutive processing machines. The variable $z_{n,m}$ is used to indicate the placement of an inspection machine between two consecutive processing machines, M_n and M_m. The first subscript of $z_{n,m}$ (i.e. n) represents the upstream processing machine, M_n; and the second subscript of $z_{n,m}$ (i.e. m) represents the downstream machine, M_m. If an inspection machine is placed at this location, $z_{n,m}$=1; otherwise, $z_{n,m}$=0. For instance, if an inspection machine is inserted between processing machines M_3 and M_6 in the assembly line of Figure 4.6, we have $z_{3,6}=1$. In this section, we also assume that an output buffer is added after an inserted inspection machine. For convenience, a $1\times(N-1)$ vector Z is used to represent a solution to the inspection allocation problem. For example, in the assembly line of Figure 4.6, candidate inspection machines may be inserted between M_1 and M_2, M_2 and

M_3, M_3 and M_6, M_4 and M_5, M_5 and M_6, and M_6 and M_7. Hence, for this case,

$$Z = \left[z_{1,2}, z_{2,3}, z_{3,6}, z_{4,5}, z_{5,6}, z_{6,7} \right] \tag{4.23}$$

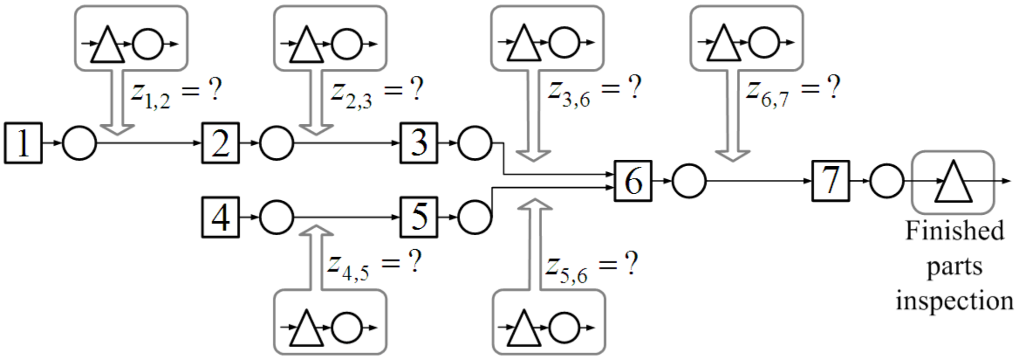


Figure 4.6. Possible locations for placing inspection machines.

The determination of inspection allocation is formulated as an optimization problem to maximize expected profit (*EP*):

Performance Enhancement Problem 4-1:

Inspection Allocation for Enhancing Quantitative and Qualitative Performance

Maximize: $EP = Revenue - Penalty\ Cost - Holding\ Cost -$

$Processing\ and\ Inspection\ Cost$

$$= Price \cdot PR^{Total} - Penalty \cdot PR^{Defect} - Cost^{WIP} \cdot WIP - EPIC \tag{4.24}$$

Subject to: $z_{n,m} = 0 \text{ or } 1$, $\quad z_{n,m} \in Z$

where PR^{Total}, PR^{Defect}, WIP, and $EPIC$, are calculated as discussed in Section 4.3.4. $EP(Z)$ will be used to represent the expected profit when inspection machines are placed as indicated by *Z*.

In this chapter, an algorithm is developed to solve this problem, and it is based on the following considerations:

- An inspection machine is placed before a processing machine that has high wasted capacity (calculated using Eqn (4.21)). This placement may prevent wasting the capacity of machines on processing defective parts, and hence may lead to an improvement in production rate.
- Additionally, an inspection machine is also placed before a processing machine that has high wasted processing cost (calculated using Eqn (4.22)). This may result in a reduction of resources spent on processing defective parts.
- An inspection machine may be unnecessary if it only detects a very small amount of defective parts, or exhibits very few rejections. Hence, inspection machines that have small reject rates are removed from the assembly line.

The proposed algorithm consists of two major procedures: inspection machine insertion and inspection machine removal. In the first procedure, inspection machines are inserted before processing machines with high wasted capacity or wasted processing cost. It should be noted that if such a processing machine is an assembly machine (e.g. M_6 in Figure 4.6), there will be more than one upstream branch. In this case, the inspection machine is inserted in the branch with the worst quality of parts (θ_k, which is calculated as discussed in Section 4.3.1). In the second procedure, inspection machines, which do not have a significant contribution for improving the quality of material flow, are removed. Once an inspection machine is removed, its output buffer is also eliminated. These two procedures are repeated until the profit cannot be further improved.

Inspection Allocation Algorithm

Step 1: Initialization.

Initialize $Z=[0,0,\cdots,0]$ (where Z is defined as Eqn (4.23)), i.e. there are no inspection machine placed between any two consecutive processing machines initially.

Step 2: Inspection machine insertion.

Step 2A: Identify inspection location before processing machines with high wasted capacity.

Step 2A.1: Place the inspection machines as suggested by Z, and calculate the wasted capacity (Eqn (4.21)) for each processing machine. Find the processing machine with the highest wasted capacity (e.g. M_c, where c is the index of such a machine).

Step 2A.2: Identify the location to insert a new inspection machine. This location is between M_c and its immediately upstream processing machine (we shall refer to this upstream processing machine as M_b). Hence, the corresponding location to insert the inspection machine is denoted by variable $z_{b,c}$. If M_c is an assembly machine, M_b is the immediately upstream processing machine of M_c with the worst output quality.

Step 2A.3: If ($z_{b,c}$=0) {($z_{b,c}$=0 indicates that no inspection machine has been placed at this location in the current solution Z. If $z_{b,c}$=1, Step 2A.3 is skipped.)

Generate a new solution (which will be referred to as Z^{New}) for testing. Z^{New} is similar to Z except that a new inspection machine is inserted before the processing machine with the highest wasted capacity (i.e. $z_{b,c}$ =1).

If ($EP\left(Z^{New}\right) > EP\left(Z\right)$) $Z = Z^{New}$. (The current solution is updated with a better solution.)

}

Step 2B: Identify inspection location before processing machines with high wasted processing cost.

Step 2B.1: Place the inspection machines as suggested by *Z*, and calculate the wasted processing cost (Eqn (4.22)) for each processing machine. Find the processing machine with the highest wasted processing cost (e.g. M_p, where *p* is the index of such a machine).

Step 2B.2: Identify the location to insert a new inspection machine. This location is between M_p and its immediately upstream processing machine (we shall refer to this upstream processing machine as M_o). This location is denoted by $z_{o,p}$. If M_p is an assembly machine, M_o is the immediately upstream processing machine of M_p with the worst output quality.

Step 2B.3: If ($z_{o,p}$ =0) {(No inspection machine has been placed at this location in the current solution *Z*. If $z_{o,p}$ =1, Step 2B.3 is skipped.)

Generate a new solution (which will be referred to as Z^{New}) for testing. Z^{New} is similar to Z except that a new inspection machine is inserted before the processing machine with the highest wasted processing cost (i.e. $z_{o,p}$ =1).

If ($EP\left(Z^{New}\right) > EP\left(Z\right)$) $Z = Z^{New}$. (The current solution is updated with a better solution.)

}

Step 3: Inspection machine removal.

Identify and remove unnecessary inspection machine.

Step 3.1: Place the inspection machines as suggested by *Z*, and calculate the reject rate (Eqn (4.8)) for each existing inspection machine. Identify the inspection machine which has the lowest reject rate (suppose such an inspection machine is between processing machines M_r and M_s, and hence denoted by $z_{r,s}$).

Step 3.2: Generate a new solution, Z^{New} for testing. Z^{New} is similar to Z except that the inspection machine with the lowest reject rate is removed (i.e. $z_{r,s}$ =0).

Step 3.3: If ($EP\left(Z^{New}\right) > EP\left(Z\right)$) $Z = Z^{New}$. (Update the current solution.)

Step 4: Termination condition.

If no inspection machine is inserted or removed in the current iteration, terminate; otherwise, go to **Step 2**.

4.5. Model Validation

In this section, the accuracy of the integrated quantitative and qualitative model is validated by comparing the analytical results with the results obtained through simulation. Six assembly lines (referred to as Cases A, B, C, D, E, and F) as illustrated in Figure 4.7 are investigated. In Cases A and C, ubiquitous inspection is implemented. In comparison, non-ubiquitous inspection is performed in Cases B and D. In Cases A to D, we investigate the balanced systems, where machines have identical processing times. The parameters for these four experiments are chosen as follows:

- Mean processing or inspection rate: 1 part/min (for each machine).
- Defective rate: 0.02 (for each processing machine).
- Probability of type I error: 0.01 (for each inspection machine).
- Probability of type II error: 0.01 (for each inspection machine).
- Failure rate: 0.001 min^{-1} (for each machine).
- Repair rate: 0.1 min^{-1} (for each machine).

For each case, the assembly line under two conditions is examined:

1) The size of each buffer is 5;

2) The size of each buffer is 10.

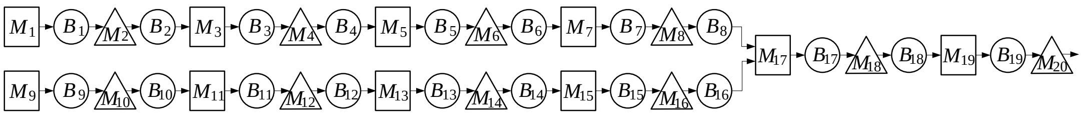
In Cases E and F, unbalanced systems are considered. The processing and inspection rates of machines in these two experiments are summarized in Tables 4.1 and 4.2. The other parameters are chosen as in Cases A to D.

Table 4.1. The processing or inspection rate for machines in Case E.

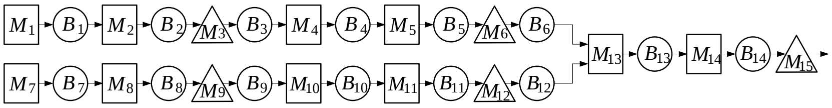
Machine Number	1	2	3	4	5	6	7	8	9	10
Processing/inspection rate	1	0.9	0.9	1.1	0.9	1	1.1	0.95	1	1
Machine Number	11	12	13	14	15	16	17	18	19	20
Processing/inspection rate	1.05	1.1	0.9	1	1.1	1.05	0.95	0.9	1	0.9

Table 4.2. The processing or inspection rate for machines in Case F.

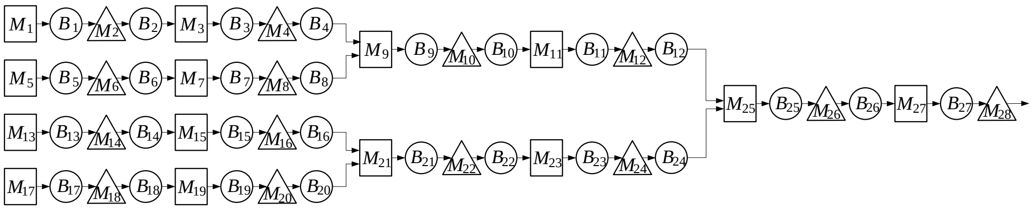
Machine Number	1	2	3	4	5	6	7	8	9	10
Processing/inspection rate	0.9	0.95	1	1.1	0.95	1.1	0.95	1.05	1.1	0.95
Machine Number	11	12	13	14	15	16	17	18	19	20
Processing/inspection rate	1	1	1.1	0.95	1	1.1	1.05	1.05	0.95	0.9
Machine Number	21	22	23	24	25	26	27	28	29	30
Processing/inspection rate	1	0.9	0.95	1	1	0.9	1.05	1	1.1	1



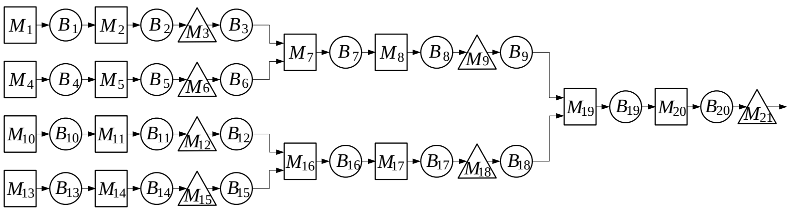
(a) Case A. Ubiquitous inspection (i.e. inspection is performed after each processing machine).



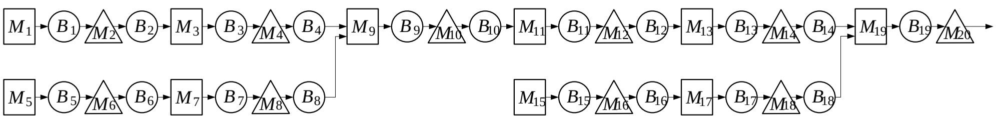
(b) Case B. Non-ubiquitous inspection.



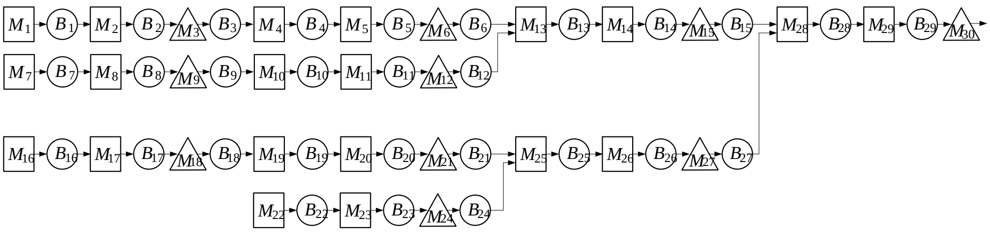
(c) Case C. Ubiquitous inspection.



(d) Case D. Non-ubiquitous inspection



(e) Case E.



(f) Case F.

Figure 4.7. The assembly lines studied in the experiments.

The numerical results obtained from the decomposition model and simulation are provided in Table 4.3. The CPU times in Table 4.3 demonstrate that the integrated quantitative and qualitative model is much more time efficient than simulation. In addition to the CPU time, Table 4.3 also lists the results of three performance measures, viz. the total production rate (including both good and defective parts), defective production rate, and *WIP*. The relative differences between the analytical and simulation results are generally less than 4%. This agreement demonstrates that the proposed model is of reasonable accuracy for both balanced and unbalanced production lines.

Comparing the results of Cases A and B (or Cases C and D), we observe that ubiquitous inspection may result in a low defective production rate. However, placing an inspection machine after each processing machine may reduce total production rate and increase *WIP*, which also undermines profit. Therefore, ubiquitous inspection may not be the most economically feasible solution in an assembly line.

Table 4.3. Comparison of results from the integrated quantitative and qualitative model (IQQ) and simulation (Sim). Diff=100%|IQQ-Sim|/Sim. Each simulation includes ten runs.

Case	Buffer size	Total production rate (parts/min)			Defective production rate (parts/min)			*WIP* (parts)			CPU time (second)	
		IQQ	Sim	Diff (%)	IQQ	Sim	Diff (%)	IQQ	Sim	Diff (%)	IQQ	Sim (per run)
A	5	0.62838	0.62712 ±0.00047	0.20	0.000131	0.000130 ±0.000002	0.46	52.567	54.446 ±0.020	3.45	0.45	195.4
	10	0.72610	0.71990 ±0.00052	0.86	0.000151	0.000157 ±0.000003	0.41	90.112	92.842 ±0.025	2.94	0.65	201.3
B	5	0.63938	0.63915 ±0.00053	0.036	0.000272	0.000271 ±0.000003	0.17	41.576	42.440 ±0.032	2.04	0.34	135.3
	10	0.74183	0.73403 ±0.00058	1.06	0.000315	0.000311 ±0.000003	1.33	71.211	73.982 ±0.057	3.75	0.52	137.9
C	5	0.61253	0.60513 ±0.00057	1.22	0.000127	0.000125 ±0.000002	1.61	81.211	83.017 ±0.017	2.18	0.68	224.3
	10	0.71664	0.70807 ±0.00062	1.21	0.000149	0.000147 ±0.000002	1.22	141.753	145.982 ±0.063	2.89	0.85	230.4
D	5	0.62085	0.62157 ±0.00063	0.12	0.000264	0.000265 ±0.000003	0.46	61.871	62.869 ±0.025	1.59	0.54	187.2
	10	0.72789	0.72004 ±0.00067	1.09	0.000310	0.000307 ±0.000001	0.98	108.368	111.103 ±0.052	2.46	0.72	195.3
E	5	0.60935	0.61395 ±0.00032	0.749	0.0001282	0.000129 ±0.000002	0.620	56.8297	57.6649 ±0.027	1.448	0.50	199.7
	10	0.706249	0.702027 ±0.00042	0.601	0.000149	0.000148 ±0.000002	0.676	104.4051	105.2412 ±0.022	0.794	0.67	205.9
F	5	0.58981	0.58815 ±0.00043	0.282	0.0002508	0.0002491 ±0.000002	0.682	89.2724	91.1624 ±0.031	2.073	0.95	240.2
	10	0.679399	0.671290 ±0.00055	1.208	0.0002889	0.00028428 ±0.000003	1.625	156.8394	160.9044 ±0.052	2.526	0.99	243.3

4.6. Comparison with the Model of Penn and Raviv (2007, 2008)

In the literature, several models have been formulated for evaluating the quantitative and qualitative performance of multistage manufacturing systems with imperfect production (Colledani and Tolio, 2006, 2009; Penn and Raviv 2007, 2008). Colledani and Tolio (2006, 2009) proposed an integrated quantitative and qualitative model for manufacturing system with persistent quality failures. By contrast, the model presented in this thesis is developed for multistage manufacturing systems with Bernoulli-type quality failures. As discussed in Section 2.5, Bernoulli-type quality failure is distinct from persistent quality failure. Manufacturing systems with these two distinct types of quality failures may behave differently. Therefore, it is not useful to compare the model in this thesis with Colledani and Tolio's model.

Penn and Raviv (2007, 2008) formulated an analytical model for serial manufacturing systems, where machines are subjected to Bernoulli-type quality failures (Montgomery, 2001). Compared with Penn and Raviv's model, the proposed model in this thesis has two advantages:

1) The proposed model is applicable for both serial and non-serial manufacturing systems, while Penn and Raviv's model is specifically for serial systems.
2) Inspection error is taken into consideration in the proposed model, which makes it possible to assess the impact of inspection errors on the system performance. By contrast, Penn and Raviv's model assumes that inspection is without error.

In order to further compare the proposed model with Penn and Raviv's model, a numerical experiment will be presented in the remainder of this subsection. In this experiment, we shall consider the condition that demand is finite. This is because Penn and Raviv had formulated their model based on the assumption that the manufacturing system is operating at a constant demand rate and this rate is smaller than the system capacity. The proposed model in this thesis can be easily extended to incorporate finite demand. This can be achieved by approximating the demand as an additional processing machine (Li and Meerkov, 2009), as illustrated in Figure 4.8. With this approximation, the proposed method in Section 4.3 can then be used to model the system with finite demand.

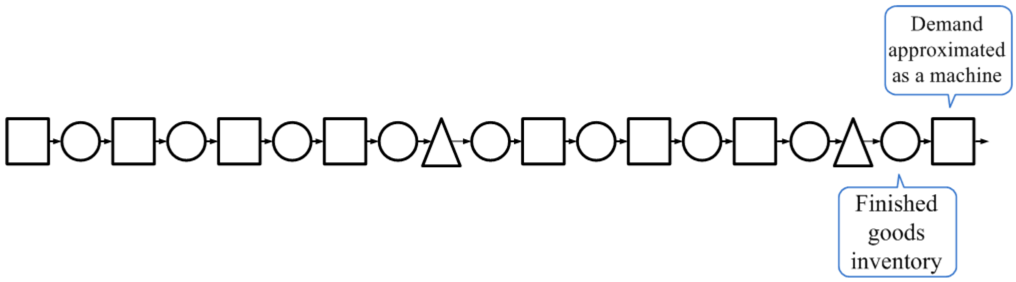


Figure 4.8. Finite demand is approximated as an additional machine.

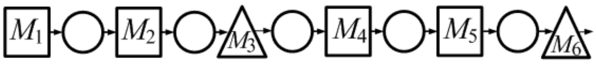
In the numerical study, the proposed model is compared with Penn and Raviv's model using two system configurations illustrated in Figure 4.9 as testbed problems. The parameters of machines in these two systems are summarized in Tables 4.4 and 4.5. Additionally, the sizes of buffers are chosen to be 5. Since Penn and Raviv's model does not consider inspection errors, we also assume that the probabilities of type I and type II inspection errors are both 0 for each inspection machine in the experiment.

Table 4.4. Parameters of the machines in Case G.

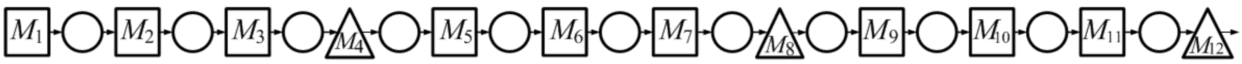
Machine Number	1	2	3	4	5	6
Processing/inspection rate	1	0.95	1.1	0.97	0.98	1.05
Failure rate	0.001	0.001	0.001	0.001	0.001	0.001
Repair rate	0.1	0.1	0.1	0.1	0.1	0.1

Table 4.5. Parameters of the machines in Case H.

Machine Number	1	2	3	4	5	6
Processing/inspection rate	1	0.95	0.98	1.1	0.94	1
Failure rate	0.001	0.001	0.001	0.001	0.001	0.001
Repair rate	0.1	0.1	0.1	0.1	0.1	0.1
Machine Number	7	8	9	10	11	12
Processing/inspection rate	1.02	1.05	1.01	1	1.02	1.04
Failure rate	0.001	0.001	0.001	0.001	0.001	0.001
Repair rate	0.1	0.1	0.1	0.1	0.1	0.1



(a) Case G

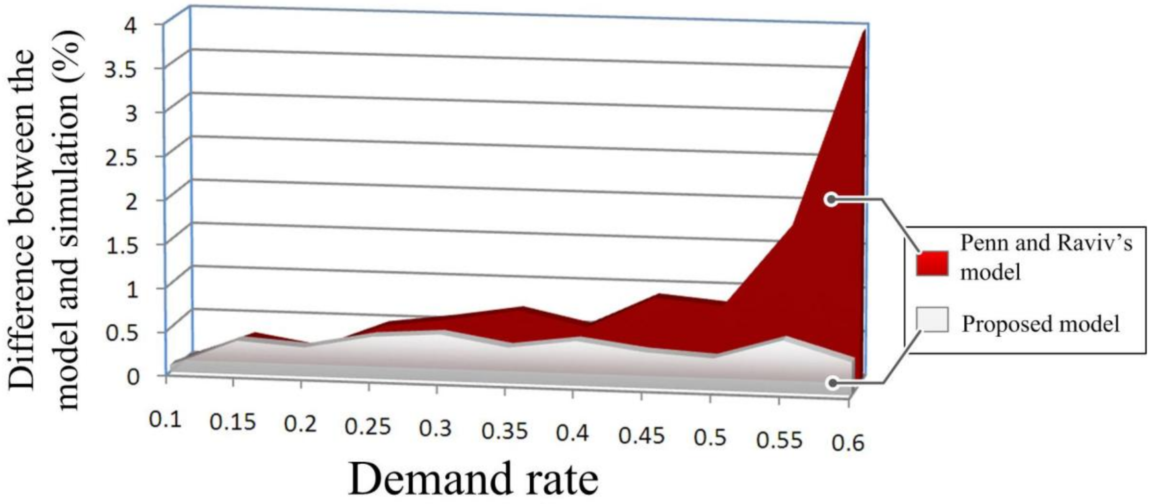


(b) Case H

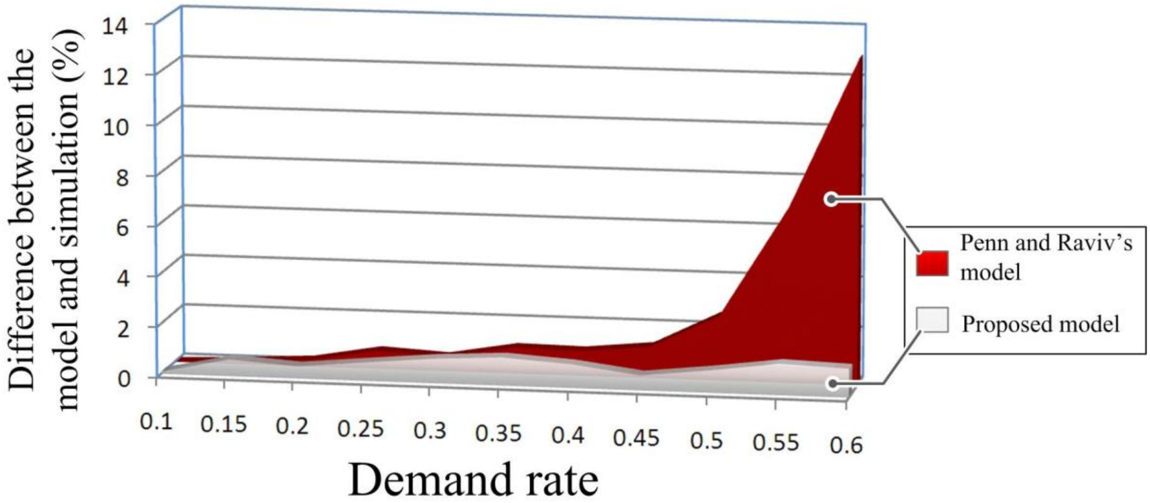
Figure 4.9. Configurations of the systems studied in the experiment.

In both experiments, the demand rate is varied from 0.1 to 0.6 in steps of 0.05. The proposed model and Penn and Raviv's model are used to estimate the total inventory in the system. The performance measures obtained using these two analytical models are subsequently compared with the simulation results.

As illustrated in Figure 4.10, the relative difference between the total inventory estimated using the proposed model and simulation is generally lower than two percent. By contrast, Penn and Raviv's model provides less accurate results when the demand rate increases. Additionally, the relative difference between their model and simulation is even more significant in the system with a larger number of machines.



(a) Case G.



(a) Case H.

Figure 4.10. The relative difference between the total inventory obtained using the analytical models and simulation. (100% |Ana-Sim|/Sim, where Ana and Sim represent the total inventory obtained from the analytical models and simulation respectively).

4.7. A Case Study for Determining the Location of Inspection Machines

In this subsection, the integrated quantitative and qualitative model is used to determine the placement of inspection machines in the following two cases:

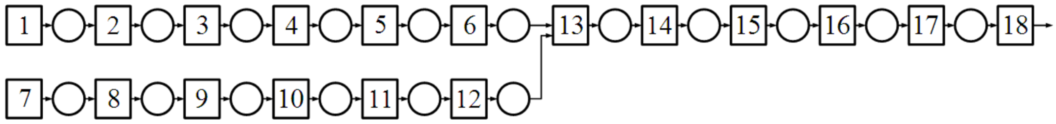
1) **Case I:** An assembly line with 18 processing machines, as illustrated in Figure 4.11(a) is considered. The parameters are chosen as:

- Mean processing or inspection rate: 1 part/min (for each machine).
- Defective rate: 0.02 (for each processing machine).
- Probability of type I error: 0.01 (for each inspection machine).
- Probability of type II error: 0.01 (for each inspection machine).
- Buffer size: 5 parts (for each buffer).
- Failure rate: 0.001 min^{-1} (for each machine).
- Repair rate: 0.1 min^{-1} (for each machine).
- Processing cost: 10 dollars/part (for each processing machine).
- Inspection cost: 5 dollars/part (for each inspection machine)
- Penalty cost: 2000 dollars/part.
- Unit price: 1000 dollars/part.
- *WIP* holding cost per part: 0.5 dollars/(part·min).

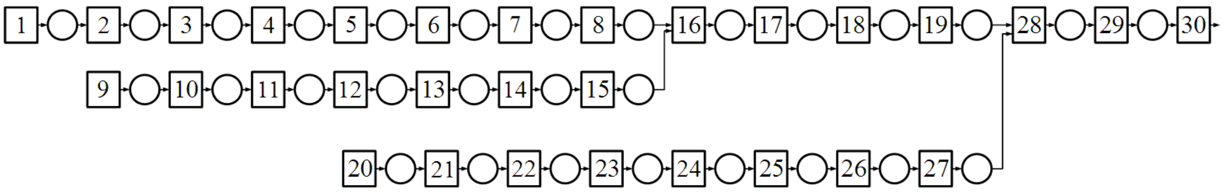
2) Case J: An assembly line with 30 processing machines (Figure 4.11(b)), where processing machines are non-homogeneous, is analyzed. The parameters are chosen as in Case I, except:

- Mean processing rate of M_{13} (the bottleneck machine) is 0.8 part/min.
- Processing cost of M_8 and M_{24} are 40 and 50 dollars/part respectively.
- Defective rate of processing machines are chosen as:

  l_1=0.01, l_2=0.005, l_3=0.02, l_4=0.01, l_5=0.001, l_6=0.05, l_7=0.01, l_8=0.001, l_9=0.02, l_{10}=0.025, l_{11}=0.02, l_{12}=0.01, l_{13}=0.005, l_{14}=0.02, l_{15}=0.01, l_{16}=0.001, l_{17}=0.05, l_{18}=0.01, l_{19}=0.02, l_{20}=0.01, l_{21}=0.025, l_{22}=0.01, l_{23}=0.02, l_{24}=0.002, l_{25}=0.01, l_{26}=0.02, l_{27}=0.001, l_{28}=0.01, l_{29}=0.005, l_{30}=0.002



(a) Case I.



(b) Case J.

Figure 4.11. The assembly lines studied in Cases I and J.

The numerical results for Cases I and J are provided in Table 4.6. The inspection allocation algorithm (IAA) presented in Section 4.4 is used to determine the placement of inspection machines. For comparison, two other methods (enumeration and Genetic Algorithms) are also used to determine the inspection allocation. Additionally, the performance measures of the assembly line under two conditions: without any inspection and with ubiquitous inspection are also provided in Table 4.6. Based on the results of Case I in Table 4.6, we observe that the solution prescribed by IAA is similar to the optimal solution obtained via enumeration. The corresponding *EP* of IAA (394.83) is only 0.28% lower than that of enumeration (395.95). On the other hand, the solution obtained using Genetic Algorithms is not as close to the optimal solution. The CPU time in Table 4.6 also demonstrates that IAA is much more computationally efficient than enumeration or Genetic Algorithms.

For Case J, the number of processing machines is 30, and there are $2^{29} \approx 5.37 \times 10^{8}$ different possible placements of inspection machines. Hence, it is infeasible to determine the inspection allocation via enumeration. For problems with a large number of processing machines, a fast method, such as

IAA as proposed in this paper, is desirable. Based on the results of Case J in Table 4.6, we notice that the solution provided by IAA may lead to a profit (243.72) that is 11.47% higher than that obtained using Genetic Algorithms (218.64). The solution obtained using IAA is also depicted in Figure 4.12. As illustrated in this figure, an inspection machine is placed before the bottleneck machine, M_{13}. This placement may avoid wasting the capacity of the bottleneck machine on defective parts, and hence improve production rate. Additionally, inspection machines are also located before processing machines with high processing cost (i.e. M_8 and M_{24}), which may prevent these machines from processing defective parts and thus reduce operational costs. Figure 4.13 illustrates the fraction of defective parts in the material flow out of each processing machine (i.e. $1-\theta_k$, and θ_k is calculated as discussed in Section 4.3.1) under three conditions: ubiquitous inspection; inspection machines are placed as prescribed by IAA; and no inspection. Compared with the condition without inspection, the percentage of defective parts throughout the assembly line is effectively reduced if inspection machines are allocated appropriately. The improvement of quality of material flow substantially reduces the waste of machine capacity and processing cost on defective parts. This may result in the simultaneous improvement of quantitative and qualitative performance. Although ubiquitous inspection assures good quality of material flow throughout the assembly line, it may result in high inspection costs, as shown in Table 4.6. Additionally, the numerical results also demonstrate that excessive inspection may also undermine effective production rate (calculated using Eqn (4.17)) of the assembly line. Therefore, ubiquitous inspection is not economically feasible.

Table 4.6. Comparison of results obtained using three methods: enumeration, inspection allocation algorithm (IAA), and Genetic Algorithms (GA). Additionally, performance measures of the systems under two conditions: no inspection and with ubiquitous inspection, are also provided for comparison.

Case	Solution method	Solution (Z) [#]	Expected profit	Effective production rate	Inspection cost	CPU time (minutes)
I	Enumeration	00100100100100000	395.95	0.5684	17.65	2169.1
	IAA	00100100010100000	394.83	0.5650	17.51	2.2
	GA	01010100101000100	376.27	0.5539	24.15	41.5
	No inspection	00000000000000000	119.79	0.4865	0	0.0038
	Ubiquitous inspection	11111111111111111	310.95	0.5291	59.96	0.0045
J	Enumeration	Not performed	Not performed	Not performed	Not performed	Not performed
	IAA	000000100001001010000001000100	243.72	0.5578	22.89	13.2
	GA	010000010001010000010100001010	218.64	0.5546	29.29	232.4
	No inspection	000000000000000000000000000000	-89.69	0.4462	0	0.0074
	Ubiquitous inspection	111111111111111111111111111111	112.01	0.5305	96.82	0.0095

[#]For Case I: $Z=[z_{1,2}, z_{2,3}, z_{3,4}, z_{4,5}, z_{5,6}, z_{6,13}, z_{7,8}, z_{8,9}, z_{9,10}, z_{10,11}, z_{11,12}, z_{12,13}, z_{13,14}, z_{14,15}, z_{15,16}, z_{16,17}, z_{17,18}]$

For Case J: $Z=[z_{1,2}, z_{2,3}, z_{3,4}, z_{4,5}, z_{5,6}, z_{6,7}, z_{7,8}, z_{8,16}, z_{9,10}, z_{10,11}, z_{11,12}, z_{12,13}, z_{13,14}, z_{14,15}, z_{15,16}, z_{16,17}, z_{17,18},$
$z_{18,19}, z_{19,28}, z_{20,21}, z_{21,22}, z_{22,23}, z_{23,24}, z_{24,25}, z_{25,26}, z_{26,27}, z_{27,28}, z_{28,29}, z_{29,30}]$

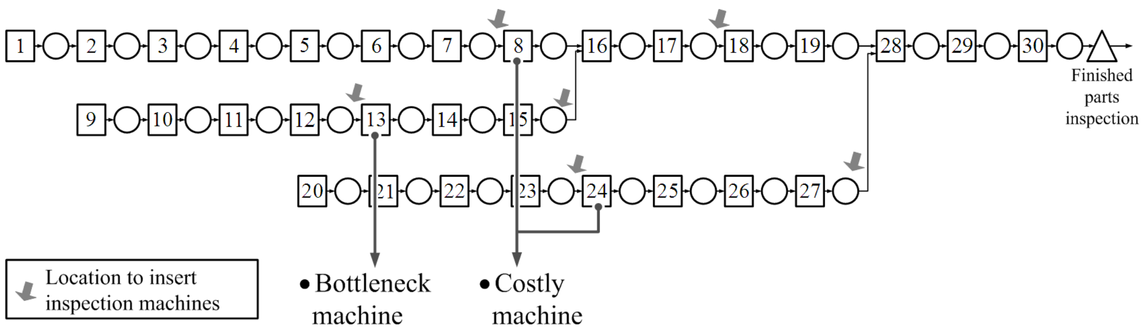


Figure 4.12. Placement of inspection machines prescribed by the inspection allocation algorithm.

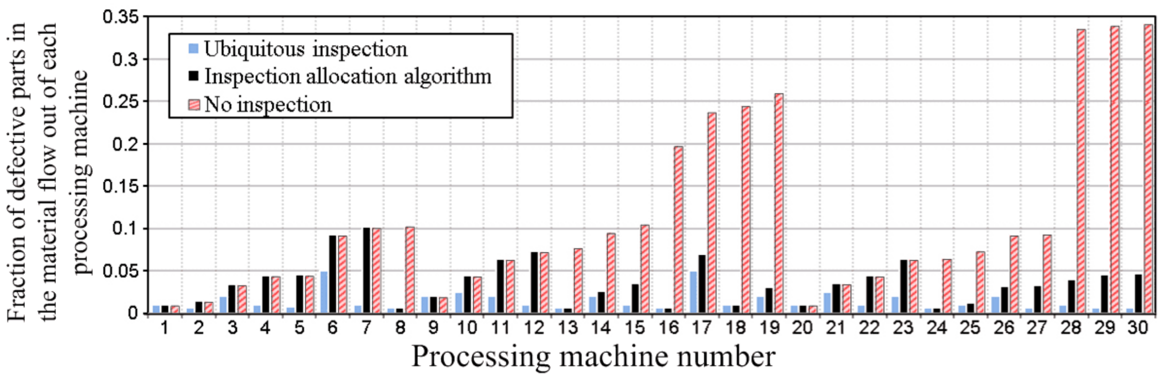


Figure 4.13. The fraction of defective parts in the flow rate out of each processing machine, under three conditions: ubiquitous inspection; inspection machines are placed as prescribed by the inspection allocation algorithm; and no inspection.

4.8. Sensitivity Analysis of the Model

The application of the integrated quantitative and qualitative model is not only limited to solving the inspection allocation problem. The model may also be used to perform sensitivity analysis of the manufacturing system (i.e. to analyze the effects of varying a parameter on the performance measures). This analysis may provide line managers with the insights for improving the control of such systems. The assembly line in Case J will be used as an example to briefly discuss this application. In this example, we shall assume that the inspection machines are placed as prescribed by the inspection allocation algorithm.

In addition to performing inspection, another way to improve the quality of material flow in the assembly line, is to reduce the defective rate of processing machines. This may be achieved by replacing processing machines with more reliable ones. Since replacing different processing machines in an assembly line may result in different levels of profit improvement, it is important to decide which machines should be replaced. A sensitivity analysis of expected profit (*EP*) with respect to the defective rate of each processing machine (l_k) may facilitate the line manager to indentify which machine should be replaced.

Expected profit of the assembly line is a function of defective rates of all the processing machines, and hence we may denote it as: $EP\left(l_1, l_2, \cdots, l_K\right)$. The percentage of profit improvement (*PPI*) when l_k ($k \in \{1, 2, \cdots, K\}$) is reduced by a small fraction (denoted as δ) can be calculated as follows:

$$PPI_1 = \frac{EP\left(l_1 \cdot (1-\delta),\ l_2,\ \cdots\ ,\ l_K\right) - EP\left(l_1,\ l_2,\ \cdots\ ,\ l_K\right)}{EP\left(l_1,\ l_2,\ \cdots\ ,\ l_K\right)} \cdot 100\% \qquad (4.25)$$

$$PPI_2 = \frac{EP\left(l_1,\ l_2 \cdot (1-\delta),\ \cdots,\ l_K\right) - EP\left(l_1,\ l_2,\ \cdots,\ l_K\right)}{EP\left(l_1,\ l_2,\ \cdots,\ l_K\right)} \cdot 100\% \tag{4.26}$$

…

$$PPI_K = \frac{EP\left(l_1,\ l_2,\ \cdots,\ l_K \cdot (1-\delta)\right) - EP\left(l_1,\ l_2,\ \cdots,\ l_K\right)}{EP\left(l_1,\ l_2,\ \cdots,\ l_K\right)} \cdot 100\% \tag{4.27}$$

For instance, if we choose δ =0.05, PPI_k ($k \in \{1, 2, \cdots, K\}$) of the assembly line in this example can be calculated using Eqns (4.25) to (4.27), and the results are as illustrated in Figure 4.14. From this figure, we observe that by reducing the defective rate of the 17th processing machine, profit of the system may be improved significantly. Therefore, if the line manager plans to replace one of the processing machines in the assembly line, this machine should be considered with higher priority than others. Additionally, if the line manager decides to replace several processing machines simultaneously, the other machines that may result in high profit improvement (such as the 19th and 14th processing machines) may also be considered.

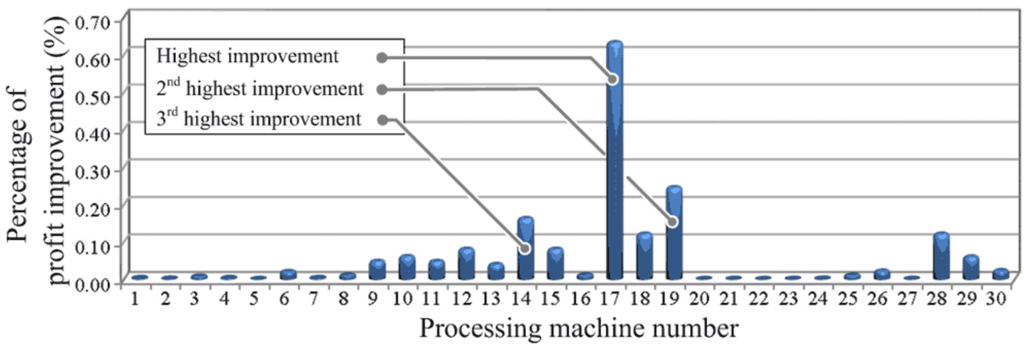


Figure 4.14. Percentage of profit improvement if the defective rate of a processing machine is reduced by 5%.

Chapter 5.

Modeling of Multistage Manufacturing Systems with Batch Operations and Generally Distributed Processing Times

5.1. Overview

In Chapters 3 and 4, the author analyzed manufacturing systems with single-item machines (i.e. each machine can process only one part at a time) and did not consider batch operations (where a machine is capable of processing several parts simultaneously). Although many analytical studies of multistage manufacturing systems in the literature also restrict the focus on single-item machines, both batch and single-item machines coexist in the production lines of many industries. For example, in garment production, fabric cutting, washing and drying processes involve batch machines whereas sewing operations and packing processes are essentially single-item operations. Other industries where batch machines are employed include electrical appliance manufacture (e.g. chemical coating processes), wafer fabrication (e.g. diffusion and oxidation processes), etc (Chen et al., 2010). The implementation of batch operations may improve the utilization of machines and production rate, as illustrated in the example of Figure 5.1. A system with batch operations may exhibit fundamentally different performance when compared with single-item systems. Therefore, it is necessary to account for the influence of batch operations in performance evaluation of the systems with such feature. Recent simulation studies (Aguirre, et al., 2008; Schmidt and Rose 2008) further support the need to explicitly account for batch operations in performance analysis.

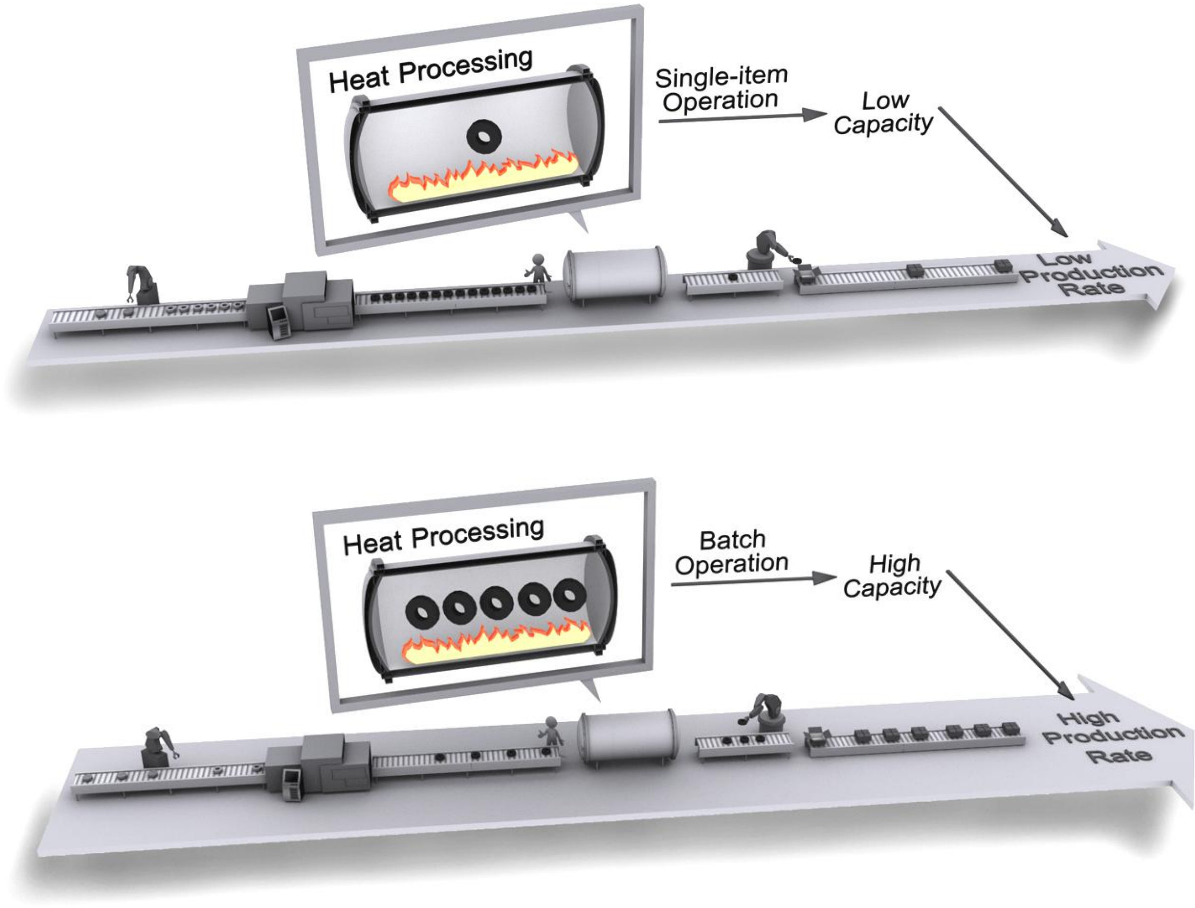


Figure 5.1. Manufacturing systems with single-item operations and batch operations. In this example, the heat processing of each part may require a relatively long processing time. If the oven processes parts item by item, production rate of the system is low. If on the other hand, the oven operates in batch, the number of parts processed in a time unit is increased, and hence production rate of the system is improved.

In order to represent a wide range of manufacturing systems, the author further assumes that the machine processing times are generally distributed. This assumption reflects the variable processing times observed in industrial batch processes such as garment washing/dyeing and material transportation processes. To account for general distributions, the processing times are represented with a phase-type distribution, in particular, the hypoexponential distribution. Using this distribution we are able to approximate processing time distributions that have a coefficient of variation (*CV*, defined as the ratio of standard deviation to mean) of less than one, which is prevalent in the majority of real cases (Li, et al., 2009).

The remainder of this chapter is organized as follows: the modeling of the multistage manufacturing system with batch operations and hypoexponential processing times is discussed in Section 5.2. Numerical experiments are provided in Section 5.3 to validate the model presented in this chapter. Additionally, in Section 5.4, a case study is presented to illustrate the use of this model in improving the control of multistage manufacturing systems with batch operations.

5.2. Modeling Multistage Manufacturing Systems with Batch Operations and Hypoexponential Processing Times

In this section, a serial multistage manufacturing system with K machines in series decoupled by $K-1$ buffers will be investigated, as illustrated in Figure 5.2. Each machine in such a system can process a specific batch size, Q_k. Single-item machines are also viewed as batch machines with Q_k =1.

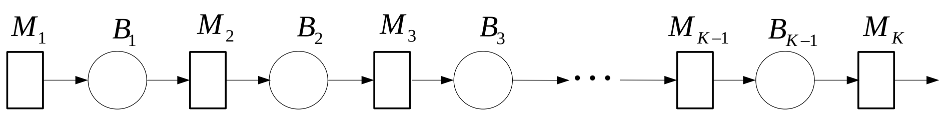


Figure 5.2. A multistage manufacturing system with batch processing machines.

The following assumptions are used in this chapter.

- M_k may commence processing only when the number of parts in its immediately upstream buffer, B_{k-1}, is equal to or larger than Q_k, i.e. a full batch of parts are available (Bolch, et al., 2006). M_k is said to be starved if it is idle and the number of parts in buffer B_{k-1} is less than Q_k.

This is common in real systems with batch operations such as heating/cooling and chemical coating processes. In many manufacturing systems, the setup cost is usually incurred for a machine to process a batch of parts (Nagaraj and Selladurai, 2002). To reduce the total setup cost, a machine does not start processing until a full batch of parts is available. As pointed out by In et al. (2003), many batch machines in wafer production lines are operated on this condition.

- Suppose M_k is not starved, i.e., there is a full batch of parts available in its upstream buffer, B_{k-1}. It will however not start processing if its immediately downstream buffer, B_k has insufficient space to unload that batch of parts. In this case, the machine is blocked.

This assumption is referred to as "blocking-before-service" in the literature (Gershwin, 1994). Many analytical models of multistage manufacturing systems have been developed based on this assumption (Kuo et al., 1997; Chiang et al., 2000; Li and Huang, 2005; Li and Meerkov, 2009; Kim and Gershwin, 2005, 2008; Colledani and Tolio, 2006, 2009). Another alternative assumption is referred to as "blocking-after-service". Under the "blocking-after-service" assumption, a machine continues processing until the finished parts cannot be delivered to its immediately downstream buffer. In this case, the machine may process an additional batch of parts even when its immediately downstream buffer is full. As discussed by Gershwin (1994), the analytical models based on the "blocking-before-service" assumption can be easily extended to the systems with "blocking-after-service". Therefore, the author formulates the model based on the "blocking-before-service" assumption.

- Machine processing times are described with a hypoexponential distribution, which is a series of J exponential distributions with rates μ_j (j=1,2,…,J). This is represented in Figure 5.3, where the batch machine is divided into a series of J virtual stages. The parts should go through all the virtual stages sequentially, and there is at most one batch of parts being processed at any time.

  In Queuing Theory, the generally distributed variables are usually characterized by two important values, viz. mean and variance (Gross and Harris, 1998). Hypoexponential distribution is a phase-type distribution, which has been widely used to approximate general distribution in terms of mean and variance (Bolch, et al., 2006). The author has performed a large number of experiments, in which production lines with processing times of different types of distributions are compared. The results demonstrate that the difference among performance measures of the various distributions is very small. These results also justify the statement by Dallery and Gershwin (1992) that approximating the mean and variance of a general distribution using a phase-type distribution is a sufficient approximation.

- We assume that the first machine in the system is never starved of raw material and the last machine is never blocked. The validity of this assumption has been discussed in Section 3.3.

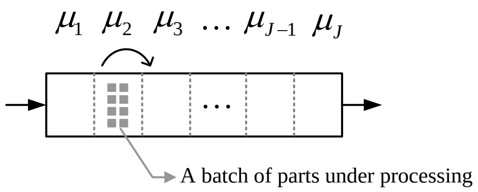


Figure 5.3. A machine modeled as a series of virtual stages.

5.2.1. Markov Model of a Primitive Line Segment

Similar to the decomposition models discussed in Chapters 3 and 4, we decompose a long production line with batch operations and hypoexponential processing times into a number of two-machine-one-buffer primitive line segments. Each primitive line segment consists of an upstream machine (M_k^u), a downstream machine (M_k^d), and an intermediate buffer (B_k), as illustrated in Figure 5.4. In addition to the notations defined in Section 3.2, the following notations will also be used in the development of the Markov model of the primitive line segment in this chapter.

Q_k^u, Q_k^d: The batch size of the upstream and downstream machines respectively.

J_k^u, J_k^d :The number of virtual stages of the upstream and downstream machines respectively.

j_k^u, j_k^d: The state of the upstream and downstream machines respectively, $j_k^u = 0,1,...,J_k^u$ and $j_k^d = 0,1,...,J_k^d$. 0 indicates that a machine is idle; and for $1 \le j_k^u \le J_k^u$ (or $1 \le j_k^d \le J_k^d$), there is a batch of parts at the (j_k^u)th (or (j_k^d)th) stage.

Furthermore, as discussed in Section 3.3.2, two variables α_k and β_k are used to denote whether M_k^u and M_k^d are "pseudo down".

The primitive line segment is shown in Figure 5.4. The system state is defined as:

$$S_k = \left(x_k, j_k^u, j_k^d, \alpha_k, \beta_k\right) \tag{5.1}$$

The balance equations are used to describe state transitions of the primitive line segment. We shall first present the balance equations for the primitive line segment in isolation, i.e., under the assumption that M_k^u and M_k^d are never "pseudo down" ($\alpha_k = 1$ and $\beta_k = 1$). In this case, the state of the system is $S_k = \left(x_k, j_k^u, j_k^d, 1, 1\right)$. For simplicity, the state is rewritten as

$$S_k = \left(x_k, j_k^u, j_k^d\right) \tag{5.2}$$

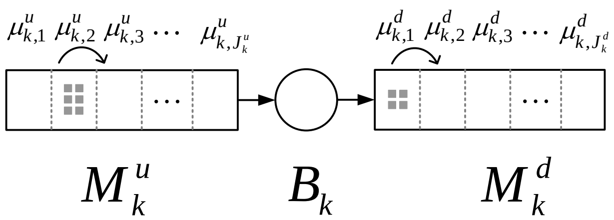


Figure 5.4. The line segment with processing times characterized as a hypoexponential distribution.

The state of the upstream or downstream machine (j_k^u or j_k^d) may be approximately categorized into the following four conditions:

- M_k^u (or M_k^d) is idle, i.e. $j_k^u = 0$ (or $j_k^d = 0$).
- M_k^u (or M_k^d) is at the first virtual stage, i.e. $j_k^u = 1$ (or $j_k^d = 1$).
- M_k^u (or M_k^d) is at the intermediate virtual stages, i.e. $1 < j_k^u < J_k^u$ (or $1 < j_k^d < J_k^d$).
- M_k^u (or M_k^d) is at the final virtual stage, i.e. $j_k^u = J_k^u$ (or $j_k^d = J_k^d$).

Based on the combination of these four conditions of both machines, we have 16 groups of balance equations. The group where M_k^u and M_k^d are both in the final virtual stages, i.e. $j_k^u = J_k^u$ and $j_k^d = J_k^d$, is selected as an example to illustrate the development of balance equations.

In this group, the states $\left(x_k, J_k^u, J_k^d\right)$ with $x_k > X_k - Q_k^u$ are transient states, i.e., the limiting probability of these states is zero. This is because, based on our assumptions, it is not possible to reach these states from any other state except from another transient state. For example, when $x_k > X_k - Q_k^u$, the buffer has insufficient space for machine M_k^u to unload a batch of parts and machine M_k^u is then blocked. In this condition, machine M_k^u will always be in state 0 ($j_k^u = 0$). The balance equations for the other states in this group can be written as Eqn (5.3) based on the transition diagrams in Figure 5.5.

$$P\left(x_k, J_k^u, J_k^d\right)\cdot\left(\mu_{k,J_k^u}^u + \mu_{k,J_k^d}^d\right) = P\left(x_k, J_k^u - 1, J_k^d\right)\cdot\mu_{k,J_k^u-1}^u + P\left(x_k, J_k^u, J_k^d - 1\right)\cdot\mu_{k,J_k^d-1}^d \tag{5.3}$$

The left side of Eqn (5.3) represents the total transition rate out of state $\left(x_k, J_k^u, J_k^d\right)$. These transitions are described as follows:

- M_k^u discharges a batch of parts to buffer B_k at the rate of $\mu_{k,J_k^u}^u$, and the resulting state of machine M_k^u has two possibilities based on the current buffer level of B_k:
    - If $x_k \le X_k - 2Q_k^u$, after B_k receives Q_k^u parts, there is still enough space to receive another batch of parts. Hence, M_k^u starts processing a

new batch of parts and these arrive into the first virtual stage of M_k^u (i.e. $j_k^u = 1$), as illustrated in Figures 5.5 (a) and (b).

  - If $X_k - 2Q_k^u < x_k \le X_k - Q_k^u$, after B_k receives Q_k^u parts, the inventory in the buffer is increased to $x_k + Q_k^u$, and there is insufficient space for unloading another batch from M_k^u. Therefore, M_k^u is blocked and j_k^u =0, as illustrated in Figures 5.5 (c) and (d).

- M_k^d sends out a batch of parts at the rate of μ^d_{k,J_k^d}, and the subsequent state of machine M_k^d has two possibilities:
  - If $x_k \ge Q_k^d$, a new batch of parts are delivered from buffer B_k to the first virtual stage of M_k^d (i.e. $j_k^d = 1$), as illustrated in Figures 5.5 (a) and (c).
  - If $x_k < Q_k^d$, M_k^d cannot obtain a full batch of parts from B_k for processing, hence it remains idle (i.e. $j_k^d = 0$), as illustrated in Figures 5.5 (b) and (d).

On the other hand, the right side of Eqn (5.3) indicates the transitions from other states into state $\left(x_k, J_k^u, J_k^d\right)$, and these can be described as follows:

- The parts at the $\left(J_k^u - 1\right)^{\text{th}}$ stage of M_k^u are transited to the $\left(J_k^u\right)^{\text{th}}$ stage with the transition rate of μ^u_{k,J_k^u-1}. Although this does not represent an actual movement of the batch within the machine, it can be seen as a progression of machine processing from one phase to the next, in this case, from phase $J_k^u - 1$ to phase J_k^u.

- The parts at the $\left(J_k^d - 1\right)^{\text{th}}$ stage of M_k^d are transited to the $\left(J_k^d\right)^{\text{th}}$ stage with the transition rate of $\mu_{k,J_k^d-1}^d$.

Likewise, the balance equations for the other groups can also be derived. By solving the balance equations and the normalization equation (i.e. the sum of all probabilities equals one), the limiting probabilities of all states, denoted as $P\left(x_k, j_k^u, j_k^d\right)$, are obtained.

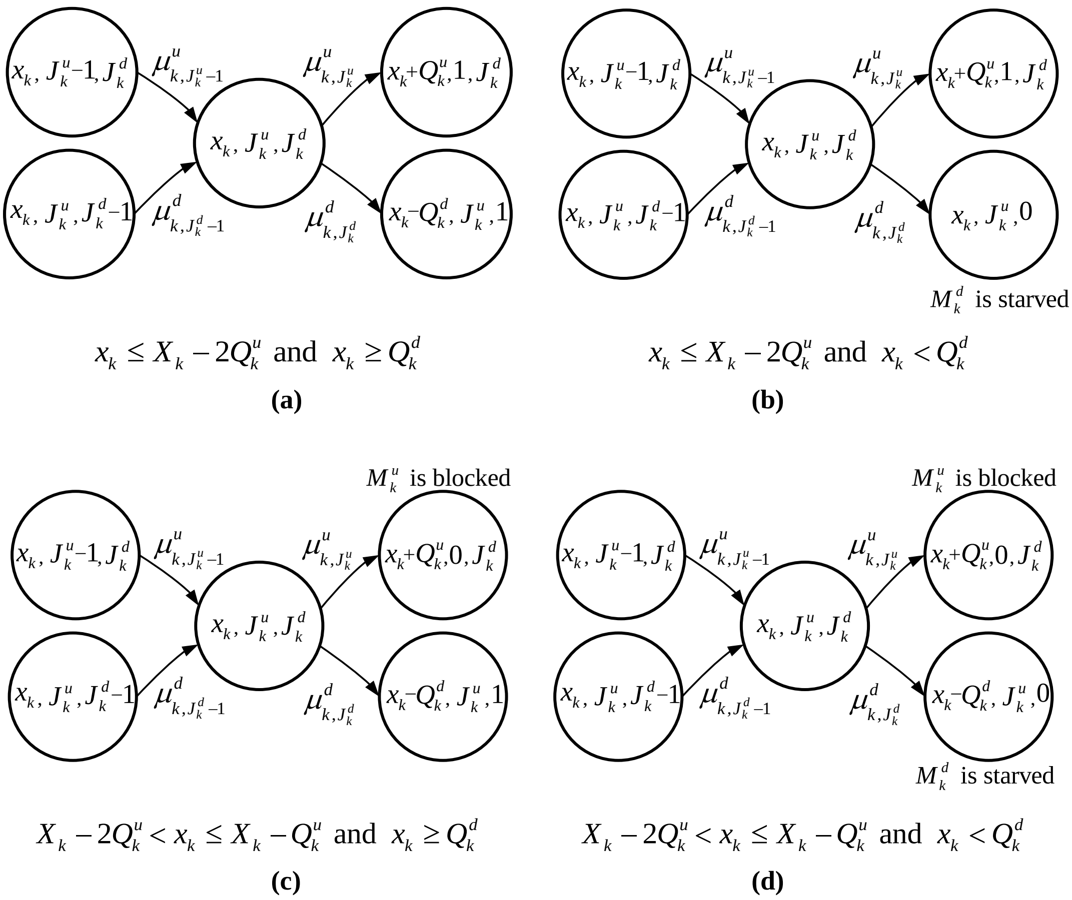


Figure 5.5. Transition diagrams for states when $j_k^u = J_k^u$ and $j_k^d = J_k^d$.

5.2.2. Incorporating the "pseudo down" state in the Primitive Line Segment

As discussed in Section 3.3, the "pseudo down" state of machines in a primitive line segment is introduced to represent the starvation and blockage due to the upstream and downstream line segments. Hence, we will revert to describing the state as $S_k = \left(x_k, j_k^u, j_k^d, \alpha_k, \beta_k\right)$. Similar to the primitive line segment in Section 3.3, four parameters (p_k^α , r_k^α , p_k^β , and r_k^β) are used to characterize the transitions of M_k^u and M_k^d between "pseudo down" and not "pseudo down" states. The calculation of these parameters is presented in Appendix B. When M_k^u finishes a batch of parts, it has the probability, p_k^α to become "pseudo down". This represents the condition that M_k^u cannot obtain a full batch of parts from the immediately upstream line segment for further processing after it completes processing a batch. If M_k^u is "pseudo down", it recovers from this condition with the transition rate of r_k^α. Similarly, when M_k^d discharges a batch of parts, it has the probability p_k^β to become "pseudo down". This represents the condition that after M_k^d delivers a batch of parts to its immediately downstream buffer, the buffer has no further space for unloading another batch. If M_k^d is "pseudo down", it recovers with the transition rate of r_k^β.

With the consideration of these additional state transitions between $\alpha_k = 1$ and $\alpha_k = 0$, and $\beta_k = 1$ and $\beta_k = 0$, the balance equations can be formulated for the primitive line segment, and then used to solve for $P\left(x_k, j_k^u, j_k^d, \alpha_k, \beta_k\right)$.

5.2.3. Performance Measures

Based on the limiting probabilities of the primitive line segments, we may calculate the performance measures of the manufacturing system. In the calculation of the following performance measures, the batch sizes of machines will be accounted for. This is a distinct feature compared with the decomposition models in Chapters 3 and 4, where only single-item machines are considered.

- Production rate (PR)

Production rate of a multistage manufacturing system is calculated based on the limiting probability of the final primitive line segment:

$$PR = \sum_{\alpha_{K-1}} \sum_{j^u_{K-1}} \sum_{x_{K-1}} P\left(x_{K-1}, j^u_{K-1}, J^d_{K-1}, \alpha_{K-1}, 1\right) \mu^d_{K-1, J^d_{K-1}} \cdot Q^d_{K-1} \tag{5.4}$$

- Work-In-Process (WIP)

The average numbers of parts in M^u_k and M^d_k are:

$$WIP^u_k = Q^u_k \cdot \sum_{\beta_k} \sum_{\alpha_k} \sum_{j^d_k} \sum_{j^u_k > 0} \sum_{x_k} P\left(x_k, j^u_k, j^d_k, \alpha_k, \beta_k\right) \tag{5.5}$$

$$WIP^d_k = Q^d_k \cdot \sum_{\beta_k} \sum_{\alpha_k} \sum_{j^d_k > 0} \sum_{j^u_k} \sum_{x_k} P\left(x_k, j^u_k, j^d_k, \alpha_k, \beta_k\right) \tag{5.6}$$

We can also obtain the WIP in the buffer, B_k , as:

$$WIP^b_k = \sum_{\beta_k} \sum_{\alpha_k} \sum_{j^d_k} \sum_{j^u_k} \sum_{x_k} \left(x_k \cdot P\left(x_k, j^u_k, j^d_k, \alpha_k, \beta_k\right)\right) \tag{5.7}$$

By summing the inventory at every line segment, we obtain the total inventory.

$$WIP = \sum_{j=1}^{K-1} WIP^u_j + \sum_{j=1}^{K-1} WIP^b_j + WIP^d_{K-1} \tag{5.8}$$

- Production lead time

According to Little's Law (Little, 1961), production lead time is calculated as:

$$PLT = WIP / PR \quad (5.9)$$

5.2.4. Unreliability of Machines

Introducing hypoexponential processing times in the decomposition model provides an alternative way to incorporate machine failures besides using additional state for representing machine failures. We may employ the concept of effective processing time (Hopp and Spearman, 2000) to incorporate machine failures into the model. In this method, the time losses due to machine failures is integrated with the processing time by lumping them into one probability distribution called the effective processing time distribution. This leads to a general distribution which we can then approximate using the hypoexponential distribution and follow the same methodology described earlier. Only the mean and variance (first two moments) of the machine processing, failure and repair time distributions are required to characterize the effective processing time. For simple distributions of machine processing, failure and repair times, the closed form expression of the probability density function for the effective processing time can be derived and this procedure is described in Appendix C.

5.3. Model Validation

In this section, the decomposition model is evaluated using ten sets of experiments. The numerical results obtained from the analytical model are validated by comparing with the results from a simulation model. Additionally, based on these experiments, the influence of batch size, probability distribution of processing times, *CV* of processing times and machine unreliability on the system performance is investigated.

5.3.1. Parameters

In the experiments, the decomposition model is applied to a ten-machine serial manufacturing system under ten different operating conditions (Cases A to J). These ten cases are organized as summarized in Table 5.1.

The parameters used in the experiments are summarized in Table 5.3. In addition, the following parameters require further elaboration.

- Mean processing times

  For Cases A to H, balanced systems are analyzed while in Cases I to J, unbalanced systems are analyzed. For balanced systems, each machine is assumed to have a mean processing rate of one part per minute. Therefore, the mean processing time of batch machines with batch sizes of Q_k is chosen to be Q_k min/batch. For unbalanced systems, the batch machines M_3, M_7 and M_9 with batch sizes Q_3, Q_7 and Q_9 are assumed to have mean processing rates of $\frac{1}{1.2}$, $\frac{1}{1.2}$ and $\frac{1}{1.1}$ parts per minute respectively while the other machines will have processing rates of one part per minute. The batch sizes (Q_k) of machines for all the experiments are summarized in Table 5.2.

- Probability distribution of processing times

  For Cases A to C, the effect of the underlying probability distribution of processing time on system performance is analyzed by comparing simulation results for the lognormal, gamma and hypoexponential distributions (with the same mean and variance) with the results obtained through the analytical method which is based on the hypoexponential distribution.

Table 5.1. Organization of Cases A to J.

Cases	For analyzing the influence of
A - C	Batch sizes and probability distribution of processing times
D - F	*CV* of processing times
G and H	Machine reliability of a balanced system
I and J	Machine reliability of an unbalanced system

Table 5.2. The batch size of each machine in the experiments.

Case	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8	Q_9	Q_{10}
A	1	1	1	1	1	1	1	1	1	1
B	2	1	2	1	2	1	2	1	2	1
C	4	1	4	1	4	1	4	1	4	1
D	1	2	3	1	4	3	1	5	1	2
E	1	2	3	1	4	3	1	5	1	2
F	1	2	3	1	4	3	1	5	1	2
G	1	2	1	1	2	3	1	1	1	2
H	1	2	1	1	2	3	1	1	1	2
I	1	2	1	1	2	3	1	1	1	2
J	1	2	1	1	2	3	1	1	1	2

5.3.2. Results and Discussion

The numerical results obtained for Cases A to J are summarized in Table 5.4. Three system performance measures, viz. production rate, total inventory, and production lead time obtained through the decomposition model are compared with those obtained from simulation. Further, the effect of increasing batch size on the average buffer levels at each stage of the system for Cases A to C is illustrated in Figure 5.6.

Table 5.3. Experimental parameters for Cases A to J.

Case	Buffer size (all buffers)	Mean processing time# (mins per batch)	*CV* of processing time (all machines)	Probability distribution of processing times in Simulation (all machines)	Machine reliability (all machines)
A	10	Q_k for $k \in \{1,\ldots,10\}$	0.6	Lognormal, hypoexponential, gamma	Reliable
B	10	Q_k for $k \in \{1,\ldots,10\}$	0.6	Lognormal, hypoexponential, gamma	Reliable
C	10	Q_k for $k \in \{1,\ldots,10\}$	0.6	Lognormal, hypoexponential, gamma	Reliable
D	10	Q_k for $k \in \{1,\ldots,10\}$	0.4	Lognormal	Reliable
E	10	Q_k for $k \in \{1,\ldots,10\}$	0.6	Lognormal	Reliable
F	10	Q_k for $k \in \{1,\ldots,10\}$	0.8	Lognormal	Reliable
G	10	Q_k for $k \in \{1,\ldots,10\}$	0.6	Lognormal	Reliable
H	10	Q_k for $k \in \{1,\ldots,10\}$	0.6	Lognormal	Unreliable (mean uptime =500 mins, mean downtime=10 mins)
I	10	Q_k (for k=1, 2, 4, 5, 6, 8, 10) 120% of Q_k (for k =3 and 7) 110% of Q_k (for k =9)	0.6	Lognormal	Reliable
J	10	Q_k (for k=1, 2, 4, 5, 6, 8, 10) 120% of Q_k (for k =3 and 7) 110% of Q_k (for k =9)	0.6	Lognormal	Unreliable (mean uptime=500 mins, mean downtime=10 mins)

The values of Q_k are provided in Table 5.2.

In all these numerical experiments, the absolute relative difference between the analytical and simulation results of each performance measure is generally less than 3%, which indicates that the decomposition model is capable of providing reliable estimates of performance measures for a multistage manufacturing system with batch processing. Based on the results in Table 5.4, the decomposition model requires only a small fraction of the CPU time required by simulation.

For Cases A to C, Table 5.4 shows the simulation results for lognormal, gamma and hypoexponential distributions of processing times (each with the same mean and variance) compared with the analytical results obtained assuming a hypoexponential distribution. It can be observed that the difference among performance measures under the various distributions is very small. These results justify the statement by Dallery and Gershwin (1992) that approximating the first two moments (mean and variance) of a general distribution using a phase-type distribution is a sufficient approximation when the coefficients of variation (*CV*) are not very large (in this chapter, $CV<1$). Therefore, we can infer that the analytical model is generally insensitive to the underlying distribution of processing times. Hence, for brevity, we only provide the simulation results under lognormal processing times for Cases D to J. To approximate generally distributed processing times with $CV>1$, the hyperexponential distribution can be assumed instead of the hypoexponential distribution. In this chapter, we only consider processing times with $CV<1$.

In the following subsections, the influence of batch sizes, *CV* of processing times and machine unreliability on the system performance is discussed.

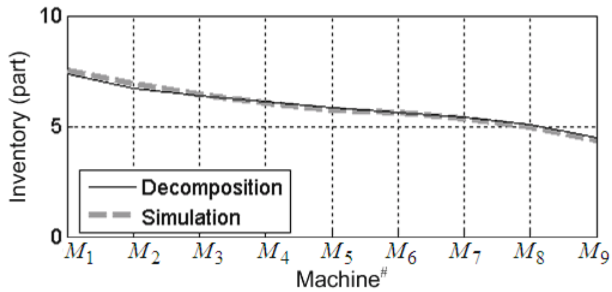
- **Influence of Batch Size**

To understand the influence of batch sizes, we first use Case A as a base case where all machines have batch sizes of one (single-item machines). In Case B, we then increase the batch sizes of machines 1, 3, 5, 7, and 9 to two, as in Table 5.2, but maintain the mean processing rate of 1 part/min by adjusting the mean processing time of these batch machines to 2 mins/batch such that the mean processing time of every machine is identical for both Cases A and B. Based on the results in Table 5.4, we observe that the system in Case B has relatively lower production rate and longer lead time than for Case A. In Case C, we further increase the batch size of these machines (machines 1, 3, 5, 7, and 9) to four, while maintaining the mean processing rate of 1 part/min for each machine. The results in Table 5.4 indicate that the deterioration of system performance increases in Case C. The deterioration of performance is mainly because of our assumption that batch machines do not commence processing until a full batch is available, which may increase the starvation and blockage of the system and thereby reduce the production rate and increase the waiting time of parts in the buffers.

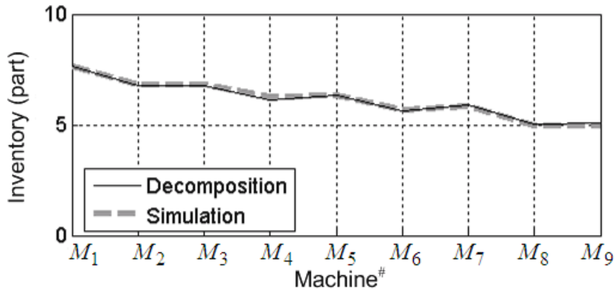
Figure 5.6 shows graphical plots of the average inventory levels in each machine and its immediately downstream buffer for Cases A to C. It may be observed that as the batch size increases, the average inventory levels tend to increase. These results show that although the mean processing rate of every machine is identical in Cases A, B and C, the increased batch sizes in Cases B and C increase the waiting time of parts in buffers and lead to increased average inventory levels.

Table 5.4. Comparison of performance measures and CPU times between decomposition model (Dec) and simulation (Sim). Diff=100%·|Dec-Sim|/Sim. Ten simulation runs were performed for each simulation.

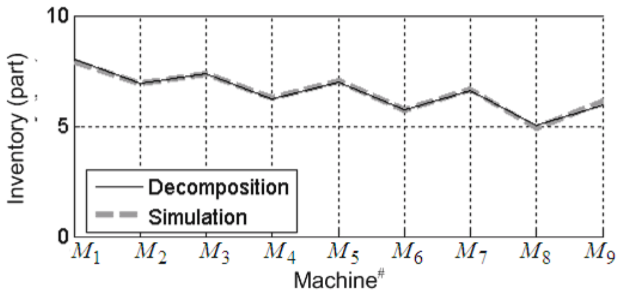
Experiment		Production rate	Total inventory	Production lead time	CPU time (per run)/sec
CaseA	Dec	0.93681	54.072	57.720	0.24
	Sim 1 (lognormal)	0.93479 ± 0.00033	54.852 ± 0.096	58.678 ± 0.101	111.32
	Diff 1 (%)	0.2155	1.4211	1.6331	
	Sim 2 (gamma)	0.93593 ± 0.00028	54.612 ± 0.102	58.350 ± 0.109	112.54
	Diff 2 (%)	0.0935	0.9879	1.0804	
	Sim 3 (hypoexponential)	0.94001 ± 0.00037	55.260 ± 0.089	58.786 ± 0.097	116.31
	Diff 3 (%)	0.3408	2.1498	1.8133	
Case B	Dec	0.89705	56.283	62.742	0.32
	Sim 1 (lognormal)	0.89673 ± 0.00048	56.615 ± 0.105	63.135 ± 0.113	119.34
	Diff 1 (%)	0.0355	0.5859	0.6213	
	Sim 2 (gamma)	0.89926 ± 0.00050	56.508 ± 0.102	62.839 ± 0.109	117.15
	Diff 2 (%)	0.2455	0.39862	0.15351	
	Sim 3 (hypoexponential)	0.90544 ± 0.00048	56.771 ± 0.097	62.699 ± 0.121	118.55
	Diff 3 (%)	0.9268	0.8591	0.0684	
Case C	Dec	0.79523	59.773	75.164	0.37
	Sim 1(lognormal)	0.80849 ± 0.00059	59.856 ± 0.106	74.035 ± 0.117	119.13
	Diff 1 (%)	1.6402	0.1397	1.5255	
	Sim 2 (gamma)	0.81215 ± 0.00055	60.016 ± 0.117	73.898 ± 0.159	120.31
	Diff 2 (%)	2.0827	0.4046	1.7138	
	Sim 3 (hypoexponential)	0.81218 ± 0.00052	60.357 ± 0.102	74.3148 ± 0.133	118.52
	Diff 3 (%)	2.0869	0.96774	1.1427	
Case D	Dec	0.87043	65.207	74.913	0.54
	Sim (lognormal)	0.88189 ± 0.00038	66.255 ± 0.105	75.128 ± 0.125	113.26
	Diff (%)	1.2998	1.5817	0.2855	
Case E	Dec	0.78518	64.063	81.590	0.42
	Sim (lognormal)	0.80177 ± 0.00045	65.107 ± 0.093	81.204 ± 0.146	117.85
	Diff (%)	2.0695	1.6033	0.47609	
Case F	Dec	0.70652	63.044	89.232	0.37
	Sim (lognormal)	0.72807 ± 0.00074	63.898 ± 0.098	87.764 ± 0.108	116.43
	Diff (%)	2.9598	1.3362	1.6715	
Case G	Dec	0.87797	56.979	64.898	0.44
	Sim (lognormal)	0.88492 ± 0.00061	57.481 ± 0.101	64.956 ± 1.112	117.33
	Diff (%)	0.7854	0.8742	0.0896	
Case H	Dec	0.76993	56.406	73.424	0.49
	Sim (lognormal)	0.79287 ± 0.00072	56.998 ± 0.106	71.889 ± 0.146	114.59
	Diff (%)	2.8932	1.0387	2.1353	
Case I	Dec	0.80667	56.408	69.927	0.42
	Sim (lognormal)	0.80571 ± 0.00031	55.978 ± 0.104	69.476 ± 0.136	120.31
	Diff (%)	0.1186	0.7676	0.6481	
Case J	Dec	0.72027	56.307	78.175	0.47
	Sim (lognormal)	0.73925 ± 0.00035	56.695 ± 0.081	76.693 ± 0.122	119.22
	Diff (%)	2.5665	0.6838	1.9323	



(a) Case A (lognormal processing times)



(b) Case B (lognormal processing times)



(c) Case C (lognormal processing times)

#A machine and its immediately downstream buffer.

Figure 5.6. The average inventory in each machine and its immediately downstream buffer for machines M_1 to M_9 in Cases A to C (with lognormal processing times). For machine M_{10}, parts that complete processing enter the finished goods buffer and are immediately removed. Hence, M_{10} will not contribute to the inventory of the system.

- **Influence of *CV***

The *CV* of processing times has a significant impact on the performance of a system. The comparison of results between Cases D, E, and F provided in Table 5.4 indicates that an increase in the *CV* of processing time in a system leads to a decrease in the production rate and an increase in the production lead time. This is due to the increased propagation of starvation and blockage in the system as a result of the increase in processing time variability.

- **Influence of Machine Unreliability**

Comparing the results of Cases G and H, and Cases I and J, we observe that machine unreliability undermines production rate and increases production lead time in both the balanced and unbalanced systems. This is because random machine failures also increase the variability in the system and hence the probability of starvation and blockage, leading to a deterioration of performance. The effects of random machine failures are therefore similar to the effects due to an increase in the *CV* of machine processing time.

5.4. A Case Study for Determining Batch Size of Machines

The performance measures obtained based on the proposed model may provide line managers with deeper insights, which may facilitate the improvement of control and configuration of the multistage manufacturing system with batch processing. This model may be used to determine the control parameters of the system. For instance, the proposed model may be applied to determine the batch sizes of machines. In some systems where consecutive processing machines are located far apart, material transfer is conducted in batches to reduce the transportation cost. Although choosing large batch sizes for such machines reduces the operation cost, it also requires longer waiting time for accumulating a full batch of parts, and this may reduce production rate of the system.

To better illustrate the application, we shall consider a simple system with a number of transportation facilities, dedicated to the material transfer at different locations of the system. These transportation facilities are modeled as batch machines, the set of indices of these machines is denoted as Γ (for example, if machines M_3 and M_6 are transportation machines, $\Gamma = \{3,6\}$). The batch sizes of the transportation machines (i.e. Q_k , $k \in \Gamma$) are the decision variables to be determined. The determination of Q_k , $k \in \Gamma$ is formulated as the following optimization problems. The first problem is intended to maximize production rate, and the second problem considers the maximization of expected profit.

Performance Enhancement Problem 5-1:

Determining Batch Size for Maximizing Production Rate (*PR*)

Maximize: $$PR = \sum_{\alpha_{K-1}} \sum_{n^d_{K-1}} \sum_{x_{K-1}} P\left(x_{K-1}, n^u_{K-1}, N^d_{K-1}, \alpha_{K-1}, 1\right) \mu^d_{K-1, n^d_{K-1}} \cdot Q^d_{K-1} \qquad (5.10)$$

Subject to: $1 \le Q_k \le Q_k^{Max}$, $k \in \Gamma$

where Q_k^{Max} denotes the maximum batch size of transportation machine M_k .

Performance Enhancement Problem 5-2:

Determining Batch Size for Maximizing Expected Profit (*EP*)

Maximize: $$EP = Price \cdot PR - Cost^{WIP} \cdot WIP - \sum_{k=1}^{K} \left(\frac{PR}{Q_k} \cdot Cost_k^{Batch} \right) \qquad (5.11)$$

Subject to: $1 \le Q_k \le Q_k^{Max}$, $k \in \Gamma$

where $Price$ is the unit price, $Cost^{WIP}$ represents the *WIP* holding cost per part, and $Cost_k^{Batch}$ denotes the operation cost per batch in machine M_k. $\frac{PR}{Q_k} \cdot Cost_k^{Batch}$ is the operation cost per unit time of machine M_k .

- **Numerical example**

In this example, an eight-machine system with two material transportation machines, M_3 and M_6, as illustrated in Figure 5.7 is considered. The other machines are single-item processing machines. The batch sizes of M_3 and

M_6 (i.e. Q_3 and Q_6) are the decision variables to be determined. The maximum batch sizes of M_3 and M_6 are both 18 parts. The size of each buffer is 20. *Price* is 1000 dollars/part and $Cost^{WIP}$ is chosen to be 1 dollar/(part·min). The other parameters are summarized in Table 5.5.

To provide an intuitive description of the influence of Q_3 and Q_6 on the performance of the system, we enumerate the combinations of Q_3 and Q_6, and obtain the corresponding performance measures based on the proposed model. The production rate and profit are illustrated in Figure 5.8. Based on the surfaces in Figure 5.8, we may identify the batch sizes that lead to the highest production rate and highest profit, as provided in Table 5.6. When Q_3=6 and Q_6=7, production rate is maximized. However, if the operation cost is considered, we may choose Q_3=9 and Q_6=11. By increasing the batch sizes of these two machines, the operation cost is reduced, which may compensate the decrement of production rate and result in a higher profit.

Table 5.5. Parameters for the application problem.

	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8
Mean processing time (min)	1	1	4	1	1	4	1	1
CV of processing time	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Batch size	1	1	**?**	1	1	**?**	1	1
Operation cost (dollar per batch)	20	20	200	20	20	700	20	20

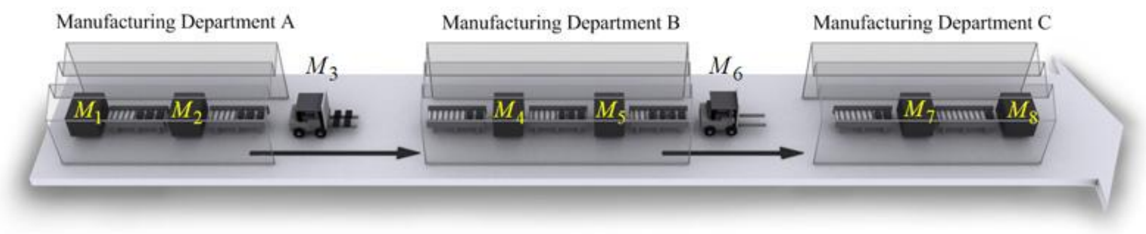
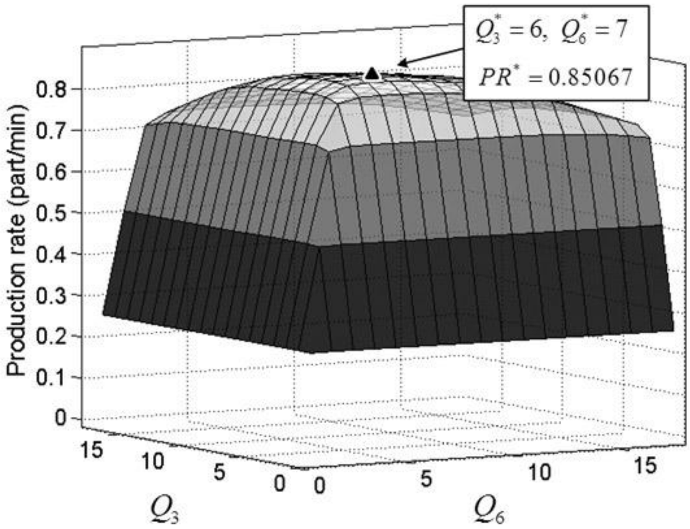
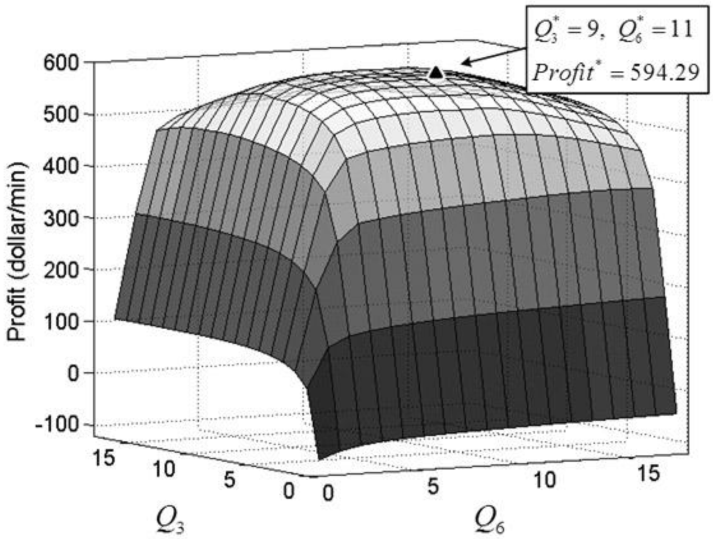


Figure 5.7. The manufacturing system studied in the example.



(a) Production rate vs. Q_3 and Q_6.



(b) Profit vs. Q_3 and Q_6.

Figure 5.8. Production rate and profit per minute under different batch sizes.

Table 5.6. Solutions for Q_3 and Q_6 .

Objective	Q_3	Q_6	Production rate	Profit
Maximize production rate	6	7	0.85067	573.92
Maximize profit	9	11	0.84657	594.29

Chapter 6.

Future Research Opportunities

6.1. Overview

Throughout the research, the author has observed the contrast between the prevalence of multistage manufacturing systems in modern industry and the lack of analytical studies on modeling and performance enhancement for such systems. The models presented in this thesis provide the mathematical tools to estimate performance measures of multistage manufacturing systems, based on which one may improve the control and configuration of the systems. The case studies of Chapters 3, 4, and 5 demonstrate that the proposed models may be used to investigate managerial problems in multistage manufacturing systems, such as determining the frequency of preventive maintenance, allocating inspection machines, and choosing batch sizes of machines. In addition to these problems, there are many research problems in this area remaining to be explored and solving these problems may require further extension of the models. In the remainder of this chapter, several promising research opportunities relevant to this study are highlighted.

6.2. Preventive Maintenance with Variable Machine State Inspection Rate

In Section 3.10, the author discussed a multistage manufacturing system where the state of each machine is inspected at a constant rate, and preventive maintenance is triggered when the maintenance operator detects that a machine has deteriorated to a specific level. This preventive maintenance strategy is also reflected in Figure 6.1. To further reduce the operation cost associated with machine state inspection, the inspection rate may be varied based on the inspection result, as illustrated in Figure 6.2. When a machine is in a relatively good condition, low machine state inspection rate is adopted. Once the maintenance operator detects that a machine has deteriorated to a worse state between two consecutive inspections, the inspection rate is increased. Similar preventive maintenance strategy is mentioned by Bloch-Mercier (2002) based on the single-machine system. However, in the multistage manufacturing system, this preventive maintenance strategy has been explored limitedly. The development of the analytical model incorporating this preventive maintenance strategy may be a potential research problem. Such a model will facilitate the maintenance operator to determine a suitable machine state inspection rate for each upstate of every machine in a multistage manufacturing system, and hence maximize profit of the system.

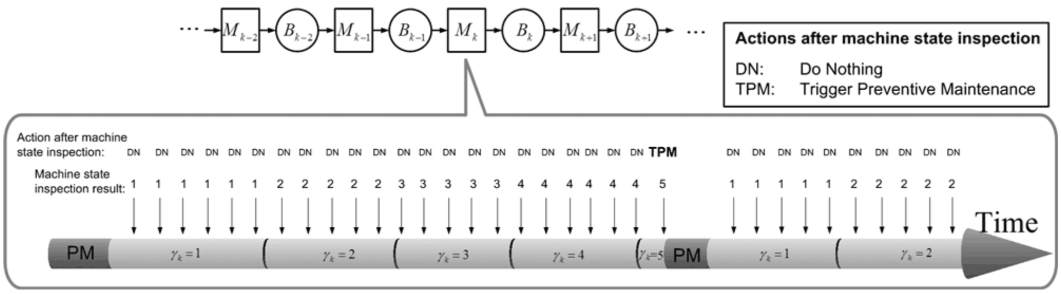


Figure 6.1. Preventive maintenance with constant machine state inspection rate. Preventive maintenance is triggered when machine inspection detects that the machine has deteriorated to a specific level.

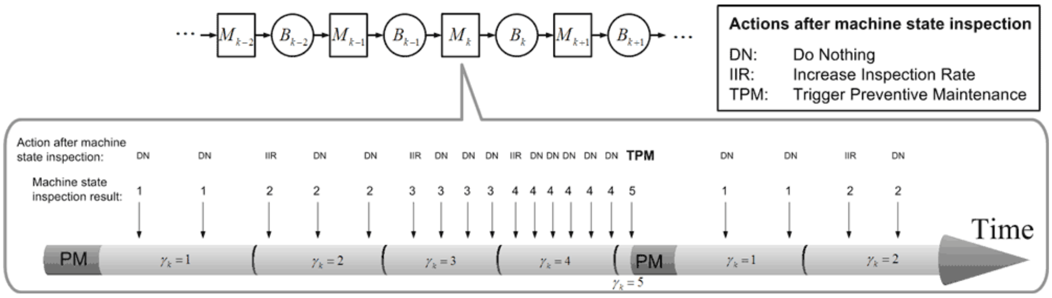


Figure 6.2. Preventive maintenance with variable machine state inspection rate. Machine state inspection rate is increased when the maintenance operator detects that the machine deteriorates to a worse level between two consecutive inspections. Preventive maintenance is triggered when maintenance operator detects that a machine has deteriorated to a specific level.

6.3. Preventive Maintenance with Consideration of Inventory

In the research reported in this thesis, the author has not accounted for the upstream or downstream inventory of a machine in the decision of preventive maintenance. Since inventory in the buffer may effectively reduce the impact of random production interruption, a preventive maintenance strategy that accounts for inventory may further improve production rate of a manufacturing system (Abboud, 2001). Performing preventive maintenance on a machine when its immediately downstream buffer has low inventory may increase the likelihood that starvation occurs in the downstream system. To minimize the starvation, preventive maintenance should be initiated on a machine only when it has high downstream inventory. Similarly, performing preventive maintenance on a machine when its immediately upstream buffer is near full may result in blockage of the upstream system. Hence, preventive maintenance under this condition should be avoided in order to reduce the likelihood of blockage. To account for inventory in the decision of preventive maintenance, the extension of the decomposition model presented in Chapter 3 is needed.

6.4. Imperfect Production and Repair or Rework of Defective Parts

The integrated quantitative and qualitative model presented in Chapter 4 is based on the assumption that defective parts are rejected. However, in some real systems, scraping defective parts is costly, and the value of defective parts may be salvaged via repair or rework (Rau, et al. 2005). Figure 6.3 illustrates a production line where repair machines are placed after the inspection machines. After a defective part is repaired, it proceeds to the next processing machine. Figure 6.4, on the other hand, illustrates a production line where defective parts are sent back for reworking. As both the repair and rework mechanisms may induce a substantial influence on the quantitative and qualitative performance of a multistage manufacturing system, the analytical model introduced in Chapter 4 may not be adequate to address repair or rework of defective parts.

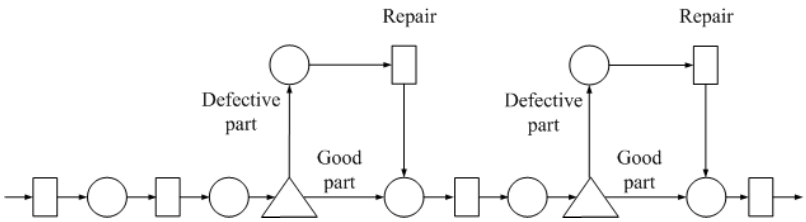


Figure 6.3. A production line where defective parts are repaired.

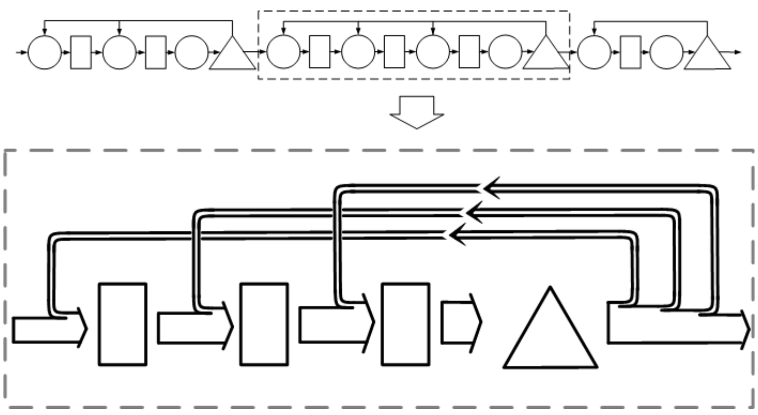


Figure 6.4. A production line where defective parts are reworked.

6.5. Performance Enhancement of other Complex Multistage Manufacturing Systems

In this thesis, the author limits the focus on two of the most common multistage manufacturing systems, namely assembly lines and serial production lines. In addition to these systems, multistage manufacturing systems of other configurations are also employed in the industry. For instance, reentrant manufacturing systems, as illustrated in Figure 6.5, have been widely used in wafer fabrication (Kumar, 1993). In a reentrant system, some downstream parts may be sent back to the upstream system, as these parts require the processing of certain machines more than once. The disassembly line (Gershwin, 1994) is another type of multistage manufacturing system, which is common in recycling industry. As illustrated in Figure 6.6, parts are disaggregated into a series of subparts in a disassembly line. Since the modeling and analysis of a manufacturing system are fundamentally influenced by its topology (Kim, 2005), the models presented in the thesis may not be directly applicable for reentrant and disassembly lines. It is therefore desirable to explore the variations of the proposed models for such systems. This will provide the mathematical tools for solving managerial problems, such as preventive maintenance, inspection allocation, etc, in reentrant or disassembly lines.

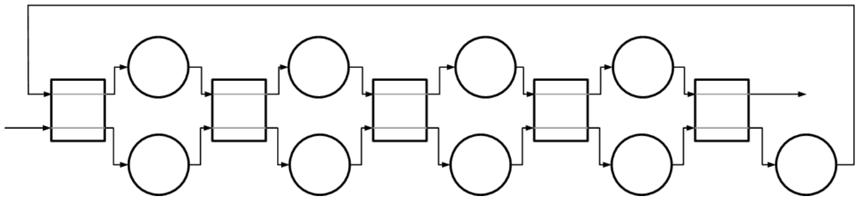


Figure 6.5. Reentrant system.

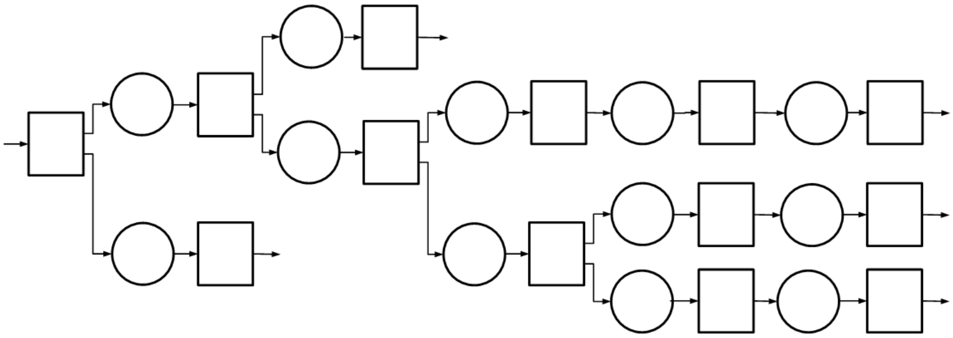


Figure 6.6. Disassembly line.

6.6. Modeling Manufacturing Systems with Uncertain Supply

The models of multistage manufacturing systems presented in Chapters 3 to 5 assume that raw materials are always available. Although this assumption is commonly used in the literature, it may not always be satisfied in practice. Since supply delay may significantly undermine the performance of a multistage manufacturing system, incorporating this uncertainty may add more rigour to the model. To incorporate material supply in the decomposition model, an additional line segment, which includes the supplier, raw material buffer, and the first machine of the production line, may be introduced. Figure 6.7 illustrates an example of such a line segment, where the raw material inventory is monitored using the (s,Q) policy. Under this policy, if the raw material inventory is equal to or lower than the replenishment point s, the supplier receives a request for replenishment, and subsequently the raw materials with the quantity of Q parts will be delivered to the production line.

The line segment illustrated in Figure 6.7 may be characterized as a continuous-time-discrete-state Markov chain, for which balance equations can also be derived. Based on this line segment, performance measures relevant to supply, such as the raw material inventory and frequency of replenishment, may be estimated. The frequency of replenishment is an important performance indicator and it can be used to estimate the transportation cost associated with raw material replenishment. As demonstrated in Figure 6.8, transportation cost consumes a significant portion of revenue in a manufacturing system. By incorporating raw material supply, the decomposition model may be used to determine an appropriate replenishment quantity for a manufacturing system, which reduces transportation cost, whilst avoiding excessive raw material inventory.

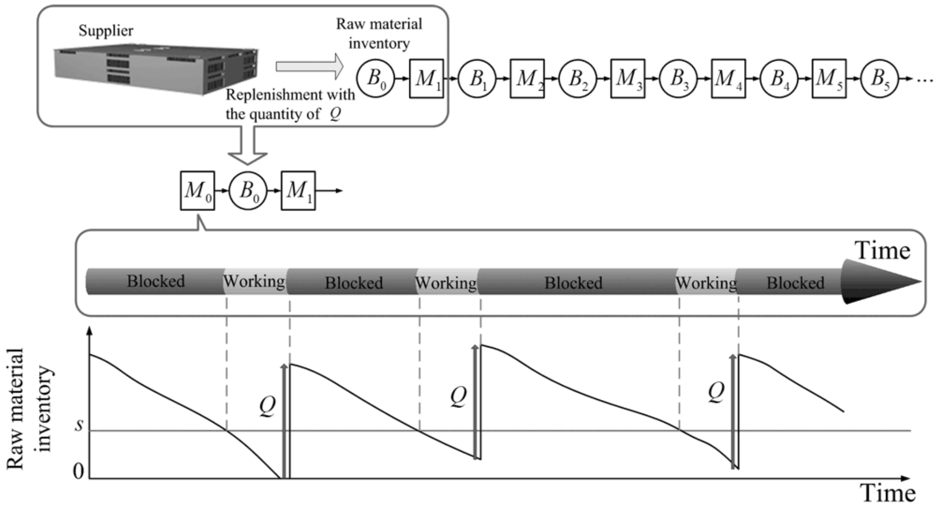


Figure 6.7. The supplier-buffer-machine line segment. The supplier is modeled as a batch machine. When the raw material inventory is lower than or equal to the replenishment point (s), a request of replenishment is sent to the supplier. Then, a batch of raw materials with the quantity of Q will be delivered to buffer B_0.

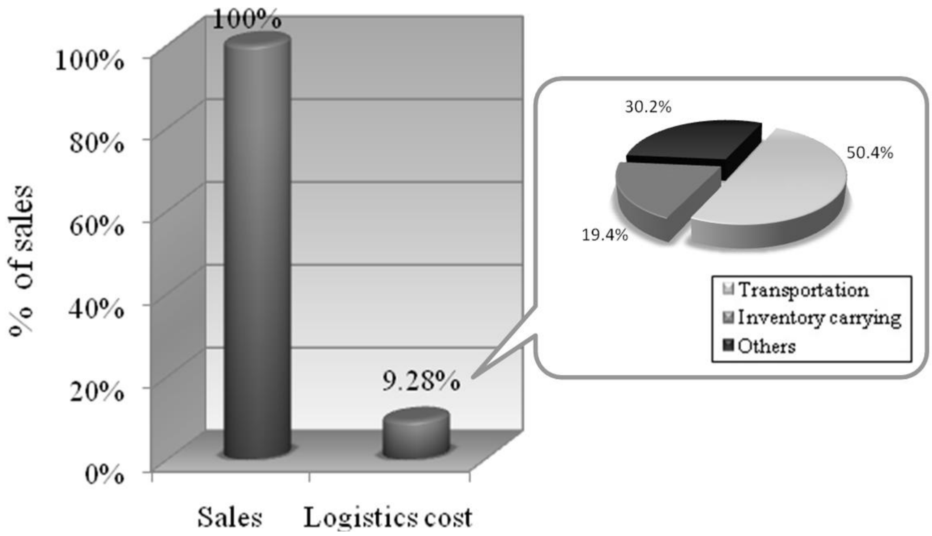


Figure 6.8. The logistics cost vs. total sales in an average manufacturing company (2008). The transportation and inventory cost 50.4% and 19.4% of the logistics investment. (Source: Logistics cost and service 2008).

6.7. Integration of Multi-factory Manufacturing Systems

In some industries, the manufacturing of a product is not completed in a single factory. Several factories collaborate closely and take charge of different parts of the manufacturing process. Materials are transferred between factories via the transportation system and hence these factories are connected to form a supply chain, as shown in Figure 6.9. In a multi-factory system, the production of each factory is influenced by its upstream or downstream partners. On the one hand, the starvation of material in an upstream factory may propagate through the supply chain and thus delay the production of downstream factories. On the other hand, the blockage of production in a downstream factory due to the overstock of inventory propagates upstream and prohibits manufacturing in other factories. To reduce inventory cost and avoid the risk of stock out for a factory, coordinating its production, inventory, and transportation frequency with upstream and downstream partners is necessary (Simchi-Levi et al., 2000). However, the uncertain characteristics of the supply chain (such as the uncertain processing times, uncertain inter-factory transportation times, random machine breakdowns, etc) make it difficult to predict the supplies of upstream partners and demand from downstream partners. To address this problem, an analytical model for the multi-factory manufacturing system may be developed based on the models presented in this thesis. This model may be used to investigate managerial problems in a multi-factory manufacturing system, such as determining the inventory in each factory for reducing holding costs and the transportation frequency between two connected factories for minimizing transportation costs.

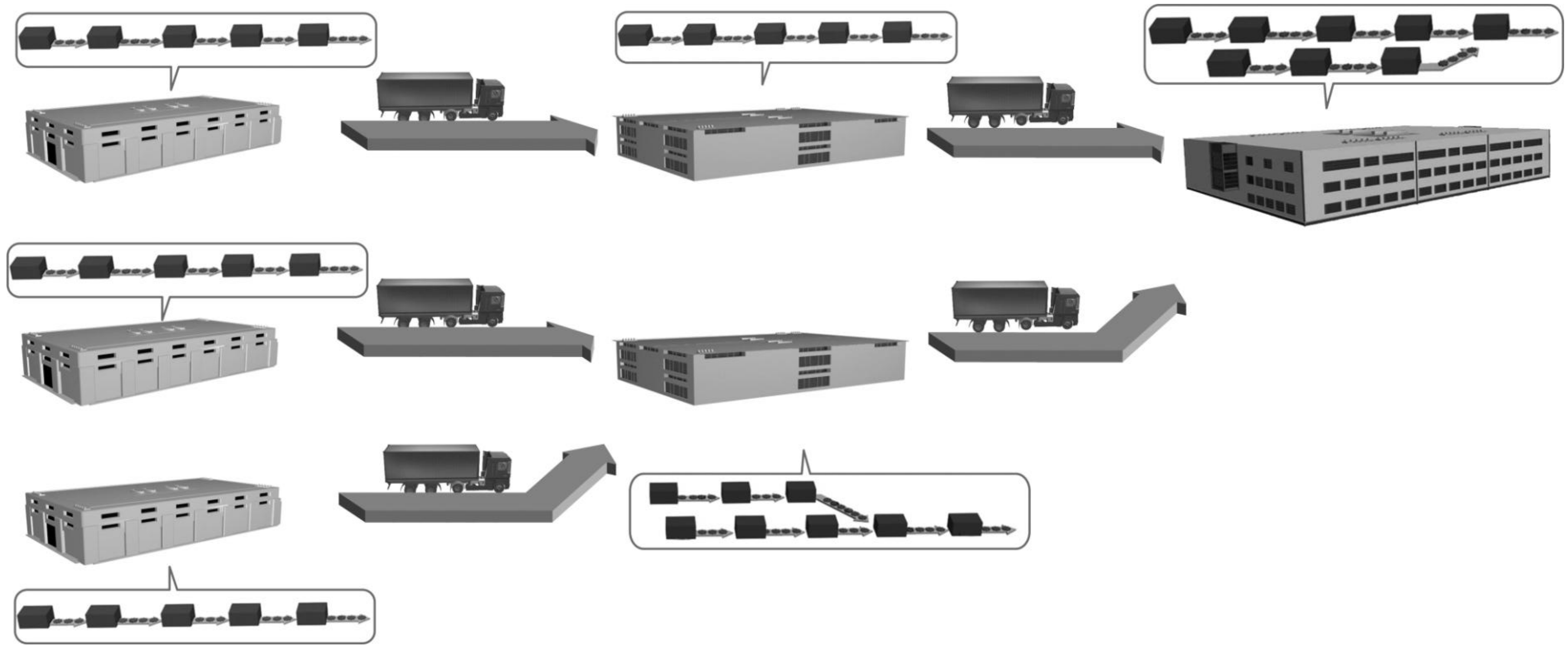


Figure 6.9. A multi-factory manufacturing system. Inside each factory, there is a multistage manufacturing system.

Chapter 7. Conclusions

In many manufacturing systems, machines with various functions are connected to form multistage networks. Machines in such systems may influence each other, which makes the quantitative and qualitative behavior of multistage manufacturing systems fundamentally different from single-machine systems. For instance, the failure of a machine may result in material starvation in its downstream machines and blockage of upstream machines, and hence prohibit their processing. To analyze the influence of machines on each other and predict the performance measures of the multistage manufacturing system, the author formulates a modeling framework based on the decomposition method (Gershwin, 1994). The modeling framework provides a mathematical tool to assess the impact of uncertainty on the performance of a multistage manufacturing system. Based on the modeling framework, one may develop efficient control or configuration scheme for the multistage manufacturing system.

The major contributions of this thesis are highlighted as follows:

1) In Chapter 3, the author investigates unreliable multistage manufacturing systems where machines are subjected to deterioration. Unlike previous analytical studies on multistage manufacturing systems with unreliable machines, the model formulated in Chapter 3 incorporates preventive maintenance. Since preventive maintenance is a strategy that has been successfully implemented in the manufacturing industry and since it

substantially improves machine reliability, incorporating this issue in the model is desirable.

Two major applications of the proposed model are discussed. First, it may be applied in the performance analysis of a multistage manufacturing system. The model provides the limiting probabilities of states of each primitive line segment, based on which production rate and *WIP* of the system can be estimated. Comparisons between analytical and simulation results in the numerical studies of Section 3.5 demonstrate that these estimates are of good accuracy. The second application of the proposed model is to determine the frequency of preventive maintenance for each machine in the multistage manufacturing system. Both insufficient and excessive maintenance results in a loss of machine capacity. The author formulates the determination of the frequency of preventive maintenance as an optimization problem to maximize production rate of the manufacturing system. An algorithm for solving this problem is also presented. The case study in Section 3.6 indicates that production rate of the system is substantially improved when the frequency of preventive maintenance as prescribed by the algorithm is adopted.

2) In Chapter 4, the author develops an integrated quantitative and qualitative model for multistage manufacturing systems with imperfect production and inspection errors based on the model presented in Chapter 3. One important feature of this model is that it can be used to estimate a variety of quantitative and qualitative performance measures. This feature distinguishes the proposed model from many previous models in the literature, which usually focus on predicting either the quantitative or

qualitative performance measures. In addition, the author also provides a time efficient algorithm for allocating inspection machines in a multistage manufacturing system. Inspection has a substantial influence on the performance of a multistage manufacturing system. Inspection may remove defective parts from the manufacturing process, resulting in a reduction of processing costs. On the other hand, excessive inspection may increase the inspection costs. Determining the placement of inspection machines in a multistage manufacturing system is a complex problem, as it may influence both the quantitative and qualitative performance measures. With the integrated quantitative and qualitative model, this influence may be evaluated comprehensively and hence a better solution to this problem can be provided. As the number of feasible solutions for the inspection allocation problem increases exponentially with an increase in the number of machines in the manufacturing system, determining the placement of inspection machines via exhaustive search may be prohibitive. Based on the proposed algorithm, the number of computations is substantially reduced and a good feasible solution for allocating inspection machines is provided.

3) The proposed modeling framework provides the flexibility and extensibility to incorporate various characteristics of multistage manufacturing systems that may be encountered in reality. For instance, in Chapter 5, the extension of the modeling framework for multistage manufacturing systems with batch operations and generally distributed processing times is investigated. Although batch operations are prevalent in industry, most previous studies of multistage manufacturing systems

have limited their focus on single-item operations. In addition, the assumption of generally distributed processing times represents the non-deterministic nature of many industrial processes due to such factors as random disturbances, operator inconsistencies etc. Such an extension may facilitate the application of the modeling framework to more complex manufacturing systems.

Bibliography

Abboud, N.E., 2001. Discrete-time Markov production-inventory model with machine breakdowns. Computers and Industrial Engineering, 39(1-2), 95-107.

Aguirre, A., Muller, E., Seffino, S., and Mendez, C.A., 2008. Applying a simulation-based tool to productivity management in an automotive-parts industry, in: Proceedings of the 2008 Winter Simulation Conference, Florida, USA, 1838-1846.

Alsyouf, I., 2006. Measuring maintenance performance using a balanced scorecard approach. Journal of Quality in Maintenance Engineering, 12 (2), 133–149.

Alden, J.M., Burns, L.D., Costy, T., Hutton, R.D., Jackson, C.A., Kim, D.S., Kohls, K.A., Owen, J.H., Turnquist, M.A., and Vander Veen, D.J., 2006. General Motors increases its production throughput. Interfaces, 36 (1), 6–25.

Alsyouf, I., 2009. Maintenance practices in Swedish industries: Survey results. International Journal of Production Economics, 121(1), 212-223.

Altiok, T., 1996. Performance Analysis of Manufacturing Systems, Springer.

Ambani, S., Li, L., and Ni, J., 2009. Condition-based maintenance decision-making for multiple machine systems. Journal of Manufacturing Science and Engineering, 131(3), p 031009 (9 pp).

Ancelin, B. and Semery, A., 1987. Calculation of the productivity of an integrated production line: CALIF, an industrial software package based on a new heuristic. Automatique Productique Informatique Industrielle, 21(3), 209-238.

Bai, D.S. and Yun, H.J., 1996. Optimal allocation of inspection effort in a serial multi-stage production system. Computers & Industrial Engineering, 30(3), 387-396.

Bao, H. and Jaishankar, D., 2008. Optimal preventive maintenance for a device submitted to random and deteriorating failures. 2008 IEEE International Conference on Industrial Engineering and Engineering Management, 491-495.

Baynat, B., Dallery, Y., Di Mascolo, M., and Frein, Y., 2001. A multi-class approximation technique for the analysis of kanban-like control systems. International Journal of Production Research, 39(2), 307-328.

Betterton, C.E. and Cox III, J.F., 2009. Espoused drum-buffer-rope flow control in serial lines: A comparative study of simulation models. International Journal of Production Economics, 117(1), 66-79.

Bhat, U.N., 1986. Finite capacity assembly-like queues. Queueing Systems, 1, 85 – 101.

Bloch-Mercier, S., 2002. A preventive maintenance policy with sequential checking procedure for a Markov deteriorating system. European Journal of Operational Research, 142 (3), 548-576.

Bolch, G., Greiner, S., Meer, H.D., Trivedi, K.S., 2006. Queueing Networks and Markov Chains, John Wiley & Sons, Inc.

Bonvik, A.M., Couch, C.E., and Gershwin, S.B., 1997. A comparison of production-line control mechanisms. International Journal of Production Research, 35(3), 789-804.

Bonvik, A.M., Dallery, Y., and Gershwin, S.B., 2000. Approximate analysis of production systems operated by a CONWIP/finite buffer hybrid control policy, International Journal of Production Research, 38(13), 2845-2869.

Britney, R.R., 1972. Optimal screening plans for nonserial production systems. Management Science, 18 (9), 550-559.

Burman, M.H., Gershwin, S.B., and Suyematsu, C., 1998. Hewlett-Packard uses operations research to improve the design of a printer production line. Interfaces, 28 (1), 24–36.

Buzacott, J.A. and Shanthikumar, J.G., 1993. Stochastic Models of Manufacturing Systems. Prentice-Hall: Englewood Cliffs, NJ.

Cao, Y.X., Subramaniam, V., and Chen, R.F., 2010. Integrated quantity and quality modeling of a flow line with assignable cause quality failures. Mathematical and Computer Modelling. Submitted to the journal and currently under review.

Carlson, J.G.H. and Yao, A.C., 2008. Simulating an agile, synchronized manufacturing system. International Journal of Production Economics, 112 (2) 714-722.

Canada/United States Manufacturing Perspective: Logistics and SCM KPI Analysis. http://www.ic.gc.ca/eic/site/dsib-logi.nsf/eng/h_pj00220. html.

Chen, A. and Wu, G.S., 2007. Real-time health prognosis and dynamic preventive maintenance policy for equipment under aging Markovian deterioration. International Journal of Production Research, 45(15), 3351-3379.

Chen, D. and Trivedi, K.S., 2005. Optimization for condition-based maintenance with semi-Markov decision process. Reliability Engineering & System Safety, 90(1), 25-29.

Chen, R.F. and Subramaniam, V., 2010. Inspection allocation in Kanban controlled assembly lines for simultaneously improving quantitative and qualitative performance. International Journal of Production Economics. Submitted to the journal and currently under review.

Chen, R.F. and Subramaniam, V., 2010. Increasing production rate in Kanban controlled assembly lines through preventive maintenance. International Journal of Production Research. In press.

Chen, R.F., Subramaniam V., and Senanayake, C.D., 2010. Performance evaluation of multi-stage manufacturing systems with batch operations and generally distributed processing times. Mathematical and Computer Modelling. Submitted to the journal and currently under review.

Chiang, S.Y., Kuo, C.T., Lim, J.T. and Meerkov, S.M., 2000. Improvability theory of assembly systems. I: Problem formulation and performance evaluation. Mathematical Problems in Engineering, 6, 321–357.

Colledani, M., Ekvall, M., Lundholm, T., Moriggi, P., Polato, A., and Tolio, T., 2010. Analytical methods to support continuous improvements at Scania. International Journal of Production Research, 48(7), 1913-1945.

Colledani, M., Gandola, F., Matta, A., and Tolio, T., 2008. Performance evaluation of linear and non-linear multi-product multi-stage lines with unreliable machines and finite homogeneous buffers. IIE Transactions, 40(6), 612-626.

Colledani, M., Matta, A., and Tolio, T., 2005. Performance evaluation of production lines with finite buffer capacity producing two different products. OR Spectrum, 27(2-3), 243-263.

Colledani, M. and Tolio, T., 2006. Impact of quality control on production system performance. 55(1), 453-456.

Colledani, M. and Tolio, T., 2009. Performance evaluation of production systems monitored by statistical process control and off-line inspections. 120(2), 348-367.

Dallery, Y. and Gershwin, S.B., 1992. Manufacturing flow line systems: a review of models and analytical results. Queueing Systems Theory and Applications, 12(1-2), 3-94.

Drezner, Z., Gurnani, H., and Akella, R., 1996. Capacity planning under different inspection strategies, European Journal of Operational Research, 89, 302–312.

Freiesleben, J., 2006. Costs and benefits of inspection systems and optimal inspection allocation for uniform defect propensity. International Journal of Quality Reliability Management, 23(5), 547-563.

Freiheit, T., Wang, W., and Patrick, S., 2007. A case study in productivity–cost trade-off in the design of paced parallel production systems. International Journal of Production Research, 45 (14), 3263–3288.

Garg, A. and Deshmukh, S.G., 2006. Maintenance management: Literature review and directions. Journal of Quality in Maintenance Engineering, 12(3), 205-238.

Gerold, J., 2004. Throughput, Productivity Top KPI List. http://www. automationworld. com /research-589.

Gershwin, S.B., 1987. Representation and analysis of transfer lines with machines that have different processing rates. Annals of Operations Research, 9, 511-530.

Gershwin, S.B. and Burman, M.H., 2000. A decomposition method for analyzing inhomogeneous assembly/disassembly systems. Annals of Operations Research, 93, 91-115.

Gershwin, S.B., 1991. Assembly/disassembly systems: an efficient decomposition algorithm for trees-structured networks. IIE Transactions, 23(4), 302-314.

Gershwin, S.B., 1994. Manufacturing Systems Engineering, Prentice Hall.

Gershwin, S.B., 2000. Design and operation of manufacturing systems: the control-point policy. IIE Transactions, 32(10), 891-906.

Gershwin, S.B., 2009. Uncertainty, variability, and randomness are the enemies of manufacturing. http://web.mit.edu/lmp/news/summit09/gershwin.pdf

Gross, D. and Harris, C.M., 1998. Queueing Theory, John Wiley & Sons, Inc.

Gurgur, C.Z. and Altiok, T., 2007. Analysis of decentralized multi-product pull systems with lost sales. Naval Research Logistics, 54(4), 357-370.

Gurgur, C.Z. and Altiok, T., 2008. Decentralized multi-product multi-stage systems with backorders. IIE Transactions, 40(3), 238-251.

Gurler, U. and Kaya, A., 2002. A maintenance policy for a system with multi-state components: an approximate solution. Reliability Engineering & System Safety, 76(2), 117-127.

Hao, Q. and Shen, W., 2008. Implementing a hybrid simulation model for a kanban-based material handling system. Robotics and Computer-Integrated Manufacturing, 24(5), 635-646.

Heredia-Langner, A., Montgomery, D.C., and Carlyle, W.M., 2002. Solving a multistage partial inspection problem using genetic algorithms. International Journal of Production Research, 40(8), 1923-1940.

Hopp, W.J. and Simon, J.T., 1989. Bounds and heuristics for assembly-like queues. Queueing Systems, 4, 137 – 156.

Hopp, W.J. and Spearman, M.L., 2000. Factory Physics: Foundations of Manufacturing Management, second ed., Irwin/McGraw-Hill, Boston, USA.

Hosoda, T.T. and Disney, S.M., 2006. On variance amplification in a three-echelon supply chain with minimum mean square error forecasting. Omega, 34(4), 344-358.

Inman, R.R., Blumenfeld, D.E., Huang, N., and Li, J., 2003. Designing production systems for quality: research opportunities from an automotive industry perspective, International journal of production research, 41(9), 1953-1971.

Kakade, V., Valenzuela, J.F., and Smith, J.S., 2004. An optimization model for selective inspection in serial manufacturing systems. International Journal of Production Research, 42(18), 3891-3909.

Kenne, J.P. and Gharbi, A., 1999. Experimental design in production and maintenance control problem of a single machine, single product manufacturing system. International Journal of Production Research, 37(3), 621-637.

Kiefer, J., 1953. Sequential minimax search for a maximum, Proceedings of the American Mathematical Society, 4, 502–506.

Kim, J. and Gershwin, S.B., 2005. Integrated quality and quantity modeling of a production line. OR Spectrum, 27(2-3), 287-314.

Kim, J., 2005. Integrated Quality and Quantity Modelling of a Production Line, Ph.D. thesis dissertation, Operations Research Center, Massachusetts

Institute of Technology.

Kim, J. and Gershwin, S.B., 2008. Analysis of long flow lines with quality and operational failures. IIE Transactions, 40(3), 284-296.

Kumar, C.S. and Panneerselvam, R., 2007. Literature review of JIT-KANBAN system. International Journal of Advanced Manufacturing Technology, 32(3-4), 393-408.

Kumar P.R., 1993. Re-entrant lines, in Queuing systems: Theory and applications: Special issues on queuing networks, 13, 87-110.

Kuo, C.T., Lim, J.T., Meerkov, S.M., and Park, E., 1997. Improvability theory for assembly systems: two component–one assembly machine case. Mathematical Problems in Engineering, 3, 95–171.

Kyriakidis, E.G. and Dimitrakos, T.D., 2006. Optimal preventive maintenance of a production system with an intermediate buffer. European Journal of Operational Research, 168(1), 86-99.

Lee, J. and Unnikrishnan, S., 1998. Planning quality inspection operations in multistage manufacturing systems with inspection errors. International Journal of Production Research, 36(1), 141-155.

Lee, W.R., Beruvides, M.G., and Chiu, Y.D., 2007. A study on the quality-productivity relationship and its verification in manufacturing industries. Engineering Economist, 52(2), 117-139.

Levantesi, R., Matta, A., and Tolio, T., 2003. Performance evaluation of continuous production lines with machines having different processing times and multiple failure modes. Performance Evaluation, 51(2-4), 247-268.

Li, J. and Huang, N., 2005. Modelling and analysis of a multiple product manufacturing system with split and merge. International Journal of Production Research, 43(19), 4049-4066.

Li, J., Blumenfeld, D.E., Huang, N., and Alden, J.M., 2009. Throughput analysis of production systems: recent advances and future topics, International Journal of Production Research, 47 (14), 3823-3851.

Li, J. and Meerkov, S.M., 2009. Production systems engineering. New York, NY: Springer.

Liberopoulos, G., Papadopoulos, C.T., Tan, B., Smith, J.M., and Gershwin, S.B., 2006. Stochastic Modeling of Manufacturing Systems: Advances in Design, Performance Evaluation, and Control Issues. Springer.

Liberopoulos, G. and Tsarouhas, P., 2002. Systems analysis speeds up Chipita's food-processing line. Interfaces, 32 (3), 62–76.

Little, J.D.C., 1961. A proof for the queuing formula $L=\lambda W$, Operations Research, 9, 383–387.

Logistics Cost and Service 2008, http://www.establishinc.com/pdfs/2008_CSCMP_ Presentation.pdf

Mandroli, S.S., Shrivastava, A.K., and Ding, Y., 2006. A survey of inspection strategy and sensor distribution studies in discrete-part manufacturing processes. IIE Transactions, 38(4), 309-328.

Manitz, M., 2008. Queueing-model based analysis of assembly lines with finite buffers and general service times. Computers and Operations Research, 35(8), 2520-2536.

Matta, A., Dallery, Y., and Di Mascolo, M., 2005. Analysis of assembly systems controlled with kanbans. European Journal of Operational Research, 166(2), 310-336, 16.

Meow, S.Y., 2001. Simulation of real-time scheduling policies in multi-product, make-to-order semiconductor fabrication facilities, MIT thesis.

Montgomery, D.C., 2001. Introduction to Statistical Quality Control. John Wiley & Sons, Inc.

Montoro-Cazorla, D. and Pérez-Ocón, R., 2006. A deteriorating two-system with two repair modes and sojourn times phase-type distributed. Reliability Engineering and System Safety, 91(1), 1-9.

Moustafa, M.S., Maksoud, E.Y.A., and Sadek, S., 2004. Optimal major and minimal maintenance policies for deteriorating systems. Reliability Engineering & System Safety, 83(3), 363-368.

Patchong, A., Lemoine, T., and Kern, G., 2003. Improving car body production at PSA Peugeot Citroen. Interfaces, 33 (1), 36–49.

Panagiotidou, S., and Tagaras, G., 2007. Optimal preventive maintenance for equipment with two quality states and general failure time distributions. European Journal of Operational Research, 180(1), 329-353.

Pavitsos, A. and Kyriakidis, E.G., 2009. Markov decision models for the optimal maintenance of a production unit with an upstream buffer. Computers & Operations Research, 36(6), 1993-2006.

Penn, M. and Raviv, T., 2007. Optimizing the quality control station configuration. Naval Research Logistics, 54(3), 301-314.

Penn, M. and Raviv, T., 2008. A polynomial time algorithm for solving a quality control station configuration problem. Discrete Applied Mathematics, 156(4), 412-419.

Radhoui, M., Rezg, N., and Chelbi, A., 2009. Integrated model of preventive maintenance, quality control and buffer sizing for unreliable and imperfect production systems. International Journal of Production Research, 47(2), 389-402.

Rau, H. and Chu, Y. H., 2005. Inspection allocation planning with two types of workstation: WVD and WAD. International Journal of Advanced Manufacturing Technology, 25(9-10), 947-953.

Rau, H., Chu, Y.H., and Cho, K.H., 2005. Layer modelling for the inspection allocation problem in re-entrant production systems. International Journal of Production Research, 43(17), 3633-3655.

Rezg, N., Xie, X., and Mati, Y., 2004. Joint optimization of preventive maintenance and inventory control a production line using simulation. International Journal of Production Research, 42(10), 2029-2046.

Sakai, H. and Amasaka, K., 2007. The robot reliability design and improvement method and the advanced Toyota production system. Industrial Robot, 34(4), 310-316.

Sandanayake, Y.G., Oduoza, C.F., and Proverbs, D.G., 2008. A systematic modelling and simulation approach for JIT performance optimisation. Robotics and Computer-Integrated Manufacturing, 24(6), 735-743.

Sandanayake, Y.G. and Oduoza, C.F., 2009. Dynamic simulation for performance optimization in just-in-time-enabled manufacturing processes. International Journal of Advanced Manufacturing Technology, 42(3-4), 372-380.

Schmidt, K., and Rose, O., 2008. Simulation analysis of semiconductor manufacturing with small lot size and batch tool replacements, in: Proceedings of the 2008 Winter Simulation Conference, Florida, USA, 2118-2126.

Sevastyanov, B.A., 1962. Influence of storage bin capacity on the average standstill time of a production line. Theory of Probability and its Applications, 7 (4), 429-438.

Shi, Z. and Sandborn, P., 2003. Optimization of test/diagnosis/rework location and characteristics in electronic systems assembly using real-coded Genetic Algorithms. Proceedings of International Test Conference.

Shi, Z. and Sandborn, P., 2006. Optimization of test/diagnosis/rework location(s) and characteristics in electronic system assembly. Journal of Electronic Testing Theory and Applications, 22, 49-60.

Shina, S. G., 2002. Six Sigma for Electronics Design and Manufacturing. Boston : McGraw-Hill.

Silver, E.A., Pyke, D.F., and Peterson, R., 1998. Inventory Management and Production Planning and Scheduling, John Wiley & Sons, Inc.

Simchi-Levi, D., Kaminsky, P., and Simchi-Levi, E., 2000. Designing and Managing the Supply Chain, McGraw-Hill Companies, Inc.

Stevenson W.J., 1992. Introduction to management science, Burr Ridge, IL: Richard D.Irwin.

Stewart, W.J., 1994. Introduction to the Numerical Solution of Markov chains, Princeton University Press, Princeton, New Jersey, USA.

Subramaniam, V., Yang, R.L., Chen, R.F., and Singh, S.P., 2009. A WIP control Policy for Tandem Lines, International Journal of Production Research, 47(4), 1127-1149.

Takahashi, K., Myreshka, Hirotani, D., 2005. Comparing CONWIP, synchronized CONWIP, and Kanban in complex supply chains, International Journal of Production Economics, 93-94, SPEC.ISS., 25-40.

Taneja, M. and Viswanadham, N., 1994. Inspection allocation in manufacturing systems: a genetic algorithm approach. Proceedings - IEEE International Conference on Robotics and Automation, 3537-3542.

Tempelmeier, H., and Burger, M., 2001. Performance evaluation of unbalanced flow lines with general distributed processing times, failures and imperfect production. IIE Transactions, 33(4), 293-302.

Tolio, T., Matta, A., and Gershwin, S.B., 2002. Analysis of two-machine lines with multiple failure modes. IIE Transactions, 34(1), 51-62.

Trichy, T., Sandborn, P., Raghavan, R., and Sahasrabudhe, S., 2001. A new test/diagnosis/rework model for use in technical cost modeling of electronic systems assembly. Proceedings of the International Test Conference, 1108-1117.

Van Volsem, S., 2002. Optimizing inspection strategies for multi-stage process chains: A case study. In: 16th Triennial IFORS Conference. International Federation of Operational Research Societies.

Van Volsem, S., Dullaert, W., and Van Landeghem, H., 2007. An Evolutionary Algorithm and discrete event simulation for optimizing inspection strategies for multi-stage processes. European Journal of Operational Research, 179(3), 621-633.

Vereecke, A., Van Dierdonck, and R. De Meyer, A., 2006. A typology of plants in global manufacturing networks. *Management Science*, 52 (11), 1737-1750.

Viswanadham, N., Sharma, S.M., and Taneja, M., 1996. Inspection allocation in manufacturing systems using stochastic search techniques. IEEE Transactions on Systems, Man & Cybernetics, Part A (Systems &

Humans), 26(2), 222-230.

Wu, J. and Makis, V., 2008. Economic and economic-statistical design of a chi-square chart for CBM. European Journal of Operational Research, 188(2), 516-529.

Yang, R.L., Chen, R.F., Subramaniam, V., and Gershwin, S.B., 2006. Setting real time WIP levels in tandem line production systems. Proceedings of International Conference on Manufacturing Science and Technology, 506-509.

Yang, R.L., 2007. Control of manufacturing systems: the DYNWIP policy for production management. Thesis. Singapore-MIT Alliance.

Yao, X., Xie, X., Fu, M.C., and Marcus, S.I. 2005. Optimal joint preventive maintenance and production policies. Naval Research Logistics, 52(7), 668-681.

Zequeira, R.I., Prida, B., and Valdes, J.E., 2004. Optimal buffer inventory and preventive maintenance for an imperfect production process. International Journal of Production Research, 42(5), 959-974.

Zimmern, B. 1956. Etude de la propagation des arrêts aléatoires dans les chaînes de production. Revue de Statistique Appliquée, 4, 85-104.

Appendix A.

Balance Equations of the 2M1B Line with Machine Deterioration and Preventive Maintenance

As discussed in Section 3.3.1, the balance equations for the 2M1B line with machine deterioration and preventive maintenance may be divided into 16 groups listed in Table 3.1. These balance equations can be derived based on the discussion in Section 3.3.1, and they are summarized as follows:

1) Group 1. $\gamma_1 = 1, \gamma_2 = 1$

$$P(x_1,1,1)\cdot(\mu_1+\mu_2+p_{1,1}+p_{2,1}+\pi_{1,1}+\pi_{2,1}) = P(x_1-1,1,1)\mu_1 + P(x_1+1,1,1)\mu_2 + P(x_1,N_1+1,1)r_1 + P(x_1,N_1+2,1)\rho_1 + P(x_1,1,N_2+1)r_2 + P(x_1,1,N_2+2)\rho_2, \quad 0 < x_1 < X_1 \tag{A.1}$$

$$P(0,1,1)\cdot(\mu_1+p_{1,1}+\pi_{1,1}+\pi_{2,1}) = P(1,1,1)\mu_2 + P(0,N_1+1,1)r_1 + P(0,N_1+2,1)\rho_1 + P(0,1,N_2+2)\rho_2 \tag{A.2}$$

$$P(X_1,1,1)\cdot(\mu_2+p_{2,1}+\pi_{1,1}+\pi_{2,1}) = P(X_1-1,1,1)\mu_1 + P(X_1,N_1+2,1)\rho_1 + P(X_1,1,N_2+1)r_2 + P(X_1,1,N_2+2)\rho_2 \tag{A.3}$$

2) Group 2. $2 \le \gamma_1 \le N_1, \gamma_2 = 1$

$$P(x_1,\gamma_1,1)\cdot(\mu_1+\mu_2+p_{1,\gamma_1}+p_{2,1}+\pi_{1,\gamma_1}+\pi_{2,1}) = P(x_1-1,\gamma_1,1)\mu_1 + P(x_1+1,\gamma_1,1)\mu_2 + P(x_1,\gamma_1-1,1)p_{1,\gamma_1-1} + P(x_1,\gamma_1,N_2+1)r_2 + P(x_1,\gamma_1,N_2+2)\rho_2, \quad 0 < x_1 < X_1 \tag{A.4}$$

$$P(0,\gamma_1,1)\cdot(\mu_1+p_{1,\gamma_1}+\pi_{1,\gamma_1}+\pi_{2,1}) = P(1,\gamma_1,1)\mu_2 + P(0,\gamma_1-1,1)p_{1,\gamma_1-1} + P(0,\gamma_1,N_2+2)\rho_2 \tag{A.5}$$

$$P(X_1,\gamma_1,1)\cdot(\mu_2+p_{2,1}+\pi_{1,\gamma_1}+\pi_{2,1}) = P(X_1-1,\gamma_1,1)\mu_1 + P(X_1,\gamma_1,N_2+1)r_2 + P(X_1,\gamma_1,N_2+2)\rho_2 \tag{A.6}$$

3) Group 3. $\gamma_1 = N_1 + 1, \gamma_2 = 1$

$$P(x_1, N_1+1,1)\cdot(\mu_2 + r_1 + p_{2,1} + \pi_{2,1}) = P(x_1+1, N_1+1,1)\mu_2 + P(x_1, N_1,1)p_{1,N_1} +$$

$$P(x_1, N_1+1, N_2+1)r_2 + P(x_1, N_1+1, N_2+2)\rho_2, \quad 0 < x_1 < X_1 \tag{A.7}$$

$$P(0, N_1+1,1)\cdot(r_1 + \pi_{2,1}) = P(1, N_1+1,1)\mu_2 + P(0, N_1,1)p_{1,N_1} + P(0, N_1+1, N_2+2)\rho_2 \tag{A.8}$$

$$P(X_1, N_1+1,1) = 0 \tag{A.9}$$

4) Group 4. $\gamma_1 = N_1 + 2, \gamma_2 = 1$

$$P(x_1, N_1+2,1)\cdot(\mu_2 + \rho_1 + p_{2,1} + \pi_{2,1}) = P(x_1+1, N_1+2,1)\mu_2 + \sum_{n=1}^{N_1} P(x_1, n,1)\pi_{1,n} +$$

$$P(x_1, N_1+2, N_2+1)r_2 + P(x_1, N_1+2, N_2+2)\rho_2, \quad 0 < x_1 < X_1 \tag{A.10}$$

$$P(0, N_1+2,1)\cdot(\rho_1 + \pi_{2,1}) = P(1, N_1+2,1)\mu_2 + \sum_{n=1}^{N_1} P(0, n,1)\pi_{1,n} + P(0, N_1+2, N_2+2)\rho_2 \tag{A.11}$$

$$P(X_1, N_1+2,1)\cdot(\mu_2 + \rho_1 + p_{2,1} + \pi_{2,1}) = \sum_{n=1}^{N_1} P(X_1, n,1)\pi_{1,n} + P(X_1, N_1+2, N_2+1)r_2 +$$

$$P(X_1, N_1+2, N_2+2)\rho_2 \tag{A.12}$$

5) Group 5. $\gamma_1 = 1, 2 \le \gamma_2 \le N_2$

$$P(x_1,1,\gamma_2)\cdot(\mu_1 + \mu_2 + p_{1,1} + p_{2,\gamma_2} + \pi_{1,1} + \pi_{2,\gamma_2}) = P(x_1-1,1,\gamma_2)\mu_1 + P(x_1+1,1,\gamma_2)\mu_2 +$$

$$P(x_1, N_1+1,\gamma_2)r_1 + P(x_1, N_1+2,\gamma_2)\rho_1 + P(x_1,1,\gamma_2-1)p_{2,\gamma_2-1}, \quad 0 < x_1 < X_1 \tag{A.13}$$

$$P(0,1,\gamma_2)\cdot(\mu_1 + p_{1,1} + \pi_{1,1} + \pi_{2,\gamma_2}) = P(1,1,\gamma_2)\mu_2 + P(0, N_1+1,\gamma_2)r_1 + P(0, N_1+2,\gamma_2)\rho_1 \tag{A.14}$$

$$P(X_1,1,\gamma_2)\cdot(\mu_2 + p_{2,\gamma_2} + \pi_{1,1} + \pi_{2,\gamma_2}) = P(X_1-1,1,\gamma_2)\mu_1 + P(X_1, N_1+2,\gamma_2)\rho_1 +$$

$$P(X_1,1,\gamma_2-1)p_{2,\gamma_2-1} \tag{A.15}$$

6) Group 6. $2 \le \gamma_1 \le N_1, 2 \le \gamma_2 \le N_2$

$$P(x_1,\gamma_1,\gamma_2)\cdot(\mu_1 + \mu_2 + p_{1,\gamma_1} + p_{2,\gamma_2} + \pi_{1,\gamma_1} + \pi_{2,\gamma_2}) = P(x_1-1,\gamma_1,\gamma_2)\mu_1 + P(x_1+1,\gamma_1,\gamma_2)\mu_2 +$$

$$P(x_1,\gamma_1-1,\gamma_2)p_{1,\gamma_1-1} + P(x_1,\gamma_1,\gamma_2-1)p_{2,\gamma_2-1}, \quad 0 < x_1 < X_1 \tag{A.16}$$

$$P(0,\gamma_1,\gamma_2)\cdot\left(\mu_1+p_{1,\gamma_1}+\pi_{1,\gamma_1}+\pi_{2,\gamma_2}\right)=P(1,\gamma_1,\gamma_2)\mu_2+P(0,\gamma_1-1,\gamma_2)\,p_{1,\gamma_1-1} \tag{A.17}$$

$$P(X_1,\gamma_1,\gamma_2)\cdot\left(\mu_2+p_{2,\gamma_2}+\pi_{1,\gamma_1}+\pi_{2,\gamma_2}\right)=P(X_1-1,\gamma_1,\gamma_2)\mu_1+\ P(X_1,\gamma_1,\gamma_2-1)\,p_{2,\gamma_2-1} \tag{A.18}$$

7) Group 7. $\gamma_1 = N_1+1$**,** $2\le\gamma_2\le N_2$

$$P(x_1,N_1+1,\gamma_2)\cdot\left(\mu_2+r_1+p_{2,\gamma_2}+\pi_{2,\gamma_2}\right)=P(x_1+1,N_1+1,\gamma_2)\mu_2+P(x_1,N_1,\gamma_2)\,p_{1,N_1}+$$

$$P(x_1,N_1+1,\gamma_2-1)\,p_{2,\gamma_2-1}\,,\quad 0<x_1<X_1 \tag{A.19}$$

$$P(0,N_1+1,\gamma_2)\cdot\left(\mu_2+r_1+\pi_{2,\gamma_2}\right)=P(1,N_1+1,\gamma_2)\mu_2+P(0,N_1,\gamma_2)\,p_{1,N_1} \tag{A.20}$$

$$P(X_1,N_1+1,\gamma_2)=0 \tag{A.21}$$

8) Group 8. $\gamma_1 = N_1+2$**,** $2\le\gamma_2\le N_2$

$$P(x_1,N_1+2,\gamma_2)\cdot\left(\mu_2+\rho_1+p_{2,\gamma_2}+\pi_{2,\gamma_2}\right)=P(x_1+1,N_1+2,\gamma_2)\mu_2+\ \sum_{n=1}^{N_1}P(x_1,n,\gamma_2)\pi_{1,n}+$$

$$P(x_1,N_1+2,\gamma_2-1)\,p_{2,\gamma_2-1}\,,\quad 0<x_1<X_1 \tag{A.22}$$

$$P(0,N_1+2,\gamma_2)\cdot\left(\rho_1+\pi_{2,\gamma_2}\right)=P(1,N_1+2,\gamma_2)\mu_2+\sum_{n=1}^{N_1}P(0,n,\gamma_2)\pi_{1,n} \tag{A.23}$$

$$P(X_1,N_1+2,\gamma_2)\cdot\left(\mu_2+\rho_1+p_{2,\gamma_2}+\pi_{2,\gamma_2}\right)=\sum_{n=1}^{N_1}P(X_1,n,\gamma_2)\pi_{1,n}+\ P(X_1,N_1+2,\gamma_2-1)\,p_{2,\gamma_1-1} \tag{A.24}$$

9) Group 9. $\gamma_1 = 1$**,** $\gamma_2 = N_2+1$

$$P(x_1,1,N_2+1)\cdot\left(\mu_1+p_{1,1}+r_2+\pi_{1,1}\right)=P(x_1-1,1,N_2+1)\mu_1+P(x_1,N_1+1,N_2+1)r_1+$$

$$P(x_1,N_1+2,N_2+1)\rho_1+P(x_1,1,N_2)\,p_{2,N_2}\,,\quad 0<x_1<X_1 \tag{A.25}$$

$$P(0,1,N_2+1)=0 \tag{A.26}$$

$$P(X_1,1,N_2+1)\cdot\left(r_2+\pi_{1,1}\right)=P(X_1-1,1,N_2+1)\mu_1+\ P(X_1,N_1+2,N_2+1)\rho_1+P(X_1,1,N_2)\,p_{2,N_2} \tag{A.27}$$

10) Group 10. $2 \le \gamma_1 \le N_1$**,** $\gamma_2 = N_2 + 1$

$$P(x_1,\gamma_1,N_2+1)\cdot\left(\mu_1 + p_{1,\gamma_1} + r_2 + \pi_{1,\gamma_1}\right) = P(x_1-1,\gamma_1,N_2+1)\mu_1 + P(x_1,\gamma_1-1,N_2+1)\,p_{1,\gamma_1-1} +$$

$$P(x_1,\gamma_1,N_2)\,p_{2,N_2}, \quad 0 < x_1 < X_1 \tag{A.28}$$

$$P(0,\gamma_1,N_2+1) = 0 \tag{A.29}$$

$$P(X_1,\gamma_1,N_2+1)\cdot\left(\mu_1 + r_2 + \pi_{1,\gamma_1}\right) = P(X_1-1,\gamma_1,N_2+1)\mu_1 + P(X_1,\gamma_1,N_2)\,p_{2,N_2} \tag{A.30}$$

11) Group 11. $\gamma_1 = N_1 + 1$**,** $\gamma_2 = N_2 + 1$

$$P(x_1,N_1+1,N_2+1)\cdot(r_1+r_2) = P(x_1,N_1,N_2+1)\,p_{1,N_1} + P(x_1,N_1+1,N_2)\,p_{2,N_2}, \quad 0 < x_1 < X_1 \tag{A.31}$$

$$P(0,N_1+1,N_2+1) = 0 \tag{A.32}$$

$$P(X_1,N_1+1,N_2+1) = 0 \tag{A.33}$$

12) Group 12. $\gamma_1 = N_1 + 2$**,** $\gamma_2 = N_2 + 1$

$$P(x_1,N_1+2,N_2+1)\cdot(\rho_1+r_2) = \sum_{n=1}^{N_1} P(x_1,n,N_2+1)\pi_{1,n} + P(x_1,N_1+2,N_2)\,p_{2,N_2}, \; 0 < x_1 < X_1 \tag{A.34}$$

$$P(0,N_1+2,N_2+1) = 0 \tag{A.35}$$

$$P(X_1,N_1+2,N_2+1)\cdot(\rho_1+r_2) = \sum_{n=1}^{N_1} P(X_1,n,N_2+1)\pi_{1,n} + P(X_1,N_1+2,N_2)\,p_{2,N_2} \tag{A.36}$$

13) Group 13. $\gamma_1 = 1$**,** $\gamma_2 = N_2 + 2$

$$P(x_1,1,N_2+2)\cdot\left(\mu_1 + p_{1,\gamma_1} + \rho_2 + \pi_{1,\gamma_1}\right) = P(x_1-1,1,N_2+2)\mu_1 + P(x_1,N_1+1,N_2+2)r_1 +$$

$$P(x_1,N_1+2,N_2+2)\rho_1 + \sum_{n=1}^{N_2} P(x_1,1,n)\pi_{2,n}, \quad 0 < x_1 < X_1 \tag{A.37}$$

$$P(0,1,N_2+2)\cdot\left(\mu_1 + p_{1,\gamma_1} + \rho_2 + \pi_{1,\gamma_1}\right) = P(0,N_1+1,N_2+2)r_1 + P(0,N_1+2,N_2+2)\rho_1 +$$

$$\sum_{n=1}^{N_2} P(0,1,n)\pi_{2,n} \tag{A.38}$$

$$P\left(X_1,1,N_2+2\right)\cdot\left(\rho_2+\pi_{1,\gamma_1}\right)=P\left(X_1-1,1,N_2+2\right)\mu_1+\ P\left(X_1,N_1+2,N_2+2\right)\rho_1+\sum_{n=1}^{N_2}P\left(X_1,1,n\right)\pi_{2,n}$$

(A.39)

14) Group 14. $2\le\gamma_1\le N_1$**,** $\gamma_2=N_2+2$

$$P\left(x_1,\gamma_1,N_2+2\right)\cdot\left(\mu_1+p_{1,\gamma_1}+\rho_2+\pi_{1,\gamma_1}\right)=P\left(x_1-1,\gamma_1,N_2+2\right)\mu_1+P\left(x_1,\gamma_1-1,N_2+2\right)p_{1,\gamma_1-1}+\sum_{n=1}^{N_2}P\left(x_1,\gamma_1,n\right)\pi_{2,n}\ ,\quad 0<x_1<X_1 \tag{A.40}$$

$$P\left(0,\gamma_1,N_2+2\right)\cdot\left(\mu_1+p_{1,\gamma_1}+\rho_2+\pi_{1,\gamma_1}\right)=P\left(0,\gamma_1-1,N_2+2\right)p_{1,\gamma_1-1}+\sum_{n=1}^{N_2}P\left(0,\gamma_1,n\right)\pi_{2,n} \tag{A.41}$$

$$P\left(X_1,\gamma_1,N_2+2\right)\cdot\left(\rho_2+\pi_{1,\gamma_1}\right)=P\left(X_1-1,\gamma_1,N_2+2\right)\mu_1+\ \sum_{n=1}^{N_2}P\left(X_1,\gamma_1,n\right)\pi_{2,n} \tag{A.42}$$

15) Group 15. $\gamma_1=N_1+1$**,** $\gamma_2=N_2+2$

$$P\left(x_1,N_1+1,N_2+2\right)\cdot\left(r_1+\rho_2\right)=P\left(x_1,N_1,N_2+2\right)p_{1,N_1}+\sum_{n=1}^{N_2}P\left(x_1,N_1+1,n\right)\pi_{2,n}\ ,\quad 0<x_1<X_1$$

(A.43)

$$P\left(0,N_1+1,N_2+2\right)\cdot\left(r_1+\rho_2\right)=P\left(0,N_1,N_2+2\right)p_{1,N_1}+\sum_{n=1}^{N_2}P\left(0,N_1+1,n\right)\pi_{2,n} \tag{A.44}$$

$$P\left(X_1,N_1+1,N_2+2\right)=0 \tag{A.45}$$

16) Group 16. $\gamma_1=N_1+2$**,** $\gamma_2=N_2+2$

$$P\left(x_1,N_1+2,N_2+2\right)\cdot\left(\rho_1+\rho_2\right)=\sum_{m=1}^{N_1}P\left(x_1,m,N_2+2\right)\cdot\pi_{1,m}+\sum_{n=1}^{N_2}P\left(x_1,N_1+2,n\right)\cdot\pi_{2,n}\ ,\quad 0<x_1<X_1$$

(A.46)

$$P\left(0,N_1+2,N_2+2\right)\cdot\left(\rho_1+\rho_2\right)=\sum_{m=1}^{N_1}P\left(0,m,N_2+2\right)\cdot\pi_{1,m}+\sum_{n=1}^{N_2}P\left(0,N_1+2,n\right)\cdot\pi_{2,n} \tag{A.47}$$

$$P\left(X_1,N_1+2,N_2+2\right)\cdot\left(\rho_1+\rho_2\right)=\sum_{m=1}^{N_1}P\left(X_1,m,N_2+2\right)\cdot\pi_{1,m}+\sum_{n=1}^{N_2}P\left(X_1,N_1+2,n\right)\cdot\pi_{2,n} \tag{A.48}$$

Appendix B.

Decomposition Algorithm

The decomposition algorithm in this appendix uses an iterative procedure to calculate the parameters of each primitive line segment (viz. p_k^α, r_k^α, p_k^β, and r_k^β, which are defined in Section 3.3.2). These parameters characterize the occurrence and disappearance of the "pseudo down" state for the upstream and downstream machines in a primitive line segment. As mentioned in Section 3.3, the "pseudo down" state of the machines in a line segment essentially reflects the starvation or blockage due to the upstream or downstream line segments. For example, in the primitive line segments of Figure B.1, the upstream machine of Line_5, M_5^u being "pseudo down" represents that M_2^d (in Line_2) or M_4^d (in Line_4) is starved. The probability that M_2^d (or M_4^d) becomes starved can be estimated using the limiting probabilities of states of Line_2 (or Line_4), as discussed later in this appendix. Similarly, the downstream machine in Line_2, M_2^d, being "pseudo down" indicates that M_4^d is starved or M_5^u is blocked. The probability that M_4^d becomes starved and M_5^u becomes blocked can be estimated based on the limiting probabilities of Line_4 and Line_5 respectively. In the algorithm presented below, the limiting probabilities of a line segment can be used to calculate the values of four additional parameters, which quantify starvation of the downstream machine M_k^d and blockage of the upstream machine M_k^u. These values are subsequently used to estimate the parameters of the adjacent line segment (p_k^α , r_k^α , p_k^β , and r_k^β).

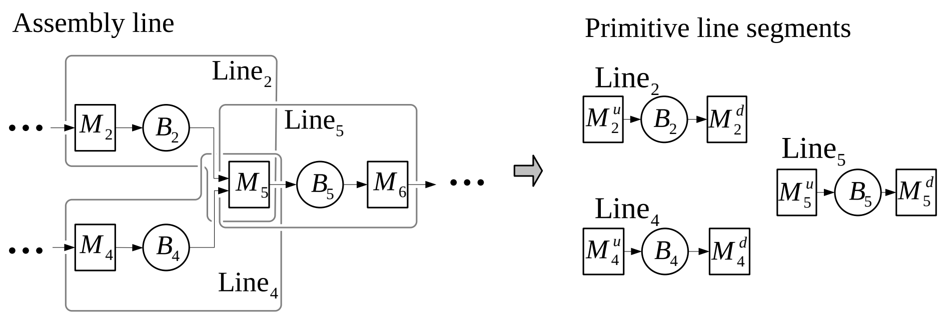


Figure B.1. A portion of the assembly line in Figure 3.5 and the corresponding primitive line segments.

In Chapters 3, 4, and 5, the decomposition models were presented for manufacturing systems under three different conditions:

- Multistage manufacturing systems with machine deterioration and preventive maintenance (Chapter 3).
- Multistage manufacturing systems with imperfect production (Chapter 4).
- Multistage manufacturing systems with batch operations and phase-type processing times (Chapter 5).

Although these decomposition models share the common framework, they differ in the details of calculation. This is because the primitive line segments used in these various models are defined differently in order to characterize different properties of the systems. In the remainder of this appendix, the author will first discuss the calculation of parameters of the decomposition model for systems with machine deterioration and preventive maintenance (presented in Chapter 3). Subsequently, the extension of the calculation to the other two decomposition models (discussed in Chapter 4 and 5 respectively) will be presented.

B.1. Calculating the Parameters of the Decomposition Model for Systems with Machine Deterioration and Preventive Maintenance

Based on the limiting probabilities of a primitive line segment ($P\left(x_k, y_k^u, y_k^d, \alpha_k, \beta_k\right)$), the following additional parameters may be calculated.

1) When M_k^d completes processing a part, there is a possibility that buffer B_k is empty, and hence M_k^d becomes starved. The probability of M_k^d becoming starved after it completes processing a part is denoted as g_k^d, and it is calculated as below:

   M_k^d continues processing if M_k^d is not "pseudo down" ($\beta_k = 1$), M_k^d is up ($\gamma_k^d \le N_k^d$), and the intermediate buffer B_k is not empty ($x_k \ge 1$). Under this condition, M_k^d completes processing parts at the transition rate of μ_k^d. Hence, the frequency of M_k^d discharging parts is:

$$FR_k = \sum_{\alpha_k} \sum_{\gamma_k^d \le N_k^d} \sum_{\gamma_k^u} \sum_{x_k \ge 1} P\left(x_k, \gamma_k^u, \gamma_k^d, \alpha_k, 1\right) \mu_k^d \tag{B.1}$$

   After M_k^d completes processing a part, x_k (number of parts in B_k) is reduced by 1. In the case where $x_k = 1$, when M_k^d completes processing a part, x_k is reduced from 1 to 0, and subsequently M_k^d becomes starved. Since the probability that M_k^d is busy and there is only one part in the line segment is $\sum_{\alpha_k} \sum_{\gamma_k^d \le N_k^d} \sum_{\gamma_k^u} P\left(1, \gamma_k^u, \gamma_k^d, \alpha_k, 1\right)$, the frequency that M_k^d completes processing a part and then becomes starved may be estimated as:

$$FR'_k = \sum_{\alpha_k} \sum_{\gamma_k^d \le N_k^d} \sum_{\gamma_k^u} P\left(1, \gamma_k^u, \gamma_k^d, \alpha_k, 1\right) \mu_k^d \tag{B.2}$$

   Thus, the probability that M_k^d becomes starved after it finishes processing a part is:

$$g_k^d = \frac{FR'_k}{FR_k}$$

$$= \frac{\sum_{\alpha_k} \sum_{\gamma_k^d \le N_k^d} \sum_{\gamma_k^u} P\left(1, \gamma_k^u, \gamma_k^d, \alpha_k, 1\right)}{\sum_{\alpha_k} \sum_{\gamma_k^d \le N_k^d} \sum_{\gamma_k^u} \sum_{x_k \ge 1} P\left(x_k, \gamma_k^u, \gamma_k^d, \alpha_k, 1\right)} \tag{B.3}$$

2) Similarly, g_k^u, the probability that the upstream machine M_k^u becomes blocked when it delivers a part to buffer B_k may be estimated as:

$$g_k^u = \frac{\sum_{\beta_k} \sum_{\gamma_k^d} \sum_{\gamma_k^u \le N_k^u} P\left(X_k - 1, \gamma_k^u, \gamma_k^d, 1, \beta_k\right)}{\sum_{\beta_k} \sum_{\gamma_k^d} \sum_{\gamma_k^u \le N_k^u} \sum_{x_k < X_k} P\left(x_k, \gamma_k^u, \gamma_k^d, 1, \beta_k\right)} \tag{B.4}$$

3) If M_k^d is starved, it recovers from starvation with the transition rate denoted as h_k^d. h_k^d is calculated as below:

Since the probability that M_k^d is starved is $\sum_{\beta_k} \sum_{\alpha_k} \sum_{\gamma_k^d} \sum_{\gamma_k^u} P\left(0, \gamma_k^u, \gamma_k^d, \alpha_k, \beta_k\right)$,

and the total transition rate that M_k^d recovers from starvation satisfies:

$$\left(\sum_{\beta_k} \sum_{\alpha_k} \sum_{\gamma_k^d} \sum_{\gamma_k^u} P\left(0, \gamma_k^u, \gamma_k^d, \alpha_k, \beta_k\right) \right) h_k^d = \sum_{\beta_k} \sum_{\gamma_k^d} \sum_{\gamma_k^u \le N_k^u} P\left(0, \gamma_k^u, \gamma_k^d, 1, \beta_k\right) \mu_k^u \tag{B.5}$$

Hence,

$$h_k^d = \frac{\sum_{\beta_k} \sum_{\gamma_k^d} \sum_{\gamma_k^u \le N_k^u} P\left(0, \gamma_k^u, \gamma_k^d, 1, \beta_k\right) \mu_k^u}{\sum_{\beta_k} \sum_{\alpha_k} \sum_{\gamma_k^d} \sum_{\gamma_k^u} P\left(0, \gamma_k^u, \gamma_k^d, \alpha_k, \beta_k\right)} \tag{B.6}$$

4) Similarly, h_k^u, the transition rate that M_k^u recovers from being blocked is

$$h_k^u = \frac{\sum_{\alpha_k} \sum_{\gamma_k^d \le N_k^d} \sum_{\gamma_k^u} P\left(X_k, \gamma_k^u, \gamma_k^d, \alpha_k, 1\right) \mu_k^d}{\sum_{\beta_k} \sum_{\alpha_k} \sum_{\gamma_k^d} \sum_{\gamma_k^u} P\left(X_k, \gamma_k^u, \gamma_k^d, \alpha_k, \beta_k\right)} \tag{B.7}$$

g_k^d , h_k^d , g_k^u , and h_k^u discussed above can be used to calculate the parameters of the primitive line segment, p_k^α , r_k^α , p_k^β , and r_k^β . For simplicity, $U(k)$ is defined as a function that returns the set of indices of line segments immediately upstream of the k^{th} primitive line segment, e.g. $U(6)=\{3,5\}$ as in Figure B.1. Additionally, $D(k)$ is defined as a function that returns the index of the immediately downstream line segment of the k^{th} primitive line segment, e.g. $D(3)=6$ as in Figure B.1.

If the upstream machine M_k^u in a line segment represents a non-assembly machine, its "pseudo down" state indicates that $M_{U(k)}^d$ (the corresponding machine in the upstream line segment) is starved. Therefore, the parameters characterizing the transitions between "pseudo down" and not "pseudo down" of M_k^u can be approximated by the parameters characterizing the transitions between the starvation and not starvation of machine $M_{U(k)}^d$:

$$p_k^\alpha = g_{U(k)}^d \tag{B.8}$$

$$r_k^\alpha = h_{U(k)}^d \tag{B.9}$$

If M_k^u represents an assembly machine, its "pseudo down" state includes all the possibilities that any of its upstream buffers are empty. Hence, the probability that M_k^u becomes "pseudo down", p_k^α , may be approximated by:

$$p_k^\alpha = \sum_{l \in U(k)} g_l^d \tag{B.10}$$

The probability that the "pseudo down" of M_k^u is caused by the i^{th} line segment ($i \in U(k)$) is:

$$\text{Prob}\left(M_i^d \text{ is starved} \mid M_k^u \text{ is "pseudo down"}\right) = \frac{g_i^d}{\sum_{l \in U(k)} g_l^d}, \qquad i \in U(k) \tag{B.11}$$

If the "pseudo down" of M_k^u is due to the i^{th} line segment ($i \in U(k)$), it may recover with the transition rate h_i^d (i.e. the transition rate that M_i^d recovers from starvation). Hence, the transition rate that M_k^u recovers from "pseudo down" (r_k^α) can be approximated by the average transition rate that the corresponding machines in all the immediately upstream line segments recover from starvation, i.e.

$$r_k^\alpha = \sum_{i \in U(k)} \left(\text{Prob}\left(M_i^d \text{ is starved} \mid M_k^u \text{ is "pseudo down"}\right) \cdot h_i^d \right)$$

$$= \sum_{i \in U(k)} \left(\frac{g_i^d}{\sum_{l \in U(k)} g_l^d} \cdot h_i^d \right) \tag{B.12}$$

If the downstream machine M_k^d in a line segment represents a non-assembly machine, its "pseudo down" state represents that this machine is blocked by the downstream line segment (i.e. the $\left(D(k)\right)^{\text{th}}$ line segment). Hence, the parameters characterizing the transitions between "pseudo down" and not "pseudo down" of M_k^d can be estimated using the parameters describing the transitions between blockage and not blockage of the $M_{D(k)}^u$, i.e.

$$p_k^\beta = g_{D(k)}^u \tag{B.13}$$

$$r_k^\beta = h_{D(k)}^u \tag{B.14}$$

If M_k^d represents an assembly machine (such as M_2^d in Figure B.1), its "pseudo down" state also includes the condition that one of the parts required for the assembly process is missing (for instance, in Figure B.1, M_2^d being "pseudo down" also includes the possibility that M_4^d is starved). This probability should be added into the calculation of p_k^β as below:

$$p_k^\beta = \sum_{l\in U(D(k))\,\&\, l\neq k} g_l^d + g_{D(k)}^u \tag{B.15}$$

Similar to Eqn (B.12), the transition rate that M_k^d recovers from "pseudo down" is the average value of transition rates that $M_{D(k)}^u$ recovers from blockage and M_i^d ($i\in U(D(k))$ and $i\neq k$) recovers from starvation. Therefore,

$$r_k^\beta = \sum_{i\in U(D(k))\,\&\, i\neq k}\left(\frac{g_i^d}{\sum_{l\in U(D(k))\,\&\, l\neq k} g_l^d + g_{D(k)}^u}\cdot h_i^d\right) + \frac{g_{D(k)}^u}{\sum_{l\in U(D(k))\,\&\, l\neq k} g_l^d + g_{D(k)}^u}\cdot h_{D(k)}^u \tag{B.16}$$

Based on the discussion above, the decomposition algorithm may be summarized as follows:

Decomposition Algorithm

- Initialize $g_k^u=0$, $h_k^u=1$, $g_k^d=0$, and $h_k^d=1$, $k\in\{1,2,...,K-1\}$. Choose a small value, ε, as the tolerance limit of the algorithm.
- **Loop**

  For($k=1$; $k\leq K-2$; $k++$), calculate p_k^α, r_k^α, p_k^β, and r_k^β using Eqns (B.10), (B.12), (B.15), and (B.16). Solve the balance equations of the k^{th} primitive line segment, and update g_k^d and h_k^d using Eqns (B.3) and (B.6) respectively.

  For($k=K-1$; $k>1$; $k--$), calculate p_k^α, r_k^α, p_k^β, and r_k^β using Eqns (B.10), (B.12), (B.15), and (B.16). Solve the balance equations of the k^{th} primitive line segment, and update g_k^u and h_k^u using Eqns (B.4) and (B.7) respectively.

- Terminate the algorithm if $Max\left(\frac{\Delta g_k^u}{g_k^u}, \frac{\Delta h_k^u}{h_k^u}, \frac{\Delta g_k^d}{g_k^d}, \frac{\Delta h_k^d}{h_k^d}, k \in \{1, 2, \cdots, K-1\}\right) < \varepsilon$ (where Δg_k^u, Δh_k^u, Δg_k^d, and Δh_k^d are the changes of g_k^u, h_k^u, g_k^d, and h_k^d in the iteration respectively); otherwise, go to **Loop**.

B.2. Calculating the Parameters of the Decomposition Model for Systems with Imperfect Production

The algorithm presented above can be extended to calculate the parameters of the decomposition model of manufacturing systems with imperfect production (refer to Section 4.3 for more detail of this model). For the primitive line segments in the decomposition model of Chapter 4, the additional parameters g_k^d, h_k^d, g_k^u, and h_k^u are calculated as discussed below (g_k^d and h_k^d characterize the occurrence and disappearance of starvation of the downstream machine in a primitive line segment respectively; while g_k^u and h_k^u characterize the occurrence and disappearance of blockage of the upstream machine respectively).

1) g_k^d, the probability that M_k^d becomes starved when it sends out a part, can be calculated as follows:

The frequency of M_k^d finishing a part is:

$$FR_k = \sum_{\alpha_k}\sum_{y_k^u}\sum_{x_k \geq 1} P\left(x_k, y_k^u, 1, \alpha_k, 1\right)\mu_k^d \tag{B.17}$$

The frequency that M_k^d sends out a part and becomes starved (x_k becomes 0) is:

$$FR'_k = \sum_{\alpha_k}\sum_{y_k^u} P\left(1, y_k^u, 1, \alpha_k, 1\right)\mu_k^d \tag{B.18}$$

Thus, the probability that M_k^d becomes starved after it completes processing a part is:

$$g_k^d = \frac{FR'_k}{FR_k}$$

$$= \frac{\sum_{\alpha_k}\sum_{y_k^u} P\left(1, y_k^u, 1, \alpha_k, 1\right)}{\sum_{\alpha_k}\sum_{y_k^u}\sum_{x_k \geq 1} P\left(x_k, y_k^u, 1, \alpha_k, 1\right)} \tag{B.19}$$

2) g_k^u represents the probability that M_k^u becomes blocked after it delivers a part to buffer B_k. Similar to Eqn (B.19), we have:

$$g_k^u = \frac{\sum_{\beta_k}\sum_{y_k^d} P\left(X_k - 1, 1, y_k^d, 1, \beta_k\right)}{\sum_{\beta_k}\sum_{y_k^d}\sum_{x_k < X_k} P\left(x_k, 1, y_k^d, 1, \beta_k\right)} \tag{B.20}$$

3) h_k^d, the transition rate that M_k^d recovers from being starved is estimated as follows:

The probability of M_k^d being starved is $\sum_{\beta_k}\sum_{\alpha_k}\sum_{y_k^d}\sum_{y_k^u} P\left(0, y_k^u, y_k^d, \alpha_k, \beta_k\right)$.

The total transition rate that M_k^d recovers from starvation satisfies:

$$\left(\sum_{\beta_k}\sum_{\alpha_k}\sum_{y_k^d}\sum_{y_k^u} P\left(0, y_k^u, y_k^d, \alpha_k, \beta_k\right)\right) h_k^d = \sum_{\beta_k}\sum_{y_k^d} P\left(0, 1, y_k^d, 1, \beta_k\right) \mu_k^u \left(1 - \eta_k^u\right) \tag{B.21}$$

Hence,

$$h_k^d = \frac{\sum_{\beta_k}\sum_{y_k^d} P\left(0, 1, y_k^d, 1, \beta_k\right) \mu_k^u \left(1 - \eta_k^u\right)}{\sum_{\beta_k}\sum_{\alpha_k}\sum_{y_k^d}\sum_{y_k^u} P\left(0, y_k^u, y_k^d, \alpha_k, \beta_k\right)} \tag{B.22}$$

4) Similarly, the transition rate that M_k^u recovers from blockage, h_k^u may be calculated as:

$$h_k^u = \frac{\sum_{\alpha_k}\sum_{y_k^u} P\left(X_k, y_k^u, 1, \alpha_k, 1\right) \mu_k^d}{\sum_{\beta_k}\sum_{\alpha_k}\sum_{y_k^d}\sum_{y_k^u} P\left(X_k, y_k^u, y_k^d, \alpha_k, \beta_k\right)} \tag{B.23}$$

Based on Eqns (B.19), (B.20), (B.22), and (B.23), we may calculate the additional parameters g_k^d, h_k^d, g_k^u, and h_k^u. These may be used to estimate the parameters of the decomposition model (viz. p_k^α, r_k^α, p_k^β, and r_k^β) for systems with imperfect production following the same methodology described in Section B.1.

B.3. Calculating the Parameters of the Decomposition Model for Systems with Batch Operations and Phase-Type Processing Times

The decomposition algorithm introduced in Section B.1 may also be extended to calculate the parameters of the decomposition model for multistage manufacturing systems with batch operations and phase-type processing times (discussed in Chapter 5). To incorporate batch operations and phase-type processing times, g_k^d, h_k^d, g_k^u, and h_k^d are calculated as follows:

1) g_k^d, the probability that M_k^d becomes starved after it completes processing a batch of parts can be estimated as discussed below.

The frequency of M_k^d finishing a batch of parts is calculated as:

$$FR_k = \sum_{\alpha_k}\sum_{j_k^u}\sum_{x_k} P\left(x_k, j_k^u, J_k^d, \alpha_k, 1\right)\mu_{k,J_k^d}^d \tag{B.24}$$

The frequency that M_k^d sends out a batch of parts and becomes starved ($x_k < Q_k^d$) is:

$$FR'_k = \sum_{\alpha_k}\sum_{j_k^u}\sum_{x_k<Q_k^d} P\left(x_k, j_k^u, J_k^d, \alpha_k, 1\right)\mu_{k,J_k^d}^d \tag{B.25}$$

Thus,

$$g_k^d = \frac{FR'_k}{FR_k} = \frac{\sum_{\alpha_k}\sum_{j_k^u}\sum_{x_k<Q_k^d} P\left(x_k, j_k^u, J_k^d, \alpha_k, 1\right)}{\sum_{\alpha_k}\sum_{j_k^u}\sum_{x_k} P\left(x_k, j_k^u, J_k^d, \alpha_k, 1\right)} \tag{B.26}$$

2) Similarly, using the limiting probabilities of the k^{th} line segment, the blockage occurring probability when M_k^u completes processing a batch of parts, g_k^u, is:

$$g_k^u = \frac{\sum_{\beta_k} \sum_{j_k^d} \sum_{x_k > X_k - 2Q_k^u} P\left(x_k, J_k^u, j_k^d, 1, \beta_k\right)}{\sum_{\beta_k} \sum_{j_k^d} \sum_{x_k} P\left(x_k, J_k^u, j_k^d, 1, \beta_k\right)} \tag{B.27}$$

3) h_k^d, the transition rate that M_k^d recovers from being starved is calculated as:

$$h_k^d = \frac{\sum_{\beta_k} \sum_{\substack{x_k < Q_k^d \ \& \\ x_k \ge Q_k^d - Q_k^u}} P\left(x_k, J_k^u, 0, 1, \beta_k\right) \mu_{k, J_k^u}^u}{\sum_{\beta_k} \sum_{\alpha_k} \sum_{j_k^u} \sum_{x_k < Q_k^d} P\left(x_k, j_k^u, 0, \alpha_k, \beta_k\right)} \tag{B.28}$$

4) Similarly, h_k^u, the transition rate that M_k^u recovers from being blocked may also be estimated as:

$$h_k^u = \frac{\sum_{\alpha_k} \sum_{\substack{x_k > X_k - Q_k^u \ \& \\ x_k - Q_k^d \le X_k - Q_k^u}} P\left(x_k, 0, J_k^d, \alpha_k, 1\right) \mu_{k, J_k^d}^d}{\sum_{\beta_k} \sum_{\alpha_k} \sum_{j_k^d} \sum_{x_k > X_k - Q_k^u} P\left(x_k, 0, j_k^d, \alpha_k, \beta_k\right)} \tag{B.29}$$

With Eqns (B.26), (B.27), (B.28), and (B.29), the parameters of each primitive line segment may be estimated using the decomposition algorithm presented in Section B.1.

Appendix C.

Using Effective Processing Times to Incorporate Machine Failures

The decomposition model in Chapter 5 can be extended to model operation dependent machine failures by utilizing the concept of effective processing times (Hopp and Spearman, 2000). In reality, the processing, up, and down times of a machine may be modeled by a variety of distributions, hence, it is difficult to derive a universal probability density function (*PDF*) of the effective processing time. However, it can be approximated from simulation data, and in some instances, the closed form expressions may be derived for simple distributions. Here, one such example is introduced, where the processing, up, and down times follow exponential distributions, with means of *P*, *U*, and *D* respectively. Thus,

$$PDF\left(x_p\right) = P^{-1}e^{-(x_p/P)}, \quad x_p \geq 0 \tag{C.1}$$

$$PDF\left(x_u\right) = U^{-1}e^{-(x_u/U)}, \quad x_u \geq 0 \tag{C.2}$$

$$PDF\left(x_d\right) = D^{-1}e^{-(x_d/D)}, \quad x_d \geq 0 \tag{C.3}$$

where x_p, x_u, and x_d denote the processing, up, and down time random variables respectively. The breakdown of a machine usually occurs far less frequently than the processing of parts, otherwise, the machine is not economically feasible. The likelihood that a machine breaks down more than once during the processing of a batch of parts is very low. Hence, the probability that a batch of parts being processed encounters machine breakdown is:

$$Prob\left(\text{breakdown occurs within } x_p\right) = \int_0^{x_p} U^{-1} e^{-(x_u/U)} dx_u$$

$$= 1 - e^{-(x_p/U)} \qquad \text{(C.4)}$$

The effective processing time z is defined as:

$$z = \begin{cases} x_p, & \text{if no breakdown occurs within } x_p \\ x_p + x_d, & \text{otherwise} \end{cases} \qquad \text{(C.5)}$$

The *CDF* (cumulative density function) of z is derived as follows:

$$CDF(z) = \int_0^z PDF(x_p)\left(1 - Prob\left(\text{breakdown occurs within } x_p\right)\right) dx_p$$

$$+ \int_0^z PDF(x_p)\left(Prob\left(\text{breakdown occurs within } x_p\right) \int_0^{z-x_p} PDF(x_d)\, dx_d\right) dx_p$$

$$= 1 + \frac{Pe^{-\frac{z}{P}}}{D-P} - \frac{D^2 Pe^{-\frac{z}{D}}}{(D-P)(DP+DU-UP)} - \frac{DUe^{-\left(\frac{z}{P}+\frac{z}{U}\right)}}{DP+DU-UP} \qquad \text{(C.6)}$$

Therefore, the *PDF* of z is:

$$PDF(z) = \frac{dCDF(z)}{dz}$$

$$= -\frac{e^{-\frac{z}{P}}}{D-P} + \frac{DPe^{-\frac{z}{D}}}{(D-P)(DP+DU-UP)} + \frac{D(U+P)e^{-\left(\frac{z}{P}+\frac{z}{U}\right)}}{P(DP+DU-UP)} \qquad \text{(C.7)}$$

Based on Eqn (C.7), the mean and variance of the effective processing times maybe shown to be:

$$Mean(z) = \frac{(U+D+P)P}{U+P} \qquad \text{(C.8)}$$

$$Var(z) = \frac{P\left(2UDP + P(P+U)^2 + D^2(P+2U)\right)}{(P+U)^2} \qquad \text{(C.9)}$$