Model for Inventory Management in Valve Manufacturing Cell at Waters Corporation

by

Bingxin Yao
B.E. in Materials Engineering,
School of Materials Science and Engineering, Nanyang Technological University, 2012

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

Master of Engineering in Manufacturing

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2013

Author ....
Bingxin Yao
Department of Mechanical Engineering
August 09, 2013

Certified by ....
Jung-Hoon Chun
Professor of Mechanical Engineering
Thesis Advisor

Accepted by ....
David E. Hardt
Professor of Mechanical Engineering
Chairman, Committee for Graduate Student

On-Time Delivery Performance Improvement in a Valve Manufacturing Cell: Design of a Pull-Type Production System with Standardized Inventory Management

by
Bingxin Yao
B.E. Materials Engineering,
School of Materials Science and Engineering, Nanyang Technological University, 2012

Submitted to the Department of Mechanical Engineering on August 9th 2013
in partial fulfillment of the requirements for the degree of
Master of Engineering in Manufacturing

Abstract

This thesis addresses the challenges of improving the on-time delivery performance of a high-volume critical part type in a high-product-mix manufacturing facility of valves. Preliminary analysis on the push-type production system of the valve manufacturing cell shows that long production lead time caused by excessive inventory queuing and accumulation as well as lack of standardized finished goods inventory management policy are the major factors that limit the on-time delivery performance. A new pull-type production system is developed with the design of a highly responsive fabrication line which enables faster material movement and an efficient inventory review framework for real-time monitoring of inventory positions. A dedicated production line with the placement of effectively controlled Work-In-Process (WIP) buffers is constructed, which is capable of reducing the production lead time by more than 80%, along with a 40% reduction in overall WIP volume. Moreover, a finished goods inventory review policy is proposed based on the (s, S) policy which significantly eliminates the possibilities of backlog and inventory explosion by the setup of both lower and upper control limits on inventory positions. The suggested policy is expected to ensure a service level of at least 95% during peak demand period, with up to 50% potential reduction in average inventory level held by the system. A Kanban system is also established to coordinate operations in the proposed pull-type production system.

Thesis Supervisor: Dr. Jung-Hoon Chun
Title: Professor of Mechanical Engineering

Acknowledgements

Here, I would like to express my deepest gratitude to the individuals who have helped in my completion of this project, inspired and provided guidance for me during this period of work.

Firstly, I would like to express my utmost gratitude to my project advisor Professor Jung-Hoon Chun for listening to our ideas, inspiring our thoughts, sharing his knowledge and leading us in the right direction over the course of the project. My grateful thanks also go to Dr. Stanley Gershwin for his valuable advice on this project. Thanks also to Professor David Hardt, Dr. Brian Anthony and Mr. Jose Pacheco for their mentorship over the course of the MEng program at MIT and Ms. Jennifer Craig for her guidance in writing this thesis.

I would also like to thank Mr. Jim McPherson and Mr. Gabriel Kelly of Waters Corporation, who have honored us this precious opportunity to gain valuable industrial experience. My thanks also go to the directors and managers at Waters for supporting the implementation of our work and the supervisors and operators who patiently answered all our questions. Thanks also to my colleagues Snegdha Gupta and Yan Zhuang for their help, support and hard work in this fruitful endeavor.

Finally, I would like to thank my parents – for their patience, encouragement and guidance at all time. Many thanks also go to Mr. Obehi Ukpebor and Mr. Vrushank Phadnis for their company during difficult times.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Waters Corporation is a manufacturer of high performance liquid chromatography (HPLC) systems, mass spectroscopy and associated products such as chromatography columns, valves sample extraction instruments and chemical reagents. The Waters facility at Milford, Massachusetts is the headquarters of the company and houses an in-house production system for various components and sub-assemblies of the HPLC systems. The current manufacturing for valves— a critical sub-assembly of HPLC systems – is carried out in a separate manufacturing cell called the valve cell that manufactures 28 different types of stators for valve assembly in batches.

The main problem associated with the valve cell is the poor on-time delivery performance of stators to the assembly department. This can be attributed to the presence of such a high mix of parts that leads to long waiting times for part types resulting in unacceptably long production lead times and inefficient finished goods inventory planning and control making it susceptible to extreme situations of stock-outs or inventory explosion. With higher expected demand in future, these problems are likely to result in much longer average lead time on all parts, backlog and excessive work in process (WIP) inventory.

1.2 Process Overview

The processing of stators utilizes the valve manufacturing cell and the milling machine from the NC milling department. The valve manufacturing cell consists of two turning machines, 4 robo-drill machines and 2 wire-EDM machines. The 4 robo-drill machines are called Robo-drill machine 2, Robo-drill machine 3, Robo-drill machine 4 and Robo-drill machine 5.The shop floor layout is shown in Fig. 1. After being machined through the valve cell, the stators are processed through a series of cleaning and inspection processes. These processes include, de-burring, lapping, passivation, VCN cleaning and critical cleaning. For instance, the entire fabrication process for 212 stators, the highest volume part type in the valve cell, involves 10 steps. These 10 steps in order of operation are turning, milling, robo-drilling, wire EDM, de-burring, lapping, passivation, vacuum cycling nucleation (VCN) cleaning, critical clean and packaging. A more detailed process description will be given in Chapter 3.

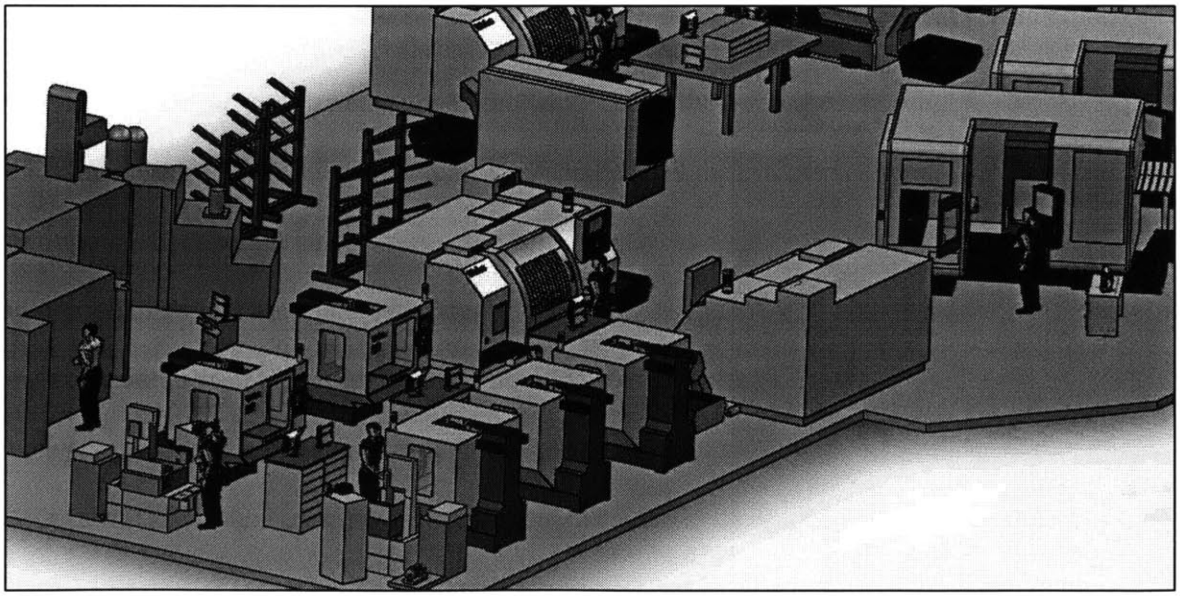


Fig. 1. Valve cell layout

1.3 Objectives

The primary objective of this project is to implement manufacturing systems and process improvements in the valve manufacturing cell of Waters Corporation at their Milford, MA facility. More specifically this translates to three main objectives:

1. Develop a responsive fabrication system for stators through lead time reduction.
2. Establish an efficient finished goods inventory review policy to achieve high service levels.
3. Determine the optimum operating conditions and tool material for throughput improvement of the bottle-neck process for additional responsiveness.

1.4 Scope and Thesis Organization

The 212 stators account for approximately 52% of the total valve cell production making it the highest volume and most critical stator. This serves as the rationale behind selecting 212 stators for studying the process flow in depth and carrying out major analyses for problem identification and improvement. Therefore, this project focusses on the lead time reduction for 212 stator fabrication, standardization of 212 stator finished goods inventory management and improvement of the robo-drill machine which is the bottle-neck in the 212 stator process flow. While the proposed solutions are associated with 212 stator fabrication, they seek to leverage the production of other stator types as well.

Inferences based on the study of the current process are described in detail in Chapter 3 while Chapter 4 to Chapter 8 discuss the proposed improvements and related implementation results.

1.5 Task Division

As mentioned above the project is divided into three main areas: lead time reduction, efficient inventory management and bottle-neck process improvement. Each team member is in charge of one area and delegates responsibility to other team members in his/her area based on expertise. Mr. Yan Zhuang is responsible for carrying out detailed statistical and machine capability analysis for bottle-neck process improvement in 212 stator manufacturing [1]. Ms. Snegdha Gupta [2] and I are responsible for the achievement of an efficient production system through a responsive fabrication line and standardized finished goods inventory management.

This thesis, in particular, focusses on the modeling and establishment of a standardized inventory review policy which ensures a service level of above 95% with potential reduction in average inventory level held by the system.

Chapter 2

Literature Review

Basic factory physics principles relating cycle time, throughput and inventory levels are discussed by Hopp & Spearman [3] and are summarized in this chapter. In addition to this, a discussion on various inventory review policies, push-pull production systems and implementing a pull planning framework is also provided.

2.1 Basic Principles of Factory Physics

If the utilization, U, of a workstation is defined as the fraction of time the workstation is not idle due to lack of parts, then U can be calculated as [3]:

$$U = \frac{r_a}{r_e} \tag{2.1}$$

where r_a is the arrival rate of parts, and r_e is the effective production rate which corresponds to the maximum average processing rate of the work-station taking into account failure rates, set-ups and other non-productive factors. It is important to note that increasing the utilization of a station without making any other changes causes a highly non-linear increase in the average WIP and cycle time.

The bottleneck rate, $r_{bottleneck}$, that can be defined as the throughput/production rate of the workstation having the highest long term utilization in a line, the raw process time, T_O, which is the average time it takes one job to traverse the entire line without waiting at any station, and the critical WIP level, $WIP_{critical}$, which is the WIP level that required to support the achievement of maximum throughput, $r_{bottleneck}$, assuming no variability in the system are related in the following way [3]:

$$WIP_{critical} = r_{bottleneck} \times T_O \tag{2.2}$$

The $r_{bottleneck}$ represents the capacity of the system which sets the control limit on the number of parts that can be released into the system or line. In other words the rate at which parts are released into the system should be less than or equal to $r_{bottleneck}$ for stability. However, in the practical world, even if efforts are made to release parts into the system at $r_{bottleneck}$, once a system reaches steady state, it will release work at an average rate that is lesser than the average capacity. This is an important factory law and helps better management decisions regarding capacity planning.

Investigating further into the relationship between WIP, process time and throughput, Little's Law [4] provides the following fundamental factory relationship (assuming no variability),

$$WIP = TH \times CT \tag{2.3}$$

where WIP, TH, CT represent the work in process, throughput and cycle time of the line, respectively. This law states that WIP is always equal to the product of throughput and cycle time at any WIP level. The Little's Law can be applied to one station, a line or an entire plant and can even be applied to lines with non- zero variability. The law provides insights such as how cycle time can be reduced by reducing the WIP for a given throughput, calculation of expected queue lengths at each work-station through the line by using the cycle time of the station. Utilization of a given station can also be deduced by knowing the queue length and number of machines at that station.

From the above, we realize that larger batch sizes result in more waiting and hence longer cycle times for a given throughput rate. Batching also has a significant impact on variability pooling and variability reduction. Therefore, batching laws are important in analyzing trade-offs regarding larger/smaller batch sizes. There are two kinds of batches – process batches and transfer batches [3].

A process batch size consists of a number of jobs of a part family that are processed on a workstation before it undergoes a set-up to process another part family. The process batch size is determined by how long it takes to change over to another part family. The rule therefore is in order to achieve a given capacity, the longer the set-up time the larger the batch size used. Also, as process batch sizes get large, cycle time increases proportionally with batch size. Other process batching laws state that the minimum process batch size to achieve a stable system maybe greater than one and that the cycle time at the station maybe minimized for some process batch size which may be greater than one.

A transfer batch size is the number of parts that are accumulated before transferring to the next station. The smaller the transfer batch size, the shorter the cycle time achieved since the wait to batch time is significantly reduced. Lot splitting is the technique of employing a large process batch to minimize utilization but a small transfer batch to reduce the cycle time. Smaller transfer batches however may result in more material handling resulting in a trade-off versus shorter cycle time. In fact, cellular manufacturing facilitates shorter cycle times through lot splitting owing to the physical compactness of the cell.

Therefore batching decisions can impact WIP, cycle time and throughput. If reducing cycle time is the main focus, then using a batch size just greater than the size that gives 100% utilization to maintain a stable system is favorable.

Another aspect of batching is variability pooling and variability reduction. Batching helps reduce variability [3]. Let t_o and σ_o be the mean and standard deviation of a random variable that describes the process time of a single part. The process time co-efficient of variation (CV) for this part is given as,

$$CV = \frac{\sigma_o}{t_o} \tag{2.4}$$

On the other hand, for a batch of n such parts the CV for a batch is given by,

$$CV(batch) = \frac{\sigma_o(batch)}{t_o(batch)} = \frac{\sigma_o\sqrt{n}}{t_o n} = \frac{CV}{\sqrt{n}} \tag{2.5}$$

Therefore batching helps in variability reduction and is especially important in sampling for quality control albeit its negative effects on cycle time and WIP levels.

Batching can help reduce variability but since variability is inherent in every manufacturing system, buffering to mitigate variability is required. Variability can in fact be buffered by some combination of inventory, capacity and time.

2.2 Inventory Policies

Two basic inventory review policies – the continuous review policy and the periodic review policy are described in detail by Simchi-Levi et al [5]. The major concepts involved in these policies are discussed in the following sections.

2.2.1 Continuous Review Policy

In a continuous review policy, inventory is monitored continuously and an order is placed whenever the inventory position reaches a particular point referred to as the reorder point [5]. In general, such a review policy leads to a highly responsive inventory management system.

In order to characterize this policy, it is important to understand the concept of inventory position. Inventory position is defined as the summed total of the actual inventory on hand plus the number of units ordered which have not arrived.

To implement the continuous review model in practice, we employ a typical approach known as the (*Q*, *R*) policy, in which an order of *Q* units would be placed whenever the inventory level drops to the reorder point *R* [6]. This concept is illustrated in Fig. 2.

The reorder point *R* here consists of two components, which covers the average inventory consumption during lead time and a safety stock to account for demand variabilities, i.e.

$$R = \mu L + SS \tag{2.6}$$

where μ is the demand rate in a specific time unit, L is the average lead time and SS represents the safety stock level.

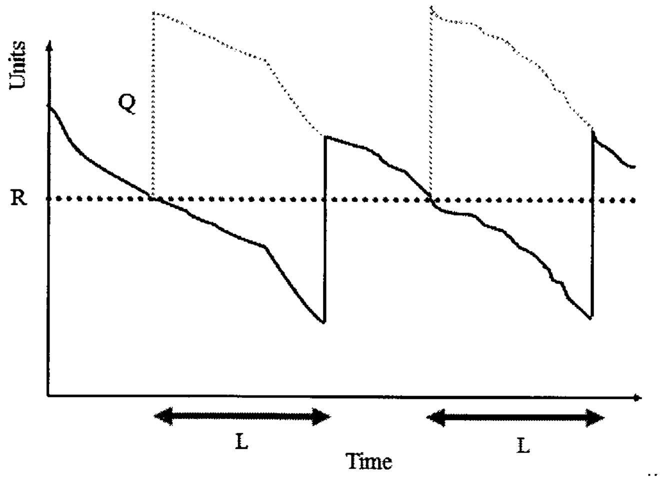


Fig. 2. Illustration for (*Q*, *R*) inventory review policy. *L* represents the lead time for an order of *Q*. The order is placed whenever the inventory position reaches *R*.[6]

The order quantity Q, on the other hand, needs to be optimized according to the nature of the distributor's operation. Two typical optimization models – the Economic Order Quantity (EOQ) Model and the Newsvendor Model are illustrated as follows.

2.2.1.1 Economic Order Quantity (EOQ) Model

The EOQ model determines the optimal order quantity based on cost minimization. It takes into consideration the tradeoff between inventory holding cost and ordering cost [7]. A small order quantity will result in a large ordering frequency, which leads to larger ordering cost, but at the same time reduces the average inventory level kept in the warehouse which, on the other hand, deceases the total inventory holding cost. In contrast, a large order quantity would require lower ordering frequency at the price of higher inventory holding cost.

Therefore, the economic order quantity Q^* is dependent on the inventory holding cost per unit h, ordering cost k as well as demand rate D. Mathematically, it is expressed as [7]:

$$Q^* = \sqrt{\frac{2k \times D}{h}} \tag{2.7}$$

2.2.1.2 Newsvendor Model

The newsvendor model aims to determine the optimal order quantity which maximizes the expected profit in a single planning period under stochastic demand [8]. The model defines an overage cost which is the difference between the original cost of an item and the salvage value of unsold inventory, as well as an underage cost expressed as lost profit due to unmet demand.

The optimal order quantity is hence established as a function of the overage and underage costs under certain demand scenario, given as [8]:

$$F(Q^*) = \frac{C_u}{C_u + C_o} \tag{2.8}$$

where $F(x)$ represents the cumulative distribution function of demand, and C_u and C_o are the underage and overage costs respectively.

2.2.2 Periodic Review Policy

In contrast to the continuous review policy discussed so far, in a period review policy, the inventory level is reviewed at fixed intervals on a regular basis and an order of appropriate quantity is placed after each review [5]. This policy is suitable to implement in systems where continuous review of inventory levels and frequent orders are inconvenient or costly.

The working mechanism of the periodic review policy is demonstrated in Fig. 3. As can be observed, this review policy is characterized by a single factor – the base stock level B. The target base stock level, along with a specific review period, r, is determined by the warehouse, the inventory position is then reviewed at these intervals and orders are placed to replenish the inventory back to the target level.

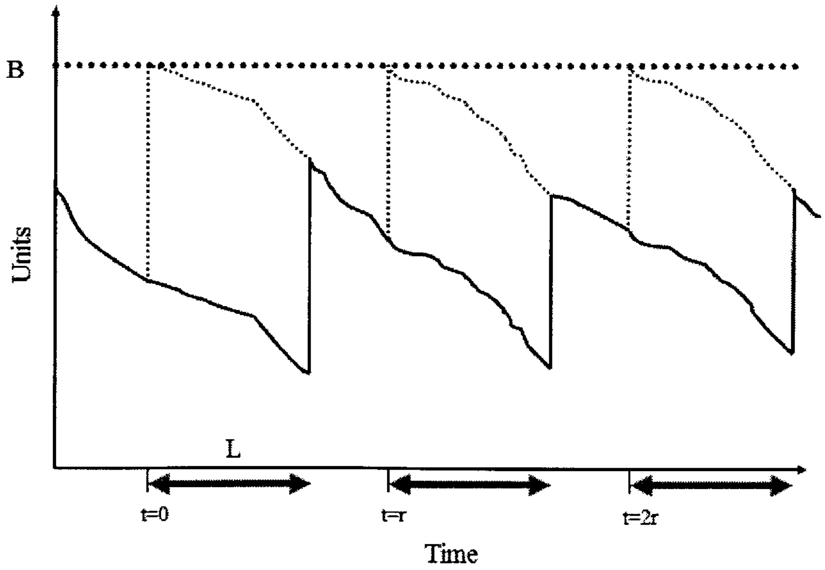


Fig. 3. Illustration for base stock inventory review policy. *L* represents the lead time for an order and *r* is the review period. An order is place after each review to raise the inventory position back to the base stock level *B*. [6]

The base stock level in a periodic review system is set up in a way that it would be enough to protect the warehouse from shortages until the next order reaches. Since orders arrive at intervals of r + L days, the base stock level should be able to cover the average demand during this period of time with a certain safety factor, taking into consideration the uncertainties in demand forecast. This is given by,

$$B = \mu \times (r + L) + SS \tag{2.9}$$

2.2.3 Vendor Managed Inventory

The two basic inventory review policies discussed previously effectively help to ensure a predetermined customer service level while at the same time keep the inventory level under control and prevent inventory explosion caused by inappropriate planning and anxiety. To further improve the coordination between the supplier and the customer and reduce overall inventory costs, strategic partnerships are usually employed. Vendor Managed Inventory (VMI) is a typical example of such partnerships in which information on demand is shared between the supplier and the customer [5]. The supplier is then given the authority to manage the inventory at the customer outlet and makes decisions on how much inventory should be kept on hand and when an order should be shipped [9]. This inventory management strategy integrates the operations at both sides and entirely eliminates the influence of variation inflation in a traditional supply chain system. The inventory cost incurred by the customer is reduced and the same would be true for the supplier in a long run due to better coordination of production and distribution.

2.3 Push-Pull Production System

A push system releases jobs for production based on a schedule which is in turn based on demand. A pull system on the other hand releases a job on the floor triggered by a signal signifying a change in the status of the line, for instance, the signaling of production upstream as parts when the process downstream requires parts as a result of a change in the downstream process [3].

Another way to look at a push system is as a make-to-order system as production is order based and not based on any signaling. However, a pull system can be seen as a make-to-stock model as production upstream is signaled by some void in the stock level downstream. In fact, the base stock model tends well to a pull system as orders can be triggered whenever the stock falls below the base stock level.

Most real world systems are however hybrid push-pull systems. For instance, if a job release is authorized by a Kanban card (feature of a pull system) but production is delayed due to anticipated lack of demand dictated by the master production schedule (feature of push system), then this results in a hybrid push-pull system [10-12].

2.4 Features and Benefits of a Pull Production System

The key feature of a pull production system is that it establishes a WIP cap and therefore avoids the production of extra WIP that do not contribute to increasing the throughput [3]. The WIP cap is established for instance using a Kanban system where the amount of WIP on the floor is limited by the number of Kanban cards used for triggering production. Also since a pull system is make-to-stock it automatically allows for the establishment of a WIP cap/level as any void in the stock level (that is whenever stock goes below a specified level) signals production to only fill up stock to the specified/base stock level. Since a pull production system helps in reducing WIP and keeping an upper bound on WIP, it helps in reducing average cycle times as well without compromising on throughput.

Therefore, the above features of a pull system result in a number of benefits like reduction of manufacturing costs since WIP will never grow beyond a pre-specified level, reduction of cycle time variability as the WIP cap also prevents any cycle time explosion (massive increase in cycle time) since it eliminates the danger of WIP explosion, pressure for quality improvements due to decreased WIP that facilitates defect detection and increased flexibility through delayed release of parts that makes engineering and priority or scheduling changes easy and ensures production of parts is authorized close to when the actual demand is realized to the maximum extent. Thus the pull system helps in developing a highly responsive customer service.

2.4.1 CONWIP and Kanban System

The easiest way to establish a WIP cap for the Pull system is through CONWIP (constant work-in-process). The WIP level is controlled by coordinating the release of a job with the departure of another job in the line (synchronized release and departure of jobs to maintain constant WIP in line) [3].

Benefits of a CONWIP system over a pure push system include the ability to observe the WIP levels directly as opposed to the possibility of WIP levels going up and down in a push system depending on the release rate determined by available capacity. Moreover, CONWIP helps achieve the same throughput with lesser WIP on average than a push system. An important factory law regarding the robustness of CONWIP states that, "CONWIP is more robust to errors in WIP level than the push system is to errors in release rates." Also, CONWIP can help in completing work ahead of schedule if circumstances permit.

Kanban system differs from CONWIP system in that it is more complex and usually involves the setting of more parameters than CONWIP [3]. For instance, CONWIP requires a single card count unlike Kanban that requires a card count for each station. However, a pure Kanban system results in a lead time of *zero* (part is available at outbound stock whenever required) whereas the lead time in a CONWIP system is always small but never zero. The Kanban system is more suited to a repetitive manufacturing environment i.e. a manufacturing environment where parts flow along a fixed sequence and at steady rates. CONWIP on the other hand shows more robustness to product mix (due to generation of work backlogs) as a result of line-specific cards. CONWIP can also adjust to a changing bottle-neck (due to product mix) because the WIP naturally accumulates in front of the bottle-neck. CONWIP also induces less operator stress as compared to the Kanban system where operators have to often wait for production signals even if they have raw materials to produce the required parts. Thus, CONWIP is more flexible than a Kanban system but unlike a pure Kanban system can never have a lead time of zero.

2.5 ARENA Simulation Software

ARENA simulation software is used for modeling dynamic processes and is a discrete event simulation software owned by Rockwell Automation [13]. Discrete-event simulation models systems as a sequence of discrete processes, each of which occurs at a particular instant of time resulting in a change in the state of the system.

ARENA simulation follows an entity based flowchart methodology and this makes it useful for documenting processes as compared to other simulation software that are either not visually compatible (purely code based) or focus more on process animation than process documentation.

Entities in ARENA (for example part types) flow through various processes that are depicted by modules that are connected to obtain a process flowchart for the system, and seize control of resource capacity as these entities are processed. This flowchart model of ARENA helps in accurately modeling and analyzing a process or system as the flowchart methodology facilitates documentation of each module or process. This results in highly detailed documentation and model development for the processes being analyzed.

Key advantages of ARENA simulation software that leverages the flowchart methodology are that it is easier to learn than other simulation tools, it is easier to validate, verify and debug, and it is easier to communicate details of complex systems or processes to others.

ARENA Academic Lab Package which is one of the ARENA academic software editions and is the academic and non-commercial version of the commercially available Enterprise Suite package which includes all available ARENA building blocks and additional features (add-ons like OptQuest for optimization, packaging, etc.) . The enterprise suite also does away with any system boundaries to analyze, model and solve.

Chapter 3

System Analysis of the Valve Manufacturing Cell

As introduced in Chapter 1, valves are critical components in liquid chromatography (LC) systems and they are strictly fabricated in-house at Water's valve manufacturing cell. This chapter provides a detailed system analysis of the current valve cell in regards to the average in-house production lead time, the conditions of the work-in-process (WIP) and finished goods inventories, as well as the overall on time delivery performance to the upstream assembly department.

3.1 Overview of the Valve Cell Production

3.1.1 Production Volume

Currently the valve cell is in charge of the production of 28 different types of valve stators. Each type of valve stator is given a specific material code such as "911000212", '"911000213" (the last three digits will be referred to in the rest part of the thesis). The total annual production of valve stators is around 31000 parts, among which 212 stator accounts for over 50% of the volume and 213 stator is about 10%. Apart from these two critical part types, 237, 230 and 251 stators are also frequently encountered part types with annual demand of over 1000 parts. The annual production volumes of the above part types, along with several other relatively high-demand parts are collected from May 2012 to April 2013 and are presented in Table 1.

Table 1. Annual production volume and percentage production of major part types at Water's valve cell

Part No.	Annual Production (Parts)	% Production
212	16355	52
213	3674	12
237	2300	7
230	1999	6
251	1533	5
250	968	3
236	809	3
215	650	2

The available machining and labor resources on the manufacturing shop floor for all these different part types include two turning machines, one milling machine shared by the NC milling department, four robo-drilling machines with different specifications, two electrical discharge machining (EDM) centers, one micro-deburring room, one lapping room and one passivation room.

Since 212 stator is the highest volume part type in the valve cell production, we will thus use 212 stator as an example for the illustration of manufacturing process flow on the shop floor.

3.1.2 Manufacturing Process Flow

The manufacturing process for 212 stators from raw material is illustrated in Fig. 4. Firstly, raw material bars are fed into the turning machine to achieve the outline profile shown in Fig. 5. The turned blanks are then sent to the milling machine which is located in NC milling department outside the valve cell. The vertical holes shown in Fig. 5, together with a curvy slot located at the bottom of the stator are made through milling. Thereafter, the milled blanks are sent back to the valve cell and the conical holes are created in the robo-drill. One point to note is that among the four robo-drills (robo-drill 2, robo-drill 3, robo-drill 4 and robo-drill 5) available on the shop floor, 212 stators are usually processed in robo-drill 4 or 5. Robo-drill 2 performs vertical hole drilling for all other part types except 212, and robo-drill 3 is a five-axis machine which mainly deals with stators that require a vertical hole at the center of the part in addition to the conical holes. Nevertheless, during high demand period, a portion of 212 stator production will also be done in robo-drill 3 given the high capability of the machine. After robo-drilling, parts are sent to the EDM center where the bottom holes indicated in Fig. 5 are drilled. The above mentioned steps comprise the machining process for the manufacturing of 212 stators. Associated machine reliability data is given in Table 2.

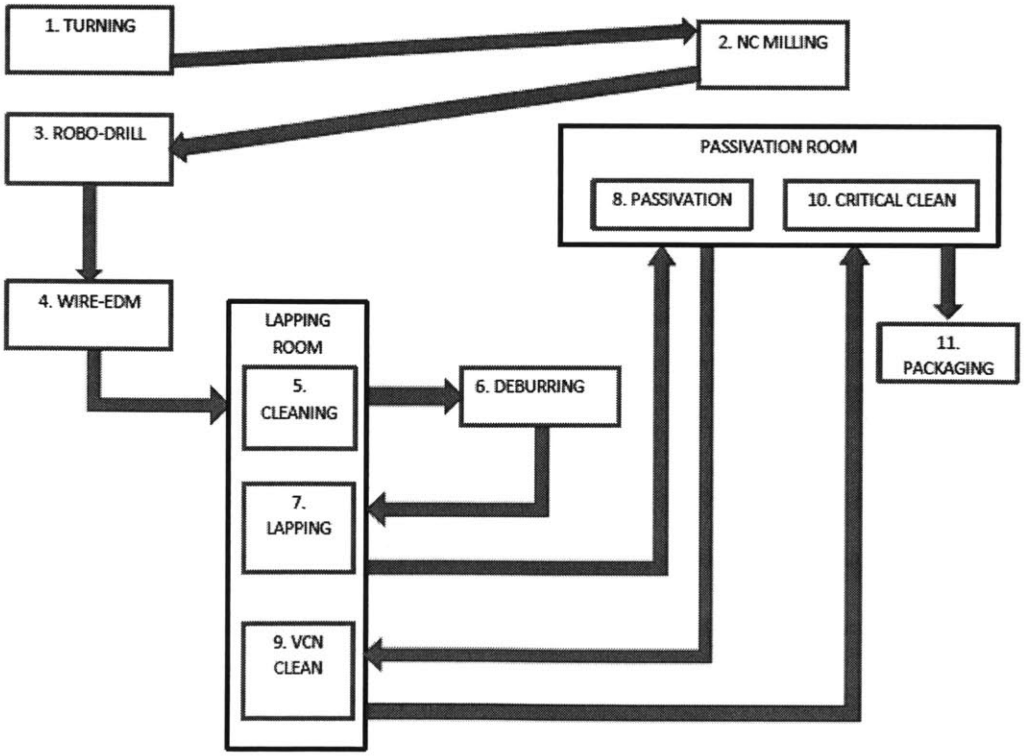


Fig. 4. Manufacturing process flow for part 212

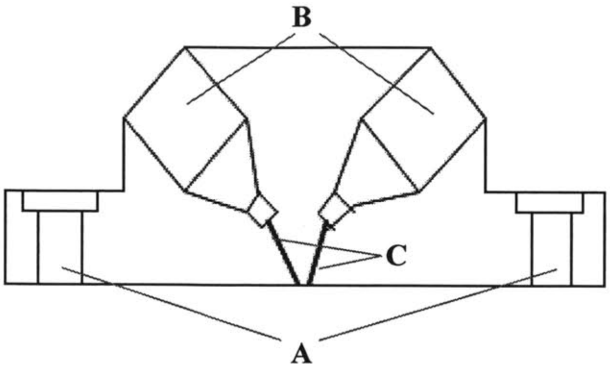


Fig. 5. Sketch of 212 stator showing: A. Vertical holes along the circumference of the part; B. Conical holes; C. Bottom holes

Table 2. Efficiency of machines in the valve manufacturing cell (MTTR – Mean time to repair; MTTF – Mean time to fail)

Machine	**MTTR （days）**	**MTTF (Days)**	**Efficiency**
Lathe	2	33	0.94
Milling	0.5	87	0.99
Robo-drill	1	49	0.98
EDM	1	22	0.96

After the machining steps, parts are sent in batches for a series inspection and cleaning processes. This is shown in Fig. 4 as Steps 5 to 10. After EDM, parts go through a brief cleaning process in the lapping room. Deburring is then performed by operators manually to remove small particles stuck in the channels of the parts. Thereafter, parts are sent back to the lapping room for a polishing process to achieve better surface finish. This is followed a passivation process in the passivation room, vacuum cycling nucleation (VCN) cleaning back in the lapping room, and finally a critical cleaning process in the passivation room.

Once all the inspection and cleaning procedures are completed, parts are packaged and placed in a stock room. Later these parts are sent to an outside vendor for coating which usually takes 18 days to accomplish. The coated parts sent back by the vendor are the final products of the valve manufacturing cell and they are stored in the finished goods inventory to fulfill orders from the upstream assembly department.

The manufacturing process flows for the rest 27 part types are very similar to that of 212 stators. The only major difference, as mentioned previously, is that the vertical holes on 212 stators are done through milling while for all the other part types, the job is performed by robo-drill 2. The cause of this differentiation is the unique curvy slot at the bottom of 212 stators which can only be created by the milling machine in NC department. Therefore, the vertical holes are made together with the curvy slot in NC department so as to save setup times.

The machining process flow and machining time per part for major stator types are presented in Table 3. The process times for the subsequent inspection and cleaning steps remain the same for all part types, as shown in Table 4. Note that the values in Table 3 and Table 4 are the maximum allowable processing times assigned to each step, thus actual operations may take shorter time.

Table 3. Machining time per part for major stator types

Process / Part No.	Turning (min)	Milling (min)	Robodrill 2 (min)	Robodrill 3 (min)	Robodrill 4/5 (min)	EDM (min)
212	3.72	7.2	-	-	16.5	4.2
213	3.72	-	9.96	-	16.5	4.2
215	3.72	-	9.96	16.5	-	4.2
230	3.72	-	9.96	-	16.5	31.2
236	3.72	-	9.96	-	16.5	4.2
237	3.72	-	9.96	-	16.5	19.8
250	3.72	-	9.96	-	16.5	4,2
251	3.72	-	9.96	-	16.5	21.1

Table 4. Process time per part from deburring to critical clean

Process	Deburring	Lapping	Passivation	VCN Clean	Critical clean
Time (mins)	2.1	3	0.6	0.06	0.18

Another important feature of the valve manufacturing process is that parts are produced based on internally generated factory orders. Order quantity is a fixed value for individual part types and can be interpreted as batch size here. For instance, 212 stators move along the production line with factory orders of 100 parts, that being said, at any station 100 parts need to be accumulated into one batch before being sent to the downstream station. The current batch sizes used for major part types are summarized in Table 5.

Table 5. Current batch sizes for major stator types

Part No.	Batch size (parts)
212	100
213	25
237	100
230	80
251	100
250	96
236	16
215	30

As can be observed, large batch sizes are used for most of the major part types. The rationale behind using large batch sizes is to reduce setup times between part changes so as to control machine utilization. However, this could also result in undesirable consequences such as excessive part queuing and inventory build-up along the production line, which will be discussed in Section 3.2.2.

3.2 System Performance of the Valve Cell

3.2.1 On-Time Delivery Performance

As the starting point of the entire LC system manufacturing line, the valve cell takes great responsibility in providing parts on time to the downstream assembly department so as to ensure no breakdown in operations. Therefore, on-time delivery performance is the most important and indeed an ultimate measure of the system performance at the valve cell.

Recall the production volume distribution shown in Table 1, the apparent high demand for 212 stators makes it the most critical part type in both the manufacturing and the assembly departments. Therefore, it would be wise to consider the on-time delivery of 212 stators as our priority and conduct a thorough analysis on that. In addition, the analysis on 212 stators, to a large extent, would also be representative of the entire system as 212 stators account for more than half of the total valve cell production.

The weekly on-time delivery performance of 212 stators over Quarter 2 of 2013 is measured and summarized in Fig. 6.

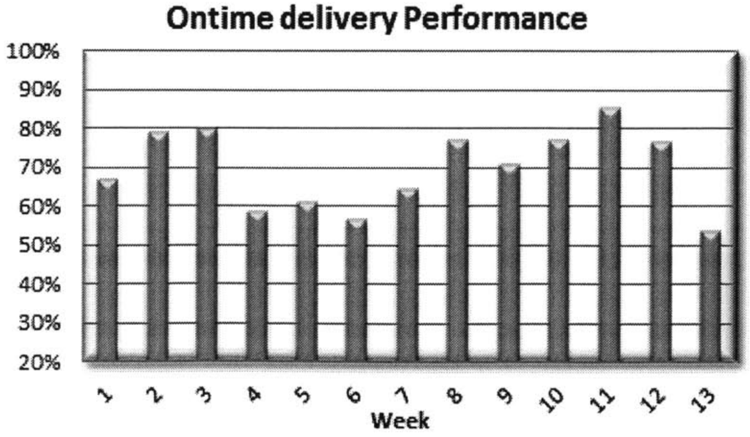


Fig. 6. Weekly on-time delivery performance of part 212 in Quarter 2, 2013

From Fig. 6, we can observe that the on-time delivery performance of 212 stators ranged from 53% to 85%. This could indicate a considerable amount of shortages in the assembly department. In fact, a serious operation breakdown did happen in Quarter 1 of 2013, during which the assembly area was shut down due to complete shortage of stators.

This result presents a pressing need for us to improve the on-time delivery performance of the valve cell. In order to do that, we need to first figure out the root causes that hinder the on-time delivery of parts. Through our analysis of the entire manufacturing system at the valve cell, two factors have been identified, namely, long production lead time in house and poor inventory management at the finished goods inventory. These will be discussed in the following sections.

3.2.2 Analysis of In-House Production Lead Time

As discussed previouly, the in-house production of 212 stators includes machining steps from turning to EDM as well as subsequent inspection and cleaming procedures up to critical clean. Parts are processed based on factory order of 100 parts per order. In order to extract the average in-house production lead time for one order of 212 stators, we tracked 33 orders in SAP system durring Quarter 2 of 2013 and obtained the production lead time for each order. The lead time distribution of these orders is shown in Fig. 7. The average production lead time is approximately 21 days per order, with a standard deviation of 6.2 days. The 95% confidence interval for lead time falls between 18.7 days to 23.1 days, which suggests that the in-house production of a 212 order will most probably take 18.7 to 23.1 days to accomplish. Since orders are usually planned to be completed within 15 days of release, such a long actual production lead time would definitely result in poor on-time delivery performance.

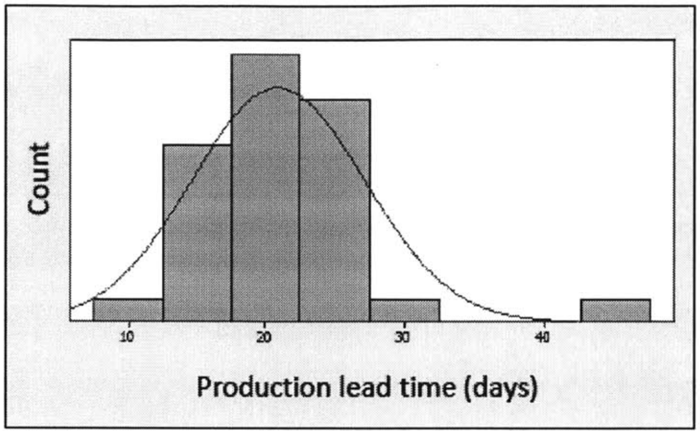


Fig. 7. Lead time distribution of 33 orders of 212 stators. The mean lead time is about 21 days with a standard deviation of 6.2 days.

However, if we consider the system capability, an order of 212 stators could indeed move through the entire system within a much shorter period of time. This can be calculated by summing up the processing times for 100 parts of 212 stators at each station, i.e.

$$\text{Production lead time} = \sum Processing\ times$$

$$= \sum(Turning\ time + Milling\ time + \cdots + Critical\ clean\ time) \qquad (3.1)$$

Substitute the processing times given in Tables 3 and 4 into Equation (3.1), we get a production lead time of approximately 4 working days for an order of 212 stators, provided 16.5 working hours/day and an additional 75min of cleaning is required after lapping.

The difference between the observed production lead time and the theoretical calculation clearly indicates that work-in-process inventories of 212 stators cannot move through the system continually. There must be significant queuing and thus inventory accumulation along the line. These are characterized as inventory waiting time and inventory volume between work stations, and are illustrated as follows.

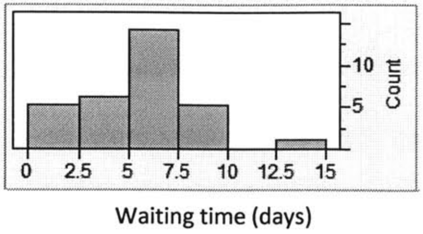
3.2.2.1 Inventory Waiting Time

To characterize inventory waiting time between stations, we further analyzed the 33 orders of 212 stators and obtained the waiting time before each work station for every individual order. The waiting time distributions in front of work stations are summarized in Fig. 8. The top left graph represents the inventory waiting time before milling, and the top right one shows the waiting time before robo-drill, and so on.

We can observe excessive inventory waiting time between milling and robo-drill, which is of no surprise as robo-drill is the bottleneck process in the production line. Apart from that, long waiting time is also observed between turning and milling, which could be attributed to the fact that milling is performed in a different department where lots of other types of jobs are performed and work coordination is more difficult. From EDM onwards, inventory waiting times are generally negligible as those processes are much faster in nature.

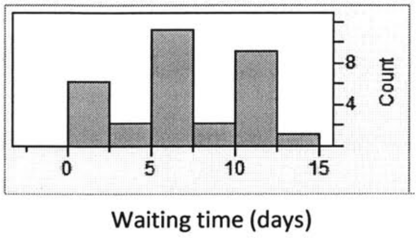
This inventory waiting phenomenon is actually the consequence of the push system based on which current production is conducted. In a push production system, there is no real-time control over WIP level and orders are released merely in accordance to planning. Therefore, parts are introduced into the system "blindly" regardless of whether there are already other parts waiting in the system. As a result, there would be high chances for queue formation along the line, especially in front of slower processes.

Milling



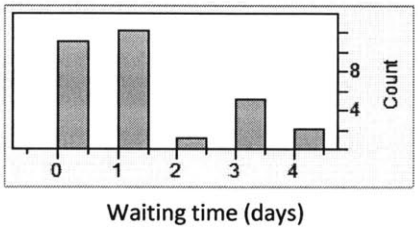
Mean: 5.2 Std dev: 2.71

Robo-drill



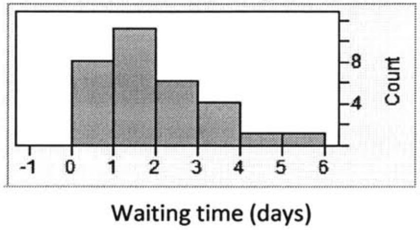
Mean: 6.8 Std dev: 3.62

EDM



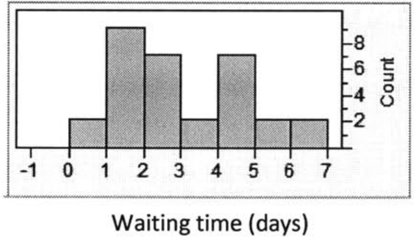
Mean: 1.2 Std dev: 1.28

Deburring



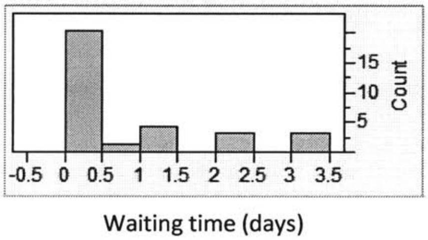
Mean: 1.4 Std dev: 1.28

Lapping



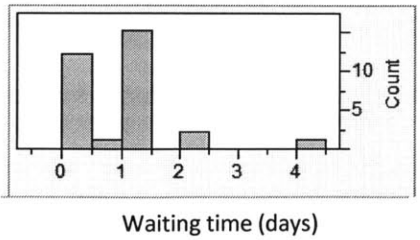
Mean: 2.6 Std dev: 1.71

Passivation



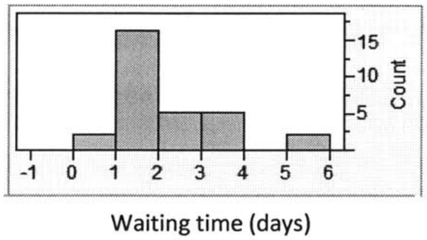
Mean: 0.6 Std dev: 1.02

VCN clean



Mean: 0.77 Std dev: 0.84

Critical clean



Mean: 1.7 Std dev: 1.24

Fig. 8. Inventory waiting time distribution in front of each work station

3.2.2.2 Work in Process Inventory volume

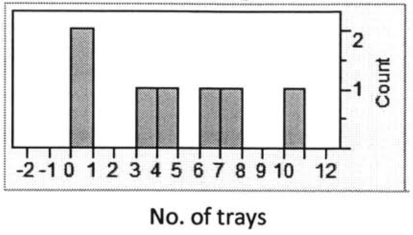
With the excessive inventory waiting times observed previously, we would also expect considerable WIP build-up along the line. To characterize the WIP level, we conducted a 10-day experiment during which the inventory levels in front of each work station were recorded at randomly selected time point. The distributions of inventory volume are presented in Fig. 9. Note that inventory volume is expressed as number of trays, where one tray usually contains 24-28 parts. In total about 20 trays were observed in the system.

Similar to the results obtained for inventory waiting time, we can observe significant inventory accumulation in front of robo-drill, whereby nearly no inventory resides between stations after EDM.

As discussed previously, this observation could also be explained by the lack of WIP level monitoring in a push based production system. In addition, the large batch size used would also contribute to the high WIP level in the system.

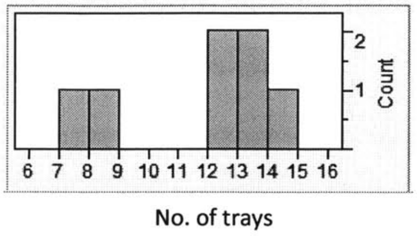
To sum up, we have identified that the current push production system at the valve cell causes WIP build-up and long inventory queuing time, which slows down material movement along the line. As a consequence, the in-house production lead time appears much longer than what the system is capable of and that in turn would contribute to the poor on-time delivery performance.

Milling



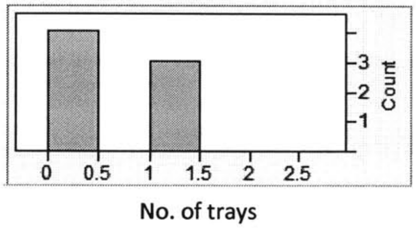
Mean:4.29 Std dev: 3.68

Robo-drill



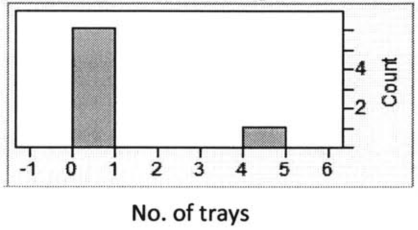
Mean: 11.29 Std dev: 2.69

EDM



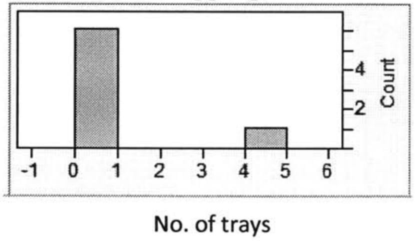
Mean: 0.43 Std dev: 0.54

Deburring



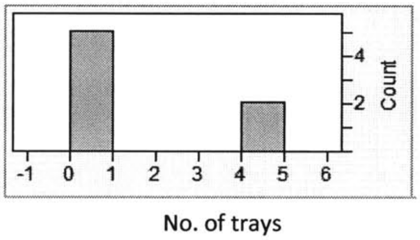
Mean: 0.57 Std dev: 1.51

Lapping



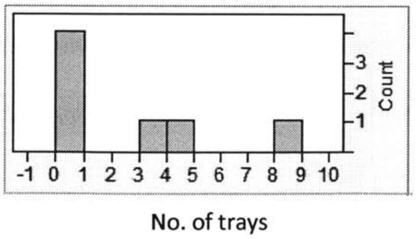
Mean: 0.57 Std dev: 1.51

Passivation



Mean: 1.14 Std dev: 1.95

VCN clean



Mean: 2.14 Std dev: 3.07

Critical clean

No inventory observed

Mean: 0 Std dev: 0

Fig. 9. Inventory volume distribution in front of each work station

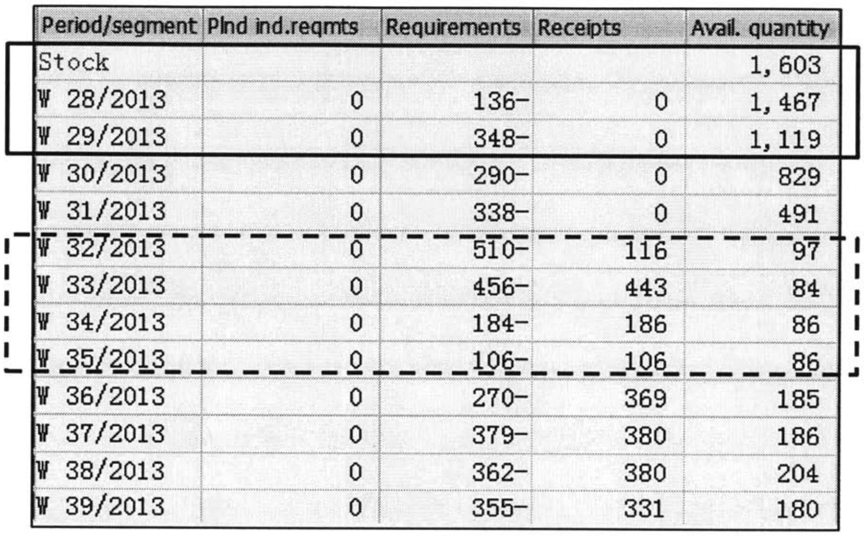
3.2.3 Analysis of Finished Goods Inventory Condition

Apart from long in-house production lead time, we have previously mentioned in Section 3.2.1 that poor finished goods inventory management is another cause of low on-time delivery performance of the valve cell. This is clearly visible after investigating the current finished goods inventory planning condition.

Table 6 shows the weekly finished goods inventory planning for Quarter 3 of 2013. We can observe from the data that planned inventory level is highly fluctuating, which goes as high as 1600 parts and drops below 100 parts for four consecutive weeks. From a planning perspective, it is rather inadequate to project such an unstable profile which would make the system vulnerable to demand uncertainties. For instance, since Weeks 32 – 35 correspond to the month of August which is usually a high demand period based on the historical data, such a tight inventory planning of less than 100 parts available in stock over the whole month would leave the system inflexible and highly endanger it to backlogs. Also, further investigation in the SAP system revealed an average inventory level of around 700 parts held in the system over a long run, which is unexpectedly high considering the poor on-time delivery performance. This is most probably a consequence of inventory explosion happened during unforeseen demand drops.

Since the finished goods inventory of valves is the last station which connects the valve cell production line to the assembly department, failure to manage it properly would directly cause shortages of parts in the assembly area or excessive inventory buildup. Therefore, it is critical to identify the shortcomings of the current inventory management mechanism and tackle them accordingly. Through further communication with the planning department, two aspects regarding the concepts of safety stock and standardized inventory review policies have surfaced.

Table 6. Finished goods inventory planning for Quarter 3, 2013 (The 1st row represents the stock level at the end of Week 27)



Period/segment	Plnd ind.reqmts	Requirements	Receipts	Avail. quantity
Stock				1,603
W 28/2013	0	136-	0	1,467
W 29/2013	0	348-	0	1,119
W 30/2013	0	290-	0	829
W 31/2013	0	338-	0	491
W 32/2013	0	510-	116	97
W 33/2013	0	456-	443	84
W 34/2013	0	184-	186	86
W 35/2013	0	106-	106	86
W 36/2013	0	270-	369	185
W 37/2013	0	379-	380	186
W 38/2013	0	362-	380	204
W 39/2013	0	355-	331	180

3.2.3.1 Lack of Safety Stock

Safety stock is the amount of inventory kept to account for demand uncertainties. In Water's current planning system, demand uncertainties are not properly taken into consideration and there is not setup for safety stocks. The planned inventory levels are calculated based on average demand forecast and are thus established to cover merely the average demand over product lead time. However, demand forecasts can never be accurate and significant uncertainties always exist [5]. A system without safety stock will not be able to respond immediately to demand surges or operation disruptions and hence backlogs will result frequently.

3.2.3.2 Lack of Standardized Inventory Review Policies

As previously introduced in Chapter 2, standardized inventory review policies facilitate close monitoring of inventory levels in stock and thus trigger production whenever needed. However, due to the push nature of the current production system at the valve cell, no proper inventory review is employed for the finished goods inventory. Production is initiated by master planning based on demand forecasts rather than being triggered by a specific real-time condition at the finished goods inventory. Such a system exhibits low flexibility and responsiveness to demand changes. Inventory tends to accumulate when unexpected demand drop happens as shown for Weeks 27 – 28 in Table 6 and on the other hand, backlogs will also result during abrupt demand surges which in turn lead to poor on-time delivery performance.

3.3 Remark

Through our analysis on the valve manufacturing cell, we have identified the improvement of on-time delivery performance of valve stators as the pressing need of the company. This could be tackled from two aspects – reduction of in-house production lead time and establishment of a standardized review policy for the finished goods inventory. Analysis in Sections 3.2.1 and 3.2.2 calls for the implementation of a pull production system instead of the current push system. Specifically, what we need is a pull system with proper inventory review at the end of the line to trigger production in time, and effective WIP control along the line that facilitates faster material movement in the system. On top of that, the concept of safety stock should be employed to further improve the robustness of the system. These points were considered during the redesign of the valve cell production line and will be demonstrated in Chapter 4.

Chapter 4

Design of a Pull Production System

Following the discussion in Chapter 3, we have identified that the fundamental approach for on-time delivery performance improvement is to establish a pull production system which effectively incorporates a WIP control mechanism along the line and a standardized inventory management policy at the finished goods inventory. This chapter presents an overview of our design of such a pull production system.

The design of a pull production system at the valve manufacturing cell is illustrated in Fig. 10. In general, production in this pull based system is initiated by triggers sent from the finished goods inventory at the end of the line when replenishment is needed. The triggering signal travels upstream and parts will be withdrawn from the nearest available WIP inventory along the line and processed. Therefore, proper WIP inventories need to be established to reduce the triggering signal path length so as to cut down the replenishment lead time. Similarly, triggers will also be generated whenever parts are withdrawn from a particular WIP inventory and hence signal production upstream.

In the context of our project, we intend to implement pull production for 212 and 213 stators which are the two highest volume products in the valve cell. The design in Fig. 10 shows two dedicated production lines (one for 212 stators and one for 213 & all other stator types) in the machining shop floor which later converge into one common line in the inspection and cleaning area. WIP buffers are setup after the turning, milling (robo-drill 2 for 213 stators)

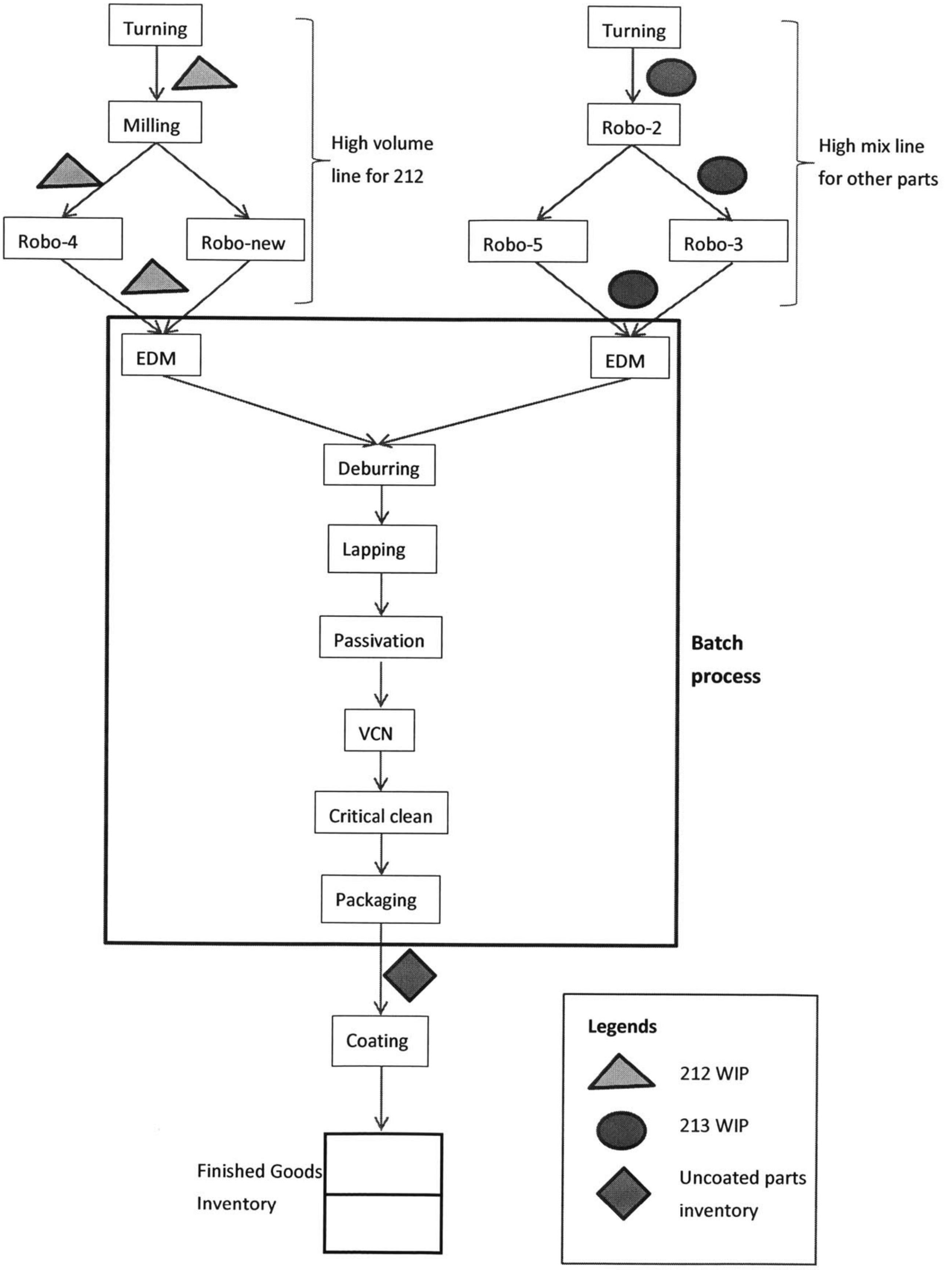


Fig. 10. Proposed design of a pull production system

and final robo-drilling stations. When a replenishment signal is generated at the finished goods inventory, it will be sent to the uncoated parts inventory for part withdrawal. The depletion of parts at the uncoated parts inventory will then result in a signal that triggers production from EDM. Parts will thus be taken from the robo-drill inventory and at the same time, another signal will be passed on to the robo-drilling station for the replenishment of this inventory. As such, signal will continue traveling upstream towards the turning station. Production at each work station is in fact triggered by the replenishment signal from downstream WIP buffers. This is a clear distinction between our new system design and the traditional scheduling based production in a push system.

The key benefit of employing such a pull system is that the service time to the assembly department is essential zero as parts can be directly withdrawn from the finished goods inventory. This, of course, could only be realized with an appropriate inventory review policy. Furthermore, the system is highly robust and responsive to demand variations and operation disruptions, owing to the fact that both the WIP and finished goods inventory levels are closely monitored such that productions could be triggered in time especially under vibrant circumstances. In addition, standardized control over inventory levels also eliminates the possibility of excessive inventory build-up or explosion, which is in consistency with the lean manufacturing culture promoted by the company.

Besides the basic concept of pull production mentioned above, the entire system could only work with proper detailed line design that optimizes responsiveness and well established inventory review policy which effectively coordinates production and distribution. These two major design criteria will be discussed in Chapter 5 and Chapter 6, respectively.

Chapter 5

Design of a Responsive Production Line

Throughout our discussion in Chapter 4, we have pointed out that a pull-type system would only work well with a production line designed for high responsiveness and a finished goods inventory managed with proper and standardized review policies. This chapter presents our design ideas and underlying rationales for a responsive production line. For more detailed mathematical modeling of the line, refer to the work of Snegdha Gupta [2].

5.1 Line Dedication

5.1.1 Description of the High Volume and High Mix Production Lines

As demonstrated in Fig. 10, we have designed two dedicated production lines in the machining area – one high volume line dedicated to the production of 212 stators and one high mix line that handles the production of all other part types. The machining resources allocated to the high volume line include one turning machine, one milling machine shared by the NC department, robo-drill 4, one new robo-drilling machine (referred to as robo-new in the diagram) which has already been purchased by the company, and one EDM machine. As for the high mix line, one turning machine, robo-drill 2, 3 & 5, as well as one EDM machine is available. After the machining process, the resources available for subsequent inspection and cleaning steps are shared by all part types, which remains the same as the current system.

The dedication of milling machine to 212 stators and robo-drill 2 to all other part types is due to the unique curvy slot located at the bottom of 212 stators which can only be made by milling. This was previously explained in section 3.1.2. Also, although robo-drill 3 has the capability to process 212 stators, it is important to allocate it to the high mix line since certain part types in this line can only be processed by robo-drill 3 due to the existence of a central vertical hole in these parts.

The rationale behind this line dedication design is essentially to reduce the inventory queuing time along the production line. According to Simchi-Levi et al [5], inventory congestion and long inventory waiting times are largely caused by the various types of low volume parts in the system. Although these parts are low in volume individually, they comprise a considerable WIP level as a whole and this hinders the material movement of 212 stators which is a much more critical part type. By allocating 212 stators in a separate production line, this inventory congestion would be avoided and 212 stators could thus move through the system faster. Similarly, parts in the high mix line would also face less congestion and be processed in a faster manner. There is, however, no need for such line dedication after EDM since no significant inventory accumulations were observed at these inspection and cleaning stations. In addition, since the production volume of 212 stators accounts for 52% of the total stator production, it appears that the required machining times from both lines would be rather balanced, which makes the idea more appealing.

Nevertheless, before moving into the detailed line design, capacity analysis on each machine is needed to actually verify the feasibility of such line dedication. The calculations for capacity analysis are shown in Section 5.1.2.

5.1.2 Capacity Analysis

5.1.2.1 Capacity Analysis for the High Volume Line

As previously presented in Table 1, the annual production volume of 212 stators is 16355 parts. This is translated into 63 parts/day on average, assuming 52 weeks/year and 5 working days/week.

Thus, the required machine time per day, T_r, for a specific machine is calculated by,

$$T_r = \mu_d \times t_p \tag{5.1}$$

where μ_d is the daily demand rate of 212 stators and t_p is the corresponding processing time per part for that machine.

Also, the ideal available operation time for machines in the valve cell (turning machine, robo-drills and EDM) is 16.5 hrs/day and that of the milling machine is 22.5 hrs/day as the NC department operates on a different schedule. However, taking into consideration the machine efficiencies listed previously in Table 2, the actual available machine time, T_{avail}, would be given by,

$$\mathrm{T}_{avail} = T_{schedule} \times \mathrm{E} \tag{5.2}$$

where $T_{schedule}$ is the scheduled operation time of a machine and E is the machine efficiency.

Therefore, the required utilization, U_r, of each machine in the high volume line could be computed as follows,

$$U_r = \frac{T_r}{T_{avail}}$$

or

$$U_r = \frac{\mu_d \times t_p}{T_{schedule} \times \mathrm{E}} \tag{5.3}$$

Referring to the unit processing time data shown in Table 3, the required utilizations of the turning machine, milling machine, robo-drilling machine and EDM are calculated using Equation (5.3) and are summarized in the Table 7. Note that the available machine time for robo-drills is 32.3 hrs/day as two robo-drills are allocated to the line.

Table 7. Required machine utilization for the high volume line

Machine	T_{avail} **(hrs/day)**	μ_d **(parts/day)**	T_r **(hrs/day)**	U_r **(%)**
Turning	15.5	63	3.9	25
Milling	22.3	63	7.5	33
Robo-drill (4+new)	32.3	63	17.3	53
EDM	15.8	63	4.4	27

Considering the fact that the milling machine is shared between the valve cell and the NC department and that it performs various other types of jobs on a regular basis, it is important to identify the required utilization on the milling machine by other jobs and make sure that the total utilization is kept below 100%. In order to obtain information on the required milling utilization by other jobs, we collected data on weekly operation time spent on other jobs from May 20th, 2013 to July 21st, 2013 as shown in Table 8 and figured out that the average milling time required by other jobs is approximately 67 hrs/week, which corresponds to a machine utilization of 50%. Therefore, the total required utilization of the milling machine is about 83%. Although this value is relatively high as compared to the other machine utilizations, it is still well below 100% and thus the machine is capable of producing required amount of parts on time.

Table 8. Required milling machine utilization by other types of jobs except 212 stators (Average values are calculated excluding the week starting July 1st due to the Independence Day holidays)

Week	Available machine time (hrs/week)	Time spent on other jobs (hrs/week)	Utilization (%)
Jul-15	134	47.7	36
Jul-08	134	61.8	46
Jul-01	134	42.0	31
Jun-24	134	69.3	52
Jun-17	134	110.2	82
Jun-10	134	93.9	70
Jun-03	134	67.7	51
May-27	134	43.3	32
May-20	134	42.8	32
	Average	**67.1**	**50**

From the above capacity analysis, we can conclude that the high volume line is capable of handling the total 212 stator production.

5.1.2.2 Capacity Analysis for the High Mix Line

The capacity analysis for the high mix line is performed in a similar manner.

The total annual production of all other part types is about 48% of the stator production, that is, 15120 parts/year, or 58 parts/day. One major difference between the capacity analysis of the high mix line and the high volume line is that machine setup times need to be considered for the high mix line as it is dealing with 27 different part types. For conservative purpose, we have assumed that one machine setup is required after the processing of each batch and that a batch size of 25 parts is used for every part type which is usually the smallest batch size observed on the shop floor. Therefore, for an average daily production of 58 parts/day, a maximum of 3 setups will be needed. A summary of the longest required setup times, t_s, for each machine is given in Table 9. Note that EDM does not require setups between batches.

Table 9. Longest required machine setup times for the high mix line

Machine	**t_s (hrs)**
Turning	1
Robo-drill 2	1
Robo-drill 3/5	0.5
EDM	0

In this case, the required machine time per day, T_r', for a specific machine is calculated by taking into account the setup times, i.e.

$$T_r' = \mu_d \times t_p + N \times t_s \tag{5.4}$$

where N is the maximum number of setups required per day on the machine.

The T_r' values for the turning machine, Robo-drill 2 and Robo-drills 3&5 are calculated following Equation (5.4). The situation for EDM is a bit more complicated due to the fact that parts 230, 237 and 251 require considerably longer EDM machining time. The average required daily machining time for these special parts can be deducted from their respective daily demands, and the total required EDM time in the high mix line would be the sum of the required times for these special parts and that of the rest 24 part types. This is illustrated in Table 10.

Table 10. Required daily EDM machining time for parts 230, 237, 251 and all other part types in the high mix line

Part No.	EDM process time (min/part)	Demand (parts/day)	Required EDM time (hrs/day)
230	31.2	8	4.0
237	19.8	9	2.9
251	21.1	6	2.1
Others	4.2	35	2.5
		Total EDM time (hrs/day)	**11.5**

Therefore, the required utilizations for all the machines in the high mix line can be calculated using Equation (5.3) and the results are presented in Table 11.

Table 11. Required machine utilization for the high mix line

Machine	T_{avail} (hrs/day)	μ_d (parts/day)	T_r (hrs/day)	U_r (%)
Turning	15.5	58	6.6	43
Robo-drill 2	16.2	58	12.7	78
Robo-drill (3+5)	32.3	58	17.5	54
EDM	15.8	58	11.5	73

The calculation above demonstrates that the capacity of the high mix line is sufficient for the entire production of the 27 low volume part types.

Therefore, the feasibility of implementing dedicated productions in the machining area is verified through the capacity analysis of both lines. The machine utilizations for the high mix line are observed to be higher than the high volume line owing to the changeover times required between different part types. Nevertheless, this analysis is done with very conservative assumptions on batch sizes and changeover times. In reality, the required machine utilizations for both lines would be more balanced and kept well below 1.

5.1.2.3 Capacity Analysis for the Inspection and Cleaning Area

As introduced previously, after going through the machining processes, the two dedicated production lines will merge together and the subsequent inspection and cleaning resources are shared by all part types. A similar capacity analysis for the inspection and cleaning area is performed based on the total average daily demand of all stators and the results are summarized in Table 12. Note that machine efficiencies are not considered in this case as most of the processes in this phase are labor intensive.

Table 12. Required labor/machine utilization for the inspection and cleaning area (two lapping machines are available for the surface finishing of stators)

Process	T_{avail} (hrs/day)	μ_d (parts/day)	T_r (hrs/day)	U_r (%)
Deburing	16.5	121	4.2	26
Lapping	33.0	121	13.6	41
Passivation	16.5	121	1.2	7
VCN clean	16.5	121	1.8	11
Critical clean	16.5	121	0.4	2

From the above table, we can conclude that the inspection and cleaning area possesses enough capacity for the processing of all valve stators. The extremely low required utilization in the passivation and critical clean area is practically reasonable as these processes are very fast and the area handles the cleaning job for all product families produced in the manufacturing department.

After verifying the proposed idea of line dedication, we will now move on to the key design features of each individual line, as presented in Sections 5.2 and 5.3.

5.2 Design for the High Volume Production Line

As discussed previously in Chapter 3, the in-house fabrication of 212 stators starts from the turning operation at the valve cell and ends at the uncoated parts inventory. To facilitate effective pull based production of 212 stators, this fabrication line needs to be highly responsive and well-coordinated with the final finished goods inventory. While this is partially achieved by the idea of line dedication, some other important in-line designs are also incorporated to enable the whole process. In general, these include the establishment of a standardized replenishment mechanism for the uncoated parts inventory as well as a push-pull production boundary on the shop floor.

5.2.1 Establishing a Replenishment Strategy for the Uncoated Parts Inventory

As illustrated in Fig. 10, the uncoated parts inventory is located right after the packaging process and it is the last station before parts being sent for coating at outside vendors. Due to the long lead time of 18 days for the coating process, the uncoated part inventory plays a crucial role in maintaining continuous operation of the entire system and ensuring on-time delivery to the assembly department. It is important for this inventory to provide parts immediately whenever a call for replenishment arises at the finished goods inventory.

In light of that, we have designed a base stock replenishment policy for the uncoated parts inventory. In our policy, the actual stock position at the inventory is monitored real-time and a signal will be generated and sent to the upstream station whenever the inventory position drops below a target base stock level. The signal triggers production which will then replenish the inventory back up to the base stock level. This target level is developed such that it is sufficient to cover the average demand of 212 stators over the replenishment lead time quoted by the upstream production line. A safety stock component is also incorporated to account for uncertainties in demand forecasts. Such a standardized review and replenishment mechanism would thus largely ensure the availability of parts at all time.

Based on the average daily demand of 212 stators derived from the total annual production volume, we have identified a target base tock level of approximately 500 parts, or 18 trays, for the uncoated parts inventory. This includes a 170-part safety stock to absorb variations. The detailed calculations are presented in Snegdha Gupta's work [2].

5.2.2 Establishing a Production Push-Pull Boundary

Once the review policy at the uncoated parts inventory is established, it is important to have an upstream production line that is capable, responsive, and stable enough to support the execution of such policy. One key point to focus on in this case is the reduction of production lead time from turning to critical clean, which would eventually enable faster replenishment to the uncoated part inventory.

Recall our discussion in Section 2.4, it is noted that the lead time for a pure pull based production system is 0 as parts can be withdrawn immediately from buffers. However, this is not practical in our system as it requires the setup of WIP buffers in between every work stations along the line, which would result in excessive total WIP volume within the system. Therefore, the problem left is to identify a strategic location for the production push-pull boundary so as to effectively cut down the production lead time of the "pull segment" of the system while maintaining a low overall WIP level.

Refer to the line design shown in Fig. 10, the push-pull boundary of our proposed system is located at EDM. The system works on a pull basis up to EDM with WIP buffers established in front of each machining station, and parts coming out from EDM are pushed through the rest of the system until reaching the uncoated parts inventory. With this design, the replenishment to the uncoated parts inventory is initiated from EDM and hence the processing time before EDM is eliminated from the total production lead time. According to Snegdha Gupta's work, the production lead time of an order of 336 parts is expected to be 2.12 days[2].

The rationale behind such push-pull boundary allocation is that the processing time of the steps before EDM is considerably larger than those after. Cutting down this processing time would make a much greater impact on the system performance. In addition, from the analysis presented in Chapter 3, we have noticed that the inventory queuing time and accumulation after EDM is negligible, which indicates smooth material flow through the inspection and cleaning processes. There is hence no pressing need to eliminate the processing time of these steps.

In addition, it is also important to emphasize that for the "pull segment" of the production line, WIP buffers established in between the turning, milling, robo-drilling and EDM work centers are critical in maintaining the proper functioning of the push-pull system. Parts should be available for withdrawal whenever production is triggered at the downstream work center, and at the same time WIP explosion should be avoided.

To closely monitor the WIP level at each buffer, we have decided to employ a similar base stock review policy. A target base stock level is determined for each buffer and replenishment back to the target level is required whenever parts are taken away from the buffer. Based on the machining time at each station as well as the average daily demand of part 212, the base stock levels were calculated [2] and summarized in Table 13.

Table 13. Targeted Base stock levels after respective work stations (1 tray = 28 parts)

Work station	Base stock level (trays)
Turning	5
Milling	6
Robo-drill	5

In this manner, the policy ensures immediate availability of parts when needed and simultaneously generates a WIP cap in the system which effectively prevents inventory build-up. The maximum allowable WIP level in the system is 16 trays, which is 24% less than the observed volume in the current system.

It is also important to note that during exceptionally high demand period where continuous production from EDM is required, since the EDM process is faster than the robo-drilling process, the robo-drill WIP buffer will eventually run out and the EDM machine will be starved for a short period of time [2]. Nevertheless, this potential starvation problem can be resolved with the reduction of robo-drilling cycle time to 14.2 min/part as introduced in Yan Zhuang's work [1].

For the "push segment" of the line, batch processing is required due to the nature of the cleaning processes. An optimal batch size of 28 was determined based on the consideration for system utilization and process cycle time. The detailed calculations are given in Snegdha Gupta's work [2].

5.3 Recommendations for the High Mix Production Line

As the focus of this project is to improve the on-time delivery performance of 212 stators which is fabricated in the high volume production line, we will thus not go deep into the design of the high mix production line. Nevertheless, the following ideas are recommended and could serve as good directions for future work regarding this line.

5.3.1 Pull Production for 213 Stators

Since 213 stators is another critical part type produced in the valve cell which accounts for approximately 11% of the total annual production of stators, we believe it is wise to implement a similar pull system for the production of 213 stators in the high mix line. The base stock review policy introduced previously would also be suitable for the real-time monitoring of the uncoated parts inventory as well as the WIP buffer levels. One major drawback of this idea is that the demand for 213 stators is much less regular and highly variable, which could result in undesirable high WIP buffer levels.

5.3.2 Push Production for Other Low Volume Part Types

For the rest 26 low volume part types, we recommend to retain the current planning-based push production system. According to Hopp & Spearman [3], a pull production system does not work well for a high mix, low volume product line. Practically, it would also be difficult to set up WIP buffers required by a pull production system for these part types on the shop floor. In fact, keeping inventories for such low volume parts usually increases the overall system costs considerably [5]. Furthermore, by isolating 212 stators into a separate production line, the inventory congestion and queuing faced by these low volume parts would be significantly reduced and we could hence expect much smoother material movement along the line. Thus, there is no need for implementing a pull production system for these parts. Rather, more work could be done from the planning perspective to seek for optimal changeover sequences so as to reduce setup times required and scraps generated.

Chapter 6

Managing the Finished Goods Inventory – Modeling and Simulation

Finished goods inventory is the last station located in the valve cell which connects the valve cell production with the assembly department. Through our analysis in Chapter 3, we have realized lack of finished goods inventory management policy as the direct cause of the poor on-time delivery performance of the valve cell. A standardized inventory policy is in need to support the pull production system proposed previously and hence effectively coordinate production and distribution. This chapter provides quantitative analysis on various inventory review policies and identifies the most suitable policy for implementation based on service standard, cost analysis and practical constraints.

6.1 Basic Inventory Policies

6.1.1 Continuous Review Policy

As the name suggests, in a continuous review policy, inventory position is monitored continuously and an order of quantity Q is placed whenever the inventory position drops down to the reorder point R. Therefore, continuous review policy is also referred to as (Q, R) policy.

It is important to note that the term “inventory position” here is defined as the sum of the actual inventory level on hand and that on order which is yet to come. In other words, an order will only be triggered when this summed value reaches the reorder point, while the real inventory level in stock might be way below. This will be illustrated more in detail in Section 6.1.1.3.

To characterize a system under the continuous reviewed policy, three important quantities need to modeled, namely, the reorder point R, the order quantity Q and the average inventory level held in the system $\bar{I}$.

6.1.1.1 Modeling for the Reorder Point

As introduced above, since orders in a continuous review policy are placed when the inventory position drops to the reorder level and it takes an average lead time of μ_l days for the order to arrive, the reorder point should thus be set to cover the average demand over μ_l and also absorb any demand variations over this period. The average demand over lead time, d_{avg}, is given by,

$$d_{avg} = \mu_d \times \mu_l \tag{6.1}$$

Also, a safety stock SS is added to account for demand uncertainties so as to ensure a projected service level. Let the standard deviation of daily demand be σ_d, according to statistical laws, the standard deviation of demand over a constant lead time, $Std(d_{avg})$, can be expressed as,

$$Std(d_{avg}) = \sigma_d\sqrt{\mu_l} \tag{6.2}$$

The safety stock *SS* required for a certain service level can be computed by multiplying the above variation term with a safety factor *z*, i.e.

$$\mathrm{SS} = \mathrm{z}\sigma_d\sqrt{\mu_l} \tag{6.3}$$

The safety factor *z* here corresponds specifically to a predetermined service level and is extracted from the demand profile. Assume the demand pattern follows a normal distribution with cumulative probability function *F(X)*, to achieve a service level of α, the safety factor is given by,

$$\mathrm{z} = F^{-1}(\alpha) \tag{6.4}$$

Therefore, combining Equation (6.3) and (6.4), the safety stock level is calculated as,

$$\mathrm{SS} = F^{-1}(\alpha)\sigma_d\sqrt{\mu_l} \tag{6.5}$$

Equation (6.5) assumed a constant lead time for orders placed. However, due to the complexity of our production system, we would expect variable delivery times to the finished goods inventory. For instance, when a replenishment call arises at the finished goods inventory, an 18-day lead time would be quoted if there are parts readily available at the uncoated parts inventory so that they could be sent out immediately for coating. However, the situation becomes worse if there are not enough parts at the uncoated parts inventory and production needs to be triggered at the EDM station to fill up this inventory first. According to Simchi-Levi et al [5], for a replenishment process with standard deviation, σ_l, in lead time, the safety stock level should be adjusted accordingly. The modified safety stock level is given by,

$$SS = F^{-1}(\alpha)\sqrt{\mu_l \times \sigma_d^2 + \mu_d^2 \times \sigma_l^2} \tag{6.6}$$

Combining Equations (6.1) and (6.6), we get the expression for reorder point *R* as,

$$R = \mu_d \times \mu_l + F^{-1}(\alpha)\sqrt{\mu_l \times \sigma_d^2 + \mu_d^2 \times \sigma_l^2} \tag{6.7}$$

6.1.1.2 Modeling for an Optimal Order Quantity

Continuous review policy employs a fixed order quantity *Q* for inventory replenishment. The determination on an optimal order quantity can be approached from two perspectives – cost minimization or expected profit maximization. These two aspects are modeled by the Economic Order Quantity (EOQ) model and the Newsvendor model, respectively.

6.1.1.2.1 Economic Order Quantity (EOQ) Model

The EOQ model considers inventory holding cost *h* and ordering cost *k*, and focuses on the minimization of total cost associated with these events. It usually assumes a long planning horizon based on relatively rough demand data and allows no backlog in the system. The expression for the optimal order quantity, Q, is given as,

$$Q = \sqrt{\frac{2k \times D}{h}} \tag{6.8}$$

where D is the average annual demand for 212 stators.

Note that in this case the equation provides a standard (constant) order quantity throughout one year's operation as the annual demand data is used.

6.1.1.2.2 Newsvendor Model

Different from the EOQ model, the Newsvendor model aims to optimize the total expected profit generated by the production activity. Planning is usually done for a short single period based on more recently updated demand. Since profit is the major concern, this model is more conservative in terms of productivity and backlogs are allowed in the system. The optimal order quantity, Q, is expressed as,

$$F(Q) = \frac{C_u}{C_u + C_o} \tag{6.9}$$

where C_u is the underage cost due to unfulfilled demand and C_o is the overage cost of an unsold item.

6.1.1.2.3 EOQ Model vs. Newsvendor Model

Based on the key features of the above two order quantity models, we have selected the EOQ model for implementation in the valve cell for the following two reasons.

Firstly, as our main focus in this project is to improve the on-time delivery performance 212 stators, the "no backlog" assumption in the EOQ model is more favorable.

Secondly, the long planning period of EOQ model makes it more suitable for practical implementation. Although the Newsvendor order quantity is theoretically more precise owing to the more accurate demand data used for each single planning period, it would be difficult to implement in our case due to the required frequent adjustments of order quantities.

6.1.1.3 Simulations for Continuous Review Policy

6.1.1.3.1 Reorder Point and Order Quantity Setup

In order to decide the reorder point R, we need to first identify the average lead time μ_l for an order of Q parts and the standard deviation σ_l of the lead time. This could be done by considering the following 5 possible replenishment scenarios.

Scenario 1 – Parts are readily available in the uncoated parts inventory and thus can be sent for coating immediately.

Scenario 2 – Parts are not available in the uncoated parts inventory and the order of Q parts needs to be produced from the EDM station, provided that enough parts are available at the robo-drill WIP buffer in front of EDM. In this case, EDM is the bottleneck process. Note that although some amount of parts might be available in the uncoated parts inventory, for conservative purpose, we have assumed an empty uncoated parts inventory here. This assumption will be used for the rest 3 scenarios as well.

Scenario 3 – Parts are not available in both the uncoated parts inventory and the robo-drill WIP buffer. The order of Q parts needs to be produced from the robo-drill station with enough parts available at the milling WIP buffer. The bottleneck process is the robo-drilling step.

Scenario 4 – Parts are not available in the uncoated parts inventory, the robo-drill as well as the milling WIP buffers. Production is started from the milling station with enough turned blanks available at the turning WIP buffer. Similar to Scenario 3, the robo-drilling process is the bottleneck for the replenishment activity.

Scenario 5 – Parts are not available in the uncoated parts inventory as well as all WIP buffers. Production has to start from the turning station with the robo-drilling process being the bottleneck.

Since the projected service level for all the WIP buffers and the uncoated parts inventory is $\alpha = 98\%$ according to Snegdha Gupta's work [2], we can thus calculate the probability of occurrence for the above 5 scenarios and the associated lead time in each case.

Let the probability of occurrence for the ith scenario be P_i, then P_i can be calculated as,

$$P_i = \begin{cases} \alpha \times (1-\alpha)^{i-1}, & 1 \le i \le 4 \\ (1-\alpha)^{i-1}, & i = 5 \end{cases} \tag{6.10}$$

assuming raw material bars are available at all time.

The production lead time for the ith scenario, L_i, can be obtained from the following equation,

$$L_i = \begin{cases} L_{coating}, & i = 1 \\ L_{coating} + T_{bottleneck-i} \times \dfrac{Q}{n} + \sum T_{non-bottleneck}, & 2 \le i \le 5 \end{cases} \tag{6.11}$$

where $L_{coating}$ represents the coating lead time quoted by the outside coating vendor and $L_{coating}$ = 18 days for 212 stators, n is the batch size employed for in-house production and $n = 28$ parts/batch as recommended in Snegdha Gupta's work [2], $T_{bottleneck\text{-}i}$ is the batch processing time for the bottleneck process in the ith scenario and $\sum T_{non-bottleneck}$ is the summation of the batch processing times for all non-bottleneck processes involved in the specific scenario. A summary of the batch processing times for all steps of the in-house production of 212 stators is shown in Table 14.

Based on the probability of occurrence and production lead time associated with each individual scenario, the mean replenishment lead time, μ_l, of an order of Q parts can be calculated as,

$$\mu_l = \sum_{i=1}^{5} P_i L_i \tag{6.12}$$

The standard deviation, σ_l, of the lead time is hence given by,

$$\sigma_l = \left(\sum_{i=1}^{5} P_i (L_i - \mu_l)^2\right)^{1/2} \tag{6.13}$$

Table 14. Batch processing times for all steps of the in-house production of 212 stators, using a batch size of 28 parts/batch [2]

Process	**Batch processing time, T (min)**
Turning	104
Milling	202
Robo-drilling	231
EDM	118
Deburring	59
Lapping	84
Passivation	17
VCN Clean	2
Critical Clean	5

Recall Equations (6.7) and (6.8) developed in Section 6.1.1.2, we can now acquire the optimal order quantity and the reorder point for the finished goods inventory of 212 stators based on the ordering and inventory holding cost data provided in Table 15 as well as the annual demand profile shown in Fig. 11. The demand profile shows an annual demand of 16600 parts with a standard deviation of 321 parts/month. This could be translated into 64 parts/day with a daily standard deviation of 59 parts/day. Note that Fig. 11 is a rough demand profile made on monthly basis. For more accurate system setup, updated demand forecasts should be derived and this will be illustrated in the Chapter 8.

Table 15. Ordering cost and inventory holding cost for 212 stators

Quantity	**Value**
k ($/order)	70
h ($/part•year)	21

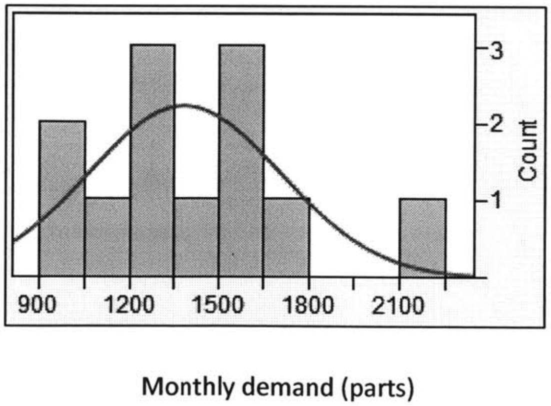


Fig. 11. Annual demand profile for 212 stators from July, 2012 to June, 2013 (For updated demand forecast and system setup for the current month, refer to Chapter 8)

For a projected service level of 98%, the results for the optimal order quantity and reorder point calculations taking into consideration different replenishment scenarios are shown in Table 16. As can be observed, an optimal order quantity of 330 parts and a reorder point of 1650 parts are established for the finished goods inventory of 212 stators under the continuous review policy. The system holds a safety stock of 499 parts.

Table 16. Calculation results for optimal order quantity and reorder point under continuous review policy, taking into consideration different replenishment scenarios

Quantity	Values	Comments
Q (parts)	330	≈12 trays (1 tray = 28 parts)
P_1	0.98	-
P_2	0.0196	-
P_3	3.9×10^{-4}	-
P_4	7.8×10^{-6}	-
P_5	1.6×10^{-7}	-
L_1 (days)	18	-
L_2 (days)	20	-
L_3 (days)	21	-
L_4 (days)	21	-
L_5 (days)	21	-
μ_l (days)	18	-
σ_l (days)	0.2	-
SS (parts)	499	≈18 trays
R (parts)	1650	≈59 trays

6.1.1.3.2 Average Inventory level in Stock and Associated Costs

Apart from service level, another important attribute that defines the performance of a policy is the average inventory level in stock as it is directly related to the cost the company would incur. This level could be derived by looking into the working mechanism of the continuous review policy.

So far, through calculations we have obtained all the important parameters for a continuous review policy. These values are illustrated in Fig. 12. The figure presents a simplified demonstration for the working mechanism of our established policy. The solid line in the figure shows the actual inventory level in stock and the dotted line represents the "inventory position (sum of inventories on hand and those on order)" of the system. As can be seen, an order of 330 parts is placed whenever the inventory position hits the reorder level of 1650 parts. This marks an immediate increase in inventory position. On the other hand, the inventory level in stock will not increase right after the placement of an order. Rather, it will keep going down until the actual arrival of an order which then raises it up by 330 parts. Therefore, although the reorder point established for the system seems high, the real inventory level held on hand is indeed much lower.

Once the system reaches steady state, it can be clearly observed from Fig. 12 that the average inventory level $\bar{I}$ is the average value of the highest and the lowest inventory levels in the system.

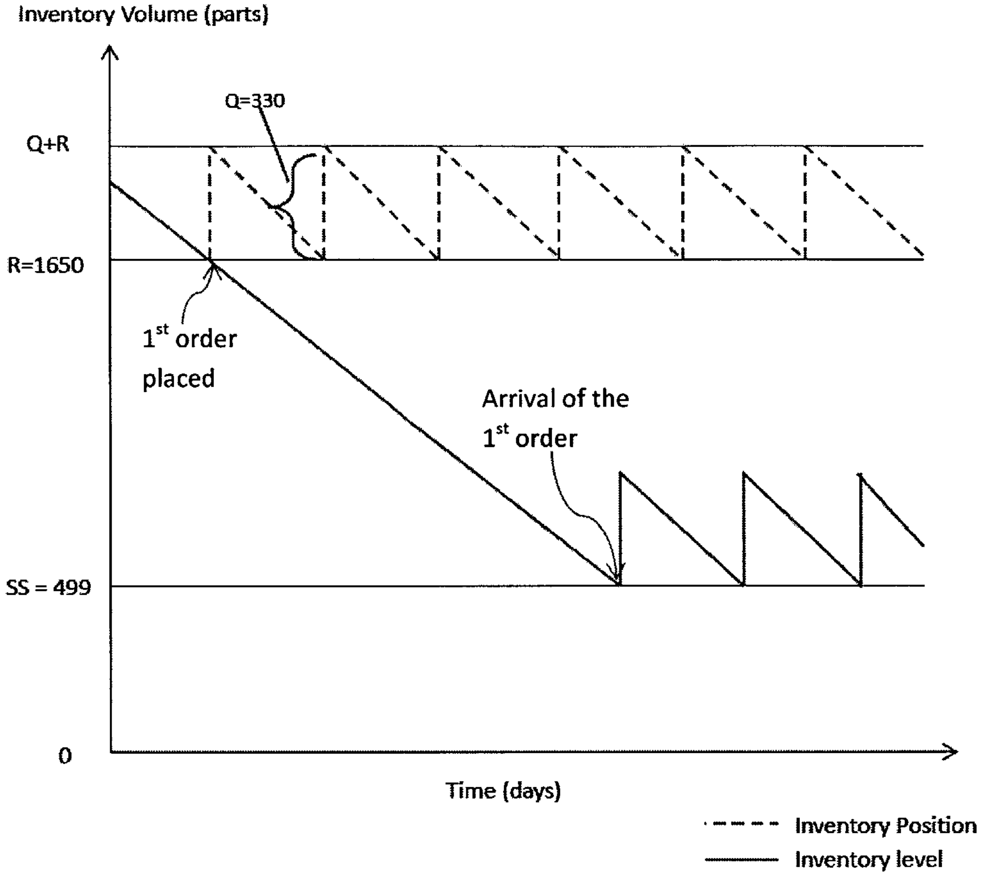


Fig. 12. Demonstration for continuous review policy based on calculated R, Q and SS

The lowest inventory level, I_{low}, is observed right before the arrival of an order, which is expected to be the safety stock level, i.e.

$$I_{low} = SS \tag{6.14}$$

Similarly, the highest inventory level, I_{high}, would be observed right after the receipt of an order, and hence is it given by,

$$I_{high} = SS + Q \tag{6.15}$$

Therefore, the average inventory level $\bar{I}$ in the system can be calculated as,

$$\bar{I} = \frac{I_{low} + I_{high}}{2}$$

or (6.16)

$$\bar{I} = SS + \frac{Q}{2}$$

Based on the average inventory level, the annual inventory holding cost, *H*, can be obtained as,

$$\mathrm{H} = \mathrm{h} \times (SS + \frac{Q}{2}) \tag{6.17}$$

Also, the total annual ordering cost, *K*, could be derived assuming constant demand, i.e.

$$\mathrm{K} = \frac{D}{Q} \times k \tag{6.18}$$

Thus, the total cost, *C*, of inventory holding and ordering incurred by the company is given by,

$$\mathrm{C} = \mathrm{H} + \mathrm{K} \tag{6.19}$$

The results for the average inventory level and associated cost calculations are presented in Table 17. We can see that under continuous review policy, the finished goods inventory of 212 stators holds an average inventory level of 664 parts in the long run with annual ordering and inventory holding cost of $17731.

Table 17. Calculation results for the average inventory level and associated costs of the finished goods inventory of 212 stators under continuous review policy

Quantity	Values	Comments
I_{low} (parts)	499	≈18 trays
I_{high} (parts)	829	≈30 trays
$\bar{I}$ (parts)	664	≈24 trays
H ($/year)	14210	-
K ($/year)	3521	-
C ($/year)	17731	-

6.1.2 Periodic Review Policy

Different from the real-time inventory monitoring mechanism of a continuous review policy, in a periodic review policy, inventory positions are reviewed at regular intervals and an order is placed after each review to replenish the inventory position back up to a target base stock level. Therefore, periodic review policy is also referred to as base stock policy.

A system under periodic review policy is characterized by the base stock level B and the review period r. The decision on these quantities in the context of the valve manufacturing cell is given in the following sections.

6.1.2.1 Modeling for the Base Stock Level

Similar to the reorder point in a continuous review policy, the base stock level in a periodic review policy should protect the inventory from backlogs until the arrival of the next order. However, unlike the continuous review policy, due to the review interval r employed, orders in a periodically reviewed system would only arrive after a period of $(r+\mu_l)$ days. Therefore, the base stock level should be established to cover the average demand over these $(r+\mu_l)$ days with an additional safety stock incorporated to account for demand uncertainties.

The average demand oven $(r+\mu_l)$ days, d'_{avg}, is given as,

$$d'_{avg} = \mu_d \times (r + \mu_l) \tag{6.20}$$

Modifying Equation (6.6) in the continuous review policy, the safety stock level SS in a periodic review policy can be calculated from,

$$SS = F^{-1}(\alpha)\sqrt{(\mu_l + r) \times \sigma_d^2 + \mu_d^2 \times \sigma_{l+r}{}^2} \tag{6.21}$$

where $\sigma_{l+r}{}^2$ is the variance of the quantity $(r+l)$, i.e.

$$\sigma_{l+r}{}^2 = Var(r + l) \tag{6.22}$$

with l representing lead time as a variable.

According to statistical laws, since

$$Var(r + l) = Var(r) + Var(l) \tag{6.23}$$

and $Var(r) = 0$ as the review period r is a constant, we have

$$\sigma_{l+r}{}^2 = Var(l) \quad \text{or} \quad \sigma_{l+r}{}^2 = \sigma_l^2 \tag{6.24}$$

Therefore, the safety stock level can be expressed as,

$$SS = F^{-1}(\alpha)\sqrt{(\mu_l + r) \times \sigma_d^2 + \mu_d^2 \times \sigma_l^2} \tag{6.25}$$

And the base stock level is thus derived by combining Equations (6.20) and (6.25), i.e.

$$B = \mu_d \times (r + \mu_l) + F^{-1}(\alpha)\sqrt{(\mu_l + r) \times \sigma_d^2 + \mu_d^2 \times \sigma_l^2} \tag{6.26}$$

6.1.2.2 Simulations for the Periodic Review Policy

In order to determine the base stock level that supports the performance of our proposed system, the review period r needs to be identified first. According to Simchi-Levi et al [5], r is usually related to the fixed ordering cost incurred by the system. A long review period would result in lower ordering cost but at the same time increase the inventory holding expenses. Therefore, an optimal review period would be the one that minimizes the overall cost of ordering and inventory holding.

The total annual ordering cost in this case can be calculated as,

$$K = k \times \frac{365}{r} \tag{6.27}$$

To obtain the annual inventory holding cost, we need to first figure out the average inventory level held in the system.

Recall our discussion in Section 6.1.1.3.2, when the system reaches steady state, the average inventory level $\bar{I}$ would be the average of the highest and the lowest inventory levels observed in the system. Same as the continuous review policy, the lowest inventory level I_{low} in this case is expected to be the safety stock level *SS*, as previously shown by Equation (6.14).

Similarly, the highest inventory level I_{high} would be reached right after the arrival of an order. In the case of a periodic review policy in which order are placed to replenish the inventory back up to a constant base stock level at intervals of *r*, under steady state condition, the expected order quantity would equal to the average demand over *r*, i.e. $r \times \mu_d$. Therefore, the highest inventory level in the system can be expressed as,

$$I_{high} = SS + r \times \mu_d \tag{6.28}$$

Thus, the average inventory level held in the system can be calculated following Equation (6.16), i.e.

$$\bar{I} = SS + \frac{r \times \mu_d}{2} \tag{6.29}$$

Accordingly, the annual inventory holding cost is thus given by,

$$\mathrm{H} = \mathrm{h} \times \left(SS + \frac{r \times \mu_d}{2}\right) \tag{6.30}$$

The total cost C of ordering and inventory holding can be obtained from Equation (6.19). Substitute Equations (6.25), (6.27) and (6.30) into Equation (6.19), C can be re-written as,

$$\mathrm{C} = k \times \frac{365}{r} + \mathrm{h} \times \left(F^{-1}(\alpha)\sqrt{(\mu_l + r) \times \sigma_d^2 + \mu_d^2 \times \sigma_l^2} + \frac{r \times \mu_d}{2}\right) \tag{6.31}$$

For a projected service level of 98%, substitute the values of k, h, μ_d, σ_d, μ_l and σ_l available in Tables 15 and 16 into Equation (6.31), and plot C as a function of r over the range of 1-20 days, the following graph in Fig. 13 is obtained.

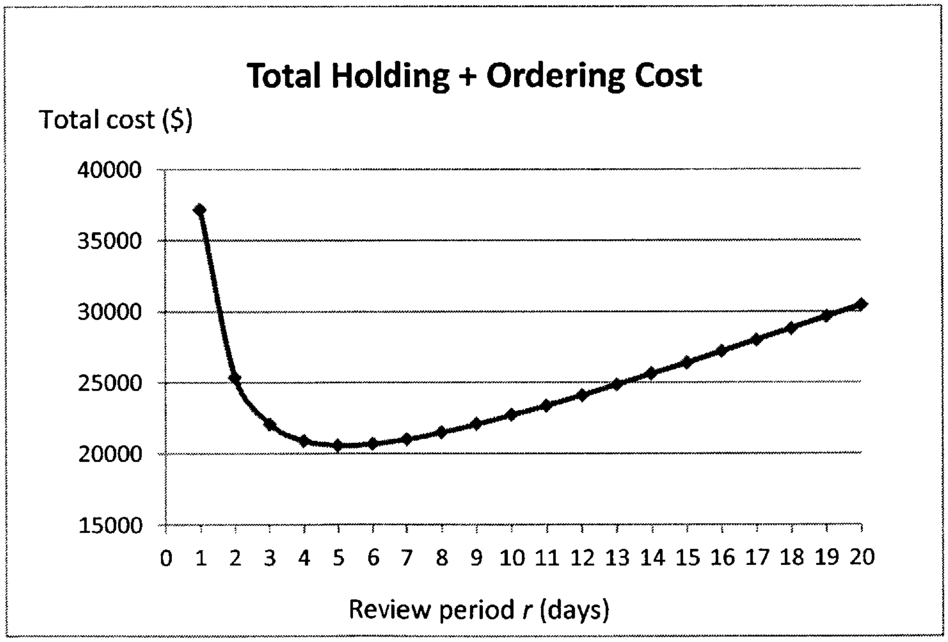


Fig. 13. The total inventory holding and ordering cost as a function of review period *r*

From the graph, we can identify that the total ordering and inventory holding cost is the lowest with a review period of $r = 5$ days. The corresponding SS, B, $\bar{I}$ and C values are calculated and summarized in Table 18.

Table 18. Calculation results for the safety stock, base stock and average inventory levels as well as associated costs of the finished goods inventory of 212 stators under periodic review policy

Quantity	Values	Comments
SS (parts)	564	≈20 trays
B (parts)	2038	≈73 trays
I_{low} (parts)	564	≈20 trays
I_{high} (parts)	884	≈32 trays
$\bar{I}$ (parts)	724	≈26 trays
H (\$/year)	15494	-
K (\$/year)	5110	-
C (\$/year)	20604	-

6.1.3 Continuous Review Policy vs. Periodic Review Policy

Through the modeling and calculations performed in Sections 6.1.1 and 6.1.2, we have now identified all the important parameters for both the continuous review policy and the periodic review policy in the context of the finished goods inventory of 212 stators. These quantities are summarized in Table 19.

Table 19. Summary of important parameters for the continuous review and periodic review policies for the finished goods of inventory of 212 stators

Parameters / Policy	Continuous Review	Periodic Review
α	98%	98%
R (parts)	1650	-
Q (parts)	330	-
B (parts)	-	2038
r (days)	-	5
SS (parts)	499	564
$\bar{I}$ (parts)	664	724
C (\$/year)	17731	20604

From the Table 19, we can clearly observe that to achieve the same projected service level, the continuous review policy requires less safety stock and consequently leads to lower average inventory volume held in the system as well as less ordering and inventory holding costs. This is as expected since the overall variability in a continuously reviewed system is lower due to real-time monitoring of inventory positions. On the other hand, under the periodic review policy inventory managers would have no knowledge on the actual inventory levels in stock between two review points, and hence a relatively larger safety stock needs to be established. Furthermore, the real-time "alarm" triggered replenishment mechanism of the continuous review policy is especially preferable in dealing with highly volatile demands in which high responsiveness is required to maintain a certain service level.

In addition, from a practical point of view, since the coating of stators is performed at outside vendors, a fixed order quantity suggested by the continuous review policy is more feasible for implementation in our case.

Therefore, we conclude that in the context of the valve manufacturing cell the continuous review policy is advantageous over the periodic review policy in terms of cost reduction, responsiveness and coordination with the outside vendors.

6.2 Implementing a Mixed Inventory Review Policy

Although the advantages of the continuous review policy has been demonstrated through our analysis so far, the actual implementation of such a policy would require the support of a powerful computerized inventory system which is not yet available in the valve manufacturing cell. This implies the "continuous" monitoring of inventory levels is not feasible under current situation. Taking into consideration this constraint, we have thus moved on to the development of a mixed inventory review policy that combines favorable features of both the continuous and the period review models.

6.2.1 The Concept of a Mix Inventory Review Policy – the (s, S) Policy

In the mixed inventory review policy, we have preserved the "triggering" based replenishment mechanism of a continuous review policy, i.e. orders will be placed whenever the inventory position is found below a predetermined level, referred to as s level here. However, unlike the continuous inventory position check in (Q, R) policy, the inventory condition here is monitored "quasi-continuously" at small intervals of one days or one shift (half day). Although some level of responsiveness in the original (*Q, R*) policy would be lost, it is still advantageous as compared to the conventional periodic review policy which usually employs a much longer review period.

Once an order is triggered, a variable quantity of parts will be ordered to bring the inventory position up to a fixed base stock level *S*. The rationale behind using a variable order quantity here rather than a fixed order quantity as depicted by the (*Q, R*) policy is that due to the possible delay of order triggering caused by the short review period involved here, the actual inventory position might have dropped below the reorder point *s* by a significant amount once the order is initiated, and hence a fix order quantity would not be able to bring the inventory position up to the projected (R+Q) level in the original continuous review model. Therefore, to avoid this depletion and obtain a steady system, it is more suitable to create a target base stock level which essentially equals to (R+Q).

To sum up, the new mixed inventory review policy works by monitoring the inventory position at short intervals of one day or one working shift, and an order is placed when the inventory position is found below the reorder point *s*. A variable order quantity is then used to bring the inventory position up to a base stock level *S*. This policy is sometimes referred to as the (*s, S*) policy in literatures [5].

6.2.2 Modeling for the Mixed Inventory Review Policy

To characterize a system under the mix inventory review policy, the reorder point *s* and the base stock level *S* should be determined.

As introduced in the previous section, similar to the continuous review policy, orders in the mixed inventory review policy are triggered by the reorder point *s*, since only a very short review period of one or half day is employed here which is negligible as compared to the long lead time, a good approximation for *s* would be to set *s* equal to the reorder point *R* in (*Q, R*) policy, i.e.,

$$\mathrm{s} = \mathrm{R}$$

or (6.32)

$$\mathrm{s} = \mu_d \times \mu_l + F^{-1}(\alpha)\sqrt{\mu_l \times \sigma_d^2 + \mu_d^2 \times \sigma_l^2}$$

Also, as illustrated previously, to account for part depletion caused by delayed order initiation and hence to achieve a steady system, the base stock level *S* should be set to the sum of *Q* and *R* in (*Q, R*) policy, that is,

$$\mathrm{S} = \mathrm{Q} + \mathrm{R}$$

or (6.33)

$$\mathrm{S} = \sqrt{\frac{2k \times D}{h}} + \mu_d \times \mu_l + F^{-1}(\alpha)\sqrt{\mu_l \times \sigma_d^2 + \mu_d^2 \times \sigma_l^2}$$

The models for the safety stock level, average inventory level and associated costs are the same as that of the (*Q, R*) policy, as given by Equations (6.6), (6.16) and (6.19), respectively. The calculation results for these quantities are shown in Table 20.

Table 20. Calculation results of important parameters under the mix inventory review policy for the finished goods inventory of 212 stators

Quantity	Values	Comments
SS (parts)	499	≈18 trays
s (parts)	1650	≈59 trays
S (parts)	1980	≈71 trays
$\bar{I}$ (parts)	664	≈24 trays
C ($/year)	17731	-

6.2.3 Discussion on Service level under the Mix Review Policy

One critical concern on implementing the mix inventory review policy is that the original projected service level by the continuous review policy might be comprised due to the delayed order initiation caused by the review period employed. To understand this impact, let us assume a review period of one day and a worst case scenario in which orders are always initiated one day late (i.e. the inventory position drops below s right after the review of Day1 and thus the order is only triggered after the review of Day 2). Refer to the discussion in Section 6.1.1.3.2, the average inventory level under the worst case scenario, $\bar{I}_{worst}$, can be calculated as follows.

In this case, due to the one-day delay in ordering, the lowest inventory level, I'_{low}, observed in the system would be obtained by subtracting one day's demand from the safety stock, i.e.

$$I'_{low} = \mathrm{SS} - \mu_d \tag{6.34}$$

Also, since the base stock level S needs to be reached by replenishment activities, the effective order quantity, Q', under such delay would be given by,

$$Q' = Q + \mu_d \tag{6.35}$$

And thus the highest inventory level, I'_{high}, observed in the system would be,

$$I'_{high} = \mathrm{SS} + \mathrm{Q} \tag{6.36}$$

Therefore, $\bar{I}_{worst}$ under this scenario can be calculated as,

$$\bar{I}_{worst} = \frac{I'_{low} + I'_{high}}{2}$$

or

$$\bar{I}_{worst} = \frac{Q}{2} + SS - \frac{\mu_d}{2} \tag{6.37}$$

Comparing Equation (6.37) with Equation (6.16), we can observe that the effective safety stock level, SS_{worst}, under the worst case scenario is expressed as,

$$SS_{worst} = SS - \frac{\mu_d}{2} \tag{6.38}$$

Denote the actual service level under the worst case scenario as α', then SS_{worst} and α' follows the relationship shown in Equation (6.39), i.e.

$$SS_{worst} = F^{-1}(\alpha')\sqrt{\mu_l \times \sigma_d^2 + \mu_d^2 \times \sigma_l^2} \tag{6.39}$$

Rearrange Equation (6.39), α'can be calculated as,

$$\alpha' = \mathrm{F}\left(\frac{SS_{worst}}{\sqrt{\mu_l \times \sigma_d^2 + \mu_d^2 \times \sigma_l^2}}\right) \tag{6.40}$$

The calculation results for the effective safety stock level and the compromised service level under worst case scenario are presented in Table 21.

Table 21. The effective safety stock level and the compromised service level under worst case scenario for the mixed inventory review policy

Quantity	Values	Comments
I'_{low} (parts)	435	≈16 trays
Q' (parts)	394	≈14 trays
I'_{high} (parts)	829	≈30 trays
$\bar{I}_{worst}$ (parts)	632	≈23 trays
SS_{worst} (parts)	467	≈17 trays
α'	97%	-

Thus, by implementing the mixed inventory review policy, the service level of the finished goods inventory is compromised from the projected 98% to 97% under the worst case scenario. This is indeed quite an acceptable performance. Moreover, the possibility of such situation happening in reality is very low, which implies that the service level provided by the mixed review policy is essentially the same as that of the continuous review policy under most circumstances. Yet it only requires one inventory position check every day which is fairly easy to realize.

Therefore, to conclude, we recommend the mixed (*s*, *S*) inventory review policy to the valve cell for the management of its finished goods inventory of 212 stators, for the high performance of this policy, and the ease of implementation.

As a comparison to the current situation at the valve cell, the (*s*, *S*) policy would be able to improve the on-time delivery performance from around 60% to at least 97%, while reducing the average inventory level held by the system from 700 parts to 664parts. In fact, this average inventory level could be further reduced if more accurate demand data rather than the rough annual demand profile were used for analysis, as will be demonstrated in Chapter 8.

6.3 Merging the Valve Cell and Assembly Inventories

After establishing the (*s, S*) policy as an effective tool for managing the finished goods inventory of 212 stators, we looked deeper into the entire manufacturing system at Waters and identified the opportunity of merging the finished goods inventory at the valve cell with the initial stator inventory at the assembly department for further responsiveness improvement and inventory reduction.

6.3.1 The Idea and Benefits of Inventory Merging

Currently the assembly department keeps a regular inventory of 300 units of finished 212 stators at the start of the assembly line. This inventory was created to ensure the immediate availability of stators for the subsequent assembly processes, and was partially due to the historical poor on-time delivery performance of the valve cell.

With a standardized inventory review policy in place at the valve cell, the need for keeping such an inventory at the assembly department is minimal, given the fact that these two departments are located very close to each other within the same building. Nevertheless, through our communication with managers and operators at the department we have realized that it is very difficult to eliminate this inventory at assembly due to the fear for operation breakdown. In order to conquer this cognitive issue and at the same time further enhance the coordination between these two departments, we have thus proposed the plan for inventory merging.

The plan works by combining the finished goods inventory at the valve cell with the starting stator inventory at the assembly department and locating the new inventory in the assembly area. The valve cell, however, would take the responsibility for managing this new inventory for both parties. The production at the valve cell would thus be directed by internal replenishment calls rather than external orders. Furthermore, information regarding demand, cost, and profit of the product should be shared completely between these two departments. With the routine real-time monitoring of inventory positions and now better knowledge of demand and activities happening at the assembly department, the valve cell would be able to coordinate production and distribution more effectively and hence optimize the performance of the entire system. Also, the assembly department would have no need to create another inventory for stators, which in turn could result in significant total inventory reduction in the system.

This idea is originally derived from the vendor managed inventory (VMI) model in which suppliers are given the authority to manage the inventory at the retailer side. The model was famously exemplified by the success of P&G and Wal-Mart, whose partnership has drastically improved the on-time delivery performance of P&G and reduced the overall inventory cost of the system [14].

6.3.2 Modeling for Inventory Merging

To characterize the system with a single merged inventory between the valve cell and the assembly department, an optimal production quantity, $Q*$, needs to be identified. Since this is a global optimization problem for overall system performance, a suitable model to use would be the Newsvendor model which aims to maximize the expected profit of a system.

Thus, $Q*$ can be calculated using Equation (6.9) presented in Section 6.1.1.2.2. In this case, the underage cost C_u in Equation (6.9) is the difference between the selling price, p, of a 212 stator and the manufacturing cost, c, incurred by the valve cell, i.e.,

$$C_u = p - c \tag{6.41}$$

And the overage cost C_o is given by the difference between the manufacturing cost and the salvage value, v, of an unsold stator, that is,

$$C_o = c - v \tag{6.42}$$

Substitute Equations (6.41) and (6.42) into Equation (6.9), $Q*$ is then expressed as,

$$F(Q^*) = \frac{p-c}{p-v}$$

or (6.43)

$$Q^* = F^{-1}\left(\frac{p-c}{p-v}\right)$$

The average inventory level held by the entire system (valve cell and assembly), $\bar{I}_{sys-merged}$, can be obtained following the discussion in Section 6.1.1.3.2, i.e.

$$\bar{I}_{sys-merged} = \frac{Q^*}{2} + F^{-1}(\alpha)\sqrt{\mu_l \times \sigma_d^2 + \mu_d^2 \times \sigma_l^2} \tag{6.44}$$

where the values of μ_l and σ_l associated with the specific order quantity Q^* can be calculated using the procedures demonstrated in Section 6.1.1.3.1.

Now let us consider the system with two separate inventories – one at the valve cell and the other at the assembly department. Although in our previous analysis we have calculated the average inventory level held in the valve cell (not including assembly) using the EOQ model order quantity, to serve as a basis for comparison, we will now calculate the average inventory level in the valve cell using the Newsvendor model and then obtain an overall inventory level for the entire system by adding the regular 212 stator inventory held by the assembly department.

To calculate Q^* for the valve cell alone, Equation (6.43) needs to be modified. The product selling price from the perspective of the valve cell is actually the manufacturer wholesale price, w, which is the value of one 212 stator when it reaches the assembly department. Therefore, the optimal production quantity in this case can be calculated as,

$$Q^* = F^{-1}\left(\frac{w-c}{w-v}\right) \tag{6.45}$$

Similarly, the average inventory level, $\bar{I}_{valve}$, held in the valve cell can be calculated using Equation (6.44).

The average inventory level in the entire system under this scenario, $\bar{I}_{sys-separate}$, is thus given by,

$$\bar{I}_{sys-separate} = \bar{I}_{valve} + \bar{I}_{assembly} \tag{6.46}$$

where $\bar{I}_{assembly}$ represents the average inventory level of 212 stators held by the assembly department.

Assume a single planning period of 1 week, using the weekly demand profile of 320 parts/week with a standard deviation of 161 parts/week, along with the cost and price information provided in Table 22, Q^* and $\bar{I}_{sys}$ values under both scenarios are calculated and presented in Table 23.

Table 22. Summary of cost and price data for the Valve cell and Assembly

Cost & Price	Valve cell	Assembly
c (\$)	67	-
w (\$)	107	-
p (\$)	-	642
v (\$)	1.09	1.09

Table 23. Calculation results of the optimal production quantity and average inventory level for both merged and separate inventories

Parameters / Scenario	Merged Inventory	Separate Inventories
Q^* (parts)	523	270
μ_l (weeks)	3.6	3.6
σ_l (weeks)	0.08	0.05
$\bar{I}_{valve}$ (parts)	-	746
$\bar{I}_{assembly}$ (parts)	-	300
$\bar{I}_{sys}$ (parts)	874	1046

Therefore, a total inventory reduction of 172 parts can be achieved by merging the valve cell and assembly inventories. This signifies 16% further reduction in total inventory volume.

6.4 Concluding Remark on the Establishment of Inventory Review Policies

In search for a suitable tool for managing the finished goods inventory at the valve manufacturing cell, we have examined the feasibility, performance and impact of the continuous review policy, the periodic review policy and a mix inventory review policy via mathematical modeling and simulations throughout this chapter. A summary for all the important parameters involved in these policies are given in Table 24. Among these three policies, the (s, S) policy is recommended to the valve cell due to the low average inventory level associated, high performance and ease of implementation.

Table. 24. Summary of important parameters for the continuous, periodic and mixed inventory review policy

Parameters / Policy	Continuous Review	Periodic Review	Mixed Review
Model	(Q, R)	Base stock	(s, S)
α	98%	98%	97%
R (parts)	1650	-	1650
Q (parts)	330	-	-
B (parts)	-	2038	1980
r (days)	-	5	1
SS (parts)	499	564	499
$\bar{I}$ (parts)	664	724	664
C (\$/year)	17731	20604	17731
Difficulty for Implementation	High	Low	Low

Comparing to the currently observed highly fluctuating inventory level of 84 – 1603 parts, the (s, S) policy effectively helps in creating a stable system with an inventory level of approximately 500 – 800 parts, thus largely eliminates the possibility of backlogs and inventory explosion. While improving the on-time delivery performance from the previous 60% to at least 97%, the policy also helps to reduce the average inventory level in the system from 700 parts to 664 parts based on rough demand estimates on a yearly basis. This level could be further reduced if a more reasonable shorter planning period (e.g. monthly, bi-weekly) were considered.

Furthermore, the opportunity for merging the valve cell and assembly inventories is also explored. Mathematical modeling demonstrates a 16% inventory reduction brought about by the merging of inventories. Although the plan is not yet mature for implementation due to coordination issues, it could definitely serve as a good direction for future work regarding cost reduction and performance improvement.

Chapter 7

Implementation of a Kanban System

In Chapter 4 – Chapter 6, we have established a pull production system that incorporates standardized finished goods inventory management with a highly responsive fabrication line. In order to implement such a system in reality, a supporting Kanban system was developed. This chapter provides a general overview of our Kanban methodology and a detailed description of the Kanban workflow in the context of the valve manufacturing cell.

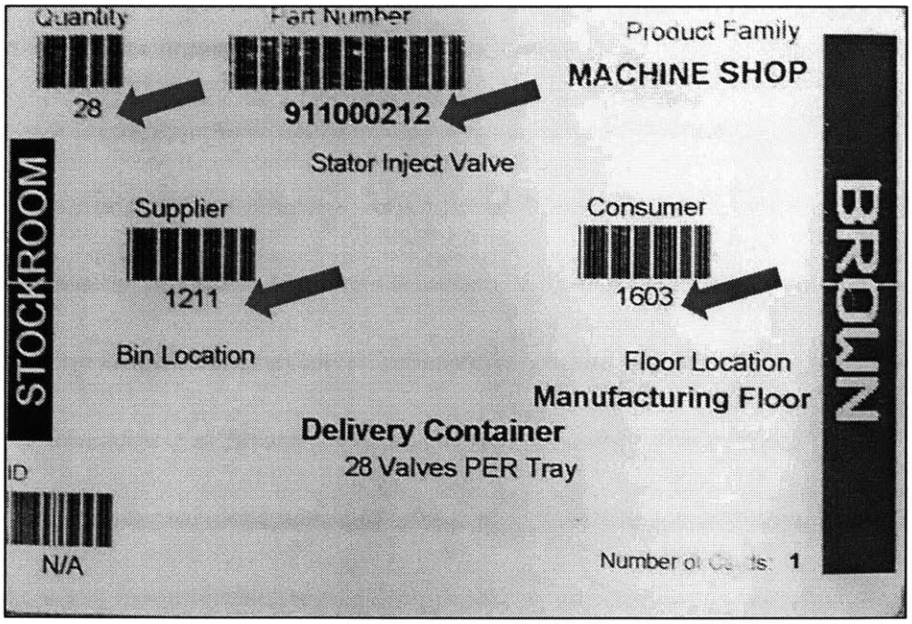
7.1 Implementing a Kanban System for the High Volume Production Line

As previously illustrated in Fig.10 in Chapter 4, our proposed system design consists of two dedicated production lines – one high volume line for the production of 212 stators and one high mix line for the rest 27 part types. The Kanban system employed in this project is to support the pull production mechanism of the high volume 212 stator production line. The rationale behind selecting a Kanban methodology for this line is that the manufacturing environment for 212 stators is highly repetitive and stable, in which parts are processed in a fixed sequence at relatively steady rates. The Kanban system is most suited for coordination and scheduling in such kind of environment. Moreover, the Kanban card count required at each station would effectively assist our proposed model in controlling the inventory levels at respective locations and hence support the realization of the entire line design.

On the other hand, for the high mix production line a CONWIP model is more suitable for implementation as it is more robust to high product mix due to the utilization of line-specific cards instead of station-specific cards in a Kanban system.

7.2 Designing the Kanban Cards

Kanban cards are a key component in the Kanban system which signals depletion of parts at a particular inventory or buffer and triggers replenishment activity when received. Several important elements need to be specified on a Kanban card so as to ensure clear communication between stations and hence effective coordination. These include the part number, Kanban size, and supplier and consumer work stations of the parts. Fig. 14 demonstrates a sample Kanban card design for the high volume 212 stator production line.



Quantity
28

Part Number
911000212
Stator Inject Valve

Product Family
MACHINE SHOP

STOCKROOM

Supplier
1211
Bin Location

Consumer
1603
Floor Location
Manufacturing Floor

Delivery Container
28 Valves PER Tray

ID
N/A

Number of Cards: 1

BROWN

Fig. 14. Kanban card for the milling WIP buffer in between the milling (1211) and robo-drilling (1603) workstations. Arrows indicate important elements of the Kanban card.

The card indicates part number “911000212” at the top and specifies the supplier work station as the milling station (1211) and the consumer work station as the robo-drilling station (1603). The Kanban size is shown at the top left corner. Referring to Snegdha Gupta’s work [2], a Kanban size of 28 parts was selected for the best combination of production lead time and machine utilizations.

According to our system design, 5 different types of Kanban cards were made – 3 for the WIP buffers in between turning and milling, milling and robo-drilling as well as robo-drilling and EDM, 1 for the uncoated parts inventory and 1 for the finished goods inventory. The detailed process flow using these Kanban cards is illustrated in the following section.

7.3 The Kanban Workflow

7.3.1 An overview of the Kanban Workflow

The Kanban signaling mechanism in the high volume 212 stator production line is shown in Fig. 15. The entire system consists of 2 loops. Loop 1 includes processes from EDM up to the finished goods inventory, and loop 2 includes all the machining steps at the beginning of the line.

The production in the system is essentially initiated from the finished goods inventory. Recall the (*s*, *S*) inventory review policy established in Chapter 6, when the inventory position in the finished goods inventory is found below the reorder level *s*, an order of 330 parts will be placed. Although the ideal order quantity according to modeling may vary a bit from 330, for

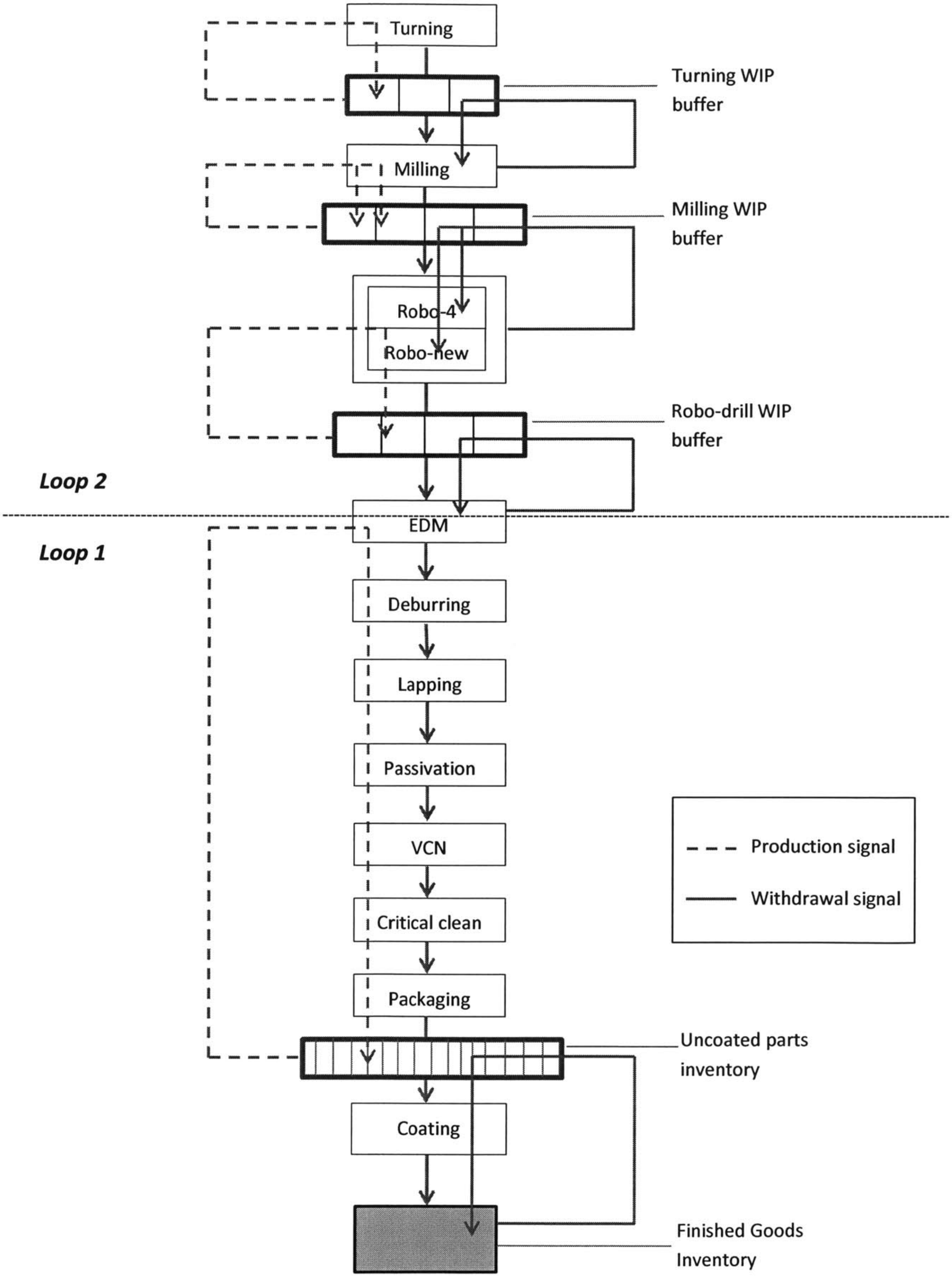


Fig. 15. Illustration of the Kanban system for the high volume 212 stator production line

ease of practical implementation, we will use 330 as the standardized order quantity. This order will be transferred as a part withdrawal signal to the uncoated parts inventory by a Kanban card of size 330, and hence uncoated parts will be withdrawn and sent for coating, which later replenish the finished goods inventory. Since a base stock review policy is used to monitor the stock level at the uncoated part inventory, the part depletion caused by withdrawal will then signal replenishment activity. This replenishment is realized by sending Kanban cards to the upstream EDM station and hence signaling production from EDM onwards. As a depletion of 330 parts would have been observed in the uncoated parts inventory, this production signal to EDM would thus require 330 parts. However, since we have decided upon a batch size (Kanban size) of 28 parts, the production of 330 parts can be translated into 12 batches which contain 336 parts. Therefore, in total 12 Kanban cards will be sent from the uncoated parts inventory to the EDM station as production signal. Using the concept of lot splitting, when EDM finishes 1 batch, 1 Kanban card will be attached to the batch and passed on to the next station, and eventually returned to the uncoated parts inventory. The above comprises the general Kanban flow in loop 1.

Once EDM withdraws parts from the robo-drill WIP buffer, the operations in loop 2 start. Since all the WIP buffers are monitored under base stock review policy, a depletion of parts due to part withdrawal by the downstream work station will immediately signal production in the upstream station. For instance, when EDM takes 1 tray (1 batch) of parts from the robo-drill WIP buffer, 1 Kanban card will be placed to the robo-drilling station upstream to trigger production. As there are 2 robo-drills available in the line, the robo-drill operators will then withdraw 2 trays from the milling WIP buffer and start processing. Once 14 parts are produced from each of the robo-drills, they will be combined into 1 tray to replenish the robo-drill WIP buffer. Also, the depletion of 2 trays of parts in the milling WIP buffer will then signal production in the milling station by sending 2 Kanban cards. Since each time 2 trays of parts are withdrawn from the milling WIP buffer, the production triggered at milling station will follow an effective lot size of 56 parts which helps to keep the utilization of the milling machine low. Similarly, production at the turning station will be triggered when the milling operator withdraws parts form the turning WIP buffer. As can be seen from Fig. 15, the activities in loop 2 are mediated by series of withdrawal and production signals.

7.3.2 Coordination between the Finished Goods and Uncoated Parts Inventories

As introduced previously, the finished goods inventory has a fixed order quantity of 330 parts from the uncoated parts inventory and when this withdrawal happens, the depletion of 330 parts in the uncoated parts inventory will be replenished by sending 12 Kanban cards to the EDM station to signal production. Under this scenario, 336 parts will be received by the uncoated parts inventory eventually, which is 6 parts more than the required quantity. In this manner, we would expect extra part accumulation in the uncoated parts inventory. To avoid excessive part accumulation, we should thus adjust the order quantity (number of Kanban cards) placed by the uncoated parts inventory to EDM in accordance to the actual inventory level. Table 25 demonstrates such kind of adjustment through a simple simulation of a series of orders.

Table 25. Simulation for part accumulation and order quantity adjustment at the uncoated parts inventory

Round 1					
Order	**1**	**2**	**3**	**4**	**-**
Quantity received from EDM	336	336	336	336	-
Quantity sent for coating	330	330	330	330	-
Total extra quantity left	6	12	18	**24**	-
Round 2					
Order	**1**	**2**	**3**	**4**	**5**
Quantity received from EDM	**308**	336	336	336	336
Quantity sent for coating	330	330	330	330	330
Total extra quantity left	2	8	14	20	**26**
Round 3					
Order	**1**	**2**	**3**	**4**	**-**
Quantity received from EDM	**308**	336	336	336	-
Quantity sent for coating	330	330	330	330	-
Total extra quantity left	4	10	16	**22**	-
Round 4					
Order	**1**	**2**	**3**	**4**	**5**
Quantity received from EDM	**308**	336	336	336	336
Quantity sent for coating	330	330	330	330	330
Total extra quantity left	0	6	12	18	**24**

As the table illustrates, in "Round 1", 6 extra parts are left in the uncoated parts inventory after each order and these accumulates into 24 parts after the 4th order. When the next order is to be placed, instead of 12 Kanban cards, only 11 cards are passed to the EDM station which corresponds to 308 parts. These, together with the additional 24 extra parts left in the inventory, will give us 332 parts in total, which is above the required replenishment quantity. In a similar manner, parts will then accumulate and after another 4 orders, the extra parts left in the inventory will again enable us to send 11 Kanban cards to the EDM station. In fact, 11 Kanban cards can be sent whenever the accumulation of parts reaches 22 parts, since a total of 330 parts are required.

With the established Kanban system, we have then moved on to the implementation phase. The implementation results will be discussed in the following chapter.

Chapter 8

Results and Discussion

In Chapter 4 to Chapter 7, we have designed and modeled a pull-type production system with a highly responsive fabrication line, a standardized inventory review policy and a supporting Kanban system. This chapter presents real-world implementation results which have proved lead time reduction in the fabrication line, as well as simulation results that verified the long-time capability of the entire system design.

8.1 Implementation Results – Lead Time Reduction for In-House Production

8.1.1 System Setup Based on Updated Demand

Noticing that the respective WIP buffer and inventory level setups given in Chapters 5 and 6 previously were calculated based the general annual demand profile, in order to implement the new system design for the current month – August 2013, the most recently updated demand profile of 212 stators was extracted and new control levels of buffers and inventories were computed accordingly. The weekly demand forecast for August 2013 is provided in Table 26. The data shows an average weekly demand of 460 parts/week with a standard deviation of 55 parts/week. This corresponds to an average daily demand of 92 parts/day, which is much higher than the average value derived from general annual demand. The forecast seems reasonable as August is normally the peak demand period according to the historical data.

Table 26. Weekly demand forecast of part 212 for August, 2013

Week	Demand (parts/week)
33	400
34	400
35	500
36	500
37	500
Weekly Average	**460**
Weekly Std. dev.	**55**

Referring to Snegdha Gupta's work [2], the new base stock and expected average inventory levels for WIP buffers and the uncoated parts inventory are summarized in Table 27.

Table 27. Base stock level setups for WIP buffers and the uncoated parts inventory based on the demand forecast for August 2013

Buffer & Inventory	Base stock level (Trays)	Expected average inventory level (Trays)
Turning WIP	3	2
Milling WIP	4	2
Robo-drill WIP	4	2
Uncoated parts inventory	14	12

For the finished goods inventory management, employing the mixed (*s, S*) policy, the safety stock level, reorder point, base stock level, as well as the average inventory level can be calculated using Equations (6.6), (6.32), (6.33) and (6.16), respectively. The calculation results based on the current demand profile are presented in Table 28.

Table 28. Finished goods inventory setups based on the demand forecast for August 2013

Quantity	No. of parts	No. of trays
SS	176	6
s	1835	66
S	2165	78
$\bar{I}$	314	11

It is important to point out that the expected average inventory level here is much smaller as compared to the value obtained in Section 6.2.2. This is due to the higher accuracy of the demand forecast made for more recent events, which has a much smaller standard deviation and hence results in a lower required safety stock level. Comparing $\bar{I} = 341$ parts to the average inventory level of 700 parts in the current system, an inventory reduction of 51% could be achieved through the implementation of the (s, S) policy.

Another important point to determine is the short review period involved in the (s, S) policy. Previously in Section 6.2.3 we have examined the compromise in service level for a review period of 1 day and proved that it is negligible. However, due to the exceptionally high demand in August, this needs to be revised. Substitute the new daily demand of $\mu_d = 92$ parts/day into Equation (6.40) for the calculation of effective service level under worst case scenario, we get $\alpha' = 93\%$, which is not a very satisfactory level. Thus, we would like to recommend a half day (1 working shift) review period under high demand situation. In that case, SS_{worst} in Equation (6.40) would be $\mu_d/4$ (instead of $\mu_d/2$ in the previous case) below SS and the compromised service level obtained from the equation is improved to 96%.

Therefore, we would expect a service level of at least 96% under this policy, meanwhile with a 51% reduction in average inventory volume as compared to the current system.

With the above new system setups constructed, we then moved on to the implementation stage. The results are presented and discussed in the following section.

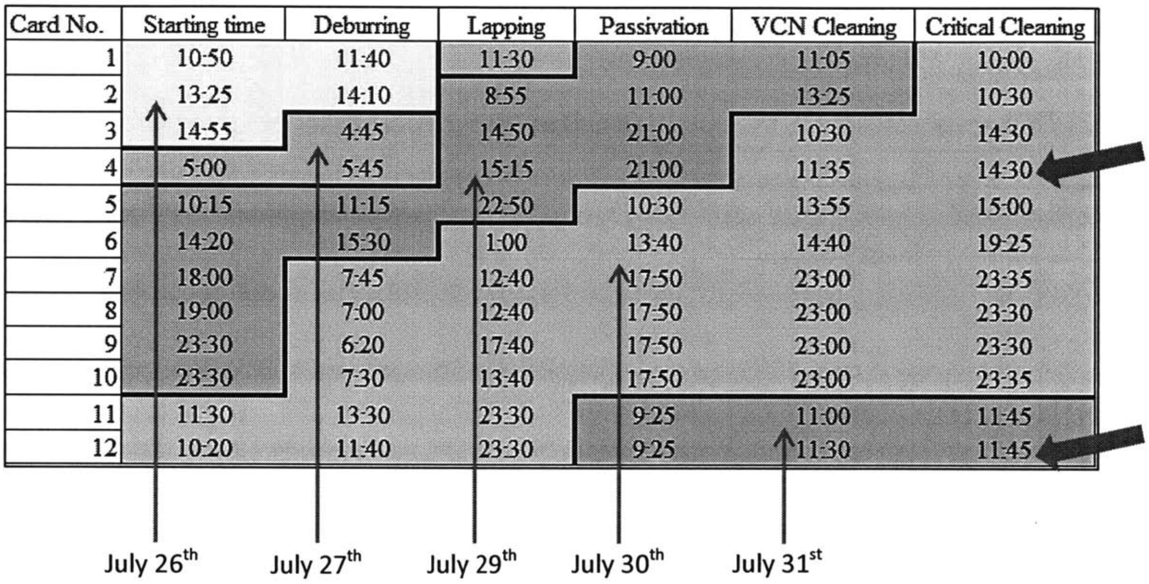
8.1.2 Implementation Results from the In-House Production Line

The implementation experiment started from Friday, July 26th, 2013. The stock position (actual inventory level + scheduled receipt of parts) then at the finished goods inventory of 212 stators is 758, which is below the designed reorder point of the system. Thus, an order of 330 parts was placed to the uncoated parts inventory and the in-house production was triggered following the mechanism described in Chapter 7. Although more parts need to be ordered to build up reasonable amount of stock in the finished goods inventory, for experimentation purpose, we have tracked and studied the response of the system during the production of this one order.

8.1.2.1 Production Lead Time

Lead time for producing 4 trays and 12 trays of 212 stators were obtained from the tracking record of process completion times for all 12 trays in the order. Table 29 shows the process completion time for each step from EDM onwards ("push segment" of the line). The time recorded is the actual clock time at which a particular process is finished. Different blocks represent different dates as indicated at the bottom of the table. From the table, we can see that the production of the 4th tray is completed at 2:30 pm on July 30th and the 12th tray came out at 11:45 am on July 31st.

Table 29. Record of process finishing times from EDM onwards for all 12 trays produced in the new system



Card No.	Starting time	Deburring	Lapping	Passivation	VCN Cleaning	Critical Cleaning
1	10:50	11:40	11:30	9:00	11:05	10:00
2	13:25	14:10	8:55	11:00	13:25	10:30
3	14:55	4:45	14:50	21:00	10:30	14:30
4	5:00	5:45	15:15	21:00	11:35	14:30
5	10:15	11:15	22:50	10:30	13:55	15:00
6	14:20	15:30	1:00	13:40	14:40	19:25
7	18:00	7:45	12:40	17:50	23:00	23:35
8	19:00	7:00	12:40	17:50	23:00	23:30
9	23:30	6:20	17:40	17:50	23:00	23:30
10	23:30	7:30	13:40	17:50	23:00	23:35
11	11:30	13:30	23:30	9:25	11:00	11:45
12	10:20	11:40	23:30	9:25	11:30	11:45

July 26th July 27th July 29th July 30th July 31st

Considering the experimentation starts from 7:30 am on July 26th with 1 working shift on that day and 2 working shifts respectively on 29th, 30th and 31st (27th and 28th are weekends and thus excluded from the analysis), the total in-house production lead time for 4 trays, $L_{4\ trays}$, is given by

$$L_{4\ trays} \approx 2\ days$$

And the production lead time for the order of 12 trays (332 parts & 4 scraps), $L_{12\ trays}$, is approximately

$$L_{12\ trays} \approx 3\ days$$

Comparing to the original lead time of 21 days for 4 trays, this is a drastic improvement. As discussed previously in Chapter 3, the root cause of such long lead time in the original system is inventory queuing and accumulation along the line due to large batch size used and the order-based production mentality. Parts could not move smoothly in the system and each order needs to be initiated from the starting point of the line. On the contrary, in our new push-pull production line, the lead time before EDM was cut down as a result of the established WIP buffers in the starting machining area. Parts were readily available for withdrawal under most circumstances. Large lots were split into smaller batches, which effectively reduced the wait-to-batch time and inventory accumulation.

Furthermore, the observed lead time of 3 days for producing 332 parts is also proved to be efficient for the current high demand period as the quantity covers 3.65 days demand on average.

Another point to note is that the expected production lead time for 336 parts as calculated in Snegdha Gupta's work is 2.12 days [2]. The observed lead time was longer than the expected value mainly due to the following three reasons.

Firstly, the ideal continuous material flow assumed in the calculation was not achievable in reality due to coordination issues and other human factors. Secondly, as the new robo-drilling machine has not yet arrived to the company, we were pulling capacity from another robo-drilling machine which had to perform a small amount of other jobs during the span of the experiment. This to some extent also led to delay of part movement in the system. Another important factor that has caused the longer-than-expected lead time is the "learning curve" effect. Operators generally found it difficult to break away from the order-based production mentality and adapt to the Kanban system. As can be observed from Table 29, relatively less amount of work was accomplished during the first 2 days (26th, 29th) of implementation. The effect is also clearly visible from the plot of the total processing time for each tray shown in Fig. 16. From the figure, we can see that in general the processing time for each tray decreased as the experiment proceeded. There were also large variations at the beginning due to confusions aroused among the night shift operators. Nevertheless, the processing time tended to stabilize around 20hrs/tray after 3 days' operation. Therefore, we would expect a shorter production lead time once all the operators are properly trained for the new system.

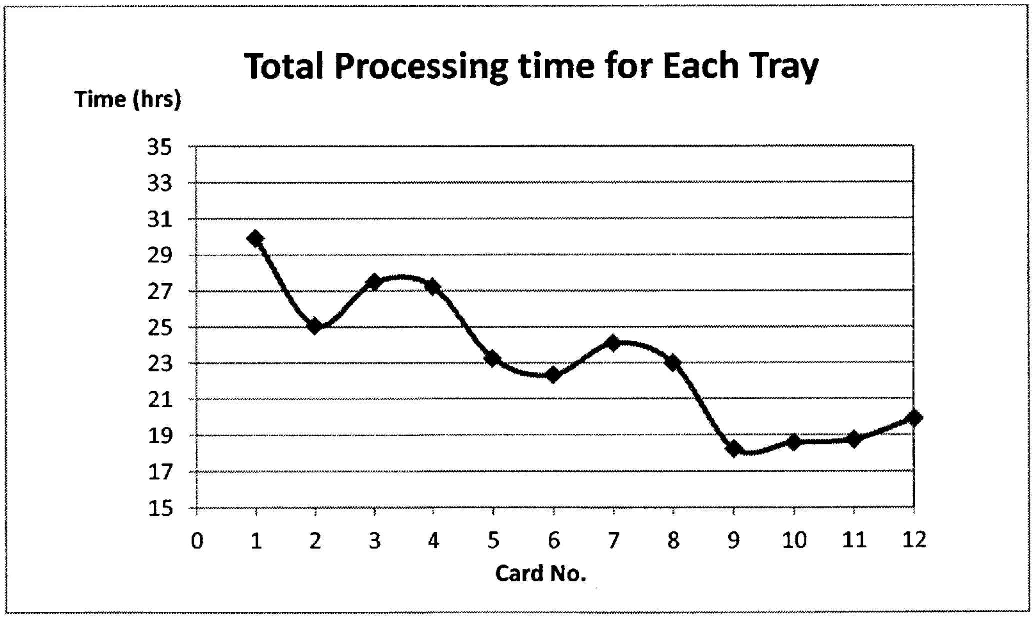


Fig. 16. Total processing time for each tray in the new system

8.1.2.2 Production Cycle Time

The production cycle time here refers to the average time interval between the arrivals of two successive completed trays. Table 30 shows the inter-arrival time record for the 12 trays in the order.

From Table 30, we get that the average cycle time in the system is approximately 1.86hrs/tray. This is very close to the expected rate of 1.96hrs/tray (EDM processing time per tray) obtained from theoretical calculation [2]. The actual cycle time is shorter than the calculated value due to the fact that the processing times used for calculation are the maximum allowable times given to each individual step, which are usually larger than the actual times required in practice.

Table 30. Time intervals between the arrivals of two successive completed trays

Tray	Time between arrivals (hrs)
1-2	0.5
2-3	4
3-4	0
5-6	0.5
6-7	4.42
7-8	4.17
8-9	0.08
9-10	0
10-11	0.08
11-12	6.67
Average	**1.86**

Also, if we look at Table 30 closely, we can observe that the inter-arrival times are highly uneven. The data indicates that tray 3 and tray 4 are finished at the same time, and the same happened to tray 8 – tray 11. The distribution of inter-arrival times is shown in Fig 17. This is most probably caused by the intentional part accumulation in front of the last critical clean station. As the critical clean process is carried out by sending trays through a series of cleaning tanks, putting one tray at a time would be cumbersome from the operator's perspective. Nevertheless, the overall system performance would not be influenced since this process happens at a very fast rate.

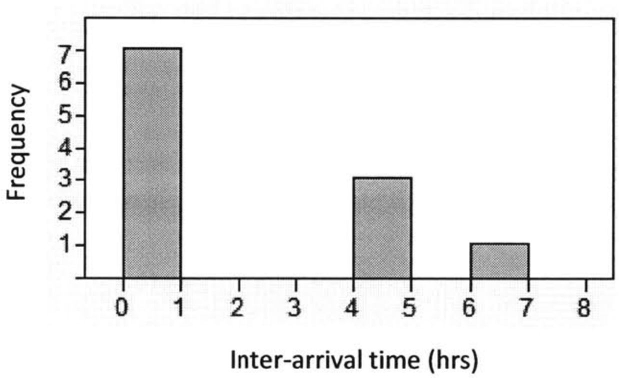
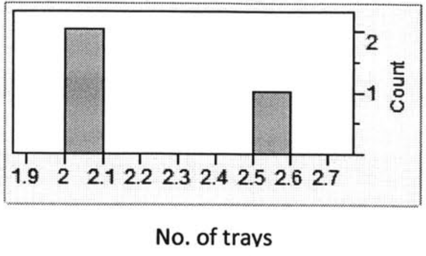


Fig. 17. Distribution of inter-arrival times for the 12 trays produced in the new system

8.1.2.3 Work-in-Process Inventory Volume

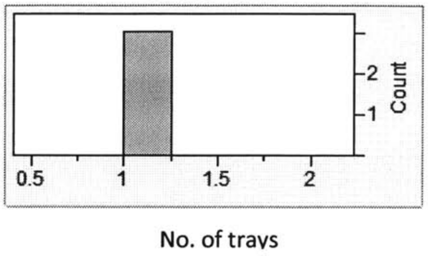
As a comparison to the WIP volume of the original system shown in Fig. 9, the WIP volume resides in the new system were collected and presented in Fig. 18.

Milling



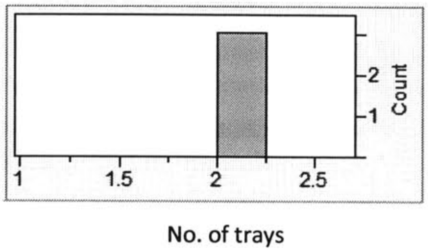
Mean:2.17 Std dev: 0.29

Robo-drill



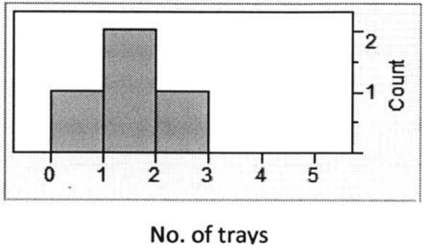
Mean: 1 Std dev: 0

EDM



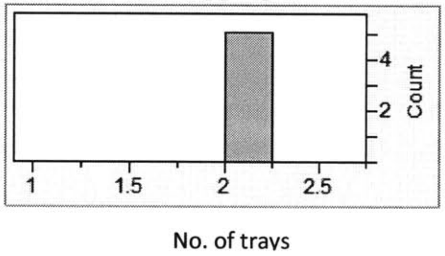
Mean: 2 Std dev: 0

Deburring



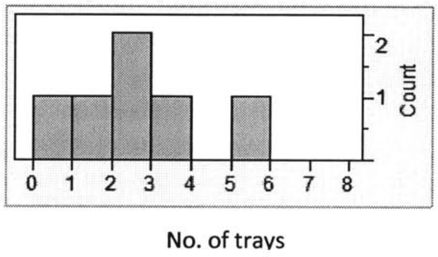
Mean: 1 Std dev: 0.82

Lapping



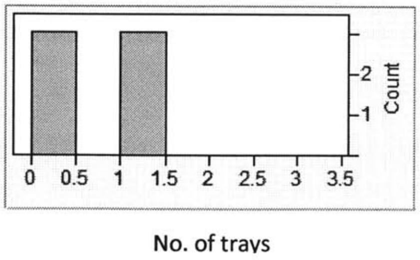
Mean: 2 Std dev: 0

Passivation



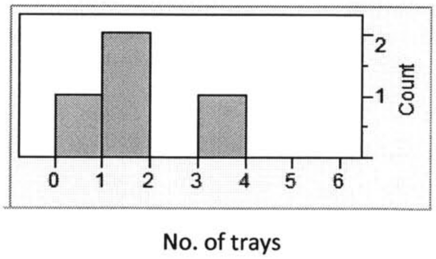
Mean: 2.17 Std dev: 1.72

VCN clean



Mean: 0.5 Std dev: 0.55

Critical clean



Mean: 1.25 Std dev: 1.26

Fig. 18. Inventory volume distribution in front of each work station in the new system

From the figure, we can derive that the total volume of in-process inventories within the system is around 12 trays. Comparing to the volume of 20 trays in the previous system, this implies a 40% reduction in WIP volume. The reduction mainly comes from elimination of part accumulation in front of the robo-drill station. Since robo-drill is the bottleneck process of the entire line, without a WIP control limit, parts will always tend to accumulate in front of this station and thus result in a large volume of unnecessary inventories. With the base stock model established in our system, this WIP explosion problem is prevented as production from milling will be stopped once inventory reaches the base stock level.

8.2 ARENA Simulation Results

Due to the limited time span of the project, the effectiveness of the inventory review polices developed for the finished goods inventory and the uncoated parts inventory could not be tested via real-world implementation. In order to verify these models as well as the long-term performance of the system design, ARENA simulation model was developed and the results are shown below.

8.2.1 Simulation Setup

Fig. 19 demonstrates the pull production system we have constructed using the ARENA simulation software. The major setup considerations are listed as follows:

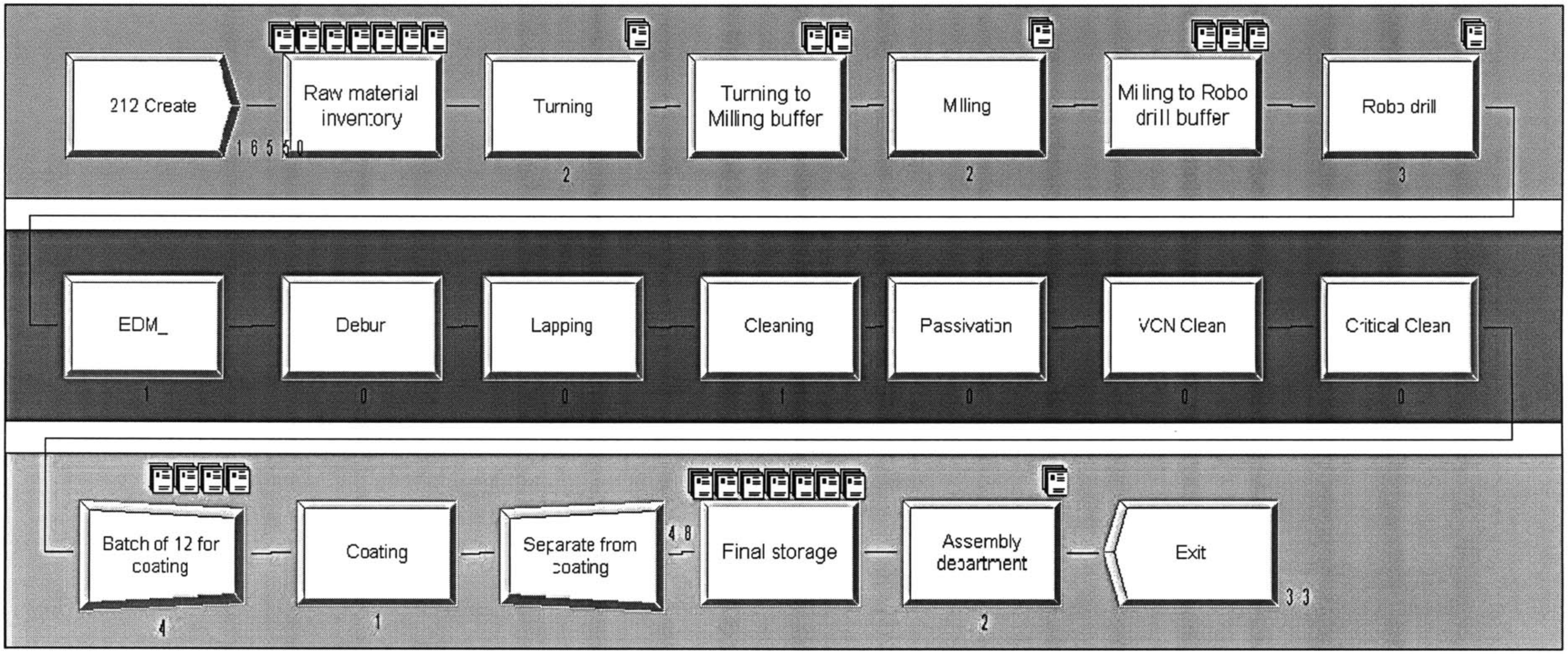


Fig. 19. An ARENA simulation model for our pull production system

1. One tray of parts is considered as one entity in the program.
2. The in-process buffers will only release parts (represent trays here) if the inventory level at the downstream buffer is below the predetermined base stock level and if the downstream machine is available (i.e. Queue = 0 in the downstream machine).
3. Parts will only be sent for coating when the inventory level in the final storage drops below the reorder point.
4. Parts must be cumulated into batches of 12 before being sent for coating.
5. Assembly is considered as a machine in the program. The production rate is equivalent to the demand rate, and the machine utilization will be considered as the service level of the entire system.

8.2.2 Simulation Results

The simulation was run for a period of 30 days in which the system was allowed to reach steady state during the first 20 days and the results were recorded for the last 10 days.

The final simulation output showed that a total of 33 trays were sent to assembly during 10 days' time. Since the demand rate of current month is 92 parts/day, which is approximately 3.3 trays/day, the result verified that the production system was capable of fulfilling the requirement of the assembly station. Note that the pull production system works in a demand-driven manner, hence only the required amount of parts will be sent to the assembly station. Therefore, the number "33" does not represent the total production quantity, rather, it serves as verification that the system was indeed able to send the required quantity to assembly.

The “machine utilization” of the assembly station was shown to be 100%. This indicated a 100% service level, which was as expected as we did not consider demand uncertainties in this model.

Therefore, with the ARENA simulation, the long term capability of our system design was verified.

Chapter 9

Conclusion and Recommendations

In this chapter a summary of the system design and implementation is presented. In addition, recommendations for future operations are provided.

9.1 Conclusion

A pull-type production system for the high volume 212 stators is designed and implemented. The objective of the proposed system was to improve on-time delivery performance of stators to the assembly department. This is achieved through a responsive fabrication line with WIP control and a standardized finished goods inventory management. Responsiveness in the fabrication of stators was achieved through line dedication for high volume 212 stators and high mix part types respectively and the placement of in-process buffers to implement a Kanban based pull-type production process that establishes a WIP cap as well. Through this line implementation, the overall lead time for 212 stators was reduced from 21 days to 3 days and this was supported with a WIP reduction of 40%. The standardization of the finished goods inventory was realized through the establishment of a mixed inventory review policy, in other words, the (s, S) policy with a re-order point that triggers production at appropriate times and a base stock level that eliminates the possibility of inventory explosion. A service level of at least 96% is expected even for a high demand month like August 2013, besides a

50% reduction in average finished goods inventory levels through the setup of an upper inventory control limit.

9.2 Recommendations and Future Work

The following are recommendations to support the proposed system design—

1. Following the successful implementation of the Kanban system for the high volume line, supporting documentation should be developed in compliance with ISO 9000.
2. A similar pull-type production process could be implemented for the 213 stators that are the second most frequently produced parts in the valve cell after the 212 stators.
3. Though the continuation of a push-type production process is recommended for all other stator types in the high mix line, a policy of releasing orders based on an optimized setup sequence is recommended to minimize waiting times in the line. This could also improve the robustness of the line during unexpected high demand periods.
4. Merging the valve cell and assembly inventories is recommended to improve visibility facilitating the elimination of unnecessary inventory that does not contribute to maintaining a high service level. This recommendation is in keeping with the corporate emphasis on developing a lean culture in operations.
5. A recommendation for developing IT capability that can continuously monitor finished goods inventory levels and production line WIP levels is made so that real-time feedback of the system can be obtained.

6. Establishment of a visual aid like a screen that keeps the operators on the floor abreast with the daily demand data is highly recommended especially for a pull based system like ours because once daily demand is satisfied, operators can work on replenishing the WIP buffers up to the respective base stock levels.

Appendix

List of Symbols

Symbol	Quantity	Unit
B	Base stock level in periodic review	parts
C	Total inventory holding and ordering cost	\$/year
C_o	Overage cost	\$/part
C_u	Underage cost	\$/part
CT	Cycle time	min
CV	Coefficient of variation	-
c	Manufacturing cost	\$/part
D	Annual demand	parts/year
E	Machine efficiency	-
H	Annual inventory holding cost	\$/year
h	Unit inventory holding cost	\$/part-year
I	Average inventory level	parts
K	Annual ordering cost	\$/year
k	Ordering cost	\$/order
p	Retailer selling price	\$/part
Q	Order quantity	parts
Q^*	Optimal production quantity	parts
R	Reorder point in (Q, R) policy	parts
r	Review period	days
r_a	Arrival rate of parts	parts/min
$r_{bottleneck}$	Bottleneck production rate	parts/min
r_e	Effective production rate	parts/min
S	Base stock level in (s, S) policy	parts
SS	Safety stock	parts
s	Reorder point in (s, S) policy	parts

Symbol	Quantity	Unit
T_{avail}	Available machine time	min
T_o	Raw processing time	min
$T_{schedule}$	Scheduled machine operation time	min
T_r	Required machine time	min
TH	Throughput	part/min
t_p	Processing time per part	min/part
t_s	Setup time	min
U	Utilization	%
U_r	Required utilization	%
v	Salvage value	$/part
$WIP_{critical}$	Critical WIP level	parts
w	Manufacturer wholesale price	$/part
z	Safety factor	-
α	Service level	%
α'	Compromised service level in (s, S) policy	%
μ_d	Daily demand rate	parts/day
μ_l	Mean lead time	days
σ_d	Standard deviation of demand	parts/day
σ_l	Standard deviation of lead time	days

References

1. Zhuang, Y., *Optimizing the Economic Efficiency by Micro-drill Life Improvement during Deep-hole Drilling in the 212-Stator Manufacturing Process*, in *Department of Mechanical Engineering*. 2013, Massachusetts Institute of Technology: Cambridge, Massachusetts, USA.
2. Gupta, S., *On-Time Delivery Performance Improvement in a Valve Manufacturing Cell: Design of a Pull-Type Production System through Efficient Line Design*, in *Department of Mechanical Engineering*. 2013, Massachusetts Institute of Technology: Cambridge, Massachusetts, USA.
3. Hopp, W.J. and M.L. Spearman, *Factory physics : foundations of manufacturing management*. 2nd ed. 2001, Boston: Irwin/McGraw-Hill. xxii, 698 p.
4. Little, J.D.C., *Tautologies, Models and Theories: Can We Find 'Laws' of Manufacturing.* IIE Transactions, 1992. **24**: p. 7.
5. Simchi-Levi, D., P. Kaminsky, and E. Simchi-Levi, *Designing and managing the supply chain : concepts, strategies, and case studies*. 2nd ed. The McGraw-Hill/Irwin series in operations and decision sciences. 2003, Boston: McGraw-Hill/Irwin. xxvii, 354 p.
6. Graves, S., C., *Lecture Material on Inventory Review Policies*, M.I.o.T.S.S.o. Management, Editor. 2013.
7. Gallego, G., *Lecture Material on the EOQ model*, C.U.T.F.F.S.o.E.a.A. Science, Editor. 2004.
8. Moutaz, K., *The single-period (news-vendor) problem: literature review and suggestions for future research.* The International Journal of Management Science, 1999. **27**: p. 17.
9. Hung, J.-S., Y.-P. Fun, and C.-C. Li, *Inventory management in the consignment system.* Production and Inventory Management Journal, 1995. **36**(4): p. 1.
10. Deleersnyder, J.L., T. J. Hodgson, R. E. King, P. J. O'Grady, and A. Savva, *Integrating Kanban Type Pull Systems and MRP Type Push Systems: Insights from a Markovian Model.* IIE Transactions, 1992. **24**(3): p. 14.
11. Suri, R., *Quick response manufacturing : a companywide approach to reducing lead times*. 1998, Portland, Or.: Productivity Press. xxx, 544 p.

12. Wight, O., *Input/Ouput Control: A Real Handle on Lead Time.* Production and Inventory management Journal, 1970. **11**(3): p. 22.

13. Kelton, W.D., R.P. Sadowski, and N.B. Swets, *Simulation with Arena.* 5th ed. 2010, Boston: McGraw-Hill Higher Education. xix, 636 p.

14. Cetinkaya, S. and C.Y. Lee, *Stock replenishment and shipment scheduling for vendor-managed inventory systems.* Management Science, 2000. **46**(2): p. 217-232.