Reduce Cycle Time and Work In Process in a Medical Device Factory: The Problem and a Proposed Solution

by

Yi Qian

Bachelor of Engineering (Electrical & Electronic)
Nanyang Technological University, 2004

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 2007

Signature of Author..
Department of Mechanical Engineering
August 21, 2007

Certified by..
Stephen C. Graves
Abraham Siegel Professor of Management, Sloan School of Management
Thesis Supervisor

Accepted by..
Lallit Anand
Professor, Department of Mechanical Engineering
Chairman, Department Committee on Graduate Students

Reduce Cycle Time and Work In Process in a Medical Device Factory: The Problem and a Proposed Solution

by

Yi Qian

Submitted to the Department of Mechanical Engineering on August 21, 2007 in partial fulfillment of the requirements for the Degree of Master of Engineering in Manufacturing

ABSTRACT

Many manufacturing firms have improved their operations by implementing a work-in-process (WIP) limiting control strategy. This project explores the application of this concept to limit WIP and reduce cycle time for the Becton, Dickinson and Company's manufacturing facility in Tuas, Singapore. BD's Eclipse Safety Needle production line is facing increasing pressure to reduce its high WIP and long cycle times. With the forecast of increasing demand, the current production control practice will sooner or later push the shop floor space to a limit. We divided the overall system into three manageable sub-systems and analyzed different strategies for each. This paper documents the approaches to Needle Assembly machine (AN) and downstream machines, as well as the inventory policy for Needle Shield (NS) and Safety Shield (SS) molded components. At AN and downstream, we can achieve significant reduction in cycle time and work in process by eliminating the unnecessary early start of production and extra delay caused from the current planning method, by reducing transfer batch sizes and by applying mixed scheduling policy at AN. We can obtain further improvement by implementing a CONWIP release rule. For the stocking of NS and SS molded parts, a simple periodic review, base stock inventory policy can effectively reduce the current inventory level by more than 40%, as well as reduce the inventory variation.

Thesis Supervisor: Stephen C. Graves
Title: Abraham Siegel Prof of Management, Sloan School of Management

ACKNOWLEDGEMENT

First and foremost, we wish to express our deepest appreciation to our thesis supervisor, Professor Stephen C. Graves from the Sloan School of Management in MIT, for his consistent guidance and encouragement throughout the project. We will always feel inspired by his enthusiasm, dedication to excellence, and careful attention to details.

We would also like to thank Associate Professor Appa Iyer Sivakumar from the School of Mechanical and Aerospace Engineering in NTU as an advisor of this project. Professor Siva helped us with exploring scheduling policies and simulation tools for this project.

Kind thanks also goes to Becton, Dickinson and Company for sponsoring this work. Specifically we want to thank Mr. Tan Kok Hui, our corporate supervisor, for giving us valuable feedback and suggestions. Many thanks go to all the staffs in the Eclipse safety needle production line in BD Tuas plant, for their strong support in the project.

I would also like to thank the other two team members of this project, Jing Yao and Kai Meng. This project could not be possible without their teamwork.

Finally, I would like to thank my family and friends for their moral support in this thesis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

CHAPTER 1 INTRODUCTION

To produce the right parts, at the right time, at competitive costs is the key success factor in the modern manufacturing world. To achieve the competitive advantage, to improve the overall performance of manufacturing operations and to obtain time and cost saving, production cycle time and work in process (WIP) limiting control strategies have become essential.

The Eclipse Safety Needle production line in BD Medical's Singapore manufacturing facility is facing increasing pressure to reduce its high WIP and long cycle times. With the forecast of increasing demand, the current WIP management and production scheduling practice will sooner or later push the shop floor space to a limit. Long cycle time will also delay the company's response to product obsolescence and quality issues. Evidently, effective strategies for reducing cycle time and limiting WIP will be necessary for BD to strengthen its competitive position.

The challenge of the task lies in the complex setting of process flow and equipment capability. Multiple product types take different routes in a multi-station production line. The only shared resource among all product types is the bottleneck in the company; yet, this piece of equipment operates faster than any other single machine. Depending on the product-mix demand distribution, other machines may experience high utilization as well. At the multi-machine molding station, there are long changeover times to switch from one product type to another. Setting up machines to produce a particular product type at maximum possible rate can help to reduce cycle time and WIP, however it also results in more changeovers. Such tradeoffs between cycle time and changeover cost further complicate the problem.

1.1 BACKGROUND

1.1.1 BD Medical and BD Tuas Plant

BD (Becton, Dickinson and Company) is a global medical technology company that is focused on improving drug therapy, enhancing the diagnosis of infectious diseases and advancing drug discovery. BD manufactures and sells medical supplies, devices, laboratory instruments, antibodies, reagents and diagnostic products. It serves healthcare institutions, life science researchers, clinical laboratories, industry and the general public.

BD Tuas plant manufactures cannula, needle, and syringe products; it supplies these products to BD's distribution centers (DC), which then supply the worldwide market. The plant is organized in value streams. There are 7 values streams (VS) producing 7 different product families in the plant. Each VS operates independently with its own equipment and work force. This project focuses on the VS that produces safety needles. This VS is internally referred to as Eclipse Value Stream.

1.1.2 The Team Project

Sponsored by Singapore-MIT-Alliance program and the company, this internship project serves as the basis for the theses for MIT's M.Eng degree in Manufacturing. A team approach is adopted, in which a group of 3 students identify and analyze the problem jointly; then each of the three students focuses on solving a sub-problem on an individual basis.

The team identified three sub-problems and provided an analysis and possible solutions for each. Details of the project are documented in the three theses. Titles of the three theses are listed below. Proposed strategies for the Needle Assembly machine (AN) and downstream packaging machines are discussed in Thesis 1 and Thesis 2. Scheduling of hub molding machines is extensively studied and documented in Thesis 3. Inventory management and production scheduling for Needle Shield (NS) and Safety Shield (SS) molding machines is presented in Thesis 1.

Thesis 1: Reduce Cycle Time and Work In Process in a Medical Device Factory: The Problem and a Proposed Solution

Thesis 2: Reduce Cycle Time and Work In Process in a Medical Device Factory: Scheduling Policies for Needle Assembly Machine

Thesis 3: Reduce Cycle Time and Work In Process in a Medical Device Factory: Scheduling of Needle Hub Molding Machines

1.2 THESIS OVERVIEW

In Chapter 2 we will provide background information of the company's operations and the cycle time and WIP problem in the Eclipse safety needle production line. In Chapter 3 we analyze the root cause of the problem and present an overall solution. In Chapter 4 we discuss strategies to improve the operation of AN and its downstream machines. Chapter 5 proposes an inventory policy for the stocking of Needle Shield and Safety Shield molded parts. Finally, we provide a conclusion in Chapter 6.

CHAPTER 2 ECLIPSE SAFETY NEEDLE PRODUCTION LINE

In this chapter, we will provide background information of the company's operations and the cycle time and WIP problem in the Eclipse safety needle production line.

2.1 THE PRODUCT

2.1.1 Eclipse Safety Needle

A needle product is a hypodermic needle connected to a syringe for hypodermic injection. It is detachable from the syringes. A conventional needle consists of a plastic needle hub, a metal cannula, and a plastic needle shield for the cannula. The needle hub is used to attach the needle to a syringe. The cannula is fixed on the needle hub by epoxy. The needle shield is a safety cover for the cannula for protection both before and after injection.

Eclipse safety needle (Figure 1) is a new product introduced by BD Medical in 2005. Different from conventional needles, the safety needle has an extra safety shield installed on its needle hub. The safety shield is designed to shield and lock the needle cannula after injection. The major reason to have the safety shield is to protect end users, like nurses, from being injured with the use of the needle (Figure 2).

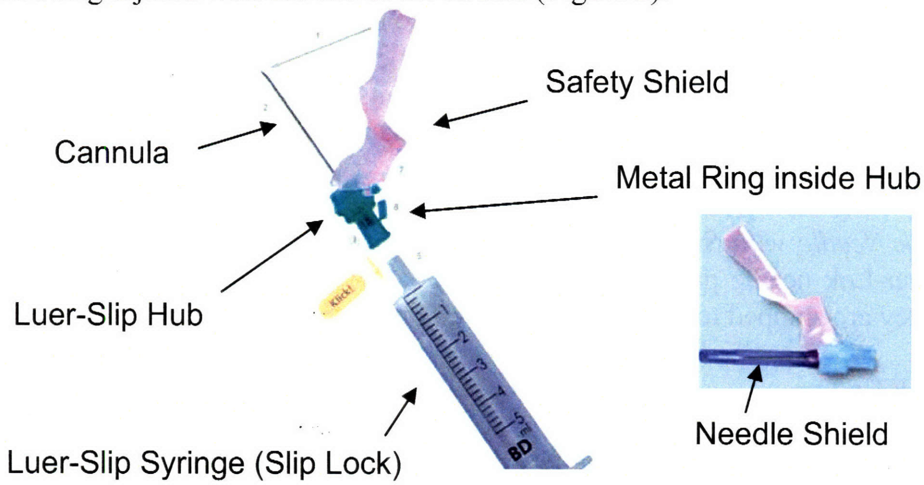

Figure 1 A Luer-Slip Safety Needle with Syringe

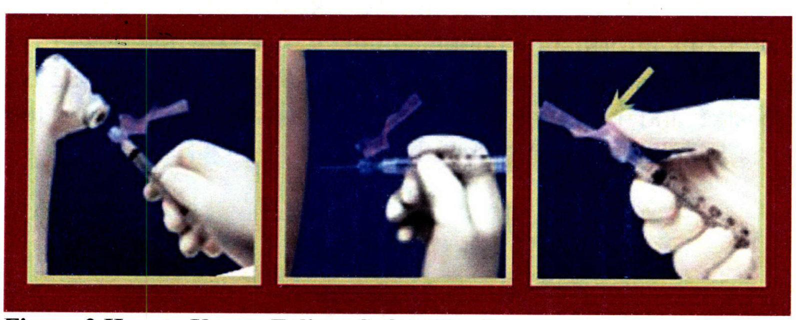

Figure 2 How to Use an Eclipse Safety Needle

2.1.2 Four Major Product Types

In the Eclipse product family there are four major types of safety needle products, which differ in packaging and needle hub design. They are

1) *Luer-Lok Needle in Individual Blister Package (LL-PN):* LL-PN products are packaged individually in blister packages. They are shipped to 3 DCs in North America, from where the products are distributed to retailers nationwide. Luer-Lok refers to a needle hub design. In a Luer-Lok design, the needle hub connects to a syringe by screw threads. The Luer-Lok design is mainly used in the North America market.

2) *Luer-Slip Needle in Individual Blister Package (LS-PN):* The same as LL-PN, LS-PN products are also packaged individually in blister packages. They are shipped to a DC in Europe and then distributed to retailers in the European countries. Similar to Luer-Lok, Luer-Slip refers to another type of needle hub design. In a Luer-Slip design, the needle hub connects to a syringe by snap fit (Figure 1). Different from Luer-Lok needles, a Luer-Slip needle has a metal clip inside its needle hub, which makes a "click" sound when a syringe is correctly attached. The Luer-Slip needles are mainly used in the European market.

3) *Luer-Lok Needle in Bulk Package (Bulk):* Bulk needles are packed in units of a few thousand in large plastic bags inside paper cartons. They are shipped to 3 DCs in North America, and then to other BD plants or pharmaceutical companies for secondary processes. Currently, the Tuas plant only produces Luer-Lok needles in bulk package.

4) *Luer-Lok Needle with Syringe in Combo Package (Combo):* A combo product is a Luer-Lok needle packaged together with a syringe in a single blister package. They are shipped to the 3 DCs in North America.

2.1.3 28 SKUs

Each of the four product families has a few SKUs. The SKUs differ in cannula gauge size and length. A difference in cannula gauge size also results in a difference in needle hub size. Hubs of different gauge sizes are molded in different colors for easy differentiation. Safety shields and needle shields are the same for all the SKUs. Table 1 shows the 28 SKUs. The product name contains information about cannula gauge size and length. Take PSN 22×1 as an example, PSN represents *Needle in Individual Blister Package*, 22 is the cannula/needle hub gauge size and is internally referred to as 22G, with an external diameter of 0.8 mm; and 1 is 1 inch, the cannula length.

Package	Product	Hub Type	Demand
	PSN 18 x 1 1/2	LL Pink	High Runner
	PSN 21 x 1 TW		
	PSN 21 x 1 1/2 TW	LL Green	High Runner
Luer-Lock	PSN 22 x 1		
	PSN 22 x 1 1/2	LL Black	
	PSN 23 x 1	LL Light Blue	High Runner
	PSN 25 x 1		
	PSN 25 x 1 1/2	LL Orange	High Runner
	PSN 20 x 1 (Luer-Slip)	LS Yellow	High Runner
	PSN 21 x 1 (Luer-Slip)		
	PSN 21 x 1 1/2 TW (Luer-Slip)	LS Green	High Runner
Luer Slip	PSN 23 x 1 (Luer-Slip)	LS Light Blue	
	PSN 25 x 5/8 (Luer-Slip)	LS Orange	High Runner
	PSN 27 x 1/2 (Luer-Slip)	LS Grey	
	PSN 30 x 1/2 (Luer-Slip)	LS Yellow	
	Bulk Needle 22 x 1 1/2		
Bulk	Bulk Needle 22 x 1 TW (ABG)	LL Black	
	Bulk Needle 23 x 1 MTW (ABG)	LL Light Blue	
	21 x 1 1/2 TW 3 ml Combo		
	21 x 1 TW 3 ml Combo	LL Green	
	22 x 1 1/2 3 ml Combo		
	22 x 1 3 ml Combo	LL Black	
Combo	23 x 1 3 ml Combo	LL Light Blue	
	25 x 1 3 ml Combo		
	25 x 5/8 1 ml Combo		
	25 x 5/8 3 ml Combo	LL Orange	
	27 x 1/2 1 ml Combo	LL Grey	
	30 x 1/2 1 ml Combo	LL Yellow	

Table 1 Available SKUs of Eclipse Safety Needle Production Line

2.2 THE DEMAND AND CUSTOMERS

2.2.1 Demand Distribution

Because the Eclipse products have been introduced to the North America market for less than 3 years and to the European market for about 1 year, there is very high demand fluctuation. BD expects the demand to increase rapidly, especially in the European market. Demand of different products types and SKUs are not evenly distributed.

In FY 2006, LL-PN contributes over 50% of the total units sold to the DCs, followed by LS-PN and Combo with about 20% each. Bulk constitutes less than 4% of the total units sold. For each major product type, demand usually concentrates in a few SKUs. For instance, SKUs with gauge size 18G, 21G, 23G, and 25G have higher demand than others in LL-PN.

A more detailed illustration of the demand distribution is shown in Table 2. We summarize the distribution based on the total units sold to the DCs in FY 2006. Due to confidentiality concerns, the real demand quantity is not disclosed. Nevertheless, to provide information on the demand fluctuation, we calculate the coefficient of variation ($c.v. = \frac{Mean}{standard\ deviation}$) based on the monthly demand of each SKU. The 4th column summarizes the demand percentage of each SKU in its major product type; and the 5th

column indicates the demand percentage of each product type in all the safety needle products in FY 2006.

Major Types	SKUs (Guage x Length)	c.v.	SKU/Type	Type/Total
LL-PN	PSN 25 x 1	0.76	14.2%	53.2%
	PSN 23 x 1	0.65	17.5%	
	PSN 22 x 1 1/2	1.01	9.4%	
	PSN 21 x 1 TW	0.77	11.7%	
	PSN 21 x 1 1/2 TW	1.21	7.9%	
	PSN 18 x 1 1/2	0.50	31.9%	
	PSN 25 x 1 1/2	0.76	5.8%	
	PSN 22 x 1	2.38	1.5%	
LS-PN	PSN 25 x 5/8 (Luer-Slip)	0.74	32.1%	19.6%
	PSN 27 x 1/2 (Luer-Slip)	0.72	6.6%	
	PSN 30 x 1/2 (Luer-Slip)	0.62	5.2%	
	PSN 23 x 1 (Luer-Slip)	0.79	5.2%	
	PSN 21 x 1 (Luer-Slip)	0.43	12.8%	
	PSN 21 x 1 1/2 TW (Luer-Slip)	0.71	16.1%	
	PSN 20 x 1 TW (Luer-Slip)	0.90	22.1%	
Combo	1 ml 25 x 5/8 Combo	0.49	16.1%	23.6%
	1 ml 27 x 1/2 Combo	0.40	11.8%	
	1 ml 30 x 1/2 Combo	0.58	2.8%	
	3 ml 21 x 1 TW Combo	0.00	1.9%	
	3 ml 25 x 5/8 Combo	0.54	19.9%	
	3 ml 23 x 1 Combo	0.57	13.3%	
	3 ml 22 x 1 1/2 Combo	0.52	26.6%	
	3 ml 21 x 1 1/2 TW Combo	1.00	1.9%	
	3 ml 25 x 1 Combo	0.73	3.8%	
	3 ml 22 x 1 Combo	0.00	1.9%	
Bulk	Bulk Needle 22 x 1 1/2	0.33	28.9%	3.6%
	Bulk Needle 23 x 1 MTW (ABG)	0.09	27.7%	
	Bulk Needle 22 x 1 TW (ABG)	0.35	43.4%	

Table 2 Demand Distribution

2.2.2 Customers and Orders

Although products from the Eclipse line are consumed by both end users and industrial customers, the production line's direct customers are the 3 DCs in North America and the 1 DC in Europe, who then provide supply to regional customers.

In the beginning of a fiscal year, the Tuas plant receives an order forecast from the 4 DCs, which contains information about the monthly order quantity for each SKU in the next 12 months. The DCs can update their forecast subject to a 45-day frozen window rule imposed by the company, in which they could only change the forecast at least 45 days in advance of an order. In other words, the forecast becomes a firm order when its shipment date is less than 45 days away. The plant can start to produce an order once it becomes firm. The forecast is updated mostly via email between the DCs and plant planner.

The time unit for order quantity commitment from the DCs is a month, which means there is not a more specific due date on an order. With the 45-day frozen window, orders for the next month are usually confirmed on the 15th of the current month. For example, if there is an order for a particular SKU in May, it is confirmed on April 15th. Its deadline will be the end of May and it could be shipped any time in May. There is no planned finished good inventory in BD as production is to order and orders are shipped once they are ready to go. This is true for all SKUs.

As a conclusion, the order for next month from each of the four DCs is confirmed on the middle of the current month. The orders for next month must be shipped by the end of the next month at the latest.

2.3 THE ECLIPSE VALUE STREAM

2.3.1 Process Flow of The Eclipse Production Line

Figure 3 shows the process flow of the Eclipse line, with the clean room boundary marked in black lines. The scope of this project is limited to the processes inside the clean room.

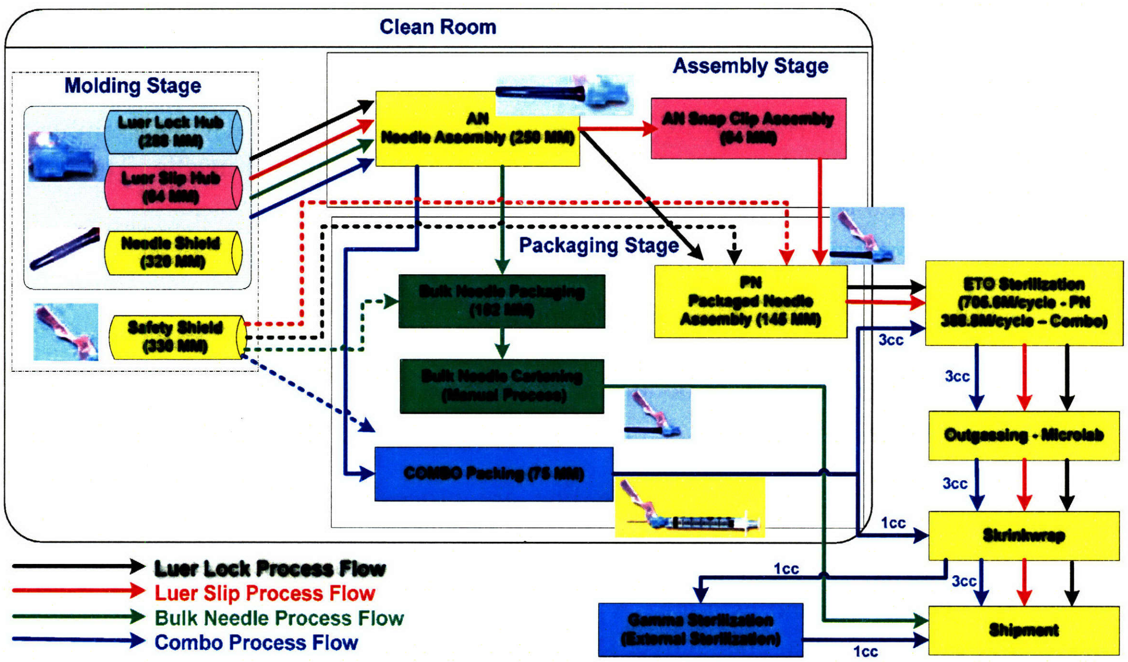

Figure 3 The Eclipse Process Flow (Modified based on BD Internal Source)

2.3.1.1 Processes inside Clean Room

The processes inside the clean room can be divided into three stages: molding, assembly and packaging. There are four groups of 33 injection molding machines in the molding stage, two machines in the needle assembly stage, and three machines in the packaging stage.

2.3.1.1.1 Molding Stage

This process produces plastic parts by injection molding; these parts are used in the assembly stage. There are four types of plastic parts being molded: Needle Shield (1 machine), Safety Shield (2 machines), Luer-Lok Hub (20 machines), and Luer-Slip Hub (10 machines). Luer-Lok and Luer-Slip hub molding machines produce 13 different types of needle hubs. Each machine can produce a few types of hubs by changing its mold inserts. The changeover typically takes 2 hours. There is only one type of needle shield and safety shield, so there is no changeover for the shield molding machines.

The production at the molding stage follows a monthly production plan. For hub molding, production capacity of a specific hub type is constrained by the number of available machines and mold inserts. Moreover, a mold insert can only run on a limited number of machines on which it has been validated. For example, there are 10 Luer-Slip Hub molding machines, but there are only 5 mold inserts for 20G hubs. These inserts can only run on 5 validated machines out of the 10. Thus, we are constrained to have at most five machines producing 20G hubs at any point of time.

2.3.1.1.2 Assembly Stage

Needle Assembly (AN) Machine

Needle assembly is performed by a complex assembly machine, which assembles the cannula, needle hub, and needle shield together into an assembled needle. The detailed steps involved in this stage are:

a) Insert cannula into the needle hub.
b) Apply epoxy between the cannula and hub.
c) Rotate the cannula position in the hub.
d) Heat the epoxy to create bond
e) Put on needle shield.

All products produced in the Eclipse line are processed through this station. As a consequence, BD perceives the AN machine as the bottleneck of the whole process flow. It is highly utilized in order to meet monthly demand. Its utilization can go as high as over 90% when demand is high. A changeover is required between assemblies of different SKUs. A typical changeover takes 30 minutes on average.

Snap Clip Assembly (AN SC) Machine

Only Luer-Slip products require this process step. It is done by a single machine, which takes assembled needles from AN and inserts a metal clip into the needle hub. A changeover is required between assemblies of different SKUs. A typical changeover takes 30 minutes on average.

2.3.1.1.3 Packaging Stage

Packaged Needle Assembly (PN) Machine

The PN machine first attaches a safety shield to the needle hub. It then seals the needle in a single blister package, and finally packages the blister packages in cartons. All three steps are performed by a single machine. Both LL-PN and LS-PN products have to route through this machine.

Bulk Needle Packaging Machine

Only one type of product, Bulk needles, use this machine, which packages needles in bulk form. Because the bulk product has very low demand (less than 4% of the total demand in FY 2006), this machine runs only a few shifts per month.

Combo Packaging Machine

This machine is very similar to the PN machine. An assembled Luer-Lok needle from AN is assembled with a safety shield, then packaged together with a syringe into a blister package, and finally packed in cartons. Syringes are manufactured and supplied by another value stream in the same plant.

2.3.1.2 Processes outside Clean Room

After being packaged in cartons, products are moved out from the clean room. Processes outside the clean room include sterilization, out gassing, shrink wrap, and shipment to customers. The total processing time of these operations takes 15 days on average. As a consequence, products need to complete all operations in the clean room 15 days before their shipment date.

2.3.2 The Eclipse Value Stream Production Planning Team

Coordinated by a Value Stream leader, the production planning team consists of a production planner, shift supervisors, technicians and material handling personnel working on the shop floor. Major decisions on scheduling and production control are made by the planner.

2.4 THE CURRENT PRACTICE

2.4.1 Planning

The production plan is generated monthly for each individual machine by a planner using Excel spreadsheets. Based on the current WIP level and demand, the planner sets the production quantity, start date and finish date for each part for each machine monthly. After these production plans are generated, they are released to the production floor. The production plans of the Eclipse line are also passed to other value streams who supply syringes and cannula, so that syringes and cannula would be delivered according to the plans.

As AN is perceived as the bottleneck, the current planning strives to minimize the changeover on AN by producing large batches. Typically the production batch size of each SKU is chosen to be the order quantity for an entire month, which can go up to 5 million pieces for high demand SKUs. The transfer batch size is always the same as the production batch size, which means any downstream production of one SKU would not start until its upstream process is finished. Moreover, the production plans build in a time buffer between successive operations; thus, according to the schedule, a production batch from upstream will complete well before the scheduled start for its next downstream operation. This scheduled queuing time varies from a few days to a few weeks. For instance, the scheduled queuing time for hubs between molding stage and assembly stage is usually around 1 to 3 weeks. The scheduled queuing time for assembled needles between assembly stage and packaging stage is usually around a few days to 2 weeks.

Because of the long queuing time, there is a long cycle time for producing any of the product types. Due to the long cycle time, the current planning practice must rely heavily on the demand forecast. Figure 4 describes an example of the current planning practice. Planning for production in May is started on 15th Apr, when customer orders for May are confirmed. Because products require 15 days for processes outside clean room, orders in May have to be completed from packaging at the latest by 15th May. As a result, products produced at the packaging stage after 15th May are for demand in June, which is still based on forecast when the production plan for May is generated. For molding and assembly stages, production starts even earlier than packaging. From our analysis, 80% of assembly and 100% of molding are planned based on forecast due to the long cycle times. It is evident that shortening the cycle time is critical to enable production to be based entirely on firm orders.

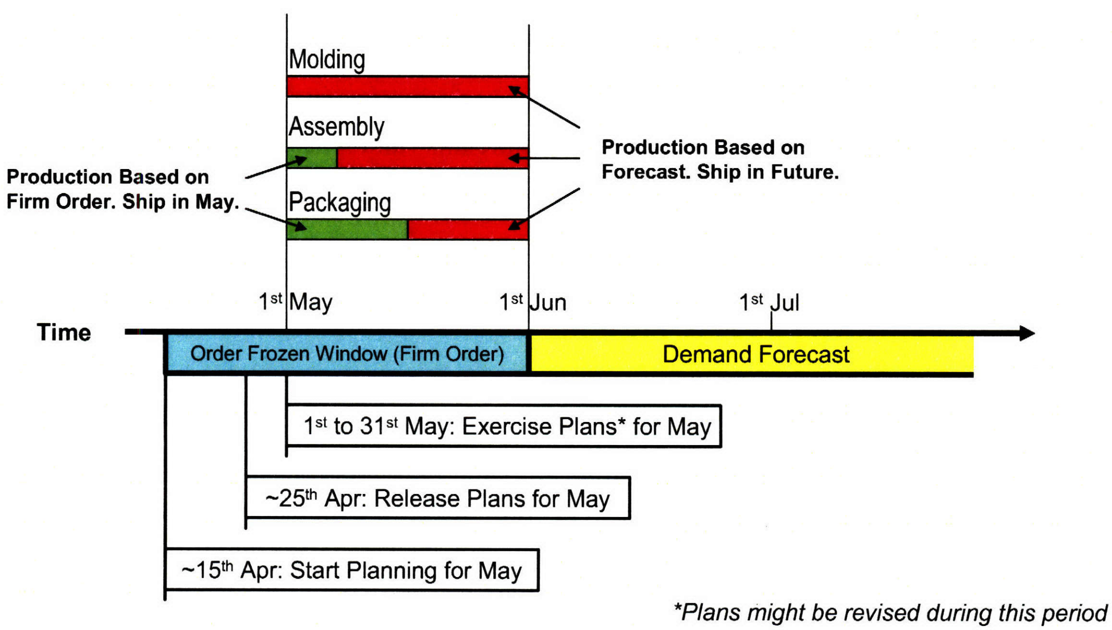

Figure 4 Current Planning Practice

2.4.2 Machine Capacity and Performance

Capacity

Currently, demand is very close to the originally designed capacity of many machines in the line, especially for the AN machine and PN machine. Since only the data from the produced quantity of the final products in FY 2006 is available, we calculated the demand for each machine based on the produced quantity, the flow paths of those products in the line, and the yield on each path, as shown in Equation 1. The quantity that machine A processes is equal to the quantity of the final products that are routed through this machine divided by the product of yield in the particular machine and downstream machines. Due to confidentiality concerns, the real demand quantity for each machine is not disclosed.

$$Demand_{machineA} = \frac{Q_{Final\ Products\ Route\ Through\ MachineA}}{\prod_{i=machineA}^{Last\ Machine} Yield_i}$$

Equation 1 Calculation of Demand of Each Machine

A shortfall could happen when the machine demand exceeds the machine capacity. There was no shortfall with the given demand data of FY 2006, because what we were given was the actual quantity produced in FY 2006.

Since all products visit the AN machine, as shown in Figure 3, this machine's designed capacity is already very close to the demand. It has to be utilized more than 90% during peak times. It is perceived to be the bottleneck. Nevertheless, the second bottleneck, PN machine, has a high utilization close to AN.

Performance

The machines on the production line are quite sophisticated yet still very reliable. Regular preventive maintenance is carried out on all machines to minimize breakdowns and excessive depreciation. The yield rates of the machines are also very high. Most machines have a yield rate of over 98%, and only AN yields a little lower at around 92%.

BD recently implemented a software suite developed by Apriso Corporation to track machine performance. Available information from the system includes production quantity, quality (yield), machine up time, planned / unplanned downtime, etc. Because the Apriso system was introduced less than one year ago, the tracking of machine data is not yet fully automated. The machine up and downtime are automatically recorded for most machines, but not the molding machines. When machines undergo planned downtime, for instance preventive maintenance, the reasons of stopping have to be manually entered from a list. Similarly when machines experience unscheduled failure, the reasons also need to be manually entered. We found many discrepancies between the performance data recorded by Apriso and that described by the production floor. For example, the reasons for machine failure are sometimes captured into wrong categories.

In spite of this, the data recorded for production quantity, quality and overall production time is reasonably accurate for most machines.

Measurements of production variability, like MTTF and MTTR, are not directly available from Apriso. In theory it is possible to estimate these metrics from the scheduled production time, unscheduled downtime and downtime-count data in Apriso. However, we met many problems when trying to do so. First of all, the recording errors in the Apriso system can make big differences in MTTF and MTTR. For example, planned downtime is sometimes recorded as unplanned downtime in the system. Because the planned downtime can sometimes be very long, it could increase MTTR dramatically if it gets recorded as unplanned downtime. Moreover, the machine downtime-count data is inaccurate in most cases. For example, the waiting time for material from upstream operations is often falsely recorded as machine downtime. Such recording errors increase downtime drastically. As a conclusion, the exact measurements of production variability are not available for our project. Because the company regards the machines as very reliable, we expect there is little loss from ignoring variability in our analysis.

2.4.3 Cycle Time and WIP Level

The WIP that we discuss in this project includes molded hubs and shields before assembly, as well as assembled needles. The locations of WIP on the process flow are indicated in Figure 5. Four major product types are labeled with different color and the light blue ellipses represent the WIP locations.

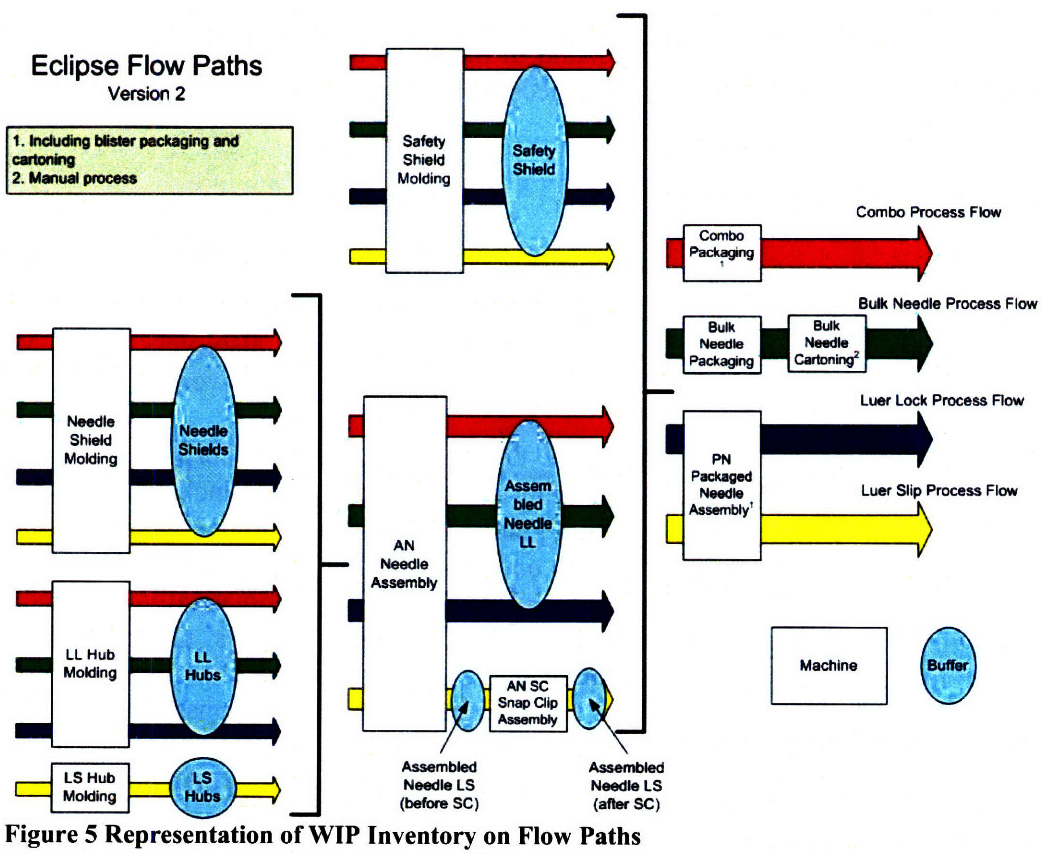

Figure 5 Representation of WIP Inventory on Flow Paths

The Apriso system is able to accurately track the real-time WIP status, but it does not store historical data. Table 3 summarizes the actual WIP quantity and cost extracted from the Apriso system, based on daily observation over the month June 2007.

Molded parts	Quantity* (K pieces)	Unit Inventory Cost ($)	Inventory Cost ($)
Luer Lock Hub	4,904	$0.013	$62,962
Luer Slip Hub	4,402	$0.018	$79,040
Needle Shield	2,003	$0.007	$14,508
Safety Shield	1,761	$0.008	$14,633
Total Molded	**13,070**		**$171,144**
Assembled parts	**Quantity* (K pieces)**	**Unit Inventory Cost ($)**	**Inventory Cost ($)**
Safety Needle - LL	8,692	$0.039	$342,765
Safety Needle - LS (before Snapclip)	6,316	$0.047	$294,323
Safety Needle - LS (after Snapclip)	4,725	$0.106	$499,108
Total Assembled	**19,733**		**$1,136,196**

* figure based on average of daily figure from 1st June to 31st June, 2007

Table 3 Actual WIP Inventory Quantity and Cost

In the month of June, an extraordinary problem with raw material supply forced the Snap Clip Assembly (AN SC) machine to stop for a month. Contamination of metal clips from the only supplier caused AN SC to starve, affecting all the Luer-Slip products. Those assembled Luer-Slip needles with the stained metal clips have to be scrapped sooner or later. The other three types of products do not route through ANSC. Upstream of the ANSC, the bottleneck AN continued to process Luer-Slip products because the company believed that ANSC would be able to catch up quickly as soon as it returned to production. As a result by early July, the WIP produced by AN kept building up in the buffer space before ANSC. Because of this rare problem, the figures in Table 3 over-state the actual inventory level for assembled Luer-Slip needles. However, the table still provides good estimate of assembled Luer-Lok needles, as well as molded components.

The average WIP inventory amounts to ~13 million pieces for molded parts, and at least ~14 million pieces for assembled needles after discounting the inflated estimate for Luer-Slip. Compared to the demand in FY 2006, the inventory levels for molded parts and assembled needles are both more than their average monthly demand. A large WIP inventory leads to long cycle time in the product line. The recorded current cycle time is about 60 days for the processes within clean room. Given the 45 days' order frozen window and the 15 days required for processes outside the clean room, the cycle time inside the clean room has to be within 30 days for the production to be fully based on firmed demand. Since the current cycle time inside the clean room far exceeds 30 days, the production is forced to be based on the demand forecast.

Reducing the cycle time may enable the production plan to be based on the firm orders. By reducing the cycle time, the factory also has the advantage of gaining flexibility [1]: The system will be more capable of very fast turnaround on individual orders, and the factory may more readily adapt to a changed order because the corresponding job may not have begun its processing. By reducing the WIP on the shop floor, the factory has the benefit of reducing inventory holding cost, faster detection of quality problems, hence less scrap or rework.

CHAPTER 3 PROPOSED SOLUTIONS TO THE OVERALL SYSTEM

After understanding the current operation practice and measuring the WIP inventory level, in this chapter we will analyze the root causes to the high WIP and long cycle time problem, and discuss our approaches to the problem.

3.1 ROOT CAUSES TO THE HIGH WIPP AND LONG CYCLE TIME PROBLEM

3.1.1 Unnecessary Early Start in a Push-Pull and Pure Push System

From the current planning practice (Section 2.4.1), we observe that the current practice is actually a mixture of push-pull and pure push approaches. In a pure push system, production is planned based on forecast. In a push-pull system, the upstream production is scheduled based on forecast, which is push, and the downstream production is based on demand, which is pull.

In the current practice, upstream machines, like the molding machines, produce completely based on forecast. It is a very typical push approach. The planning of the downstream machines, which are the assembly and packaging machines, is more complicated. For packaging machines, they produce based on demand in the first half of a month and based on forecast in the second half. As a result, they change from pull to push in the middle of each month. The situation for assembly machines is similar, but they transform from pull to push earlier (See Section 2.4). When we look at the production line as a whole, it is a push-pull system in the first half of a month and a pure push system in the second half. When it is a push-pull system, a push-pull boundary exists between molding and assembly at first, and then quickly moves to between assembly and packaging. As a result, a large amount of WIP is built up at the push-pull boundary, before and after the assembly, especially between assembly and packaging.

The major problem with this mixture of push-pull and pure push is that it does not work well with a make-to-order system, which the Eclipse line is designed to be. A push-pull approach could be used in a make-to-order system, but a pure push system is definitely not. However, how could the current practice survive with a pure push system? The secrets are the relatively accurate forecast and the planner's ability to communicate with customers frequently to further reduce forecast errors.

Even though the current practice seems to work well with meeting demand, it can increase cycle time and WIP. There are two reasons.

First of all, push starts production too early in either pure push or push-pull scenarios. The current practice simply starts pushing production of molded parts about one and half months earlier than demand, which directly increases cycle time and WIP.

Secondly, because the production line is supposed to be make-to-order, there is no finished goods inventory. If there is an error in forecast, like a cancellation of an order, the last stage (packaging stage), would not produce the cancelled order. The already partially produced order from upstream would sit in the buffers as WIP and wait until the next order of the same SKU arrives.

3.1.2 Unsynchronized Production Flow

Under the current practice, a monthly production plan is generated for all stations before the month starts. The plan specifies the production quantity, start and end date for each batch of material on all machines in the month. It is equivalent to say that the time that each batch visits each machine on its route is pre-determined. Because it is difficult to predict the exact time when the batch will be ready from the upstream station, the plan usually gives a more conservative schedule by requiring the upstream to complete the batch earlier. This planning method causes a lack of synchronization of flow for each product type and introduces an extra delay between stations on top of the queuing time.

3.1.3 Rationales behind the Root Causes

In summary, the two root causes which explain the high WIP and cycle time are

1) Unnecessary early start of production in a push-pull and pure push system.
2) Unsynchronized production flow caused by the over-detailed production plans.

After we find the root causes of our problem, it is not hard for us to understand the rationales behind them. First of all, because the AN machine is the bottleneck of the production line as a whole, the best interest of planning would naturally be to prevent starvation at the AN station. To prevent starvation, the plan requires molded parts to always be available in front of the AN machine. In order to do so, molding starts much earlier than assembly, the earlier the better AN is protected from starvation. However, a problem starts to develop when molding starts earlier and earlier. The make-to-order system gradually transforms from a pull system to push-pull system, or even pure push system. Because of the nature of push, products are manufactured based on forecast. In the case when the partially finished products do not have actual demand, they have to stay in the shop floor because there is no finished goods inventory. As a result, the last stage of the line has to be scheduled after the orders become firm. This is exactly what happens in the current practice. The last stage operates on pull according to its production plan and the other stages push material to the downstream. In the end, overly detailed production plans introduce extra delays to the production flow and further increase cycle time and WIP.

3.2 OUR APPROACHES TO REDUCE CYCLE TIME AND WIP

After understanding the root causes, our basic approaches to the problem became very clear.

First of all, we need to eliminate the unnecessary early start. Our goal is to gradually transform the system back to a pure pull system, in this case make-to-order manufacturing. However, if the cycle time can not be reduced to the quoted lead time to customers, a push-pull system is still necessary. Our goal would then be to move the push-pull boundary as far upstream as possible.

Then, we also need to achieve better synchronization of the production flow for each product family. Downstream machines could use FIFO policy instead of detailed production plans, if a pure pull system is used. If a push-pull system is required, only the most upstream machines and the machines right after the boundary have to be scheduled. Other machines can still use FIFO. Extra delays would be automatically eliminated with FIFO.

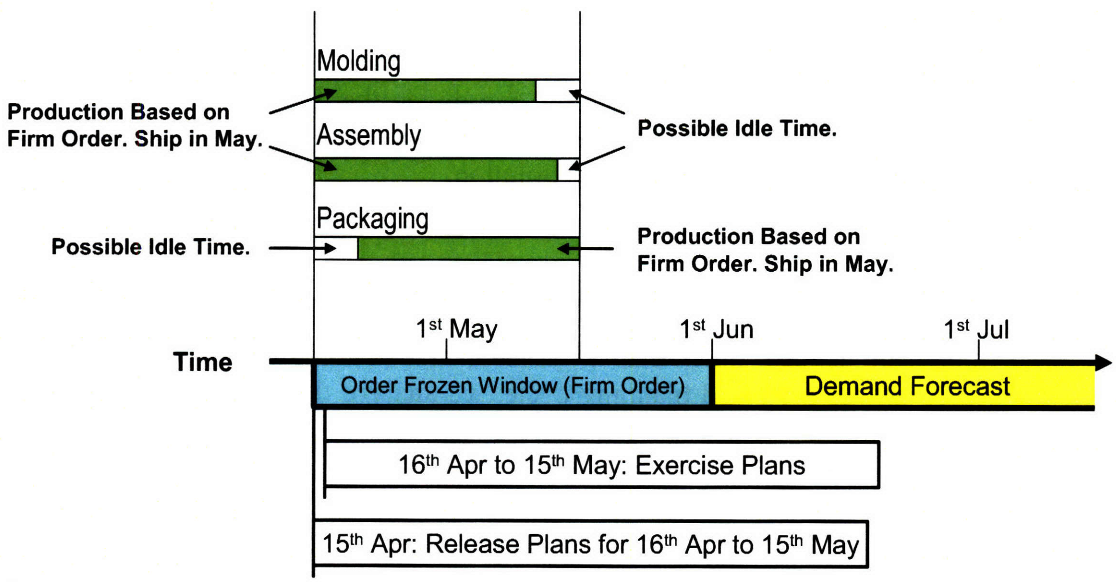

Figure 6 An Example of the Proposed Planning Cycle

These two basic approaches require some changes of the current planning cycle. The current planning cycle is from the beginning of a month to the end of it. To transform the system to a demand-based make-to-order system, we change the cycle to be from 16th of every month to 15th of the following month, which is just the first 30 days of the order frozen window. It is important to note that, in our new planning cycle, we just want to produce the orders which are confirmed on the 15th of that month. It means planning for each cycle is done just one day before the cycle starts based on actual demand.

Figure 6 shows an example of our proposed planning cycle. On 15th April, demand for May is confirmed. We release the production plans for the period from 16th Apr to 15th May on that day, based only on actual demand in May. During the production period, 16th Apr to 15th May, we run the machines to satisfy demand in May. An optimal outcome is that all products for May leave the clean room before 15th May, so that they can proceed to processes outside the clean room and be shipped before May 31. A new planning cycle for June then starts on 15th May.

There are two important questions to be addressed in the above example.

First of all, can the demand of May be satisfied, if we only start production on 16th Apr? If the answer is yes, then what we have described is a make-to-order or pull system which meets demand. If the answer is no, we need to further adjust the pull system to be push-pull, which means upstream of the production line has to start before 16th Apr.

Secondly, the current practice requires 10 days to do planning, is it possible to plan everything in just one day now? The answer might be yes. As we have already discussed, we propose to use FIFO when possible. No planning would be required when a machine produces based on FIFO. If the answer is no, our proposed solution would be hard to implement.

Now we can see that the two important questions are actually two requirements for our proposed solution. In the rest of our theses, we divide the problem into three parts and elaborate our solution in detailed steps. Meanwhile, we also show that the two requirements could actually be met by our solution.

The three parts are

1) Scheduling AN and downstream machines. In this part, we do not consider the molding machines other than to assume that molding will not starve the AN. Based on the proposed planning cycle, we analyze the machines downstream of molding in great details. Results are shown in both this thesis and the second thesis.

2) Scheduling hub molding machines. With the findings in the first part, we then consider the whole line including the molding machines. These results are discussed in the third thesis.

3) Scheduling safety shield and needle shield machines. We use a different approach for these two machines. This is discussed in this thesis.

CHAPTER 4 PROPOSED SOLUTIONS TO AN AND PACKAGING MACHINES

4.1 STRATEGIES FOR NEEDLE ASSEMBLY (AN) AND PACKAGING MACHINES

Following the overall proposed solution presented in Section 3.2, we will now discuss production control strategies for AN and downstream machines in detail.

Firstly, in order to address the root cause of the problem, we eliminate the unnecessary early start of production at the needle assembly machine as well as the extra delays between stations.

Then, we reduce the transfer batch size to further improve the system performance.

Finally, we analyze two scheduling strategies for AN to achieve a pure pull system with these machines:

- o Mixed Dispatching Rule
- o CONWIP Release Policy

In this chapter, we will discuss the first two strategies in great detail and briefly talk about the mixed dispatching rule and CONWIP policy. More details of the AN scheduling policies are discussed in the second thesis.

4.1.1 Early Start and Extra Delay Elimination

In Section 3.1 we have observed that the current planning method introduces unnecessary early start and extra delay by scheduling every single machine. This section discusses an approach to eliminate this delay by scheduling only AN, and then making the downstream packaging machines process work once material from upstream is available, based on FIFO policy. By properly controlling the release schedule of jobs to AN, we will show that this method can reduce the cycle time.

To evaluate the effectiveness of this method, we conduct simulations in the SIMUL8 10.0 package. SIMUL8 is a powerful analytic tool for Discrete Event Simulation, which allows visual models to be easily created by drawing objects and results to be displayed interactively on the screen.

Since this chapter focuses on the AN and packaging machines, we confine the simulation models to these machines, as shown in Figure 7. We simulate two scenarios. The first one is to schedule every machine; AN machine starts much earlier than the downstream machines, and extra delays between stations are present. This is close to the real case under current practice and considered as the baseline case in simulation. The other one is the proposed method (see Section 3.2), in which only the AN is scheduled (according to real demand) and packaging machines at downstream process work without delay, based on FIFO. In both cases, we do not split any batches. In other words, transfer batch sizes are the same as the production batch sizes, with both set equal to the monthly demand for the SKU. This assumption is made in this analysis to isolate the impact of the early start and the elimination of delays from batch size reduction. We will analyze the benefit of batch size reduction in the subsequent section.

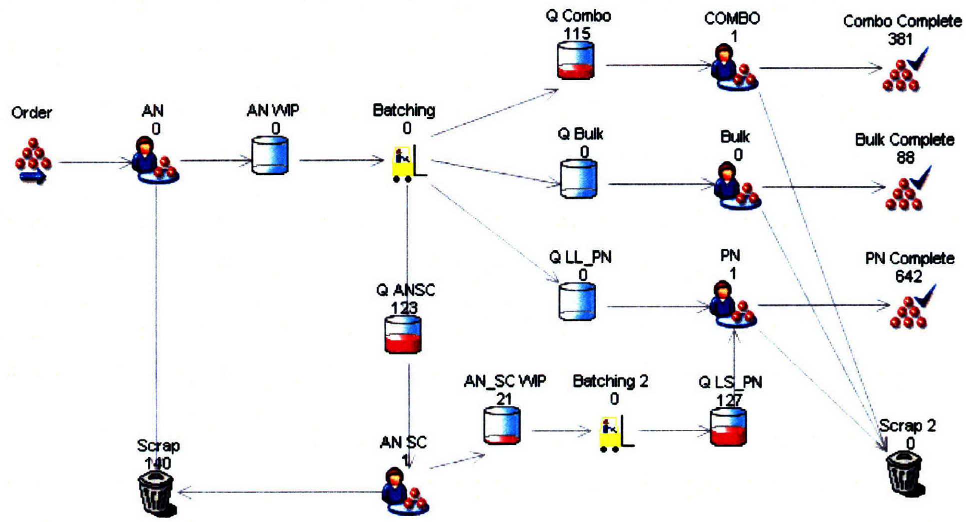

Figure 7 Simulation model for AN and Packaging machines

We first state the assumptions for this model:

1) AN machine never starves, i.e. parts supplied from the upstream molding machines are always available. Because AN is perceived as the bottleneck in the system, it is the planner's best interest to protect AN from starvation. In practice, it is almost never starved. We will explore the strategies to schedule the upstream machines to minimize starvation of AN in Chapter 5 and Thesis 3.

2) The simulation was run over a 9 months' time span: Oct 2006 – Jun 2007. The beginning of the period represents the start of the fiscal year. No backorders from the previous year were allowed to carry forward to this period. Thus we assume that production of this period starts from the orders for Oct 2006.

3) In the simulation of the base case with presence of the early start of assembly machines and extra delay between stations, individual monthly production plans are generated for each machine. We assume that AN machine starts about 10 days earlier than the downstream machines because in the monthly production plan, 10 days is the typical period for which production at the AN leads the production at its downstream machines. In addition, we have estimated the delay before the AN SC, COMBO and BULK machines based on the actual production plans. In the simulation, these delays are set to be normally distributed with mean of one work shift (8 hours) and standard deviation of one shift. The delay before PN is assumed to

follow a normal distribution with mean of half a shift (4 hours) and standard deviation of half a shift.

4) There are 28 available work days in each month.

5) In comparison, in the proposed method with unnecessary early start and extra delays eliminated, we perform scheduling only on the AN machine. The schedule follows the real monthly order quantity in the past 9 months. The downstream machines start to work as soon as the batch arrives at the front of the queue. Downstream does not wait except for the usual changeover time, 30 minutes on average, between two SKUs.

6) The release of material follows the current production practice, in which one major product type is released, followed by another. Although the detailed monthly production plans from Oct 2006 to Apr 2007 are not available for us to duplicate the real scenario, we have verified that assumption (3) represents the production practice from the production plan in May 2007 and Jun 2007. The monthly production cycles through the products one at a time. Luer-Lok Needle in Individual Blister Package (LL-PN) products are released first, followed by COMBO products, BULK products, and finally, Luer-Slip Needle in Individual Blister Package (LS-PN) products.

7) We estimate the process time at each machine based on the average time to process 10,000 pieces. The process time for a single part in any machine is very small; hence, it is easier for us to work with the time for 10,000 parts. 10,000 pieces is also 1 storage unit for WIP in the line. In the simulation we assume for each machine that the time to process 10,000 pieces follows an exponential distribution. We estimate the expected process time from the total output quantity in the past 9 months. The expected process time is calculated as the total production time over the total output quantity in 10,000 pieces. We express this relationship in Equation 2.

$$\textit{Expected Process Time} = \frac{\textit{Total Production Time}}{\textit{Total Output Quantity}(\text{in } 10\text{K})}$$

Equation 2 Calculation of Process Time in Simulation

8) Based on the machine data from the past 9 months, we assume that each machine has a normally distributed changeover time with mean of 30 minutes and standard deviation of 10 minutes.

9) Each machine has a constant yield rate which we approximate by the average yield observed in the past 2 months.

In the model shown in Figure 7, the label named 'Orders' determines the type, quantity and sequence of work released to the AN machine. We set 'Orders' based on the actual monthly orders that were completed in the period from Oct 2006 to June 2007. If AN completes the monthly order ahead of time, it stops operation and waits till the next month. On the other hand, if shortfall takes place in a month, AN carries the unfinished workload to the next month.

The workstation labels named 'AN', 'AN SC', 'COMBO', 'Bulk' and 'PN' represent the 5 machines in the Eclipse line. Scrapped parts from each machine are put into two bins named 'Scrap' and 'Scrap 2'. The cylindrical containers are the buffers for WIP. The red triangles at end of the line named 'Combo Complete', 'Bulk Complete' and 'PN Complete' represent finished products from the clean room. The numeric values marked with each object (machine or buffer) represent the quantity of work items in the current object.

'Batching' and 'Batching 2' (with yellow forklift images) are two dummy machines with fixed processing time of zero. They are created to hold partially completed batches.

Table 4 summarizes the simulation results for Average WIP level in June for the two scenarios, as well as the actual WIP level. The simulated baseline deviates from the real situation because of the discrepancy between demand forecast and actual demand. In the actual production AN operates based on forecast for 80% of the time, while in the simulation AN is scheduled according to the real demand data. Moreover, because of the conservative approach taken to address yield loss in the current practice, planning usually schedules more quantity than the downstream machine would actually consume. The increment contributes to WIP on the floor. In spite of the slight deviation from the actual case, the simulation model is considered reasonably accurate

	Baseline	System of Baseline	Simulation of System after Eliminating Early Start & Delay
Avg WIP in June (K)	~ 14,000	11,108	3,046

Table 4 Comparison of Average WIP level of June under different scenarios

The results from Table 4 show that by eliminating AN's early start and extra delays between stations, the proposed method effectively reduces the average WIP level of June by nearly 8.1 million, about 73% from the baseline. The reason for this improvement is apparent: both early start and delay elimination reduce queuing time on the shop floor.

Figure 8 compares the WIP level over the 9 months' simulation horizon under the two scenarios. Each point in the plot represents a daily ending inventory level. In the baseline case, since AN starts production about 10 days earlier than the downstream machines, WIP starts to accumulates from the middle of September. Over the course of time, although WIP may drop a little when demand is low (e.g. in late February), for most of the time the WIP level exceeds 6 million. After AN completes the orders for June in the middle of the month, it starts to work on the orders for July. The WIP level does not drop even if the monthly orders are completed ahead of time.

In comparison, in the proposed model the production of AN only starts at October. Even if AN machine has already started processing part of the demand in September, less production time will be planned in October. In the subsequent 8 months, the machines only work on orders for the current month. If the current month's orders are completed before the end of the month, AN will stop to wait until the next month's demand becomes firm. For example, after AN completes the orders for June in the middle of the month, it stops production. Since WIP is no longer created and it is consumed by the downstream stations, the WIP level drops during this period. In this way, there is much less accumulation of WIP on the shop floor. We note, however, that in the proposed model, orders for each month may not be completed on time; in this case, these orders are carried forward to the following month. Nevertheless, when we apply additional strategies (to be discussed in 4.1.2 and 4.1.3) together with the proposed model, we will see that the system is able to meet monthly demand on time. This will be discussed in Thesis 2.

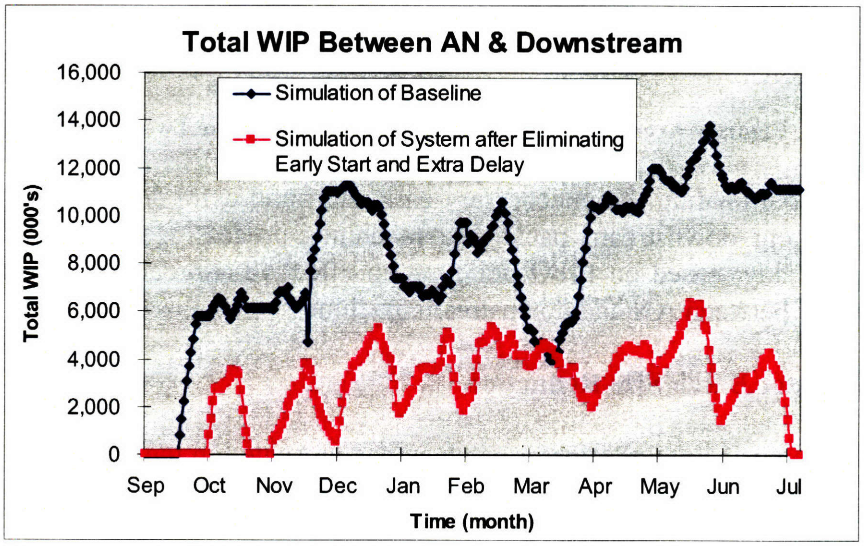

Figure 8 Evolution of WIP level between AN and Packaging Machines

A comparison of average cycle time from AN to packaging machines under the two scenarios is shown in Figure 9. The cycle time for each product type is estimated from the average cycle time of each SKU over the 9 months' period. By eliminating the unnecessary early start and extra delay between adjacent stations, the proposed method reduces the cycle time for all product types, especially for LL-PN and LS-PN products. Compared to Combo and Bulk products, LL-PN experiences longer cycle time because it possesses the highest demand distribution in the product mix (~51% of total demand). LS-PN's longer cycle time is attributed to its extra process step through the AN SC machine, as well as its second highest demand distribution (~20%). Significant cycle time reduction on both LL-PN and LS-PN can be explained by a better utilized PN machine when the extra delay is eliminated in the proposed method.

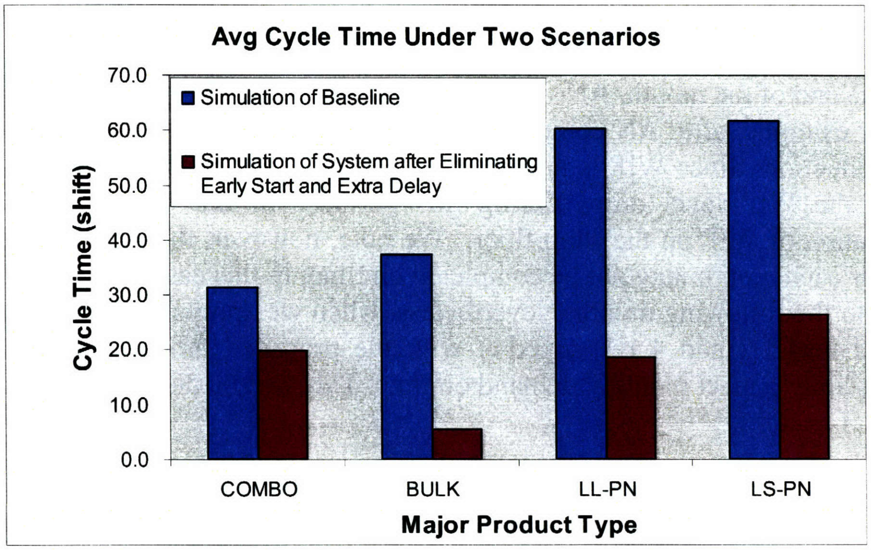

Figure 9 Average Cycle Time Comparison Under the Two Scenarios

In summary, eliminating the unnecessary early start and extra delay by scheduling only AN and making downstream packaging machines process work once material from upstream is ready, based on FIFO policy is an effective approach to reduce cycle time and inventory between AN and downstream machines.

4.1.2 Transfer Batch Size Reduction

The underlying principle of transfer batch size is presented in the Move Batching Law of factory physics [2].

Law (Move Batching): Cycle times over a segment of a routing are roughly proportional to the transfer batch sizes used over that segment, provided there is no waiting for the conveyance device.

Although the practical system is more complex, the Move Batching Law reveals the importance of reducing transfer batch size for the reduction of the cycle times.

To examine the impact of transfer batch size in this problem setting, we developed a simulation model similar to what is shown in Figure 7. The model again takes demand of the past 9 months as input, and simulates scenarios under full transfer batch size, and under smaller batches of 200K, 150K, 100K, 50K and 10K.

The simulation model with full transfer batch size is exactly the same model as the one with extra delay eliminated described in 4.2.1. In other words, unnecessary early start and extra delay have been removed by using the method proposed in 4.2.1, and the transfer batch size follows the current practice of equating the transfer batch to the production batch. This acts as the base model in this analysis; additional benefits from reducing transfer batch size can be seen from comparing the models with smaller transfer batch sizes with the base model.

The simulation models for smaller transfer batch sizes have the same assumptions and similar configurations to the base model except for the setting of the two dummy machines 'Batching' and 'Batching 2'. These two machines have fixed processing time of zero and they serve to hold partially completed batches. They are configured so that batches are only moved when the quantities reach the preset levels, or the transfer batch sizes. Take batch size of 200K as an example: batches are transferred to the downstream buffers when the WIP level reaches 200K. But for low demand SKUs whose monthly demand quantities are below 200K, batches are transferred when they are finished instead of waiting for 200K parts. In the simulation of 10K, parts would be transferred immediately to downstream since 10K is the minimum production size within the simulation.

Figure 10 shows the simulation results of average WIP levels in June under different transfer batch sizes. We see that, when we reduce the batch size from the baseline to 200K, the WIP level drops significantly by almost 1.2 million, or 31%.

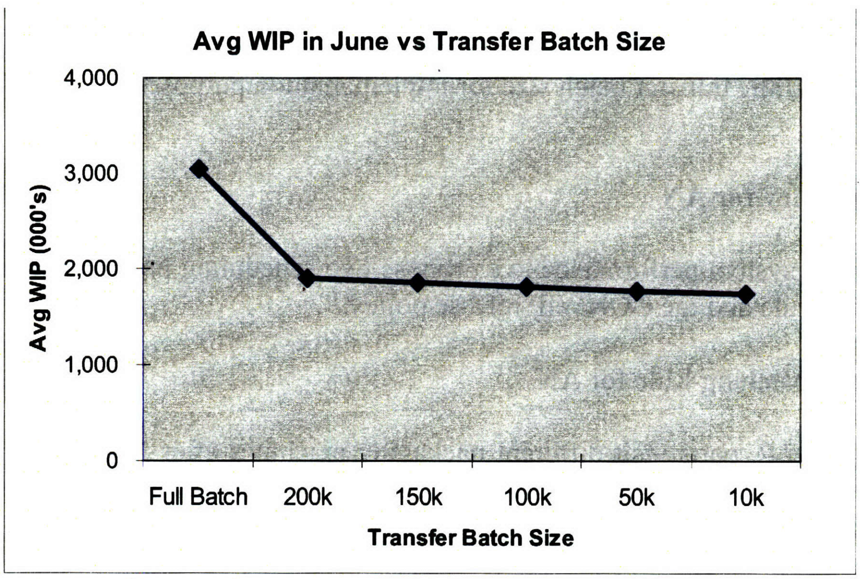

Figure 10 Average WIP in June vs Transfer Batch Size

We can see the impact of transfer batch sizes on the average cycle time from AN to packaging stations from Figure 11. For all 4 product types, cycle time reduces with transfer batch sizes. The reduction is more significant in LS-PN (~11.5 shifts) and LL-PN (~10 shifts) products. This is due to the fact that LL-PN and LS-PN originally had the highest demand with the most number of large batches. When the transfer batch size decreases to 200K, most production batches can be split and utilization at PN can be improved. At 200K and smaller batch sizes, cycle time of LS-PN is even better than COMBO products because of the faster processing rate of PN (180K/shift) than COMBO (80K/shift) at packaging station. Similar reason holds for LL-PN products.

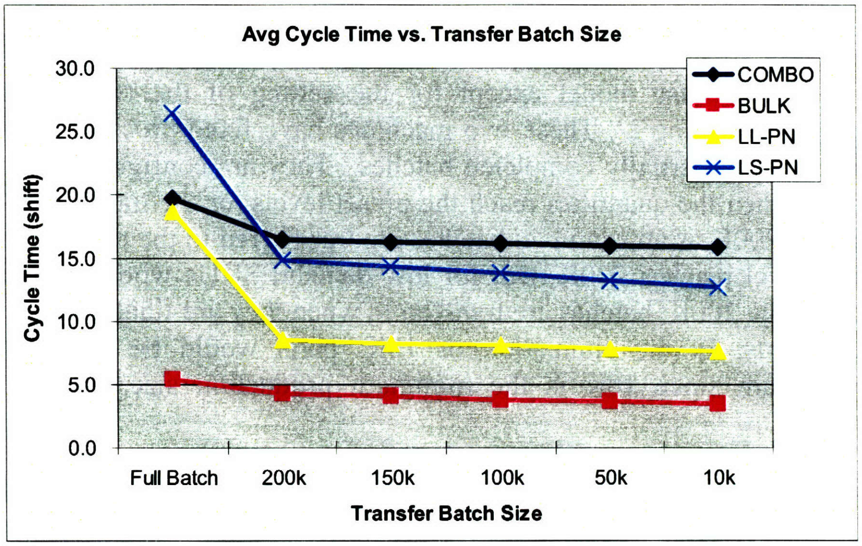

Figure 11 Average Cycle Time vs. Transfer Batch Size

The system benefits most when the transfer batch size is reduced from the original full batch to 200K. Reducing to even smaller transfer batches requires more material handling, but does not lead to significant improvement in cycle time and WIP. Thus we recommend to adopt a 200K transfer batch size for implementation purposes.

4.1.3 Scheduling Policy for AN

To further improve the system performance, we propose two scheduling policies for AN: the mixed dispatching rule and the CONWIP release policy.

4.1.3.1 The Mixed Dispatching Rule for AN

As described in Chapter 2, parts visit different packaging machines after AN depending on their product types. Although the combined processing rate at packaging station is higher than AN, each individual packaging machine is slower than AN. As assumption (3) in 4.2.1 states, under the current practice, we apply a simple schedule rule at AN where all SKUs from the same major product type are processed before switching to another product type. Under such structure, a queue will form in front of one packaging machine because AN is faster. However, at the same time, the utilization at the other two packaging machines is likely to be low.

Such problem can be alleviated if we choose a mixed dispatching rule instead. The mixed rule staggers different product types released to AN, which allows the multiple downstream routes to be utilized simultaneously. In this way we can achieve a reduced cycle time and WIP after AN.

Moreover, the schedule at AN has an implication on the upstream hub molding machines because it determines the demand for hubs. We perform a comprehensive analysis in Thesis 3 and derive a feasible mixed scheduling policy at AN. In this thesis, unless otherwise stated, all studies using mixed dispatching rule at AN will be based on the rule proposed in Thesis 3.

Thesis 2 compares such mixed dispatching rule with the simple scheduling rule under the current practice. We perform the analysis on the model that already incorporates the solution from 4.2.1, i.e. early start and extra delay eliminated, and is designed to gauge the additional benefit brought by mixed scheduling policy. Results from Thesis 2 show that at the recommended transfer batch size of 200K, a mixed scheduling policy further reduces the average WIP by ~0.94million, or 44%, from that of the simple scheduling rule in the month of June. Average cycle times are reduced by 20%-50% from the various product types.

It is evident that the solution provided by the mixed scheduling policy is effective in terms of WIP and cycle time reduction. Readers are referred to Thesis 2 for the detailed results.

4.1.3.2 The CONWIP Release Policy

A CONWIP system can provide the factory with good opportunity in WIP control between AN and packaging machines. In a CONWIP release policy, we will release a new job to AN only when another job completes production and departs from the system. The total WIP level in the system is kept below a target inventory level, or the CONWIP level.

This project focuses on a single CONWIP level system, in which the total WIP quantity from all SKUs are monitored and kept below the CONWIP level. Whenever a batch completes its production in the line, the CONWIP triggers a release of a new batch according to the scheduling policy at AN.

The CONWIP level plays an important role in the CONWIP system. An extremely low CONWIP level causes frequent blockage to AN and hence poor throughput. On the contrary, an extremely high CONWIP level has no impact on the system because it can hardly be reached. By properly sizing the CONWIP level, the system can effectively reduce cycle time and WIP without resulting in late shipment.

Thesis 2 discusses in detail about the choice of CONWIP level. We show by simulation that, a CONWIP level of 2.5 million between AN and packaging station in conjunction with the mixed dispatching rule at AN, will enable the system to meet its desired demand. The resulting average WIP level in June is 1.26 million based on the past 9 months' demand data, compared to 1.29 million before applying CONWIP (3% reduction).

4.2 RECOMMENDATION FOR IMPLEMENTATION

In 4.1 we discuss four strategies to reduce cycle time and WIP between AN and packaging machines. For implementation purpose, a feasible sequence of steps could be:

1) Eliminate unnecessary early start and extra delay between stations.

2) Reduce transfer batch sizes.
3) Apply mixed dispatching rule at AN.
4) Adopt CONWIP release policy.

The above steps are sequenced based on their ease of implementation and impact on the system. The first three are relatively easy to implement because they require less changes from the current infrastructure. The first step addresses the root of the problem, while the subsequent two steps address two important issues. Risks associated with these changes are relatively low. An extra step can be taken to benchmark the results after each change is made. One disadvantage of the above solutions is more material handling is required when we reduce the transfer batch size.

The CONWIP solution is more complicated to implement. Firstly, the target CONWIP level needs to be carefully determined. If the CONWIP level is set too low, the AN machine has the risk of frequent blockage. Throughput would be affected in such cases. Although we show that the proposed CONWIP level of 2.5 million meets the throughput requirement in simulation, a pilot run could start with a higher CONWIP level and gradually decrease whenever possible. This project examines system performance using demand data up to the end of FY 2007, thus the CONWIP level also needs to be adjusted if the average demand rate changes drastically in the future. The second challenge is in the requirement for a WIP monitoring mechanism. An infrastructure has to be in place to constantly monitor the inventory status and control the release at AN. Since the Apriso system already has the capability of tracking inventory level, a possible method would be building functions to generate WIP reports from Apriso at regular intervals and trigger material release based on the reports.

Figure 12 summarizes the impact of the proposed solutions on the average WIP level in the month of June, based on the simulation results obtained in Chapter 4 and Thesis 2. It shows that by implementing Step (1) and (2), a significant WIP reduction of 9.2 million (~83%) can potentially be achieved. Implementation of all 4 steps could reduce the WIP by 9.8 million (~88%) from the baseline case. The CONWIP policy, or step (4), seems to have insignificant impact on the WIP reduction in the month of June. This is due to the fact that the demand in June is too low for CONWIP to take effect. Nevertheless in other months with higher demand, CONWIP will be more effective in cycle time and WIP reduction.

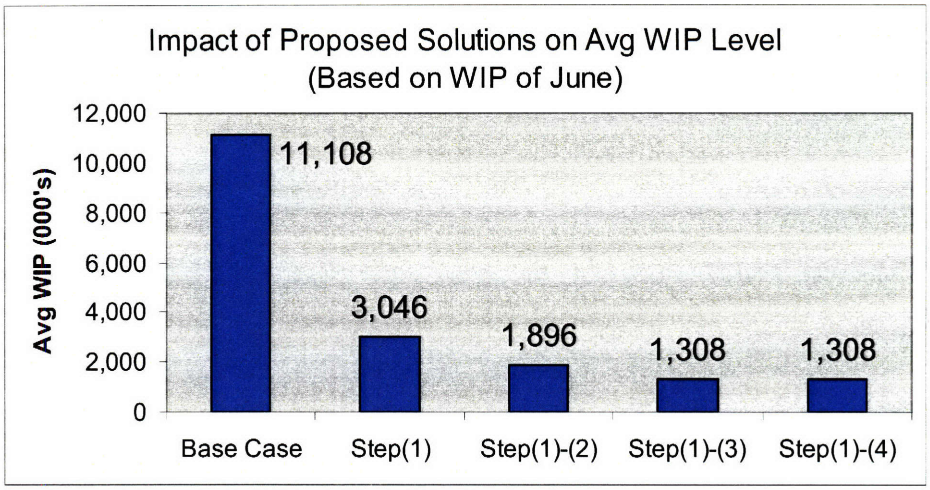

Figure 12 Impact of Proposed Solution on WIP

Moreover, two challenges in the overall problem were brought up in Section 3.2 The first one about whether monthly demand can be met if production starts just 45 days prior to the due date, the answer is yes and thus a pure pull system could be achieved for AN and downstream machines. Detailed analysis is included in Thesis 2 and Thesis 3. The second one is whether planning activities can be completed in a short period of time. The answer is very likely, because the proposed solution to AN and its downstream only requires the scheduling of AN machine. In contrast to the current practice, the packaging machines do not have to be scheduled as they would operate on FIFO policy. This simplified structure will allow significant time saving on planning activities.

CHAPTER 5 INVENTORY CONTROL FOR NEEDLE SHIELD AND SAFETY SHIELD MOLDED PARTS

5.1 CURRENT NEEDLE SHIELD AND SAFETY SHIELD PRODUCTION AND INVENTORY

Each product requires one Needle Shield (NS) and one Safety Shield (SS) to be assembled with the needle. As described in Chapter 3, these two are common parts across all SKUs and each part is molded on dedicated machines. Because of the pooling effect, demand for these two parts is more stable.

Under the current practice, NS and SS are produced only once a month, where the production quantity is set to the requirements for one month. Typically the molding machines for these two parts are scheduled to run in the first three weeks of a month to build sufficient stock to meet the demand for the month. For the remaining one week in a month, these machines are often idle. As a result, the WIP inventory level can vary quite a bit over the month. Moreover, safety stock for both shields kept on the shop floor is very large, although the reason for the excessive inventory is not understood. The average WIP level in a month is about 2.0 million pieces for NS and 1.8 million for SS, which slightly exceeds the demand for one month.

5.2 A PERIODIC REVIEW INVENTORY MODEL

Because both the NS and SS parts are common parts for all SKUs, a simpler inventory policy such as a periodic review, base stock model is possible. Although it may not represent the optimal inventory control system, we show that a periodic review policy can provide a reasonably good solution in this environment. Compared to continuous review, a periodic review system has the advantage of reducing the inventory review cost. It also allows reasonable prediction of the level of workload because replenishment decision is made at a known time.

Under a periodic review, base stock policy, an inventory review interval (r) is determined. During each review period, inventory is replenished up to the target level, the base stock (B) level. The choice of review period r is analogous to the determination of EOQ expressed as time supply, except that it is imposed by physical constraint. In this case, the minimum possible review period r has to be one day.

The optimum review period (r^*) that minimizes the total replenishment cost and inventory holding cost is given by

$$r^* = \sqrt{\frac{2K}{\lambda h}}$$

where λ is the daily demand rate of parts,
h is the inventory holding cost per piece and
K is fixed cost incurred in each replenishment cycle.

Equation 3 Calculation of Optimal Review Period

λ, the demand rate, is the average rate of parts being consumed by downstream machines. It is also adjusted by a quality factor to account for extra parts required to compensate for scrap or rework. *h*, the holding cost, can be obtained from the Apriso system which BD uses to monitor inventory status. The holding cost is $0.01 SGD per piece per day for both NS and SS. *K*, the fixed cost, includes the costs of setting up machines and reviewing inventory. No data is directly available but a reasonable estimate of the fixed cost would be in the range of $50-$500 SGD.

We summarize the values for the above parameters in Table 5, where the optimal review period *r** is calculated from Equation 3 based on a fixed cost of $100. Because the fixed cost is only estimation, a sensitivity analysis is performed to determine the impact of the estimation error in fixed cost on *r** for both NS and SS. The sensitivity plot in Figure 13 demonstrates that, *r** will be less than one day even if the real fixed cost is on the higher end of the possible range. Because of implementation constraint, it is reasonable to choose $r = 1$ day as the review period for both NS and SS.

	NS	SS
λ (pieces/day)	759,563	767,274
K ($)	100	100
h ($)	0.01	0.01
r* (days)	0.16	0.16

Table 5 Choice of Optimal Review Period

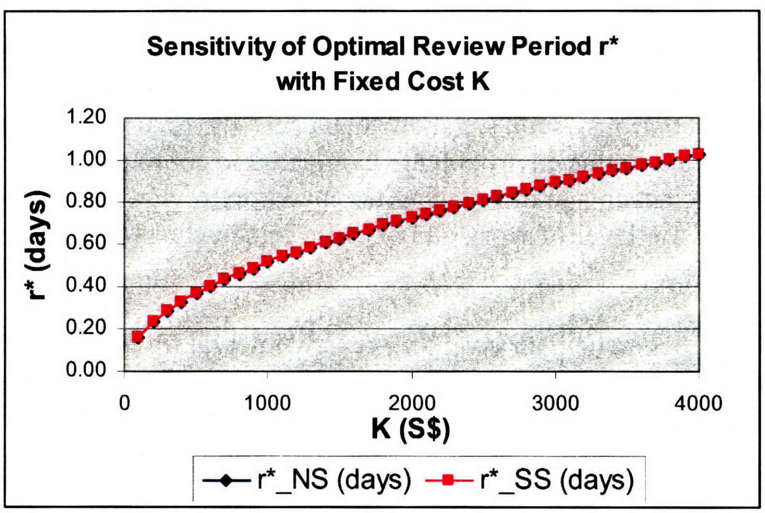

Figure 13 Sensitivity of Optimal Review Period with Value of Fixed Cost

The lead time of the replenishment (*L*) is the time taken for shield molding machines to produce the daily demand quantity as described by Equation 4.

$$L = \frac{Daily\ Demand\ Quantity}{Daily\ Production\ Quantity\ from\ Shield\ Molding\ Machines}$$

Equation 4 Calculation of Leadtime

Assume that molding machines producing NS and SS are reliable and seldom experience long failures. The production rate does not deviate much from its average value. These assumptions are validated from the observations on the shop floor. Based on average daily demand rate and the average production rate of the NS and SS molding machines, the lead time for NS is 0.65 day and that for SS is 0.77 day. We will take $L = 1$ day for both parts to buffer for any variation in daily demand or daily production.

For the deviation of daily demand, a coefficient of variation of 0.2 is assumed. This is a relatively conservative estimate because the variance of the downstream machines' production rates is often low.

A summary of the model parameters is presented in Table 6. The base stock (B) and expected inventory level can be obtained from Equation 5 and Equation 6 [3].

$$B = \lambda(r + L) + z\sigma\sqrt{r + L}$$

Equation 5 Calculation of Base Stock

$$Exp\ Inventory = \frac{r\lambda}{2} + z\sigma\sqrt{r + L}$$

Equation 6 Calculation of Expected Inventory

The WIP inventory pattern resulting from the proposed periodic review model over 20 working days is plotted on Figure 14 and Figure 15 overlaid with the actual inventory level in June from daily observation. We have shown that the average inventory level can be reduced by 49% for NS and 41% for SS. The model is also able to reduce the variance of the daily WIP level, because inventory replenishment is done on a daily basis. Compared to the current practice where one month demand is produced at one time, the periodic review model is able to smooth out the production quantity over the entire month.

		NS	SS
Model Paramters			
Demand Rate	λ (K pieces/day)	760	767
Review period	r (days)	1	1
Leadtime	L (days)	1	1
Demand C.V.		0.20	0.20
Demand Stdev	σ (K pieces/day)	152	153
Safety Factor	z	3	3
Stock Calculation			
Base stock	B (K)	2,164	2,186
Cycle Stock	(K)	380	384
Safety Stock	(K)	645	651
Expected Inventory	(K)	1,024	1,035
Actual Inventory		2,003	1,761
% Reduction		49%	41%

Table 6 Model Parameter and Inventory Stock Calculation

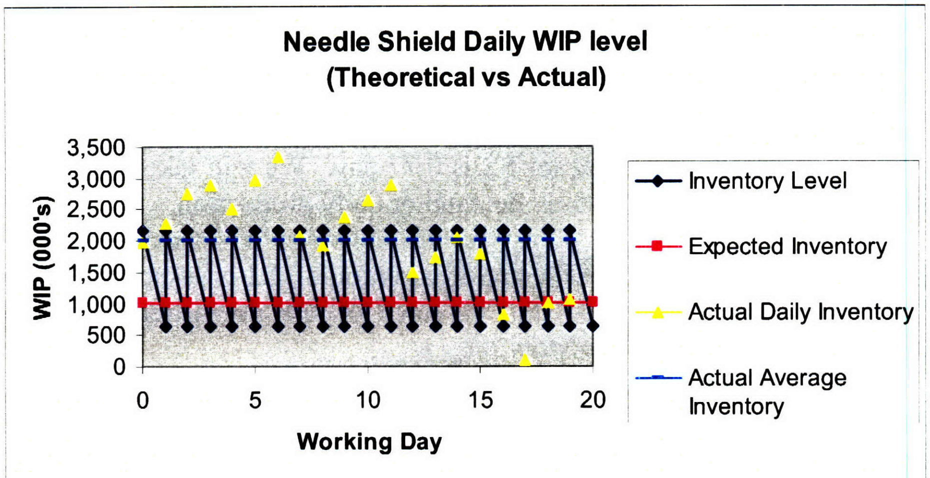

Figure 14 Needle Shield Daily WIP Level

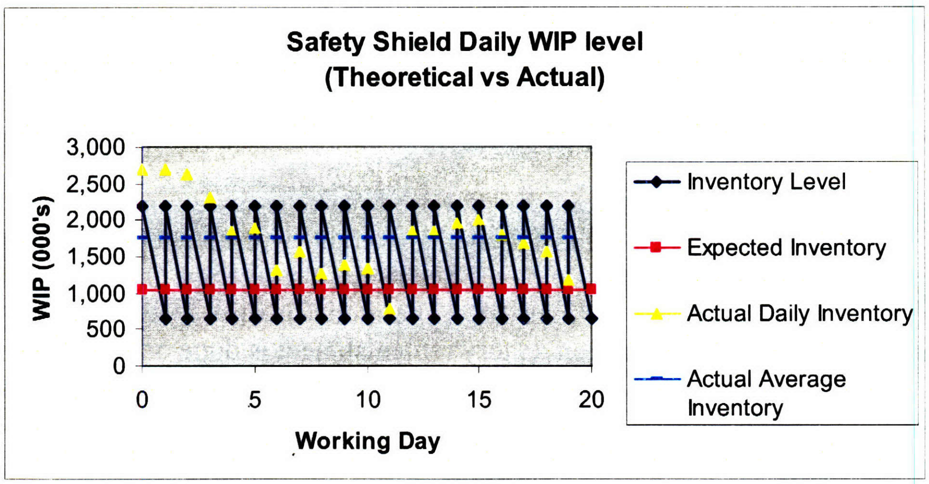

Figure 15 Safety Shield Daily WIP Level

5.3 RECOMMENDATION FOR IMPLEMENTATION

In the previous section we propose a periodic review, base-stock inventory policy with a 1 day review period for both NS and SS. With a base stock level set at 2.163 million and 2.186 million for NS and SS, we are able to keep the stock-out probability to less than 0.1%.

In the execution of this policy, the production floor has to review the WIP inventory level each day from the Apriso system. The decision on the daily production quantity is made based on the current inventory status and the base stock level. In the long run, the base stock level needs to be adjusted when the demand rate changes or when more accurate estimate of demand variance is available.

CHAPTER 6 CONCLUSION

This project examined the high cycle time and WIP problem in BD from a manufacturing system and product scheduling perspective. The overall system was divided into three sub-systems and analyzed using different approaches: Needle Assembly machine (AN) & its downstream packaging machines, Hub molding machines, and Needle Shield (NS) & Safety Shield (SS) molding machines.

This paper first introduced the current practice and identified the root causes of the problem. The high cycle time and WIP are attributed to the unnecessary early start of production, as well as the unsynchronized production flow caused by the over-detailed production plans. Four strategies were proposed to address the cycle time and WIP problem from AN to downstream machines:

1) Eliminate unnecessary early start and extra delay by scheduling only AN, and leaving packaging machines to produce according to FIFO policy.
2) Reduce the transfer batch size to 200K.
3) Apply the mixed dispatching rule to AN. This policy allows multiple routes to be utilized simultaneously.
4) Apply CONWIP release policy. A CONWIP level of 2.5 million will enable the system to meet desired throughput and product mix.

The results from simulation have shown that, by incorporating simple changes like eliminating early start and extra delays, reducing transfer batch sizes, or applying mixed dispatching rule at AN, the system can have large benefit from reducing cycle time and WIP at AN & downstream. Further improvement can be achieved by implementing a more complex solution like CONWIP system.

In the analysis for NS and SS parts, as they are common across SKUs, a simple periodic review, base stock inventory policy was suggested. By choosing a review interval of 1 day and setting the base stock level of 2.16 million and 2.19 million to NS and SS respectively, we can reduce the average inventory for both parts by more than 40% from the current level. Daily inventory replenishment also allows the shop floor to benefit from more smooth production and results in reduction of inventory variance.

For the implementation of the proposed solutions, we recommend to start pilot runs with more conservative measures. Over time, the models need to be adjusted to reflect the changes in the real world.

REFERENCES

[1] Wein, L. M. 1990. Scheduling Networks of Queues: Heavy Traffic Analysis of a Two-Station Network with Controllable Inputs. *Opns. Res.* **38**, 1065-1078

[2] Hopp and Spearman, Factory Physics, 2nd edition, Irwin, 2000. (HS)

[3] Simchi-Levi, Kaminsky and Simchi-Levi, Designing and Managing the Supply Chain, Irwin McGraw Hill, 2nd edition, 2003 (SKS)