MODEL DEVELOPMENT OF A CONWIP SYSTEM FOR PRODUCTION CONTROL IN MULTI-STAGE MULTI-PRODUCT MANUFACTURING ENVIRONMENTS CONSISTING OF HIGH-RUNNER AND LOW-RUNNER PRODUCT FAMILIES

by

JOSHUA A/L JAYA PRAKASH

Thesis submitted in fulfillment of the requirements

for the degree of

Doctor of Philosophy

December 2013

ACKNOWLEDGEMENT

All praise and glory to my Savior Jesus Christ for the many blessings undeservingly bestowed upon me. Never in my life have I imagined of coming this far in my education. I must also take this opportunity to thank the following people that have made this thesis possible.

To my supervisor, Dr Chin Jeng Feng, thank you for your guidance, your encouragement and most importantly, for shaping the way I think as a researcher. To the staff in Continental Malaysia Sdn Bhd and Spirit Aerosystems Malaysia Sdn Bhd that have assisted me in the case study portion of the research, thank you for welcoming an outsider like me and supporting my work.

To my mother, my father, my sisters and my brothers, thank you for your encouragement and your faith in me. I hope I have made all of you proud. To my friends and colleagues in campus, thank you for making my experience in campus so much better and for making it a comfortable workplace.

To everyone else I have missed, thank you for your contribution in the completion of this thesis. I am grateful for you help. I would like to extend my gratitude to the Higher Education Ministry of Malaysia for providing me with the necessary funding to carry out my research.

I dedicate this thesis to my mother, whose love I could not do without.

‘I have fought the good fight, I have finished the race, I have kept the faith.’ 2 Timothy 4:7

TABLE OF CONTENT

LIST OF TABLES

Page

LIST OF FIGURES

LIST OF ABBREVIATIONS

PFIFO	push system with FIFO dispatch rule
CFIFO	CONWIP system with FIFO dispatch rule
PCFIFO	parallel CONWIP system with FIFO dispatch rule
PCHL	parallel CONWIP system with HL dispatch rule
PCLH	parallel CONWIP system with LH dispatch rule
NS	no internal setup
S	setup at every changeover
NB	no machine breakdown
B	breakdown in the bottleneck machine
M	no machine sharing
MM	full flexibility machine sharing
BI	bottleneck at S_2 and S_3
BE	bottleneck at S_4
LS	lot size
RT	ratio of HR orders to total orders
KC	total card quantity
KH	HR card quantity
KL	LR card quantity
AWIP	average WIP level
AFT	average flow time per product
BU	bottleneck utilization
SLHR	HR service level
SLLR	LR service level

LIST OF SYMBOLS

t_1	production control system and dispatch rule
t_2	machine setup
t_3	machine breakdown
t_4	machine sharing
t_5	bottleneck position
A_n	setup time of a machine in S_n (s), n = 1, 2, 3 or 4
B_n	breakdown rate of a machine in S_n (d^{-1}), n = 1, 2, 3 or 4
C_n	number of machines in S_n, n = 1, 2, 3 or 4
D_n	cycle time of a machine in S_n (s), n = 1, 2, 3 or 4
n_{HR}	total HR product completed
n_{LR}	total LR product completed
a_{HRi}	time HR product i arrives in B_1, i=1,2… n_{HR}
a_{LRi}	time LR product i arrives in B_1, i=1,2… n_{LR}
b_{HRi}	time HR product i is completed, i=1,2… n_{HR}
b_{LRi}	time LR product i is completed, i=1,2… n_{LR}
h_{HR}	lead time of HR product (s)
h_{LR}	lead time of LR product (s)
T	duration of analysis (s)

PEMBANGUNAN MODEL SEJENIS SISTEM CONWIP UNTUK PENGENDALIAN PRODUKSI DALAM PERSEKITARAN PEMBUATAN PELBAGAI PERINGKAT PELBAGAI PRODUK TERDIRI DARIPADA PRODUK KELUARGA PELARI TINGGI DAN PELARI RENDAH

ABSTRAK

Sistem pengendalian produksi boleh diklasifikasikan sebagai sistem dorong atau sistem tarik. Dalam sistem dorong, produksi dimulakan pada masa tertentu, sementara dalam sistem tarik, produksi dimulakan apabila isyarat diterima. Walau bagaimanapun, kedua-dua sistem mempunyai kelemahan masing-masing. Sistem dorong dikawal dengan memerhatikan output yang memerlukan anggaran kapasiti sistem. Penganggaran yang tidak tepat akan menyebabkan kerja yang sedang dijalankan (WIP) meningkat di luar had kawalan. Sistem tarik memerlukan pengekalan WIP dalam kuantiti yang kecil bagi setiap jenis produk keluarga. Namun, aneka produk yang tinggi akan menyebabkan tahap WIP yang tinggi. Matlamat tesis ini adalah untuk mereka dan menyelidik satu sistem tarik yang baru, dikenali sebagai sistem WIP tetap (CONWIP) selari (terdiri daripada beberapa variasi). Dalam sistem ini, produk keluarga dibahagikan kepada dua kelas (pelari tinggi dan pelari rendah) berdasarkan permintaan aneka produk. Setiap kelas menggunakan sistem CONWIP, dimana produksi dimulakan dengan pengalihan produk lengkap.

Kaedah penyelidikan yang digunakan ialah kajian deskripsi, dimana sistem CONWIP selari dikaji melalui simulasi peristiwa diskret serta analisis statistik. Dalam fasa pertama (persekitaran pembuatan tiruan), sistem pengendalian produksi penanda aras, parameter, pemboleh ubah serta prestasi diukur diperoleh melalui kajian literatur dan pengetahuan proses pembuatan. Dalam fasa kedua (persekitaran

pembuatan kajian kes), kesesuaian sistem CONWIP selari dengan aplikasi dikaji berdasarkan perbandingan prestasi dengan sistem pengendalian produksi yang sedia ada. Sejumlah 90 model telah dibina (25110 larian). Dalam kedua-dua fasa, ANOVA menggambarkan kewujudan kesan signifikan di antara prestasi diukur dan pemboleh ubah, sementara regresi digunakan untuk menganggarkan magnitud hubungan ini untuk membandingkan sistem CONWIP selari dengan sistem pengendalian produksi penanda aras secara graf. Dalam fasa kedua juga, parameter operasi optimum bagi sistem pengendalian produksi terpilih diperoleh menggunakan metodologi permukaan sambutan (RSM).

Dalam kedua-dua fasa, hasil analisis ANOVA menunjukkan terdapat kesan yang signifikan di antara parameter operasi dan permintaan aneka produk dalam sistem CONWIP selari. Graf menunjukkan sistem CONWIP selari lebih baik dari sistem dorong dan sistem CONWIP dari segi jumlah WIP purata dan masa pemprosesan purata per produk. Penyelidikan ini manunjukkan sistem CONWIP selari bergerak balas lebih baik terhadap perubahan permintaan aneka produk berbanding sistem CONWIP, serta menunjukkan prestasi lebih baik berbanding sistem dorong dan sistem CONWIP dalam persekitaran pembuatan kurang sempurna.

MODEL DEVELOPMENT OF A CONWIP SYSTEM FOR PRODUCTION CONTROL IN MULTI-STAGE MULTI-PRODUCT MANUFACTURING ENVIRONMENTS CONSISTING OF HIGH-RUNNER AND LOW-RUNNER PRODUCT FAMILIES

ABSTRACT

Production control systems can be generally categorized as push or pull systems. In a push system, production is initiated at scheduled times, whereas in a pull system, production is initiated when a signal is received. However, there are limitations to each system. Push systems are controlled by observing throughput, which requires an estimation of system capacity. Inaccurate estimates can cause work-in-process (WIP) to increase beyond the limit. Pull systems require maintaining a small amount of WIP for each product family. Nevertheless, a large product mix may still result in a high WIP level. The aim of this research is to develop and investigate a new pull system (made up of several variants) known as a parallel constant work-in-process (CONWIP) system. In the systems, product families are classified into two classes (high-runner and low-runner) based on the demand of the product mix. Each class uses a CONWIP system, where production is initiated upon withdrawal of finished goods.

The research method adopted is descriptive research, where parallel CONWIP systems were studied through discrete event simulation and statistical analysis. In the first phase (artificial manufacturing environments), benchmark production control systems, parameters, variables and performance measures were established from literature review and knowledge of manufacturing processes. In the second phase (case study manufacturing environments), the suitability of parallel

CONWIP systems for application is examined based on performance comparison with the existing production control system. A total of 90 simulation models were constructed (25110 runs). In both phases, ANOVA estimated the significant effect of variables on the performance measures, while regression estimated the magnitude of this relationship for graphical comparison between parallel CONWIP and the benchmark/existing production control systems. In addition for the second phase, response surface methodology (RSM) obtains optimal operating parameters for the selected production control system.

For both phases, the ANOVAs indicate that in parallel CONWIP systems, the relationship between the operating parameters and the demand of the product mix has a significant effect on the performance measures. The graphs reveal that parallel CONWIP systems are superior to push and CONWIP systems in terms of average WIP level and average flow time per product. This research shows that parallel CONWIP systems respond better to demand changes than CONWIP system and performs better than push and CONWIP systems in less pristine environments.

CHAPTER 1

INTRODUCTION

1.1 Introduction

This chapter provides the overview of this thesis which includes the background, problem statement, research objectives, research limitations and thesis structure.

1.2 Background

Today, many manufacturing industries are offering a larger product variety. This arises from factors such as globalization of markets, growing customer preference and changes in trading structure (MacDuffie et al., 1996). Bils and Klenow (2001) studied the increase in product variety offered by various industries and have concluded that on average, product variety increases by 1% each year. The product variety growth rate is in turn dependent on factors such as distance from the frontier of export variety and intensity of research and development (Addison, 2003). Manufacturing facilities have therefore made several alterations in its production process to support this increase, such as a reduction in product development time, minimization of resource consumption and utilization of technologies capable of product customization (Alford et al., 2000; Denton et al., 2003). These efforts have reduced the cost in offering a broad product variety, thereby enabling firms to offer greater customer value and increase profit (Gao and Hitt, 2004).

One aspect of production that facilitates the shift toward an increasing product variety is production control, which is the set of material handling activities involved in manufacturing, taking into account the demand, resource availability, cost and capacity constraints (McKay and Wiers, 2004). The objective of production control is to achieve the required quality and quantity of products at the time needed. The term ‘multi-stage multi-product’, coined in Villa (1989), is a formalization that is used to address existing production control architectures consisting of more than one processing stage and more than one product variety. This term has been applied in many studies, including Hui and Gupta, (2000), Prasad and Maravelias (2008) and Altiparmak et al. (2009). Generally, production control systems can be classified into two categories: push and pull (Karmarkar, 1991). In a push system, an upstream workstation begins processing without a request from a downstream workstation. In a pull system, an upstream workstation begins processing when it receives a request from a downstream workstation.

The major challenge of a production control system in a multi-stage multi-product environment is fulfilling the demand of the product variety while maintaining low inventory. Pull systems are keys in maintaining low inventory, in particular, work-in-processes (WIPs), as it regulates movement of the WIPs. However, the use of a pull system in a multi-stage multi-product environment is no simple task. This is because there are various methods to limit the work-in-process (WIP) level, each with its advantages and disadvantages. In addition, the method must facilitate the degree of product variety offered by the manufacturing facility. In other words, there is no single pull system that is suitable for all manufacturing environments (Framinan et al., 2003). Therefore, the chosen pull system in a multi-stage multi-product environment must regulate the right quantity and the right product variety to meet the demand while maintaining a low WIP level.

Existing pull systems in multi-stage multi-product environments include constant work-in-process (CONWIP) systems (Spearman et al., 1990) and mixed pull systems (Lane, 2007). In a CONWIP system, the required product quantity and mix are set in the correct sequence when production is initiated. The sequence of the product quantity and mix depend on the demand of each product family (Spearman et al., 1990). In a mixed pull system, the product variety is grouped into separate classes based on the demand of each product family. For the group with high and medium demand, a common kanban system is used, whereas for the group with low demand, a push system is used (Lane, 2007).

The demand of product families in a multi-stage multi-product manufacturing environment ranges from low to high. For product families with high demand, demand forecasts are generally available as it can be estimated from historical data. WIPs in this category are kept for frequent consumption and replenishment to ensure demand is continuously met (Smalley, 2009). For product families with low demand, demand forecasts may not be available. WIPs in this category are usually kept in small quantity as less frequent consumption and replenishment may result in its deterioration over time (Smalley, 2009). The distinction between high and low demand product families is crucial to ensure that high demand products can be produced during slack period, and this will in turn reduce the need for a higher flexibility in the manufacturing system during production of low demand products (Köber and Heinecke, 2012).

1.3 Problem statement

The success in the application of a pull system in a multi-stage multi-product environment requires fulfillment of several prerequisites. First, a large number of product families should not lead to a state where the WIP level is high (Suri, 1998), especially for product families with high demand. Second, a sudden surge in production quantity should not potentially cause the expansion of production flow time (Christopher and Towill, 2001), especially for product families with low demand. Third, the conditions for the chosen pull system to work should be fulfilled. For example, kanban system requires a product-type layout, short setup time and standardization of work (Panneerselvam, 2012).

The third prerequisite warrants further explanation. Firstly, a product-type layout is needed to facilitate pulling WIPs between workstations, hence may require the relocation of machines from an existing layout (Harrison and Petty, 2002). If these machines are shared by multiple product families, the production line may need to be reconfigured, resulting in the stoppage of production during this transformation period. Secondly, the setup time must be reduced to offset the increased number of setups (Swamidass, 2000). However, setup time reduction initiatives are typically approached in an unstructured manner, resulting in the lack of sustainability (Mileham et al., 1999). Thirdly, pure standardization of work, which involves coordinating and implementing the precise work sequence for a task, is not likely to be achieved because the preconditions (limited equipment downtime and lack of quality problems) cannot be met (Katō and Smalley, 2011).

Spearman and Zazanis (1992) proves that by limiting the WIP level within a boundary of workstations (rather than pulling WIPs between workstations), the conditions for the pull system can be relaxed and its benefits can still be realized. With the exception of CONWIP systems, pull systems in multi-stage multi-product environments do not emphasize on limiting the WIP level within a boundary of workstations. In addition, the literature review reveals that existing CONWIP systems in multi-stage multi-product environments focus either on maintaining a single WIP level for all product families, irrespective of the demand of each product family, or maintaining the WIP level for each product family, based on their individual demand: there is no CONWIP system that addresses product families with high and low demand separately.

1.4 Research objectives

The aim of this research is to develop and investigate a new modified CONWIP system with several variations that addresses separately product families with high and low demand in multi-stage multi-product environments. This is achieved through the following objectives:

1. To develop a new modified CONWIP system with several variants based on a literature review of existing production control systems in multi-stage multi-product environments.
2. To investigate the advantages of the new modified CONWIP system and its variants in comparison to relevant production control systems in artificial and case study manufacturing environments to derive a generalized set of behavior.

1.5 Research limitations

1. The manufacturing environments studied were of a multi-stage multi-product type, and is not applicable to a continuous or single-product type. Therefore, several production control systems pertaining to a single-product type are not cited, and this include drum-buffer-rope and hybrid push-pull, which are only applicable in repetitive manufacturing (Fernandes and Filho, 2011).

2. The new CONWIP systems were developed from a comparison between existing production control systems (CONWIP, POLCA, mixed pull and push) to identify limitations and gaps in research. The development of a new system from limitations and gaps of existing systems was also adopted by Spearman et al. (1990) and Suri (1998).
3. The study requires that raw materials of products under investigation are consistently available. The study of the supply chain is excluded because it is dictated by the market position of the suppliers, where there is restricted control (Mares, 2010).
4. The investigation of the new CONWIP systems in case study manufacturing environments was confined to only two cases, due to the substantial amount of time attributed to each case. It was ensured that the two cases were significantly different in terms of the application environment to avoid diminishing the credibility of the findings.
5. The manufacturing simulation software used was WITNESS 2008. The results obtained from this software were not verified using other software. In fact, the use of WITNESS 2008 as the only simulation software is applied in many researches, which include Akers (1997), Calinescu (2002), Ochwa (2007) and Lu (2009).
6. The statistical analysis software used was MATLAB 2008. The results obtained using this software was verified using Minitab 16. The use of MATLAB as a statistical analysis software is employed in many researches, which include Li et al. (2008), Kwong et al. (2008), Miguez et al. (2010) and Asiltürk and Çunkaş (2011).

1.6 Thesis structure

The structure of this thesis is presented in Figure 1.1. The first chapter (Introduction) provides a description of the research background and identifies the problem with existing systems. The second chapter (Literature review) presents a contemporary literature found within the research topic. The third chapter (Methodology) outlines the methodology adopted in analyzing the new systems. The fourth chapter (Development of models and results) introduces the construction of the models from artificial data for analyzing the new systems and presents the simulation results. The fifth chapter (Case studies) analyzes the new systems in two case study manufacturing environments and presents the simulation results. The sixth chapter (Discussion) discusses and compiles the findings from the fourth and fifth chapter. The seventh chapter (Conclusions and directions for future research) concludes the research and proposes potential directions for future research.

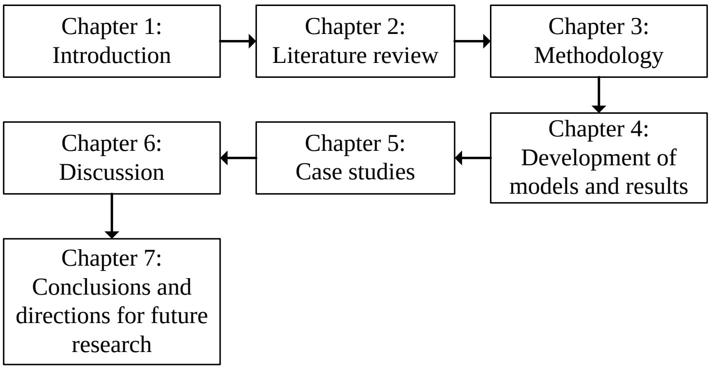


Figure 1.1: Illustration of thesis structure

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter explores existing production control systems found in literature. The flow of the literature review is shown in Figure 2.1. Firstly, the purpose of a production control system is outlined. Secondly, the characteristics of traditional and lean production practices are outlined and for each practice, a literature review of relevant production control systems in multi-stage multi-product environments is presented. Thirdly, the findings from the review are discussed. Fourthly, the findings are summarized, and fifth, an additional review that will be used in Chapter 4 is provided.

2.2 Production control in multi-stage multi-product environments

The rapidly evolving technological advances, increase in global competition and increasing customer buying power have contributed to an overall increase in the product variety offered by industries such as automotive, pharmaceutical and banking (Ramdas, 2003). Increasing product variety helps to narrow the gap between customer preferences and offered products (Van Iwaarden and Van Der Wiele, 2012). Generally, key decisions that are involved in product variety expansion can be classified into product variety creation decisions and product variety implementation decisions (Lancaster 1990). These decisions have to be addressed to avoid operational inefficiencies during product switching and escalation in product development, material and holding cost (Benjaafar et al, 2004).

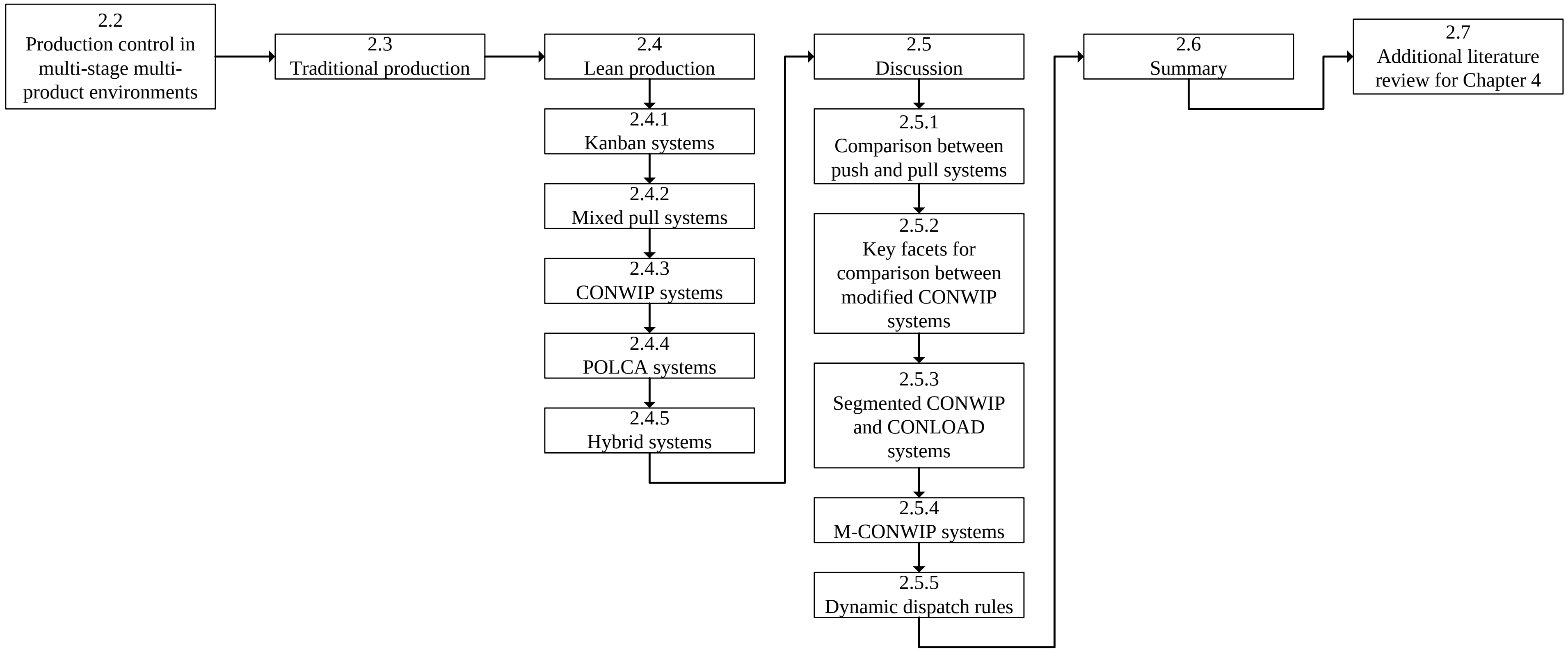


Figure 2.1: Literature review structure

Product variety creation decisions focus on the quantity, variety and timing for product introduction to the target markets, whereas product variety implementation decisions focus on the strategy employed by a manufacturing organization's delivery process to implement the variety creation decisions (Ramdas, 2003). One crucial product variety implementation decision is the installation of a suitable production control system to ensure the right quantity and quality of products in a multi-stage multi-product environment are produced at the right time (Garg and Lee, 1999). There are two paradigms of production control systems, namely traditional and lean production (Macduffie et al., 1996). These production control systems are described, illustrated and reviewed in the subsequent subsections.

2.3 Traditional production

Traditional production is based on two principal objectives: a) no resource should be left idling, and b) production runs should be as economical as possible (Burke and Wilks, 2006). At the heart of traditional production is a materials requirement planning (MRP) system. An MRP system works backwards from the production schedule of finished goods to derive schedules for raw materials. Fundamentally, an MRP system functions as a 'push' trigger as it pushes products downstream based on the schedules of the raw materials (Hopp and Spearman, 2001). Succinctly, a push system is based on the premise that raw materials are released at scheduled times and pushed from one workstation to another based on that schedule (Lyons et al., 2012).

Three common MRP systems are materials requirement planning (MRP), manufacturing resource planning (MRP II) and enterprise resource planning (ERP) (Burke and Wilks, 2006). In MRP, every production area is scheduled independently, hence the absence of a linkage between the planned and actual production as well as between production in one workstation and other workstations (Dolcemascolo, 2006). MRP II overcomes the limitations of MRP by tracking inventory while managing the schedule (Schutt, 2004). MRP II grew to become enterprise resource planning (ERP), which integrates the financial, material, asset and resource aspects of a manufacturing organization (Schutt, 2004). A reorder point (ROP) system is frequently used in tandem with MRP systems to replenish less expensive and commonly used components when their inventory hits a minimum. However, it soon became clear that MRP systems were not the definite planning approach in many multi-stage multi-product environments.

A push system is controlled by observing throughput, which is in turn controlled with respect to capacity (Schutt, 2004). As such, the capacity estimates (integrated into the MRP system) must include details such as cycle time, setup time, random outages, operator efficiency and rework (Spearman and Zazanis, 1992). By incorrectly estimating the capacity, there will be error in the raw material release rate and the WIP level can increase beyond the permitted limit, resulting in long production flow time and additional storage and transportation costs (Lyons et al., 2012). In the subsequent sections, the operations of several production control systems are illustrated. Diagrams following the examples from Krieg (2005) are referred to extensively in this thesis because there is a common representation for production flow and ease of understanding. Table 2.1 provides a description of the elements used in a diagram and Figure 2.2 illustrates the symbols used in a diagram.

Table 2.1: Description of elements used in a diagram

Elements	Description
Card	Representation of signal (electronic or physical) travel.
Batch	Contains a predefined lot size of products.
Production stage	Contains one or several machines.
Buffer	Contains storage for card and WIP/raw material.
Workstation	Contains one production stage and one buffer.
Cell/Segment	Contains two workstations in a sequence that supports processing operations.
Operation sequence	Two forms of numbering are used:
	a) *jk*, used for flow within workstations/cells/segments,
	j = sequence number
	j = 1, 2, 3….
	k = workstation/cell/segment number
	k = *i*, *ii*, *iii*
	i = workstation/cell/segment 1
	ii = workstation/cell/segment 2
	iii = workstation/cell/segment 3
	b) *l*, used for flow between workstations/cells/segments
	l = sequence number
	l = 1, 2, 3…

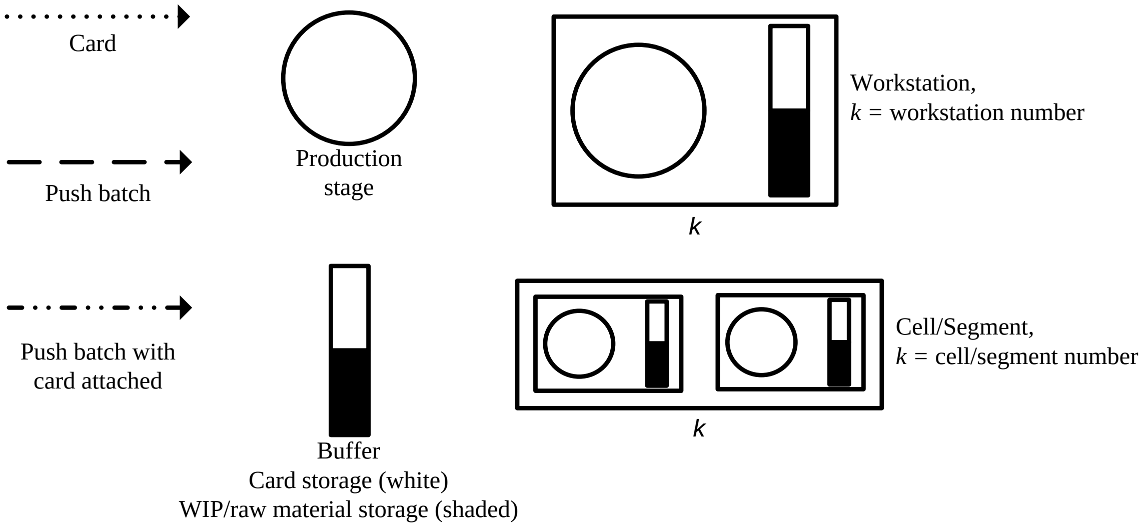
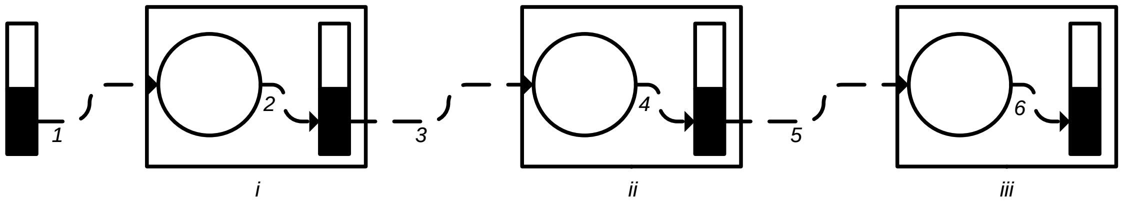


Figure 2.2: Symbols used in a diagram

Figure 2.3 illustrates and summarizes the flow of material in a push system. In a push system, one batch of raw material from the input buffer is released to the first workstation, where processing begins. Upon completing the process, the batch is pushed to the immediate downstream workstation for subsequent processing. If this workstation is occupied, it waits until the workstation is vacant. This flow progresses until the batch reaches the end of the line where a final product is ready.



Sequence	Description
1, 3, 5	The batch is pushed to the immediate downstream production stage
2, 4, 6	The batch is pushed to the immediate downstream buffer

Figure 2.3: Flow of material in a push system

There is ample literature discussing push systems in multi-stage multi-product environments. Robinson et al. (2005) investigate the effect of manual, semi-automated and fully automated replenishment strategies in decentralized and fully coordinated supply chains, observed from two construction firms. Upon replenishment triggered on the shop floor, the replenishment strategies and supply chains are analyzed via mathematical modeling to determine the total cost. The paper suggests that using semi-automated replenishment and decentralized supply chain generates the highest cost savings.

Lee et al. (2009) present a modified MRP system using computational grid to resolve capacity constraint issues on a shop floor. By applying a simple heuristic known as the longest first tail rule, the duration from production plan generation to production completion is reduced significantly. The authors further described the implementation procedure of the proposed system. One issue highlighted is the importance of predicting the speed up variance based on the routing data.

Barba-Gutiérrez and Adenso-Díaz (2009) formulate an algorithm for planning the disassembly of discrete products for component recovery. Known as the reverse MRP system, the algorithm (reformulated from Gupta and Taleb (1994)'s work using fuzzy logic approach) plans the correct quantity to disassemble to fulfill the demand of the required component. The analysis shows that at a given disassembly lead time, the reverse MRP system accumulates lower inventory than traditional MRP system. However, the reverse MRP system is only suitable for products with simple structures.

While Barba-Gutiérrez and Adenso-Díaz (2009) analyze the benefits of fuzzy logic in disassembly, Mula and Poler (2010) examine its benefits in production by incorporating fuzziness into the MRP system. A linear programming model is the basis in evaluating the production lead time, inventory accumulated and total cost in a capacity constrained environment. The paper presents a general function which serves as a constructive block for complex models of multi-stage multi-product environments.

Kanet and Stößlein (2010) resolve capacity constraint issues on a shop floor by taking into account the resource capacity before exploding to lower level components. Known as the capacitated ERP system, a framework for integrating the proposed MRP system to the supply chain is introduced. The framework is tested in a facility manufacturing aircraft engines. Aside from a reduction in inventory, the capacitated ERP system also provides the facility with knowledge of capacity requirements in the master production schedule.

Mohammaditabar et al. (2012) propose a model for separating raw materials into distinct classes to determine components fit for replenishment using reorder point (ROP). Simulated annealing is used to determine the component classes based on criteria such as annual cost, analytical hierarchical procedure weighted score and optimal inventory. The paper shows that ROP is able to attain low inventory and low cost.

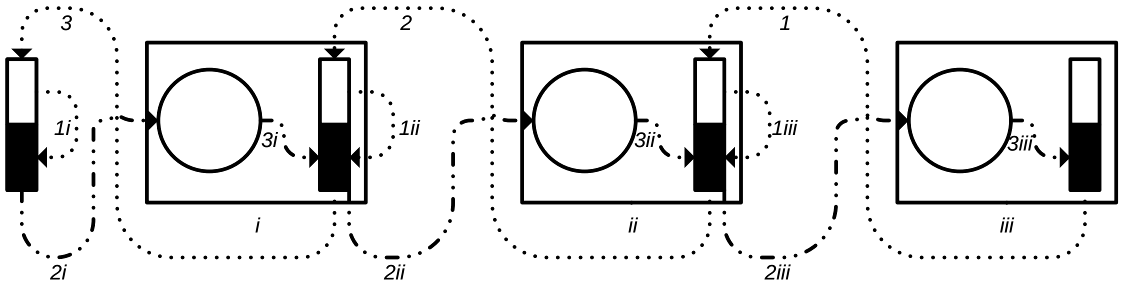
2.4 Lean production

Lean production is based on two principal objectives: a) continuous waste elimination, and b) value creation for the customer (Hansen et al., 2009). Lean production does more with less labor, less space, less inventory and delivers products in less time (Jacobsen, 2011). One important tool in lean production is pull systems, where production control is directly linked to customer demand (Gahagan, 2008). Pull systems dictate when resources should work in order to reduce the WIP level (Liker, 2004). Although the operation of different pull systems vary in nature, one common characteristic is the initiation of processing operations only when a signal is received (Bonney et al., 1999). The advantage of pull systems over push systems is attributed to the predictable (hence controllable) production flow time (Wang, 2010). The result is a shorter production flow time, lower holding costs and quicker defect detection. Several pull systems found in multi-stage multi-product environments are kanban, mixed pull, constant work-in-process (CONWIP), paired-cell overlapping loop of cards with authorization (POLCA) and hybrid system. The workings of each system are illustrated as follows.

2.4.1 Kanban systems

Figure 2.4 illustrates and summarizes the flows of material and information in a kanban system. In a kanban system, first described by Monden (1983), each batch of product is attached with a card. When one batch of finished goods is withdrawn, the card is detached and transferred to the immediate upstream workstation. In this workstation, the received card triggers production. However, before processing can begin, a card is detached from a batch of product and transferred to the immediate upstream workstation, while the card received previously is attached to this batch.

The transfer of cards from one workstation to an immediate upstream workstation progresses until a card is received by the first workstation, where it is attached to a batch of raw material. Generally, production is initiated through the transfer of cards to upstream workstations, but the circulation of cards between two consecutive workstations controls the WIP level between them (Wang, 2010).



Sequence	Description
1, 2, 3	Card is detached from the batch and transferred to the immediate upstream buffer
1i, 1ii, 1iii	Additional card is attached to the batch
2i, 2ii, 2iii	The batch is pushed to the immediate downstream production stage
3i, 3ii, 3iii	The batch is pushed to the immediate downstream buffer

Figure 2.4: Flows of material and information in a kanban system

Literature discussing kanban systems in multi-stage multi-product environments is commonly analytical. Krieg and Kuhn (2008) applied a decomposition-based approximation method in a two-stage model of a kanban system to evaluate the service level and WIP level. Although the results suggest that the approximation can be extended for a large variety of manufacturing environments, the assumptions made during the model construction does not reflect actual production environments. Gurgur and Altiok (2008) also discussed the use of an approximation method to evaluate the service level and WIP level of kanban system in a multi-stage model. The difference is that the assumptions made during model construction do reflect actual production environments.

Kumar and Panneerselvan (2007) provide a review of kanban systems and trends in kanban systems research. The review finds that kanban systems require the product variety to be kept at a minimum due to the repetitive nature of production. A kanban system will not work for make-to-order (MTO) products, whereby production can only start once demand is received (Hill, 2000). This is evident in a pull system that incorporates the kanban mechanism, known as the hybrid kanban-CONWIP system, where only a single product family is studied. In hybrid kanban-CONWIP system (Bonvik et al., 1997), kanban system provides localized WIP control and CONWIP system limits the WIP level in the line simultaneously. A mixed pull system provides a more practical approach, where a kanban system is only used for make-to-stock (MTS) products, whereby production is based on demand forecasts (Smalley, 2009).

2.4.2 Mixed pull systems

In a mixed pull system, expounded by Smalley (2009), the product variety is classified into three classes, which are high-runner (HR), medium-runner (MR) and low-runner (LR). HR, MR and LR contain product families with high, medium and low demand respectively. In addition, the product families in HR, MR and LR have low, medium and high variability in quantity respectively. Each class uses a separate production control system: HR and MR use a common kanban system while LR uses a push system. Mixed pull system combines the advantages of kanban system (WIP control for MTS products) and push system (dynamic production for MTO products) (Smalley, 2009).

Literature discussing mixed pull system in multi-stage multi-product environments focuses on implementations and case studies. Gates (2004) delineates a set of lean production tools to complement value stream mapping, which is used to identify wastes. One of the tools proposed is a mixed pull system in order to improve visibility and accountability on the shop floor. The system, supported by discrete event simulation, is tested in a facility manufacturing aircraft assembly parts. The simulation anticipated $350000 of cost saving and a reduction in production flow time by 60%.

Skelley (2004) outlines several lean production tools to be used in a facility manufacturing motion and control technologies. The facility has been facing difficulty in meeting due dates and requires the assessment of the root causes behind them. At the time of implementation, tools such as layout reconfiguration are already in use. The implementation of other tools such as standardization of work tasks and mixed pull system has improved the service level by 10%.

Bar (2006) describes a lean production framework for a facility manufacturing helicopter blades. The facility, previously using push system, is to be relocated to a new site, hence requiring the assessment of the production state for performance improvement. Tools such as task prioritization, team organizing, mixed pull system and progress tracking are discussed. Through implementation of a mixed pull system, the throughput is increased by 50%.

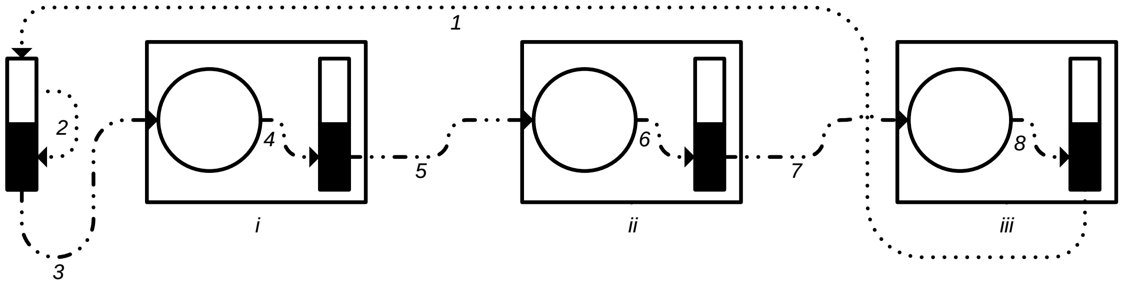
Horbal et al. (2008) explain a lean production framework for a facility manufacturing over 1000 types of valves. Tools such as work standardization, layout reconfiguration and mixed pull system are explained in detail. The improved system shows an increase in throughput by 33% and a reduction in space occupied by 50%. The paper finally presents a general lean production framework for multi-stage multi-product environments with low and high product variety.

Serrano et al. (2008) evaluates the applicability of value stream mapping by implementing a mixed pull system in the flow lines of six case study companies with various logistic problems. Several discrepancies between theoretical concepts from the value stream mapping and actual implementation are found. The discrepancies led to several conclusions pertaining to value stream mapping, such as the importance of communication between practitioner and resources as well as the formal definition of the theoretical concepts during implementation.

Saurin et al. (2011) assess the applicability of lean production tools in manufacturing cells. The paper distinguish various tools that are commonly used from existing literature, one of which is mixed pull systems, and produces a framework that identifies the relationship between these tools based on a survey with lean production experts. The framework is tested in a facility manufacturing automotive parts to identify possible shortcomings. One such shortcoming, which is the knowledge of the practitioner, is highlighted.

2.4.3 CONWIP systems

Figure 2.5 illustrates and summarizes the flows of material and information in a CONWIP system. In a CONWIP system, conceived by Spearman et al. (1990), a card is attached to each batch of finished goods. When one batch of finished goods is withdrawn, the card is detached and transferred to the first workstation. In the first workstation, the received card triggers production: the card is attached to a batch of raw material and processing begins. Upon completing the process, the batch is pushed to the immediate downstream workstation for processing. If this workstation is occupied, it waits until the workstation is vacant. This flow progresses until the batch reaches the end of the line where a final product is ready. Generally, kanban system provides localized WIP control through card circulation between consecutive workstations, while CONWIP system limits the WIP level in a line through card circulation between the first and last workstation (Spearman et al., 1990).



Sequence	Description
1	Card is detached from the batch and transferred to the input buffer
2	Additional card is attached to the a new batch of raw material
3, 5, 7	The batch is pushed to the immediate downstream production stage
4, 6, 8	The batch is pushed to the immediate downstream buffer

Figure 2.5: Flows of material and information in a CONWIP system

Literature addressing CONWIP systems in multi-stage multi-product environments is more extensive than mixed pull systems, ranging from analytical studies to implementation frameworks. In the literature cited hereafter, the term modified CONWIP system is introduced. This refers to the modification made to the operation of the original Spearman et al. (1990)'s CONWIP system in order to cater to a specific manufacturing environment. The details of these modified CONWIP systems are discussed further in Section 2.5.2.

Duenyas (1994) explores various order releases and dispatch rules in a multi-station multi-class queueing network in order to meet a desired throughput level. The paper compares the performance of a modified CONWIP system that uses static dispatch rules with other complex order releases that uses dynamic dispatch rules obtained from Wein (1992). In the modified CONWIP system, known as multi-loop CONWIP (M-CONWIP) system, separate cards are allocated to each product family.

The simulation results reveal that the M-CONWIP system using static dispatch rules is more effective than Wein (1992)'s, which is attributed to its simplicity.

Rubio and Wein (1996) consider a modified CONWIP system in a manufacturing facility where order release occurs when the WIP level drops to minimum, instead of during withdrawal of finished goods. The proposed system is essentially an open-loop queueing network. The objective is to find a function that relates the holding costs to the inventory level at the finished goods buffer. The function is further used to obtain optimal control parameters that attain highest service level.

Lee and Chen (1997) propose a dynamic dispatch rule in a wafer fabrication facility using a CONWIP system in order to attain low WIP level, high service level, high machine utilization and high throughput. The dispatch rule combines move control, weighted due date concepts and genetic algorithm based on critical real-time production information. The simulation results show that the proposed dispatch rule outperforms six conventional dispatch rules in all performance measures. However, the paper does not consider batching and setup time, which are important factors in wafer fabrication.

Kelkar (1999) also studies a modified CONWIP system to minimize the production cost of an agricultural equipment manufacturing facility. The modified system, known as the segmented CONWIP system, incorporates kanban system for segments of the line where products are MTS and CONWIP system for segments of the line where products are MTO. Using mixed integer linear programming and other cost minimization plans, the segmented CONWIP system is found appropriate for the problem in consideration.

Rose (1999) introduces another modified CONWIP system for a wafer fabrication facility. In the proposed system, known as constant load (CONLOAD) system, a new batch of raw material is released when one batch of product at the bottleneck is withdrawn. The simulation model does not include workstations downstream of the bottleneck as performance measures of interests are the WIP level and the utilization at the bottleneck. The simulation results construe that in comparison to a CONWIP system, CONLOAD system is found more effective due to the smaller variance resultant in production flow time.

Ryan et al. (2000) test the performance of M-CONWIP system in a job shop where products make multiple visits to the same workstation. The objective is to attain a fixed overall WIP level and its allocation to each product family. A simple heuristic that obtains these values is progressively derived from an optimization problem of an open-queueing network. Although the proposed method achieves high service level across all product families, cost for holding inventories of finished goods increases.

Wang et al. (2000) examine the effect of various order release strategies and dynamic dispatch rules in a facility manufacturing micro electro mechanical system in order to minimize the production flow time, minimize the WIP level and maximize throughput. A visual interactive simulation model is constructed for each combination of order release strategy and dynamic dispatch rule to assess the performances. The simulation results reveal that the combination of CONWIP system with the shortest remaining processing time dispatch rule is most effective.

In Rose (2001), CONWIP, M-CONWIP and CONLOAD systems are assessed in a wafer fabrication facility in terms of their abilities to cushion against variability in the bottleneck utilization and the WIP level. The simulation is modeled to consist solely of the bottleneck. The simulation results show that although M-CONWIP and CONLOAD systems are able to buffer the variation in the bottleneck utilization and WIP level, it also comes with an increase in the total WIP level and the production flow time.

Ryan and Vorasayan (2005) use nonlinear programming to approximate performance optimization for the allocation of a fixed number of CONWIP cards in an M-CONWIP system. Numerical examples from the nonlinear programming compared with simulation show that the results are negligibly different for all performance measures. In addition to having a fixed number of cards, the paper also presents a variant of the model that minimizes the total number of cards to achieve a targeted throughput.

Bahaji and Kuhl (2005) investigate the influence of push and CONWIP systems that use various dynamic dispatch rules on the variance of production flow time and service level in two wafer fabrication facilities. Following the simulation modeling and statistical analysis of the results, a composite dispatch rule is formed, and is demonstrated to perform well in the CONWIP system. The success of the composite dispatch rule in the CONWIP system is attributed to its robustness to adopt changes in the production flow time and throughput.

Mönch (2005) studies the performance of push, CONWIP and CONLOAD (explained in Rose (1999) in page 22) systems combined with a distributed shifting bottleneck heuristic in a wafer fabrication facility. The bottleneck heuristic is intended to solve the limitations of static dispatching rules, and is computed using a decomposition approach integrated with a disjunctive graph. The simulation results divulge that the CONWIP and CONLOAD systems outperform the push system and the bottleneck heuristic is more effective than the static dispatching rules, but only in the job shop with high product demand.

Wang and Prabhu (2006) develop a card setting algorithm for an M-CONWIP system subjected to routing and throughput requirements in order to minimize the WIP level. The proposed algorithm searches the WIP space iteratively and increases the step size until a minimum WIP level is reached. Parallel operation of the algorithm attains the minimum WIP level of each product family. The algorithm is tested in three simulation models with up to 20 product families. However, the paper makes no comparison with other production control systems.

El-Khouly et al. (2009) study a wafer fabrication facility, where products make multiple visits to the same workstation in a flow shop. CONWIP systems with various static dispatch rules are simulated to determine a suitable dispatch rule that is able to minimize the WIP level and to maximize the throughput simultaneously. The simulation results confirm that CONWIP system with the earliest due date dispatch rule drastically improves both performance measures.

Slomp et al. (2009) investigate the use of lean production tools in an MTO job shop, where the product variety is high and the production volume is low. In addition to the implementation of a CONWIP system, the implementation of tools such as ‘Takt’ time and production leveling in a facility manufacturing switchgear component is explained. The implementation led to a reduction in production flow time and increase in service level. Although the implementation was successful, the system requires continuous attention, as technicians can deviate from the rules of lean production.

Li (2009) describes the implementation procedure of lean production tools in a facility manufacturing pumps and pressure gauges. Two procedures discussed in