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

by

Kai Meng

M.S. Electronic Engineering
Tsinghua University, 2006

B. S. Electronic Engineering
Tsinghua 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 of Mechanical Engineering
Chairman, Department Committee on Graduate Students

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

by

Kai Meng

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. At Needle Assembly machine (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 by the current planning method, and by reducing the transfer batch sizes. In this paper, we refine further these approaches to AN and packaging machines with consideration of a mixed dispatching rule and a CONWIP release rule. The mixed dispatching rule reduces WIP level of the system by enhancing the total throughput of the four production routes after the bottleneck (AN machine). The CONWIP release rule further reduces WIP by controlling the total amount of inventory in the system. With these four proposed strategies, we can have a pure pull system within AN and downstream machines and achieve significant reduction in cycle time and WIP.

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 Sivakumar Appa Iyer 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 Yi Qian. This project could not be possible without their teamwork.

Finally, I would like to thank my parents, my brother, and my friends, for their great support throughout this project.

Table of Contents

Abstract 3

Acknowledgement 5

Chapter 1 Introduction 11

1.1 Background 11

1.1.1 BD Medical and BD Tuas Plant 11

1.1.2 The Team Project 12

1.2 Thesis Overview 13

Chapter 2 Eclipse Safety Needle Production Line 14

2.1 The Product 14

2.1.1 Eclipse Safety Needle 14

2.1.2 Four Major Product Types 15

2.1.3 28 SKUs 16

2.2 Demand and Customers 16

2.2.1 Demand Distribution 16

2.2.2 Customers and Orders 18

2.3 The Eclipse Value Stream 19

2.3.1 Process Flow of The Eclipse Production Line 19

2.3.2 The Eclipse Value Stream Production Planning Team 22

2.4 Current Practice 22

2.4.1 Planning 22

2.4.2 Machine Capacity and Performance 23

2.4.3 Cycle Time and WIP Level 26

Chapter 3 Proposed Solutions to the Overall System 29

3.1 Root Causes to the high WIP and long Cycle Time Problem 29

3.1.1 Unnecessary Early Start in a Push-Pull and Pure Push System 29

3.1.2 Unsynchronized Production Flow 30
3.1.3 Rationales behind the Root Causes 30
3.2 Our Approaches to Reduce Cycle Time and WIP 31
3.3 Strategies for Needle Assembly (AN) and Packaging Machines 34

Chapter 4 Overall Scheduling Policy for AN 35

4.1 Literature Review for Scheduling Policies 35
4.1.1 Workload Regulating (WR) Policy 35
4.1.2 Drum-Buffer-Rope (DBR) Scheduling 35
4.1.3 Kanban Release Policy 36
4.1.4 CONWIP Release Policy 37
4.2 Proposed Scheduling Policies for AN 38

Chapter 5 Mixed Dispatching Rule for AN 39

5.1 Introduction 39
5.2 The Mixed Dispatching Rule for AN 40
5.3 Simulation for the Mixed Dispatching Rule 42
5.3.1 Simulation Model Description 42
5.3.2 Simulation Results for Two Dispatching Rules 45

Chapter 6 CONWIP Release Policy 49

6.1 Introduction 49
6.2 The Simulation Model 49
6.3 Simulation Results for CONWIP 50
6.3.1 The Effect of CONWIP 50
6.3.2 The Discussion of CONWIP Level 52
6.3.3 The Implementation of Proposed Strategies for AN 54

Chapter 7 Conclusion 56

References 58

List of Figures

Figure 1 A Luer-Slip Safety Needle with Syringe 14
Figure 2 How to Use an Eclipse Safety Needle 15
Figure 3 Eclipse Process Flow (Modified based on BD Internal Source) 19
Figure 4 Current Planning Practice 24
Figure 5 Representation of WIP Inventory on Flow Paths 26
Figure 6 An Example of the Proposed Planning Cycle 32
Figure 7 The illustration of 4 routes and production rate of each machine 39
Figure 8 SIMUL8 Model for AN and Packaging Machines 43
Figure 9 June Average WIP Level under Two Dispatching Rules 45
Figure 10 WIP Level of Each Working Day under Two Dispatching Rules 46
Figure 11 Average Cycle Times under Two Dispatching Rules 48
Figure 12 System WIP Lever under Different CONWIP Level 51
Figure 13 Average WIP in System vs. Different CONWIP Level 51
Figure 14 Average Cycle Time of 4 Major Product Types under Different CONWIP Level 52
Figure 15 Average Cycle Time Comparison before and after Using 2.5M CONWIP 54
Figure 16 Average Cycle Time Comparison of Proposed system and Current practice .. 55

List of Tables

Table 1 Available SKUs of Eclipse Safety Needle Production Line 17
Table 2 Demand Distribution 18
Table 3 Actual WIP Inventory Quantity and Cost 27
Table 4 Production Order for AN under the Mixed Dispatching Rule 41
Table 5 Production Order for AN under the Non-Mixed Dispatching Rule 42
Table 6 Simulation results for June average WIP level 46
Table 7 Working Days Needed to Finish Each Month's Order under different CONWIP Level 53
Table 8 Average WIP Comparison before and after Using 2.5M CONWIP 53
Table 9 June Average WIP Comparison of Proposed System and Current Practice 54

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

Chapter 2 will provide background information of for the company's operations and the cycle time and WIP problem in the Eclipse safety needle production line. Chapter 3 analyzes root causes of the problem and present an overall solution. Chapter 4 reviews the popular scheduling policies in literature and proposes for a mixed dispatching policy for the Needle Assembly machine for the Eclipse production line. Chapter 5 evaluates the effectiveness of the proposed dispatching policy for AN by using simulation. Chapter 6 discusses a CONWIP policy that can further limit the WIP and reduce the cycle time between AN and its downstream machines. Finally, conclusion is made in Chapter 7.

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).

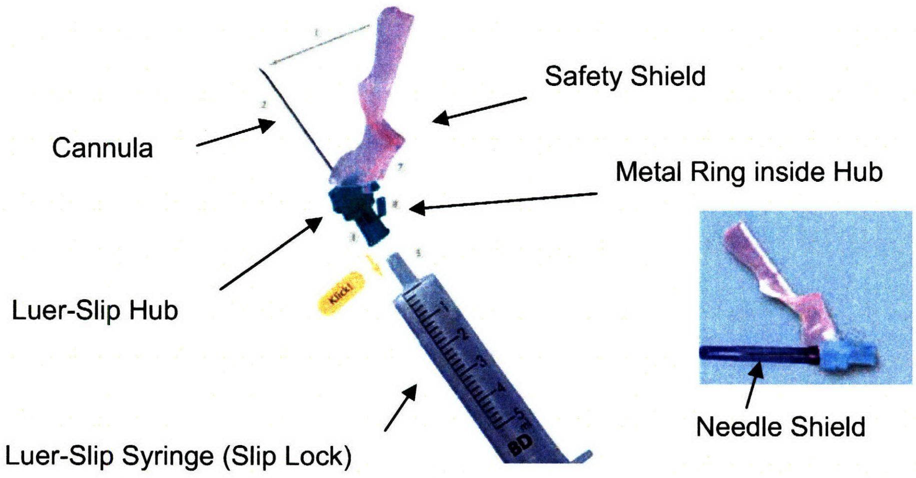


Figure 1 A Luer-Slip Safety Needle with Syringe

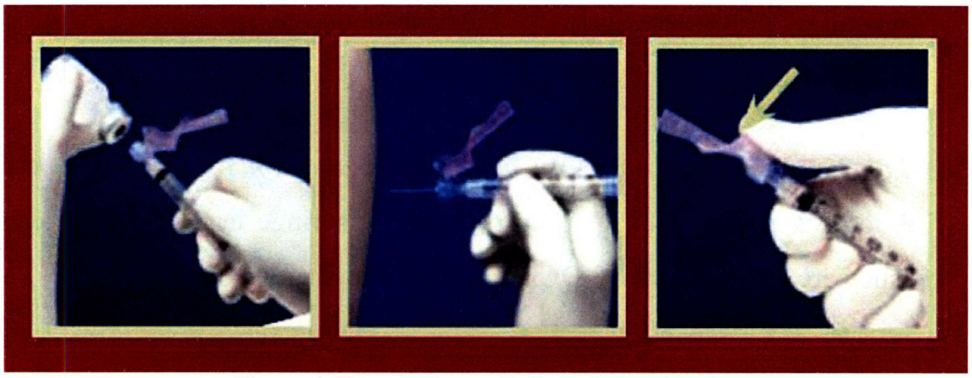


Figure 2 How to Use an Eclipse Safety Needle

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 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.

2.2 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.

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

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 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.

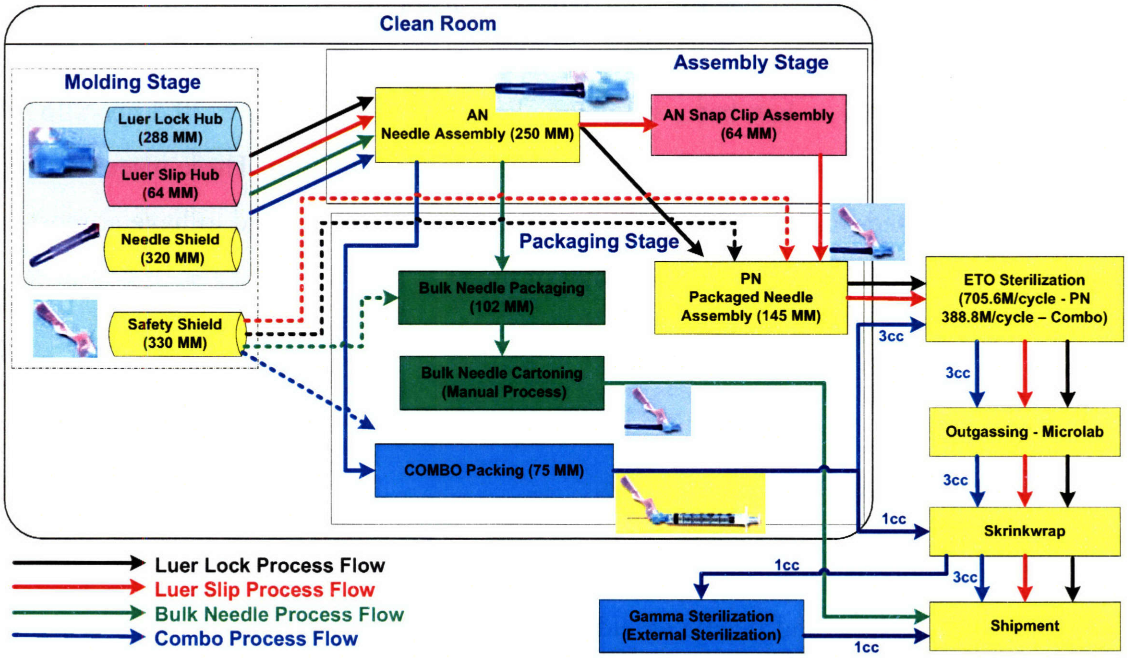


Figure 3 Eclipse Process Flow (Modified based on BD Internal Source)

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 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.

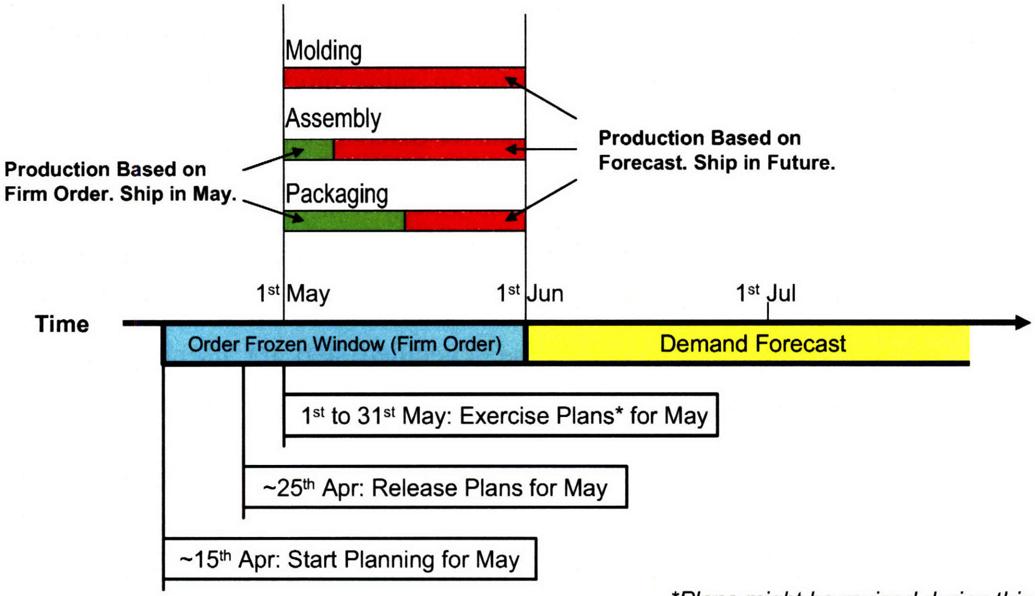
As AN is perceived as the bottleneck, the current planning strives to minimize the changeover on AN by producing large batches. Typically the production batch size of each SKU is chosen to be the order quantity for an entire month, which can go up to 5 million pieces for high demand SKUs. The transfer 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.

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



*Plans might be revised during this period

Figure 4 Current Planning Practice

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.

$$Demand_{machineA} = \frac{Q_{Final\ Products\ Route\ Through\ MachineA}}{\prod_{i=machineA}^{Last\ Machine} Yield_i} \qquad (1)$$

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.

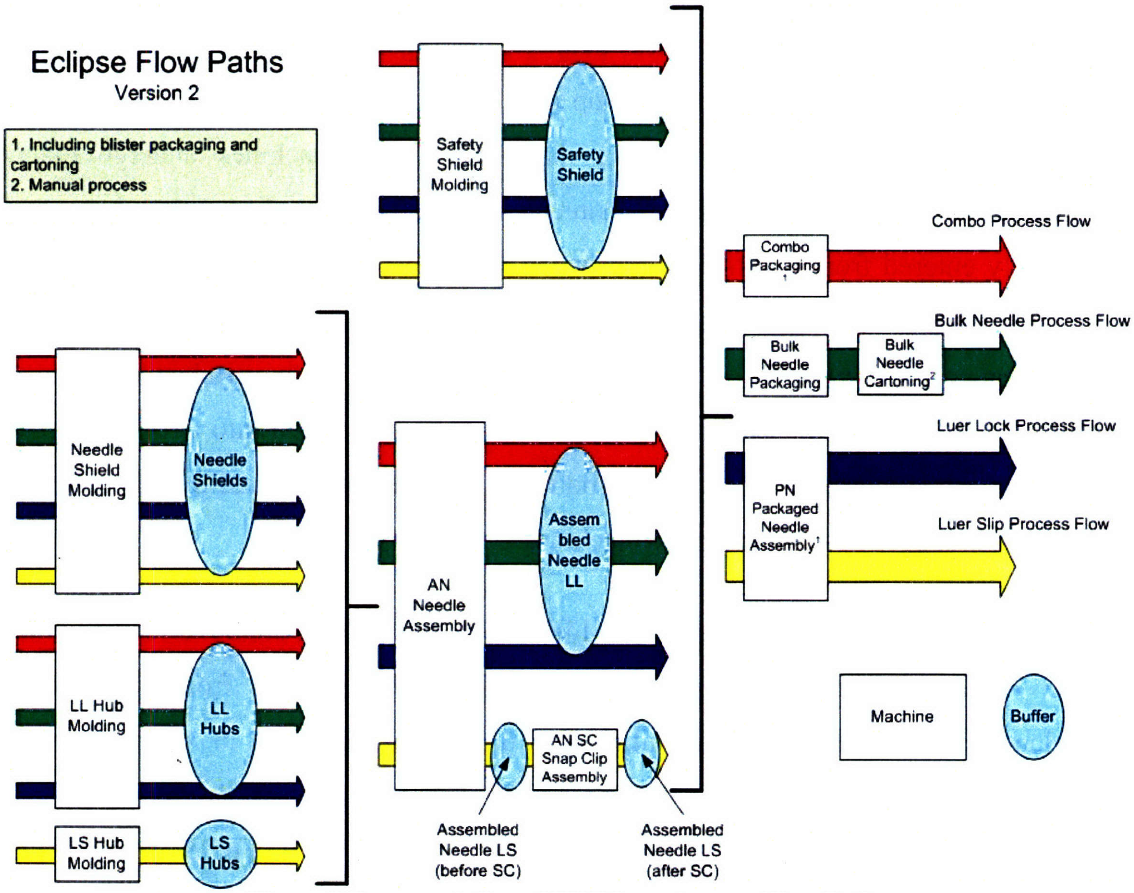


Figure 5 Representation of WIP Inventory on Flow Paths

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 WIP 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.

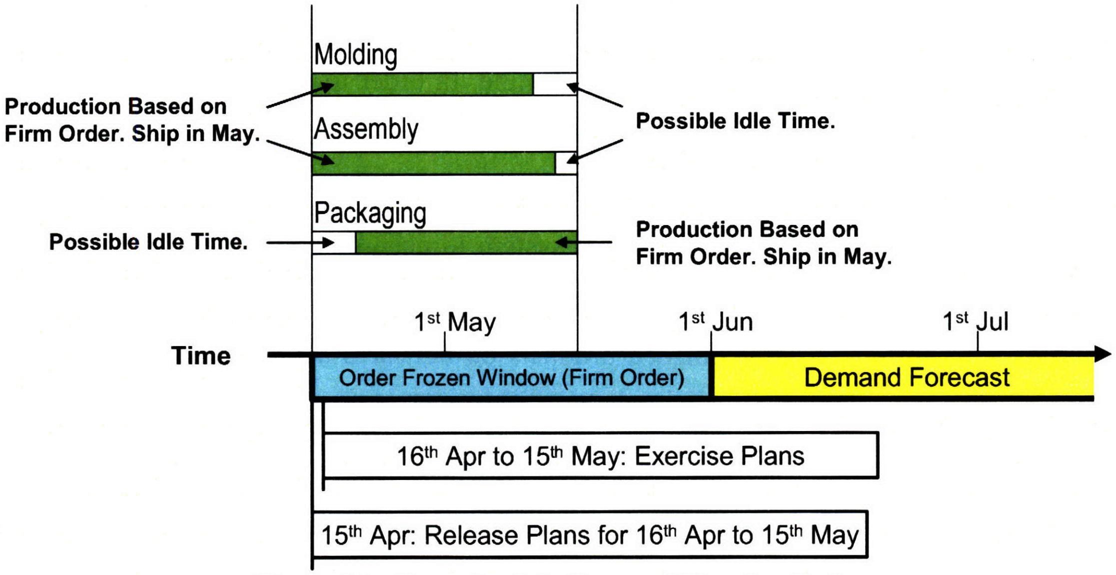


Figure 6 An Example of the Proposed Planning Cycle

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.

3.3 Strategies for Needle Assembly (AN) and Packaging Machines

Following the overall proposed solution presented, we will now discuss strategies for AN and packaging machines in details.

Firstly, in order to address the root cause of the problem, we eliminate the unnecessary early start of assembly machines as well as the extra delays between stations by scheduling only AN, and make downstream packaging machines process work once material from upstream is available, based on FIFO policy. The simulation results in the first thesis show that this approach can effectively reduce the average WIP level of June drop by 8 million, or 72.6%, from the case with early start and extra delays. Meanwhile, the average cycle time of the 4 major product types also have a reduction of 37%~86%. Thus this is a very effective approach to reduce cycle time and inventory between AN and packaging machines.

Secondly, we propose to reduce the transfer batch size to further improve the system performance. The simulation results in the first thesis show that by reducing transfer batch size from full transfer batch size to 200K, the average WIP level in June drops significantly by about 1.15 million, or 38%; and the average cycle times of 4 major product types reduces by 16%~54%. Thus it is recommended to adopt a 200K transfer batch size for implementation purposes.

Finally, to achieve a pure pull system with these machines, we analyze two scheduling strategies for the AN machine in details in the following chapters:

- Mixed Dispatching Rule
- CONWIP Release Policy

We also built simulation models to evaluate the effectiveness of these two proposed scheduling policies.

Chapter 4 Overall Scheduling Policy for AN

4.1 Literature Review for Scheduling Policies

Strategies for reducing cycle time and limiting work in process usually rely heavily on an effective input control and a priority sequencing mechanism. A rich literature exists in the context of job release and scheduling in a multi-station, multi-class queuing network. This section briefly discusses several popular approaches in the literature, and makes a recommendation on the methodology applicable to the current setting of BD's Eclipse safety needle production line.

4.1.1 Workload Regulating (WR) Policy

In a multi-station, multi-class queuing network, the decision involves when to release new jobs to the network, as well as how to sequence jobs at each machine in the network to meet the desired throughput rates and to minimize cycle time. Wein (1990, 1992) addresses the problem by developing an effective heuristic, the workload regulating (WR) policy [1].

The workload regulating (WR) policy outperforms the traditional input releasing and scheduling rules; however, the work release rules are complex and the sequencing rules are dynamic [2]. The resulting input and sequencing policies require the knowledge of the queue lengths of all stations at all times. The dynamic nature of the sequencing rule requires a highly sophisticated information and computation system; hence it may not be practical for implementation purposes. For this reason, we do not consider further the WR rule for this project.

4.1.2 Drum-Buffer-Rope (DBR) Scheduling

Another approach is Drum-Buffer-Rope, developed by Goldratt [3]. DBR is the Theory of Constraints production application, which assumes that there is one or, at the most, a limited number of scarce resources within the system across the aggregate product mix. The scarce resource is called a constraint or bottleneck and it is the limiting factor that determines the actual amount of throughput the system is able to achieve. Under DBR a 'drumbeat' is maintained for the rest of the plant by sequencing work to be done at the bottleneck. The Theory of Constraints focuses on protecting the constraint to best maintain the throughput of the system.

A potential dilemma exists because inventory has to be built before the bottleneck to keep it running, while at the same time limiting the overall amount of inventory. DBR breaks the dilemma by allowing the bottleneck to pull work into the system via a scheduling 'rope'. This limits the amount of inventory in the system because the non-bottleneck resources have sufficient capacity to ensure that they can quickly pass work on to the bottleneck. Thus in theory the queue only exists in front of the bottleneck. The queue or 'buffer' provides protection for the constraint as well as provides necessary information about upstream processes.

DBR finds its popular use where there is an identifiable bottleneck in the system. However, in BD's Eclipse safety needle line, the bottleneck is not unique. Depending on the demand distribution across the aggregated product mix, the bottleneck floats between the Needle Assembly (AN) machine and Packaged Needle Assembly (PN) machine. This also explains why the WIP level on the shop floor is high not only before AN, but also between AN and PN. Thus DBR is not well-suited for this setting due to the lack of a distinct bottleneck in the system.

4.1.3 Kanban Release Policy

Kanban is a pull production control system that uses simple, visual signals to control the movement of materials between work stations as well as the introduction of new material into the production system. Generally there are two types of kanban circulating between stations, namely, a production-kanban that dictates the need to produce more material and a conveyance-kanban that denotes the need to deliver material to the next station [4]. Parts are not allowed to be produced or moved unless authorized by kanbans.

Because each work station only produces and delivers parts when they are needed, there is no storage of excess inventory on the shop floor. Moreover, kanban limits the amount of WIP by acting as an authorization to produce more parts. Because the need for more parts or components stem from the customer order, the need for production or delivery is 'pulled' to the production line.

However, the kanban solution has a few drawbacks that limit its application in real world. Firstly, in product lines with many SKUs, as is the case in BD, it is impractical to keep standard containers for each SKU present. A possible consequence will be unnecessarily high WIP. Secondly, kanban is not well-suited in a system where frequent changeover is costly [5]. Thirdly, job sequencing has to be passed to the shop floor instead of the planning personnel. Sequences may need to be controlled when jobs have different priorities. For example, if kanban were introduced between every two adjacent work stations in BD, AN machine would receive kanbans from 4 different downstream routes for different product types. If these production kanbans arrive at AN together, the sequencing decision would have to be made on the shop floor.

A better alternative to kanban, a CONWIP system, is discussed in the following section.

4.1.4 CONWIP Release Policy

The CONWIP system possesses the benefit of pull, and has been successful in a wide variety of manufacturing environments. CONWIP limits total inventory on the shop floor by allowing job to be released into the system when another job departs. The sequence of job release is dictated by the backlog. Because the sequencing of work release is explicitly done, it is possible to incorporate the implication of changeovers into the planning process. Moreover, sequencing can be controlled by planning personnel in a CONWIP system whereas sequencing has to take place on the shop floor in the kanban case.

Spearman et. al. (1988) has shown that CONWIP results in lower WIP levels than a kanban system under the same throughput [6]. This is especially apparent with systems with a severe bottleneck. In a CONWIP system, WIP tend to accumulate at the bottleneck.

However in a kanban system, WIP at work stations upstream from bottleneck will be present at all times.

There are two schemes in setting the inventory target levels in CONWIP systems. The first one is to maintain a single WIP level for all product types, denoted as S-CLOSED. The second one is to set WIP levels for each SKU and to authorize the release of a new job of a particular SKU only when a job of that SKU has been completed, denoted as M-CLOSED. In this project, a S-CLOSED release policy is preferred because of the difficulty and complexity in determining the WIP level for each SKU to meet the desired throughput under an M-CLOSED case. In addition, raising the WIP level in an S-CLOSED system leads to an increase in total throughput. However, because WIP levels in M-CLOSED systems are separately maintained, raising the WIP level only increases throughput for the particular SKU but not necessarily the overall system throughput. Thus S-CLOSED has the advantage of more predictable product throughput [2].

4.2 Proposed Scheduling Policies for AN

For the complex problem setting involving multiple products and multiple routes, few insights are directly available. To obtain a reasonable and practical solution for BD's Eclipse line, we propose a mixed dispatching rule for the AN machine. This will be presented in Chapter 5.

Moreover, the S-CLOSED CONWIP system discussed in 4.1 has several advantages over other scheduling policies. It may not represent the optimal solution to the problem but it has been shown to produce reasonably satisfactory results in BD's setting. It is also desirable due to its ease of implementation amongst several rules. Thus the S-CLOSED CONWIP release policy is considered to further improve the system performance. It will be discussed in Chapter 6.

Chapter 5 Mixed Dispatching Rule for AN

5.1 Introduction

As described in Chapter 3, parts visit different packaging machines after AN depending on their product types. An illustration of the four routes and the production rates is shown in Figure 7, in which "VisualCenter" is a dummy work center with zero processing time that determines the route to take for the parts from AN. Red route represents for COMBO, blue for BULK, orange for LL-PN and green for LS-PN, respectively. The production rate of each machine is indicated in the red box. Although the combined processing rate (340K/shift) at the packaging stage is higher than AN (250K/shift), each individual packaging machine is slower than AN.

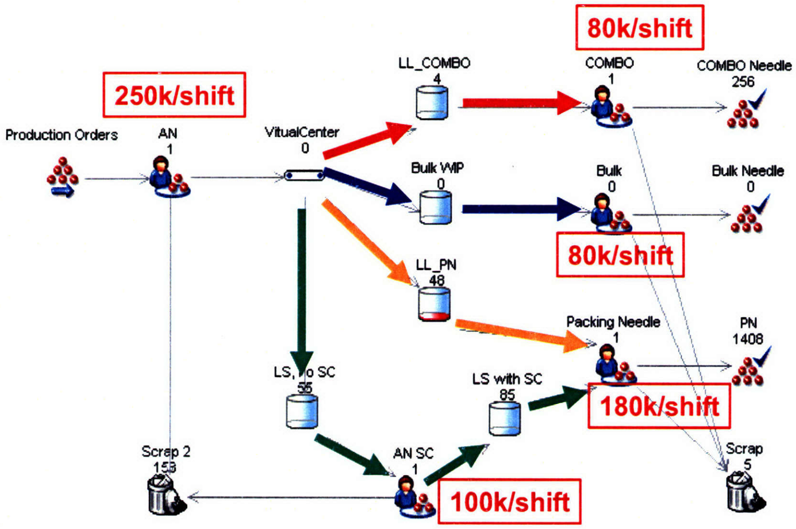


Figure 7 The illustration of 4 routes and production rate of each machine

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 major product type. In such a 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. For example, all LL-PN and LS-PN products need to route through the PN machine for packaging and they constitute 73% of the product mix. This leads to a high utilization in PN. When PN is already busy processing previous jobs, any new job has to wait even if the other two packaging machines - Combo and Bulk - are idle. One result from this simple scheduling policy at AN is a relatively long cycle time and large WIP.

The situation can be even worse if a simple scheduling rule is used in conjunction with a CONWIP system. For the previous example, if AN keeps working on LL-PN and LS-PN, the buffers between AN and PN build up, and blockage to AN might happen if the CONWIP level is reached.

We can alleviate this problem if we apply a mixed dispatching rule instead. A mixed policy 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 reduced WIP after AN.

5.2 The Mixed Dispatching Rule for AN

We have developed a mixed dispatching rule for AN to meet the monthly production order, based on the average monthly demand of FY 2006. Table 4 shows the production sequence. For each month, the assembly machine produces SKUs according to the release order in the first column. We set the production quantity equal to the demand for that month. For convenience, we denote this rule by “mixed dispatching rule” in this thesis.

Several guidelines are followed to derive the dispatching rule: the four major product types, LS-PN, LL-PN, COMBO, and BULK, are mixed in the release order; the release order allows each molding group to produce a hub type only once in a month; AN gets hubs from the three molding groups alternatively and starts with LS SKUs. A comprehensive analysis on such guidelines is performed in the third thesis.

Release Order	Product Type	Hub Size
PSN 20 x 1 TW (Luer-Slip)	LS-PN	LS 20G
PSN 21 x 1 (Luer-Slip)		LS 21G
PSN 21 x 1 1/2 TW (Luer-Slip)		
1 ml 25 x 5/8 Combo	Combo	LL 25G
3 ml 25 x 5/8 Combo		
3 ml 25 x 1 Combo		
PSN 25 x 1	LL-PN	
PSN 25 x 1 1/2		
3 ml 22 x 1 Combo	Combo	LL 22G
3 ml 22 x 1 1/2 Combo		
PSN 22 x 1	LL-PN	
PSN 22 x 1 1/2		
Bulk Needle 22 x 1 TW (ABG)	Bulk	
Bulk Needle 22 x 1 1/2		
PSN 23 x 1 (Luer-Slip)	LS-PN	LS 23G
PSN 25 x 5/8 (Luer-Slip)		LS 25G
PSN 27 x 1/2 (Luer-Slip)		LS 27G
PSN 30 x 1/2 (Luer-Slip)		LS 30G
PSN 18 x 1 1/2	LL-PN	LL 18G
3 ml 23 x 1 Combo	Combo	LL 23G
PSN 23 x 1	LL-PN	
Bulk Needle 23 x 1 MTW (ABG)	Bulk	
1 ml 27 x 1/2 Combo	Combo	LL 27G
1 ml 30 x 1/2 Combo		LL 30G
PSN 21 x 1 TW	LL-PN	LL 21G
PSN 21 x 1 1/2 TW		
3 ml 21 x 1 TW Combo	Combo	
3 ml 21 x 1 1/2 TW Combo		

Table 4 Production Order for AN under the Mixed Dispatching Rule

In comparison, a simple dispatching rule which follows the current production practice is shown in Table 5. In this rule we release the major product types one by one each month: LL-PN products are released first, followed by COMBO products, BULK products, and finally, LS-PN products. Although the hub types are mixed under this rule, this does not cause many changeovers to the molding machines because they start production much earlier than AN and store enough needle hub inventory to prevent AN's starvation, as described in Section 3.1. We use "non-mixed dispatching rule" to represent this simple dispatching rule.

Release Order	Product Type	Hub Size
PSN 25 x 1	LL-PN	LL 25G
PSN 23 x 1		LL 23G
PSN 22 x 1 1/2		LL 22G
PSN 21 x 1 TW		LL 21G
PSN 21 x 1 1/2 TW		
PSN 18 x 1 1/2		LL 18G
PSN 25 x 1 1/2		LL 25G
PSN 22 x 1		LL 22G
1 ml 25 x 5/8 Combo	Combo	LL 25G
1 ml 27 x 1/2 Combo		LL 27G
1 ml 30 x 1/2 Combo		LL 30G
3 ml 21 x 1 TW Combo		LL 21G
3 ml 25 x 5/8 Combo		LL 25G
3 ml 23 x 1 Combo		LL 23G
3 ml 22 x 1 1/2 Combo		LL 22G
3 ml 22 x 1 Combo		
3 ml 21 x 1 1/2 TW Combo		LL 21G
3 ml 25 x 1 Combo		LL 25G
Bulk Needle 22 x 1 1/2	Bulk	LL 22G
Bulk Needle 23 x 1 MTW (ABG)		LL 23G
Bulk Needle 22 x 1 TW (ABG)		LL 22G
PSN 20 x 1 TW (Luer-Slip)	LS-PN	LS 20G
PSN 27 x 1/2 (Luer-Slip)		LS 27G
PSN 30 x 1/2 (Luer-Slip)		LS 30G
PSN 23 x 1 (Luer-Slip)		LS 23G
PSN 21 x 1 (Luer-Slip)		LS 21G
PSN 21 x 1 1/2 TW (Luer-Slip)		
PSN 25 x 5/8 (Luer-Slip)		LS 25G

Table 5 Production Order for AN under the Non-Mixed Dispatching Rule

5.3 Simulation for the Mixed Dispatching Rule

To evaluate the effectiveness of the mixed dispatching policy to the production system, we built a simulation model using 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.

5.3.1 Simulation Model Description

Since this thesis focuses on the AN and packaging machines, we confine the simulation models to these machines, as shown in Figure 8.

The input of the model is the actual demand data of the past 9 months (Oct 2006 – Jun 2007). We compare the two scenarios - mixed and non-mixed dispatching rules – for

AN's monthly production. Under each scenario, we simulate transfer batch sizes of full transfer batch, 200k, 150k, 100k, 50k and10k.

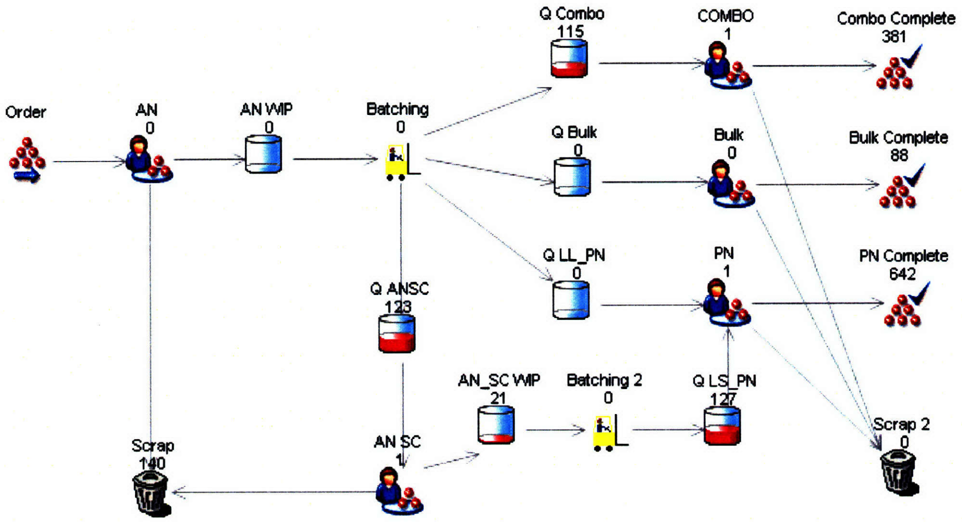


Figure 8 SIMUL8 Model for AN and Packaging Machines

We first state the assumptions for this model:

1) AN machine never starves, i.e. parts supplied from molding machines at upstream 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.

2) We eliminate the unnecessary early start of the assembly machine as well as the extra delays between stations by performing scheduling only on the AN machine. The schedule follows the real monthly order quantity in the 9 months. The downstream machines start to work as soon as the batch arrives at the front of the queue. The downstream machines do not wait except for the usual changeover time, 30 minutes on average, between two SKUs. .

3) The simulation run starts from an empty shop. The reason to do so is related to our proposed solution. In Section 3.2, we have discussed that the proposed solution produces next month's demand starting from a production cycle on the 16th of the current month. If we can satisfy demand, production should finish by the end of the

production cycle, and there should be no job left inside the clean room. Furthermore, this assumption has been verified by simulation in Thesis 1.

4) The simulation is run over a 9 months' time span: Oct 2006 – Jun 2007. There are 28 working days available each month, i.e. a working cycle is 28 days.

5) 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).

$$\textit{Expected Process Time} = \frac{\textit{Total Scheduled Production Time}}{\textit{Total Products Produced}\ (\text{in } 10\text{k})} \qquad (2)$$

6) 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.

7) 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 8, the work entry point named 'Orders' determines the type, quantity and sequence of work released to the AN machine. We set these 'Orders' based on the actual completed monthly orders in the period from Oct 2006 to Jun 2007. If AN completes the monthly order ahead of time, it stops operation and waits till the next month. On the other hand, if a shortfall takes place in a month, AN carries the unfinished workload to the next month.

The workstations 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 the finished packaged needles in the clean room. The numeric values marked with each object (machine or buffer) represent the quantity of work items in the current object.

Besides, 'Batching' and 'Batching 2' (with yellow forklift images) are two dummy machines with fixed processing time of zero, and they are created to hold partially completed batches. When the quantity of upstream WIP reaches a certain level, such as 200K, the batching machines transfer the WIP to downstream buffers. For full transfer batch, transfer batch sizes are the same as production batch sizes, with both set equal to the monthly demand for a certain SKU.

5.3.2 Simulation Results for Two Dispatching Rules

Under the mixed and non-mixed dispatching rules, we present in Figure 9 the simulation results, which show the average system WIP levels for June for different transfer batch sizes. The results show that the mixed dispatching rule can reduce system WIP regardless of the transfer batch size.

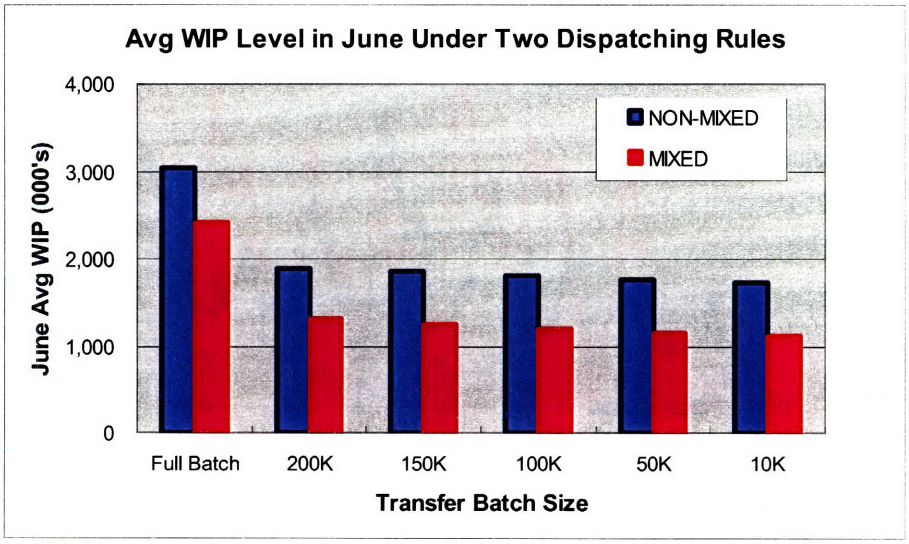


Figure 9 June Average WIP Level under Two Dispatching Rules

Table 6 summarizes the average WIP levels under the mixed and non-mixed dispatching rules. Under the transfer batch size of 200K, we can decrease the average WIP level in June by 31% (or almost 0.6 million) when changing from the non-mixed dispatching rule to the mixed dispatching rule.

	Full Batch	200K
NON-MIXED (K)	3,046	1,896
MIXED (K)	2,416	1,308

Table 6 Simulation results for June average WIP level

Figure 10 compares the WIP level between AN and Packaging machines under the two dispatching rules over the simulation period of 9 months, with the transfer batch size of 200K. Each point in the plot represents a daily ending inventory level.

As shown in Figure 10, we can reduce the WIP level in most months under the proposed dispatching rule compared to the non-mixed dispatching rule. This is due to the fact that under the mixed dispatching rule, the multiple downstream routes are utilized simultaneously. Within each month, WIP first builds up when AN starts processing work, then drops when AN completes the monthly quantity and stops, while the downstream machines are still processing the WIP.

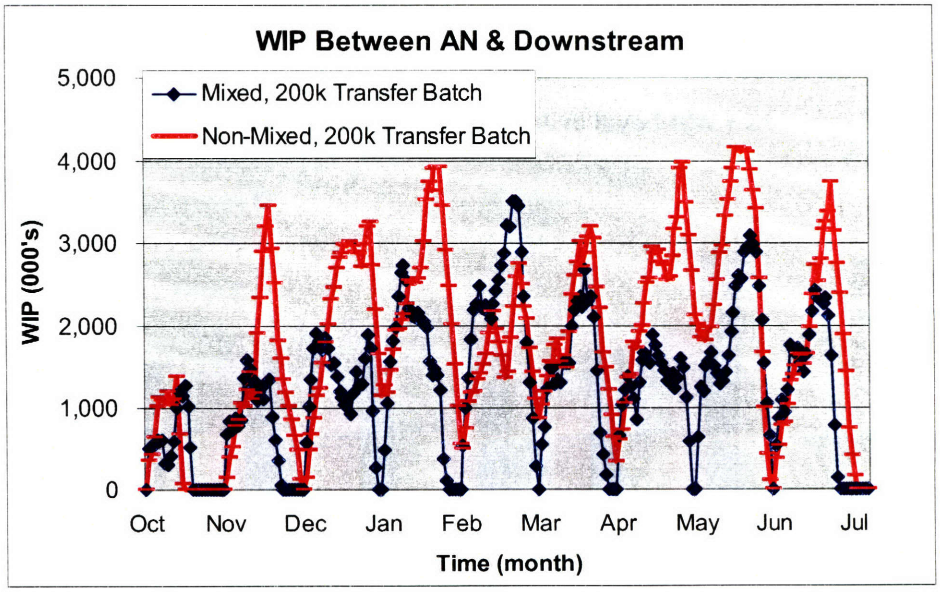
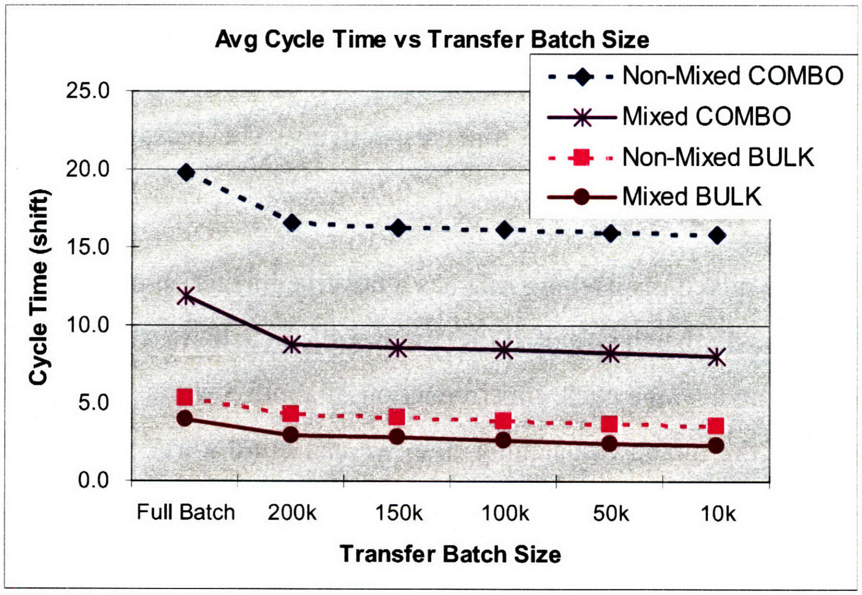


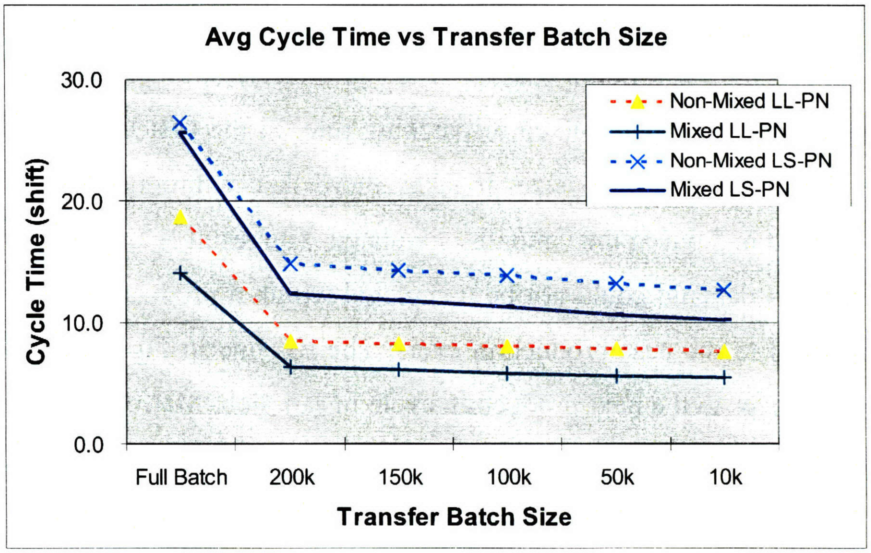
Figure 10 WIP Level of Each Working Day under Two Dispatching Rules

The most important contribution of this mixed dispatching rule is that it reduces cycle time and WIP in the system by feeding the different routes after the bottleneck (the AN machine) so that they will run in parallel. The significance of this is that we can satisfy a higher monthly demand with mixed dispatching rule. In the first thesis, we have shown that with 200K transfer batch size, it takes more than a production cycle to satisfy monthly demand in 5 out of the 9 months simulated. The result is also shown in Figure 10. When a month's demand could not be completed within a production cycle, the WIP is brought forward to the next production cycle. The red line in Figure 10 shows that the WIP is carried over to the next production cycle in Jan, Feb, Mar, Apr, and Jun. With the introduction of the mixed dispatching rule, we can complete the monthly demand on time in each production cycle. As a conclusion, a pure pull system (discussed in Section 3.2) is possible for the AN and downstream machines with the adoption of the mixed dispatching rule, elimination of early start and extra delay, and 200K transfer batch size.

Figure 11 compares the average cycle time for the 4 major product types under the two dispatching rules at various transfer batch sizes. From this figure we see that at all choices of the transfer batch size, the cycle time for each major product is less under the mixed dispatching rule. We can achieve a shorter cycle time with a smaller transfer batch size.



(a) COMBO and Bulk



(b) LL-PN and LS-PN

Figure 11 Average Cycle Times under Two Dispatching Rules

Comparing the mixed dispatching rule to the non-mixed rule, we see that the average cycle time of COMBO products drops more (about 8 shifts) compared to Bulk products (about 1 shift), as shown in Figure 11 (a). Although the COMBO and Bulk machines have the same production rates (80K/shift), the larger demand (23% of the total demand) of COMBO products, compared to Bulk products (4% of the total demand), makes the COMBO machine utilization higher than Bulk. Thus the mixed scheduling policy has a larger impact on cycle time.

From Figure 11 (b), the LL-PN and LS-PN products also show about 3 shifts' and 2 shifts' reduction in average cycle time respectively. The reason for this reduction is similar to the COMBO products.

In summary, the mixed dispatching rule has a positive impact on cycle time and WIP. For the proposed transfer batch size of 200K, we are able to achieve a WIP reduction of about 0.6 million and an average cycle time reduction of about 3 shifts. Most importantly, with the adoption of the mixed dispatching rule, we can build a true pull system within AN and packaging machines.

Chapter 6 CONWIP Release Policy

6.1 Introduction

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

As discussed in Section 4.1, this project considers 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.

Moreover, the CONWIP level plays an important role in the CONWIP system. An extremely low CONWIP level causes frequent blockage of the AN and hence a reduced 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.

6.2 The Simulation Model

To explore the effectiveness of the CONWIP release policy, we built one simulation model in SIMUL 8. This model is similar to the model described in Section 5.3.1, except that the total WIP in the system is controlled at the CONWIP level. When the CONWIP level is reached, AN does not start a new batch unless another batch completes production and departs from the system, otherwise it is blocked.

The model is based on the same assumptions as those shown in 5.3.1. We set the input data as the actual completed monthly orders in the period from Oct 2006 to Jun. We choose a transfer batch size of 200K and the mixed dispatching rule for AN in the simulation, because we have demonstrated that these tactics improve the cycle time (see Section 3.3 and 5.3).

We simulate CONWIP levels from 0.5 to 4 million, with an increment of 0.5 million.

6.3 Simulation Results for CONWIP

6.3.1 The Effect of CONWIP

Figure 12 shows the total WIP in system at each working day, under CONWIP levels of 0.5, 1.5, 2.5 and 4 million. From this figure we can see how the CONWIP release policy, takes effects. Take CONWIP level of 1.5 million (yellow line) as an example. Since the system starts from an empty shop, AN is blocked once the CONWIP level is reached; then we release work to AN only when some material finishes production and exits from the system. Systems operating at or near its full CONWIP level are represented by flat curves in the plot. For a system operating under a high CONWIP level, e.g. 4 million, it behaves as if there were no CONWIP because the target level will seldom be reached.

From Figure 12, we can also see that under a higher CONWIP level such as 4 million, the WIP level rises when AN starts production every month, and drops when the monthly orders are completed at AN, even though the downstream machines are still processing material.

However, under a low CONWIP level such as 0.5 million, it takes longer for the AN to finish the orders for the month. In fact, the orders can not always be completed by the end of each month. The reason is that the throughput under 0.5 million CONWIP level (equivalent to only 2.5 transfer batches) is rather low because releases to AN are frequently blocked.

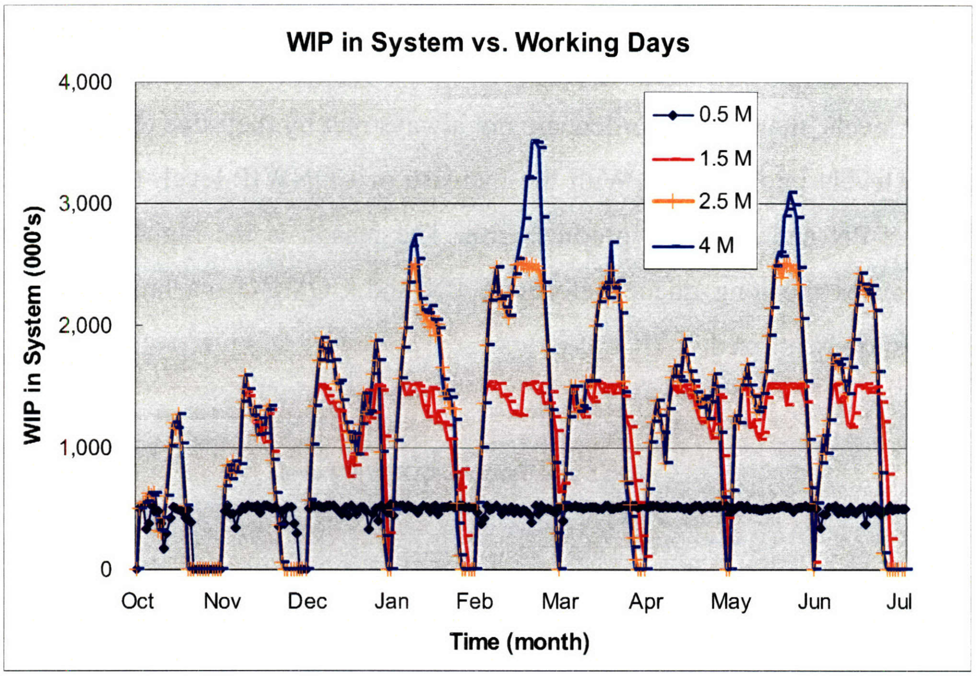


Figure 12 System WIP Lever under Different CONWIP Level

Figure 13 shows the average WIP level in the system over the 9 months under different CONWIP levels. With the increase of CONWIP level, the average WIP level in the system also rises. Although the CONWIP level is pre-set, it is not always reached in the real system.

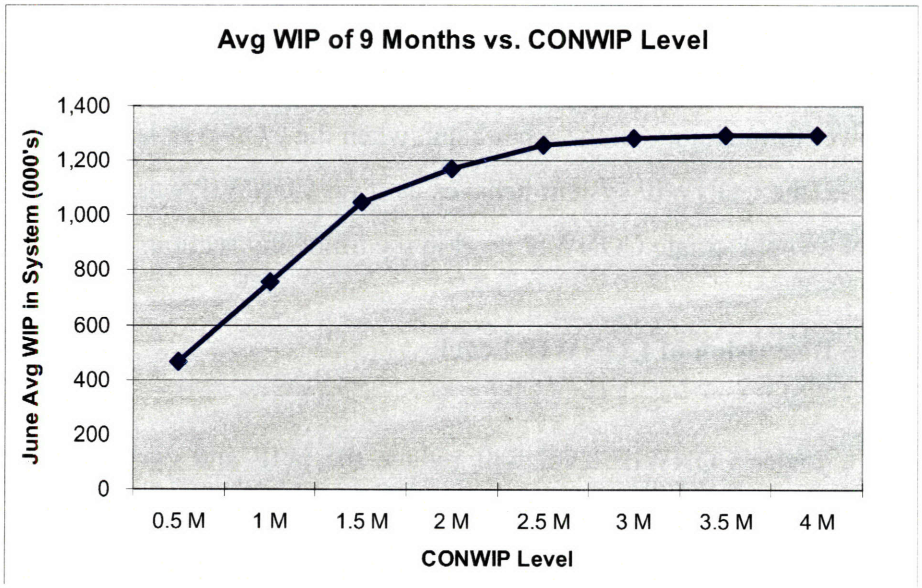


Figure 13 Average WIP in System vs. Different CONWIP Level

With lower CONWIP levels, there is less WIP in the system and less queuing time, and hence a shorter cycle time for each SKU. The average cycle time of the 4 major product types under different CONWIP levels is shown in Figure 14. Since under 0.5M and 1 M CONWIP levels, the monthly orders are not always met by their due date, we do not take these two levels into account. With the increase of CONWIP level, the cycle times for LL-PN, LS-PN and COMBO products rise. The reason is the higher demand of each product causes the long queue before the PN and COMBO machine, while the Bulk machine is lowly utilized.

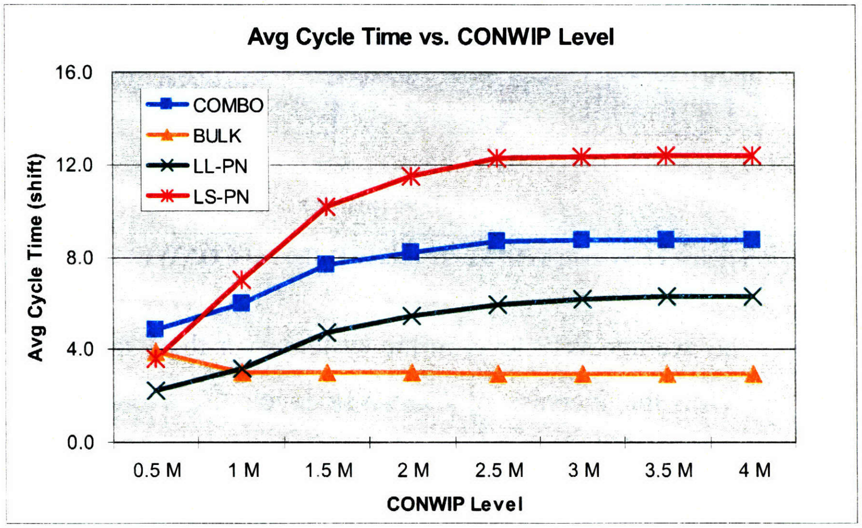


Figure 14 Average Cycle Time of 4 Major Product Types under Different CONWIP Level

In summary, lower CONWIP levels results in less WIP and a shorter cycle time in system but also lower throughput. On the other hand, when the CONWIP level is raised beyond a certain value, the CONWIP system behaves as if no CONWIP was used. We discuss the decision for an appropriate CONWIP level in the following section.

6.3.2 The Discussion of CONWIP Level

Although a lower CONWIP level will reduce the WIP and cycle time, it has to be properly set so that the desired monthly demand can be met for each product.

Table 7 summaries the work days required to complete the orders in each month under different CONWIP levels. In assumption (4) of our simulation model, 28 working days are available for production each month. When the required work days exceed the available work days in a month (28 days), customer orders cannot be met by their due date and their delivery will be delayed. This is highlighted in orange in Table 7.

	0.5 M	1 M	1.5 M	2 M	2.5 M	3 M	3.5 M	4 M
Oct, 06'	18	16	16	16	16	16	16	16
Nov, 06'	25	20	20	20	20	20	20	20
Dec, 06'	>28	>28	27	26	26	26	26	26
Jan, 07'	>28	>28	25	23	23	23	23	23
Feb, 07'	>28	>28	>28	>28	27	27	27	27
Mar, 07'	>28	>28	>28	26	25	24	24	24
Apr, 07'	>28	>28	27	26	26	26	26	26
May, 07'	>28	>28	>28	>28	27	27	27	27
Jun, 07'	>28	>28	25	23	23	23	23	23

Table 7 Working Days Needed to Finish Each Month's Order under different CONWIP Level

Under CONWIP levels of 0.5 million and 1 million, most months have delayed order completion because of the relatively lower throughput of AN. Under the CONWIP level of 2 million, only the two months with highest demand cannot complete orders on time. To meet the monthly demand requirement in all months, the minimum CONWIP level is 2.5 million. It also means when CONWIP levels is larger than 2.5 million, we can have a pure pull system within AN and downstream machines.

Table 8 compares the average WIP level over the 9 months' period in a system without CONWIP and a system with a CONWIP level of 2.5 million. Both cases already incorporate the proposed solutions from previous discussions: early start and extra delay elimination, mixed dispatching rule and transfer batch size of 200K. The result shows that the 2.5 million CONWIP system has 3% less WIP on average compared to the system without CONWIP.

	System without CONWIP	2.5M CONWIP System
Avg WIP in system (K)	1,292	1,255

Table 8 Average WIP Comparison before and after Using 2.5M CONWIP

Meanwhile, Figure 15 shows the comparison of average cycle times before and after using 2.5M CONWIP. We can see that the 2.5M CONWIP can reduce the average cycle times a bit.

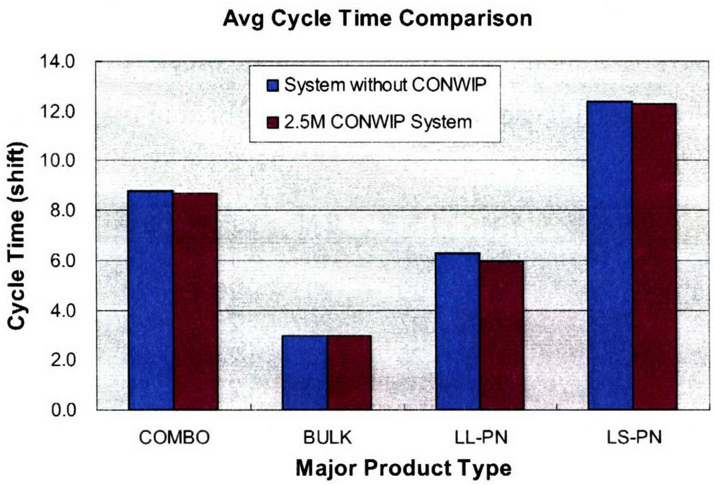


Figure 15 Average Cycle Time Comparison before and after Using 2.5M CONWIP

6.3.3 The Implementation of Proposed Strategies for AN

To reduce cycle time and WIP between AN and packaging machines, 4 feasible strategies are proposed:

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.

To prove the effectiveness of these strategies, we use two simulation models. One is the simulation model of the current practice, presented in the first thesis, with early start and extra delays present between stations, non-mixed dispatching rule and full transfer batch. The other is the 2.5 million CONWIP system presented above, with early start and extra delay elimination, mixed dispatching rule and transfer batch size of 200K.

Table 9 shows that we can reduce the June average WIP in system by 88% (9.8 million) after the 4 proposed strategies are incorporated.

	Real Case	Simulation Model of Current Practice	Simulation Model of Proposed System
Avg WIP in June (K)	~14,000	11,108	1,308

Table 9 June Average WIP Comparison of Proposed System and Current Practice

Figure 16 shows the average cycle time comparison between the proposed model and current practice. The results are obtained from simulation. The proposed model gives greatest cycle time reduction of 54 shifts, or by 90%, for LL-PN products because of their highest demand. LS-PN products also benefit significantly, reducing by 80% (49 shifts) from the proposed strategies. The reason is that they pass one more machine (AN SC) and thus experience the longest cycle time.

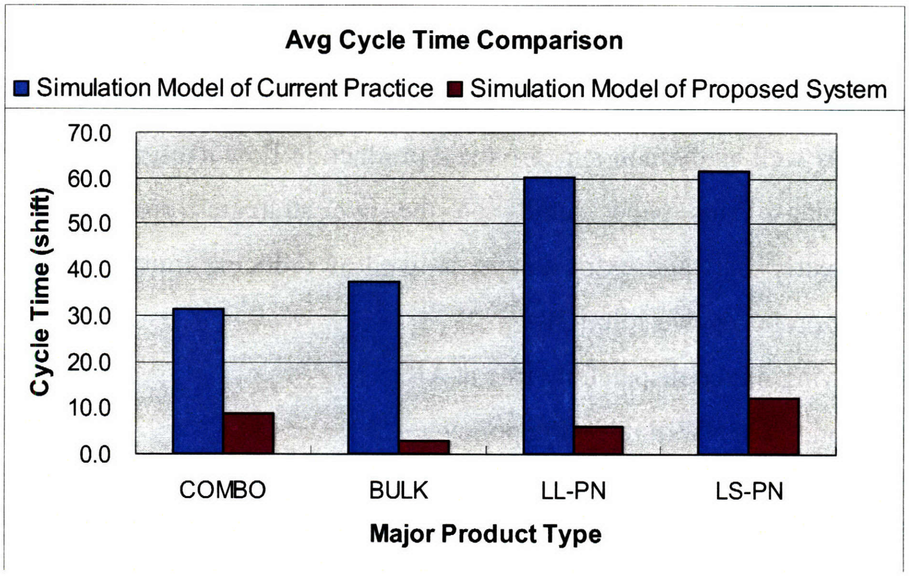


Figure 16 Average Cycle Time Comparison of Proposed system and Current practice

In summary, the four proposed strategies for AN and packaging machines could effectively reduce the WIP and cycle time from the current practice.

Chapter 7 Conclusion

This project examined the high cycle time and WIP problem in the safety needle production line from a manufacturing system and product scheduling perspective. The overall system was divided into three parts and analyzed using different approaches: Needle Assembly machine (AN) and its downstream packaging machines, Hub molding machines, and Needle Shield (NS) and 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 non-synchronized production flow caused by the over-detailed production plans. This paper builds on the two strategies proposed by Thesis 1: Eliminating early start and extra delays, as well as reducing transfer batch sizes. This paper further proposes two solutions:

- Applying mixed dispatching rule at AN
- Adopting CONWIP releasing policy

We demonstrate that the mixed dispatching rule has an advantage over the non-mixed rule from the current practice by enhancing the total throughput of the 4 production routes after the bottleneck (AN machine). With the adoption of the mixed dispatching rule, we show by simulation that we can have a pure pull system within AN and downstream machines. In such a system, the average cycle time could be reduced by 3 shifts and average WIP level could be reduced by 0.6 million.

The CONWIP release rule further reduced WIP by controlling the total amount of inventory between AN and packaging machines. The simulation results show that under the CONWIP level of 2.5 million, the system is able to meet monthly demand requirement and reduce the average WIP by 3%.

Overall, the implementation of the four proposed solutions can reduce the cycle time effectively by 7~18 working days, and reduce the WIP by 9.8 million, or 88%, in comparison with the current practice.

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] I. Duenyas, "A Simple Release Policy for Networks of Queues with Controllable Inputs" Operations Research, 42 No:6, j(1994): 1162-1171.

[3] E. M. Goldratt, "The Goal: A Process of Ongoing Improvement" New York: North River Press, 1992.

[4] M. Bertolini, M. Bevilacqua, A. Grassi, 2005, "Advanced Manufacturing Control Systems: a Simulation Comparative Analysis",

[5] M. L. Spearman, D. L. Woodruff, W. J. Hopp, "CONWIP – A Pull Alternative to Kanban", INT. J. PROD. RES., 1990, VOL. 28, NO. 5, 879-894

[6] M. L. Spearman, M. A. Zazanis, 1988, Push and Pull Production Systems: Issues and Comparisons, Technical Report 88-24, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL.