Optimizing Lot Sizes and Establishing Supermarkets in a Multi-Part, Limited-Capacity Manufacturing System

by

Arjun Subramani Chandar

Bachelor of Science in Mechanical Engineering and Business, Economics and Management,
California Institute of Technology, 2013

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 2014

©2014 Arjun Subramani Chandar. All rights reserved.

The author hereby grants MIT permission to reproduce and distribute publicly paper and electronic copies of this thesis document, in whole or in part, and to grant others permission to do so.
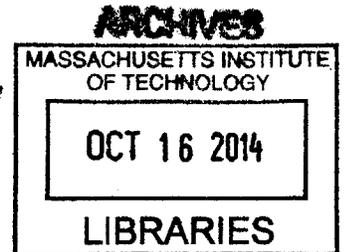
Signature redacted

Author...

Arjun Subramani Chandar
Department of Mechanical Engineering
August 15, 2014

Signature redacted

Certified by.....................................

Stanley B. Gershwin
Senior Research Scientist
Thesis Supervisor

Signature redacted

Accepted by.......................

David E. Hardt
Ralph E. and Eloise F. Cross Professor of Mechanical Engineering
Chairman, Committee for Graduate Students

This page is intentionally left blank.

Optimizing Lot Sizes and Establishing Supermarkets in a Multi-Part, Limited-Capacity Manufacturing System

by

Arjun Chandar
Bachelor of Science in Mechanical Engineering and Business, Economics and Management,
California Institute of Technology, 2013

Submitted to the Department of Mechanical Engineering on August 15th, 2014,
in partial fulfillment of the requirements for the degree of
Master of Engineering in Manufacturing.

Abstract

This thesis addresses the value of optimizing lot sizing to meet part demand within the limits of machine capacity, focusing on a method for improving productivity within the CNC turning and CNC milling departments at the Waters Corporation Machining Center in Milford, Massachusetts. A detailed study of the machining center revealed problems with low machine utilization in turning and milling, low on-time delivery performance and a need for day-to-day adjustments to the production schedule. These problems were attributed to inefficient data collection and use of data, poor production scheduling, and lot sizing that the system's capacity could not handle, causing frequent occurrence of redundant part setups and enabling delays in turning and milling which cascaded to downstream processes. This thesis addresses the latter problem by implementing optimized lot sizing for ten selected part types going through the turning and milling departments; the system design called for increases to the lot sizes of five of these part types. In order to prevent increased lot sizes from causing unforeseen problems in downstream processes, the project further implemented a supermarket for these selected parts at the end of milling operations, in order to decouple turning and milling from other processes.

The lot sizing methodology focused on parts going through a turning operation followed by a milling one, with selected part types being machined on one of two machines in each department. In order to limit increases to work-in-progress inventory caused by increased lot sizes, the supermarket was designed to be managed by a Kanban-based pull system using a modified (Q, R) inventory policy with an expected weekly service level of 95%. Over the course of the implementation period, the lot sizing methodology saved an estimated 36.75 hours of setup time for parts made in 407.25 productive hours of run time. Moreover, a simulation over a year-long period estimates that with the new lot sizes, the machines in question will achieve an aggregated increase in productive hours of nearly 10%.

Thesis Supervisor: Stanley B. Gershwin

Title: Senior Research Scientist

This page is intentionally left blank.

Acknowledgements

I would like to give special thanks to everyone who contributed to the success of this thesis and this project:

To begin, I would like to thank my project partners, AJ Perez and Greg Puszko, for working as an excellent team with me for the past six months, and for weathering various kinds of adversity to produce a result we can all be proud of.

I would next like to acknowledge our supporters at MIT. Thanks to our advisor, Dr. Stanley Gershwin, for encouraging us to always be skeptical, and for providing insightful guidance while allowing us the freedom to explore our chosen avenues for our projects. Thanks to Professor David Hardt for his mentorship and leadership throughout the MEngM program and to Dr. Brian Anthony for signing off on bringing me to MIT. Thanks to Jose Pacheco for his career advice and program administration during my time here. Finally, thanks to Jennifer Craig for helping to make this thesis a coherent, organized document.

Next, I would like to express gratitude to the people at Waters Corporation who encouraged this project. Thanks to David Terricciano, Jim McPherson and Dan Welch for giving us the opportunity to work at Waters and encouraging us to pursue the projects of our choice. Thanks to all of the people at Waters, from the finance department to planning to the machine shop, whose excitement about our project and willingness to be open allowed us to make informed decisions; in particular, thanks to Wayne Wentworth for his insightfulness and enthusiasm. And, of course, thanks to our project supervisors, Gabriel Kelly and Kerry McNamara, for their constant support, encouragement, and diligence in granting us the necessary resources to be successful.

Finally, I would like to thank all of the friends and family who have been invested in my success since the beginning, at MIT, at Caltech, back home in Miami or anywhere around the world. Thank you for inviting me over for lunch, helping me with my most difficult decisions, or making me laugh during my most stressful days. In particular, thanks to the people I grew up with: my brother and sister, whose own success keeps pushing me to be better; my grandmother, who is still going strong as she reaches the age of 90; and last but not least, my parents, who have always done whatever it takes to let me follow my desired path in life.

This page is intentionally left blank.

Table of Contents

9

List of Figures

List of Tables

Definition of Key Terms

Note: manufacturing environments tend to use these terms to refer to different concepts, so these definitions should be considered only within the context of this thesis.

Bottleneck- Machine with the lowest production rate in a production process, limiting the production rate of the process as a whole.

Capacity- Available hours for manufacturing operations, including time for setup, operations, maintenance and repair. Total capacity is the sum of capacity of each machine in a department. For the purposes of this thesis, the capacity for each machine considered is 135 hours per week.

Cycle Time- Average length of time between completion of two successive units in a process.

Lot/Batch- Both refer to a group of parts being produced together, and may be used interchangeably for the purposes of this thesis.

Machining Time/Process Time/Operation Time/Run Time- All refer to the total amount of time required to complete a lot or batch of parts in a production process, including productive operations but excluding setup time.

Machine Utilization- Percentage of time that a machine is performing productive operations.

Manufacturing Lead Time- Total time from the moment an order is placed until it is delivered to the customer. Depending on the department, the "customer" may refer to the product's end user, the distribution department, the finished goods inventory buffer or the next department in a part's production process.

Part- A single unit produced during a production process.

Part Family- A group of part types with similar setups and process plans. The setup time of transitioning from one part type in a family to another is much less than that of transitioning to non-family part types in the system.

Part Type- A set of (theoretically) identical parts repeatedly produced over time, referred to with the same part material number or SKU number.

Productive Hours- Time spent performing productive operations, such as cutting, polishing or finishing. Excludes hours devoted to setup, maintenance and repair of machines.

Setup- Set of steps taken to prepare a machine for a production run, such as a tooling change, fixture change, material change or machine calibration. Unlike productive hours, setup hours can be considered fixed with respect to lot size.

Stock-Keeping Unit (SKU)- Waters Corporation terminology referring to independent part types, used interchangeably with "part type" in this thesis.

Utilization- Fraction of available time that parts are being produced. For the purposes of this thesis, this is measured by fraction of available time that a machine is up and its spindle is drawing power.

1   Background and Project Motivation

1.1   Introduction and Statement of Purpose

The purpose of this thesis is to demonstrate the value of implementing optimized lot sizing and supermarkets for high-volume parts in a manufacturing environment, and to describe a practical method for doing so implemented at Waters Corporation in Milford, Massachusetts.

This thesis is based upon a project conducted by Alfonso Perez [1], Greg Puszko [2] and the author, representing the Massachusetts Institute of Technology, at Waters Corporation between February and August of 2014. Portions of the first chapters of each of these theses will be similar due to the authors' collaboration throughout the project.

As part of its operation, Waters Corporation manufactures high-precision, high-performance liquid chromatography parts and part families at its Machining Center in Milford. In 2013, the Milford facility reported approximately $18 million worth of internal accounting credits from manufacturing. The heads of global manufacturing and finance set 2014 production targets for the Milford manufacturing facility to approximately $21 million of internal accounting credits from manufacturing. The overarching purpose of the project was to determine a scalable method to increase productivity or otherwise provide value to Waters through manufacturing system improvements. Additionally, Waters management sought a continuous improvement plan to help increase the Milford manufacturing facility internal accounting credits from $18 million to $21 million.[1] These objectives were achieved through a series of improvements in the two primary upstream departments in the facility, CNC turning and CNC milling. The improvements focused on reducing total setup time for ten part types: through optimized lot sizing and a supermarket, as explained in this thesis; through improved scheduling policies [2]; and through implementation of an automated data collection system [1].

---

[1] Personal communication, Vice President of Global Manufacturing, Waters Corporation, March 14th, 2014.

The lot sizing methodology saved an estimated 36.75 setup hours in turning and milling over the course of 407.25 hours of productive run time for a four-week period, and is estimated to generate additional productive hours equal to about 10% of existing productive hours if maintained over a year (see Chapter 7). Moreover, the supermarket decoupled the turning and milling operations from downstream processes and is expected to provide further benefits in terms of inventory control and improved on-time delivery.

1.2   Background Information on Waters

Waters Corporation is an analytical instruments company that develops test equipment used in pharmaceutical, industrial, and academic research laboratories. Their two main product divisions are their biochemical and chemical analysis division, based in Milford, Massachusetts, and their physical testing division, based in Manchester, England and Wexford, Ireland. The biochemical and chemical analysis division produces liquid chromatography instruments, which comprise about a $6 billion global market and are the largest source of revenue for Waters[2], as well as high-end mass spectrometry instruments. The physical testing division produces thermal analysis, rheology and calorimetric instruments. Each division develops and manufactures the standalone products as well as all the consumables, chemicals, and accessories to feed or support their particular instruments. Waters also maintains a global support network of authorized service centers around the world that manage local installation services, training, technical support, repair, and replacement part services.

The manufacturing system improvements described in this thesis pertain to the Machining Center at the chemical analysis division based in Milford, although variations of the lot sizing methodology can be implemented in other production systems.

---

[2] Personal communication, Director of Strategic Sourcing, Waters Corporation, January 28th, 2014.

1.3   Waters' Expansion into Contract Manufacturing

Before 2001, 100% of the production and 100% of the assembly operations for products in the chemical analysis division were done in the Milford facility. In 2001, Waters began expanding their production of components to local and overseas contract manufacturers, primarily in Singapore. This expansion, however, was done without downsizing the production in the Milford facility. The objective was to use contract manufacturers in order to manufacture and complete the less critical, low-value-added components and operations, while the Machining Center in Milford would focus on the critical, high-precision, high-value-added parts and operations. This was done in order to increase the total capacity of the chemical analysis division and generate the supply necessary to meet growing demand for new products while still maintaining the same quality and control over manufacturing processes. Today, 85% of production is done by local and overseas contract manufacturers, while 15% is done in Milford. Since 2006, Waters has realized a growth of nearly 1000% in sales revenue in the chemical analysis division due to this expansion into contract manufacturing[3]. The Milford facility still operates most of the assembly steps for their products, including all major assembly operations.

While the Machining Center in Milford focuses its operations on about 1500 distinct products (internally referred to as stock-keeping units or SKUs), they have the ability (with the machines they have in the facility) to produce most of the components used in their products that are currently being fabricated by contract manufacturers. Parts are selected to be made either in-house or by a contract manufacturer based on their cost to produce, the quality/tolerances required for the part, the raw material (all parts produced in the Milford facility are metal), and the current stage of the product life cycle of the instrument using the part.

Under Waters' current system, there is a constant shifting of components from buy (having the component produced by contract manufacturers and then purchased by Waters), to make (producing the component in-house at the Milford facility), and vice versa. This shift occurs due to a variety of reasons: for example, a newer product line that has grown out of its initial release production and requires higher output to meet demand would see a shift from make

---

[3] Personal communication, Vice President of Global Manufacturing, Waters Corporation, March 6th, 2014.

to buy; a product at the end of its life would move to a make-to-order policy for its components (a shift from buy to make); a product which may be experiencing quality issues would induce Waters to investigate the problem or take stronger control of its production (buy to make); or management would deem a part cheaper to outsource due to its uniqueness (make to buy).

Management wishes to increase the revenue per share for the organization through increased output of the machining center, and it plans to use this ability to shift from buy to make in order to do so. Waters wants to increase the finished-goods output of the Milford facility from 2013 by 18% in 2014, without increases in labor or capital or decreases in on-time delivery to customers. This increase in finished goods production (and conversely, the increase in revenue) from the machining center will be realized through a combination of increased demand for liquid chromatography products and buy-to-make shifting of products from contract manufacturers back to the Milford facility. The desire for this shift forms the basis for the productivity improvement goal specified at the outset of the project.

1.4   The Waters Advanced Manufacturing Center

The Waters Advanced Manufacturing Center houses a 50,000 square-foot Machining Center of Excellence which produces 2.7 million parts annually covering 28,000 SKUs, a 29,000 square-foot Advanced Instrument Assembly & Accessory Kitting Operation facility which produces over 130,000 finished goods assemblies, spare parts, and accessory kits, and an 8,500 square-foot Class 10,000 Clean Room for optics, micro valves, and critical parts. The project of interest focused on the operations of the Machining Center of Excellence (referred to as the "Machining Center" or "machining center"), which produces precision-machined metal components for the final assembly of instruments and consumables.

17

Figure 1.1: The Waters Machining Center of Excellence. The photo above comprises the column cell and model shop, about one-third of the total machine shop space.

The Machining Center currently operates for 24 hours per day, 6 days per week, with their standard for full machine utilization being 22.5 hours per day, 6 days per week[4]. The Center is divided up in a job shop format, where machines that are of the same type (lathes, NC mills, lapping machines, etc.), or produce a very specific family of components (i.e. a certain line of consumables for an instrument) are grouped together in the same location. The main departments in the job shop are: CNC turning; CNC milling; the valve cell (which produces a line of check valves used in the pumps for the liquid chromatography instruments), which consists of lathes, mills, EDM wire drillers, and lapping machines; and the column cell (which produces a line of consumables for the LC instruments), which consists of specialized CNC lathes with automatic long-stock feeders. Each of these departments also has its own utility area, consisting of deburring machines and simple cleaning machines used to perform secondary operations on parts produced in the department. The Machining Center also houses a model shop which maintains its own CNC and manual mills and lathes, a fused deposition modeling 3D printer, and micro-machining capabilities, but the model shop is used by the New Product Division in Waters to prototype parts currently undergoing development, and it not used for the production of products to customers.

---

[4] Personal communication, Head of Machining Center, Waters Corporation, June 24th, 2014.

18
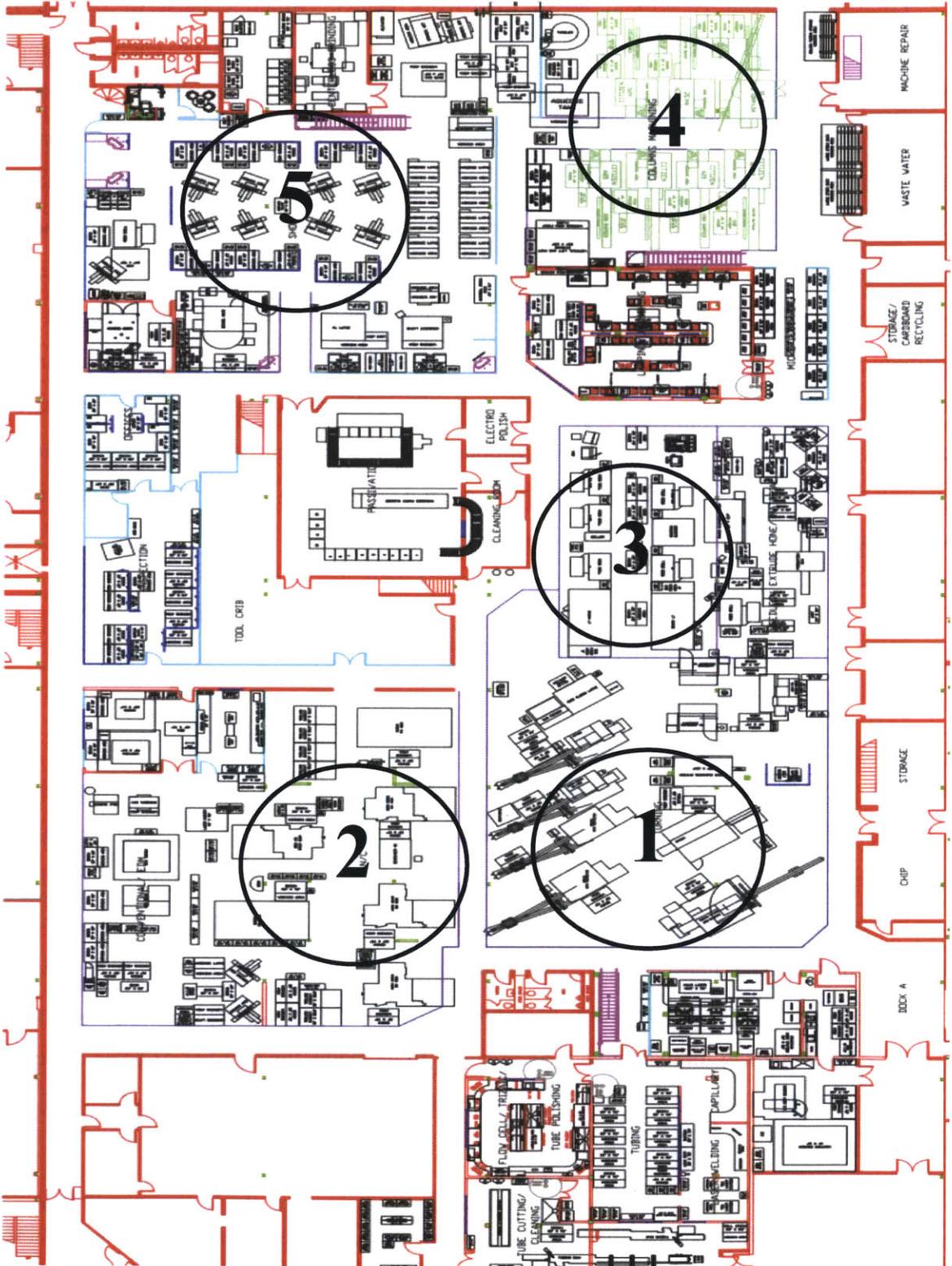
Figure 1.2: Schematic of Machining Center.
1 = CNC turning, 2 = CNC milling, 3 = valve cell, 4 = column cell, 5 = model shop.

Each department is broken down into work centers which consist of one or more machines. When performing operations or undergoing setups or teardowns, each machine will be staffed by one machinist, who is not tied to a particular machine but instead is able to run any machine within his or her department (with a few exceptions). Directly supervising the machinists in each department are the section leaders. Along with the typical duties machinists have of operating the machines and producing parts, section leaders develop the production schedule for their department on a day-to-day basis based on the requirements of the Schedule and Planning (SAP) system, assign each job to a particular machine and machinist in the department, and run debriefing meetings at the end of each shift. Each department will have one or two section leaders working per shift. Directly above the section leaders are the department supervisors, in charge of managing the operations of their specific department and making sure that production is on schedule. The supervisors are not tasked with machining, but work directly with the machinists and section leaders on a day-to-day basis.

1.5   Typical Flow of Production and Parts through Machining Center

A majority of parts produced by the Machining Center will first go through the turning department, so consider the example of a typical part flowing through turning. The Schedule and Planning (SAP) system Waters uses plans out the production schedule for all parts produced by the Machining Center. The start of production for a certain part will be triggered by the SAP system either when the inventory level of a made-to-stock part is expected to drop below some minimum value (based upon expected demand), or a made-to-order part is placed into the system. This will trigger the creation of a job, which is physically represented by a process sheet. This process sheet indicates what SKU needs to be made, the number of parts to be made, the start date of the job, the expected end date of the job (to stay on schedule), and all the operations the part must undergo before being completed.

When the job first begins, the process sheet will be delivered to the department where the first operation will take place. The section leader will take all the process sheets at the beginning of the day and determine which parts should be completed on what machine, with what operator, and in what order. When production of a particular part is to begin, the machinist will take the

process sheet, go to his or her assigned machine, and set up the machine for the particular job, which can take anywhere between 10 minutes and 8 hours.

The machinist will then pull the raw material for the job from the raw materials inventory, positioned in a central location in the turning department, and will load the raw material into the machine and set the machine to begin the operation. A job or machine may or may not need constant supervision; this is usually dependent on the raw material that is placed into the machine. For example, long stock will utilize an automatic feeder so that the machine will output the finished part when it is complete, and then autonomously reset the stock into the chuck and continue. On the other hand, pre-cut blanks have to be individually loaded into the chuck and then unloaded after the operation is finished. Typically, if a machine or job does not require constant supervision, the machinist will use this time to set up or perform basic maintenance on other machines in their department.

The machinists are tasked with inspecting the finished parts coming off the machine and determining which are good and which do not fall within the specification ("nonconforming" or "scrap" parts). They will mark these numbers (of good and bad parts) on the process sheet. Nonconforming parts are marked as scrap, or "red-tagged". The supervisor will then also inspect the red-tagged parts, and determine whether or not they are acceptable, need to be reworked, or must be abandoned. Depending on the part, this basic inspection by the machinist will take place either after each part comes out of the machine, or will be put off until the entire batch is complete. Along with this basic inspection, most parts also undergo a critical inspection at some point during their production, and this step will be patently included on the process sheet.

After a batch is complete, the machinist will deliver the parts and the process sheet to the utility area. In self-contained departments such as the valve cell or column cell, the operators themselves are responsible for delivering batches to downstream operations. However, for the higher-volume, higher-mix areas of turning and milling, the only person who moves the parts out of the utility area is the utility operator, who is in charge of the utility area for the particular department. The utility operator will input the completed operation and how many good and scrap parts were produced into the SAP system. From here, the process sheet will indicate whether or not a part needs to undergo a secondary operation such as a deburring process or a

simple cleaning. If it does, the utility operator will perform these operations and then deliver the finished parts and process sheet from the utility area to the next department, where they sit in an incoming goods buffer. If, however, a part does not need a secondary operation, then the finished parts and the process sheet will still be delivered by the utility operator from the utility area to the next department, where they sit in an incoming goods buffer. In this way, the utility operator can be thought of as another machine in the production line (see Figure 1.3). The process sheets for those parts will then be placed into the stack of incoming process sheets for that department, and the section leader will decide when the part should continue its operations, on what machine, with what operator. This process continues until the part is complete.
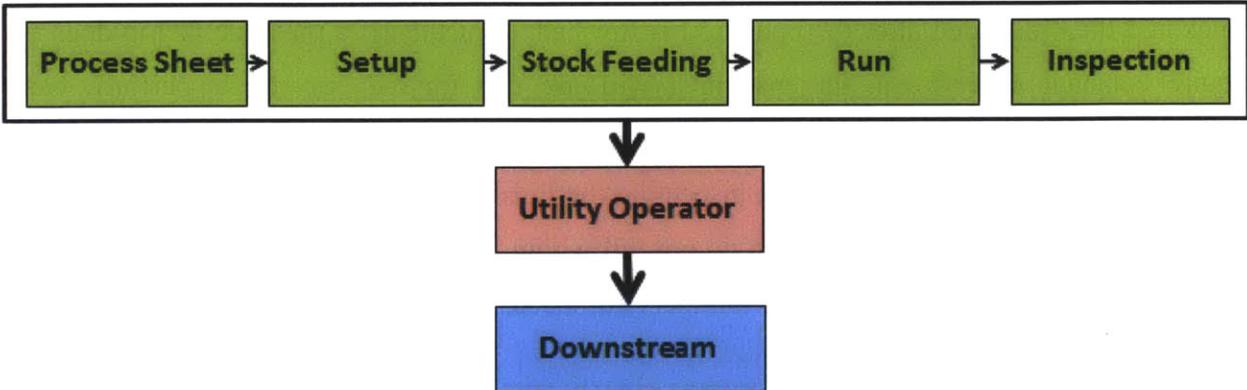
Figure 1.3: Map of operation steps for production process in CNC turning and CNC milling. Consider initial operation, utility operator station, and downstream as three separate machines.

1.6   Problems with the Waters Manufacturing System

The production flow as described in Section 1.5 creates several problems which can hamper Waters' ability to meet its target of an 18% increase in productivity. Issues can start at the beginning with inaccurate demand forecasting, creating either a surplus or a backlog of parts or raw material inventory and forcing day-to-day changes to the production schedule output from SAP. Since section leaders are ultimately responsible day-by-day for prioritizing parts, machines and machinists, adjustments to the SAP schedule will be unsystematic, subject purely to section leaders' intuition rather than to written protocols for optimizing production. Over even a short period of a week or less, this can result in conflicts in the scheduling of parts. Requiring all parts in turning or milling to go through their respective utility operator prior to leaving essentially

places them at a bottlenecked machine for up to 16 hours, since each department has only 1 utility operator working an 8-hour shift. Moreover, production scheduling is contingent on accurate data being input to SAP about how long it takes to make parts, and any inaccuracies can cause discrepancies between order schedules and the reality of day-to-day needs in the Machining Center. Finally, the production of parts can require long machine, fixture or tool setups, limiting the amount of time machines can be operating to produce parts. Sizing production lots to have too many setups over a given period will cut down productivity.

1.7   Thesis Outline

Chapter 2 will cover initial investigations and the resulting problem statement and diagnoses. Chapter 3 will review the manufacturing systems literature that informed subsequent methodologies for improvements. Chapter 4 will describe that methodology for optimizing lot sizes, and Chapter 5 will describe the methodology for setting up and sizing a supermarket after milling operations. Chapter 6 will discuss the logistical details and challenges of implementing these changes within the Waters Machining Center. Chapter 7 will show the results of the implementation as well as a simulation estimating results over a year of implementation. Chapter 8 will conclude the thesis and make recommendations for continuing the project, as well as recommendations for future work.

2   Problem Statement and Diagnoses

Upon the project team's arrival at Waters Corporation, the original problem formulation was to improve the flow of parts in the shop to aid in increasing productivity from $18 million to $21 million annually. The major functional requirements of the project are detailed in Table 2.1 below; chief among them is the fact that improvements had to be made without adding machines or floor space and without adding labor, and that any implementation would be scalable (and therefore reproducible) beyond the time constraints of this thesis.

Table 2.1: List of client functional requirements from Waters.

Client Functional Requirements	Priority Level (1=highest)
No new machines/floor space	1
Limit line downtime during implementation	1
Payoff; demonstrate break even point, NPV, cost of capital	1
Low risk profile	1
System-wide, scalable, expandable	2
Management: $18M→$21M 2014 goal (~18% capacity increase)	2
Reproducible	3
Fits culture	3

Following discussions with project coordinators and Waters executives, the team agreed that this could be best accomplished beginning with a factory analysis, which was conducted between February and May of 2014.

2.1   Factory Analysis Methodology

Analysis of the factory involved three distinct steps. It began with an analysis of existing production data logged in the Scheduling and Planning (SAP) system; however, due to the steep learning curve necessary to use SAP, this method did not provide immediate insights into the flow or scheduling of parts. Instead, it revealed inconsistencies and issues with data fidelity. SAP analysis did, however, ultimately provide the primary basis for determining the demand of all parts moving through the Machining Center.

After that initial step, factory analysis involved conducting interviews with key stakeholders at all levels of the company. This included, among others, the heads of manufacturing operations, demand planning, and finance, machining department supervisors and section leaders and machinists on the shop floor. The bulk of the characterization of the factory came about due to the results of these interviews.

The final method used for factory analysis was personal observation of practices in the Machining Center, which provided some clues into the root causes of observed problems.

2.2   Overview of Problems Discovered

The factory study revealed a number of problems in the Waters production system, from inaccurate demand forecasting[5] to poor communication and sharing of practices between departments.[6] However, within the scope of the Machining Center and time constraints of the project, the team considered the following problems most critical:
1. Low machine utilization.
2. Low on-time delivery of parts to assembly.
3. Significant day-to-day adjustments made to the production schedule.

---

[5] Personal communication, Head of Master Demand Scheduling, Waters Corporation, March 21st, 2014.
[6] Personal communication, Manager of Shipping and Receiving, Waters Corporation, April 11th, 2014.

2.2.1 Low Machine Utilization

Through interviews with the supervisors of various departments, the project team found that although Waters finds itself unable to meet its production demands with its current capacity, many of its machines suffer from utilization (percentage of time spent making parts; see Chapter 3.1) considerably lower than the company target of 80%.[7] Power monitoring data demonstrated that the problem of low utilization was particularly glaring in CNC turning and CNC milling, where utilization rates for machines producing high-volume parts were below 70% and below 50%, respectively. Figure 2.1 provides an example from July, 2014, in milling:
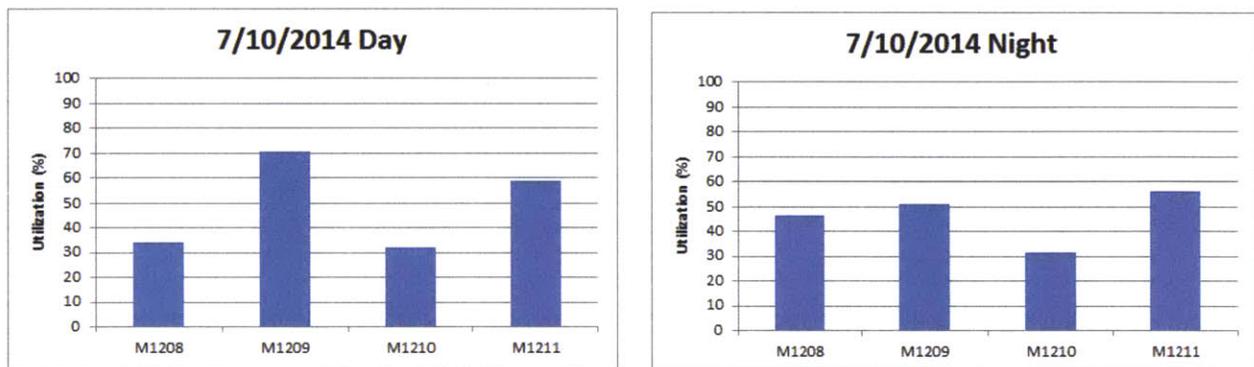
Figure 2.1: Machine utilization for four work centers in CNC milling for July 10th, 2014. Aggregate utilization of the machines for that week was about 52%.

2.2.2 Low On-Time Delivery of Parts to Assembly

Interviews with machine shop and assembly supervisors revealed a consensus that on-time delivery rates for all parts are much lower than desired between the machining area and the downstream process of assembly. The machine shop seeks a service level of 95% to assembly, as measured by on-time delivery of available inventory. However, those percentages, when broken down by department, are consistently under 60% and, in some weeks, have dipped below 20%. Again, this problem is especially prevalent in the main upstream processes of turning and milling, as Table 2.2 demonstrates for the second quarter of 2014.

---

[7] Personal communication, Supervisor of CNC Turning and Milling, Waters Corporation, March 21st, 2014.

Table 2.2: On-time delivery statistics for CNC turning and CNC milling for the second quarter of 2014[8].

On-Time Delivery	Turning		Milling	
Q2	OTD	Avg. Days Late	OTD	Avg. Days Late
week 14	35%	4.65	17%	8.22
week 15	14%	5.86	5%	8.10
week 16	17%	5.39	15%	8.90
week 17	8%	7.57	18%	11.15
week 18	16%	7.22	17%	11.17
week 19	17%	6.82	12%	12.19
week 20	10%	7.50	13%	13.72
week 21	26%	5.45	5%	12.84
week 22	22%	5.12	15%	9.71
week 23	30%	6.13	22%	7.73
week 24	34%	5.08	33%	7.07
week 25	32%	4.52	24%	5.54

This slow upstream delivery translates to overall low on-time delivery to assembly.

2.2.3 Significant Day-to-Day Adjustments to Production Schedule

The team learned through staff interviews and monitoring of daily production schedules that the schedule output by the SAP-generated demand forecast changes on a daily basis. As orders come in and are classified as high-priority, production of previously scheduled parts gets moved back and put behind schedule. The same problem occurs when a previously completed job needs rework due to high amounts of nonconforming parts, or in the case of machine failures or other unforeseen delays.

---

[8] Personal communication, Machining Center Head of Financial Accounting, June 24th, 2014.

2.3   Diagnoses of Problem Causes

The factory analysis showed the problems detailed above had many random causes, such as tool failures or sudden orders, but that several systematic causes existed as well:

1. Inefficient data collection and use.
2. Poor production scheduling and lot sizing.
3. Inability to produce to demand within capacity constraints due to redundant part setups.
4. System-wide delays originating with delays in turning and milling.

2.3.1   Inefficient Data Collection

The first issue noticed during the factory study, and one that exacerbates the others that follow, is the lack of efficient data collection and use in the Machining Center and elsewhere at Waters. The project team learned through staff interviews that many legacy data collection practices found in the finance and accounting department cause issues in production planning. The finance and accounting staff indicated that company policy and practice is to perform a cost roll (review of all cost measures for accounting purposes) once per five years. The data collected during the cost roll are average process cycle time, average machine tool set up time, and average machine tool change over time. These data are collected manually for a limited set of part types and processes using handheld stop watches as timing mechanisms. The data collected are then used to calculate the expected number of machinists needed and used to perform cost analyses for additional machines and shop space. This cost roll methodology not only records data company-wide too infrequently, but makes broad comparative assumptions. For example, the finance department assumes that the process time and setup times are the same for a specific part family (set of part types with similar setups and process plans).

The machine shop requires more up-to-date data and so records improvements to setup and cycle times as they are made; however, this information is not fed back into SAP for the benefit of financial planning, order scheduling or machine or labor allocation. Moreover, one of the universal issues addressed by most staff is that the machine shop data on these processes is collected manually by machinists and fed into a Filemaker program only accessible on a few

computers. In order for a machinist to capture machine utilization data, he or she must power down the machine. This is an issue on the production floor because the collection of utilization data requires the machine to go down for a minimum of 1 minute per production run, leading to unnecessary stoppage of production. This stoppage costs Waters in terms of machinist and department time, which Perez estimates to be greater than $85,000 annually [1].

In order for the utility operator or machinist to record data on the completion of a process on a batch of parts, he or she must manually record the SKU number, quantity of parts completed, and the process performed. This process is prone to human errors in the form of time delays (the operator waited too long to input data) and incorrect data input (for example, the operator recorded information on the wrong SKU).

The final issue with the existing data collection methodology is the lack of feedback it provides toward decision-making. Setup and cycle times are just lumped into product lead time rather than being used to improve scheduling or reorganize part flow to maximize machine utilization. Since collected data is not being used at all, it is unclear to some people on the shop floor what the useful purpose of collecting it is.

As a result, one long-term project stemming from the factory analysis involved finding an efficient, automated data collection system [1] which could be used to accurately develop methodologies such as the author used for this thesis. In order to understand, model, and optimize the Waters production floor it is necessary to know the exact production factors that govern each step of each process. In an ideal system, a manufacturing engineer would know how the production factors correlate to individual machinists in order to prioritize the production plan. A description of necessary production factors for long-term systems engineering is provided by Perez [1]. For the purposes of lot sizing, the project team centered its efforts on collecting process time and setup time data and using it to inform the lot size calculations.

2.3.2 Poor Production Scheduling and Lot Sizing

Production Planning staff indicated that company policy is to take quarterly forecasts and divide by the number of weeks to smooth out demand evenly, which increases setups and does not account for the reality espoused by demand schedulers[9], who claim that demand always spikes near the end of quarters. The Production Planning group also works under the assumption of infinite daily production capacity (in Waters' case, the true capacity of a machine is the amount of time a machine is available during the week, which is 22.5 hours a day, 6 days a week), which means it schedules orders beyond the ability of the machine shop to produce them and calculates lot sizes for parts independently of other parts going through any given machine. This causes severe issues because SAP does not prohibit the "time machine effect" from occurring: immediate orders have pre-programmed lead times, and if an order's due date relative to the time of order comes earlier than the recorded lead time, the order gets registered as having been made in the past. This causes the production floor to fall behind schedule before even receiving orders[10]. Furthermore, research indicates that the safety stock numbers in SAP used to calculate inventory thresholds for legacy SKUs are often ignored by the planning department[11], which can cause an abundance of obsolete inventory for old parts or a backlog of orders for high-volume parts. This leads directly to the need to alter the production schedule daily to manage crises, as well as low on-time delivery. Moreover, the lack of all-encompassing scheduling in SAP lowers machine utilization because it does not take advantage of parts with similar setups [2] or size lots appropriately.

2.3.3 Redundant Part Setups given Capacity Constraints

The problems resulting from the infinite capacity assumption intensified several years ago when Waters instituted a company-wide lean manufacturing initiative. Lot sizes for high-volume parts, in particular, were reduced in order for to reduce inventory according to the "just-in-time" manufacturing model [3]. However, reducing lot sizes increases the total number of

---

[9] Personal communication, Head of Master Demand Scheduling, Waters Corporation, March 21st, 2014.
[10] Personal communication, Acting Head of Production Planning, Waters Corporation, April 4th, 2014.
[11] Personal communication, Head of Procurement, Waters Corporation, March 21st, 2014.

machine, tool and fixture setups necessary to run the same number of parts, since a part will require a new setup and teardown of old setups every time it gets produced (see Chapter 3.1 for a description of setups). This caused existing machines to overrun their capacity to handle both setups and production of parts[12] [13].

In addition to the capacity problem, this practice decreases production of those parts per unit time. Moreover, although Waters pushed to reduce inventory, interviews with executives in various departments indicated that no one tracks the cost of holding inventory prior to its finished good stages. Given that high-volume parts produced in Milford generally have low obsolescence risk (liquid chromatography machines still sell parts 20 to 25 years after first release), even conservative estimates of holding costs would be much lower than the costs of performing extra setups, since temporarily holding more inventory for high-volume parts only costs Waters in terms of the space it takes up. Thus, although inventory has a cost and adding inventory goes against accepted lean manufacturing dogma, Waters can benefit from taking on additional inventory for high-volume part types in order to reduce setup time.

The lot sizing methodology described in this thesis, therefore, sought to adjust lot sizes to account for the capacity of the machines being investigated and reduce total setup time.

2.3.4 Delays in Turning and Milling Cascade to System-Wide Delays

As indicated in Chapter 1, CNC turning and CNC milling are the first two upstream processes in the manufacturing of nearly all of Waters' critical SKUs. These two departments also tend to feature the longest setup times for batches of parts and tend to be the busiest machines in the shop, which causes them to bottleneck the manufacturing process for all parts (see Chapter 3.1 for a description of bottlenecks). These upstream delays result in reduced on-time delivery further downstream in the manufacturing process; in fact, delays in turning and milling exacerbate downstream delays due to the "bullwhip effect" [4] (see Chapter 3.2).

---

[12] Personal communication, Supervisor of CNC Turning and Milling, Waters Corporation, May 23rd, 2014.
[13] Personal communication, Section Leader, CNC Turning, Waters Corporation, June 19th, 2014.

2.4 Problem Statement Summary and Team Roles

Based upon the issues mentioned in this chapter, the team worked to help Waters improve productivity through reduction of setups, decoupling of turning and milling from downstream processes, and improved production scheduling, all tracked by an improved and automated data collection system. Perez was primarily responsible for monitoring the logistics of data collection [1]; Puszko was responsible for improving production scheduling [2]. The author of this thesis focused on productivity improvements related to adjustment of lot sizes and decoupling of lot sizes between turning, milling, and downstream processes.

2.5 Selection of Departments and SKUs

The project aim was to realize an improvement in productivity to help Milford increase annual output from $18 million to $21 million, and to produce an implementation which could be scaled to incorporate the entire shop. As a result, the team targeted one area of the shop and one small subset of parts produced, enacting changes scalable to other SKUs or departments.

For shop location, the CNC turning and CNC milling areas were selected. The reasoning behind this decision was twofold. For one, the turning and milling areas are most subjected to the problems the team noticed in the shop: high setup times resulting in long lead times for orders, low machine utilization, and low on-time delivery of parts to customers or to the assembly area.

For example, consider a Waters flow cell, 405008774, whose process sheet is shown in Table 2.3. Waters' SAP standards indicate a total of nine hours in its production cycle are spent with turning and milling machine, tooling and fixture setups, with all other operations combined totaling only one setup hour. In addition, machining process times are much higher in turning and milling than downstream, indicating that, given a single order for the part type, downstream machines can get starved waiting for turning and milling to finish operations:

32

Table 2.3: Process sheet for part 405008774. Both setup and machining times are much higher for turning and milling compared to other operations.

Material	OpAc	Operation short text	Work ctr	Std Setup	Std Lbr/Mc	WrkCtr Q Tm	Std. queue time
405008774	0010	CNC TURN	M1302	7	0.473	4	0
	0030	N/C MACHINE	M1210	2	0.625	0	0
	0035	N/C SECONDARY AND DEGREASE	M1204	0	0.01	4	0
	0040	MICRO DEBURR'	M1100	0	0.08	4	0
	0050	WEDM	M1005	1	0.129	4	0
	0060	IINSPECT	M9999	0	0	8	0
	0070	PASSIVATE	M1101	0	0.02	4	0
405008774	0080	CRITICAL CLEAN	M1105	0	0.005	4	0

In general, the CNC turning and CNC milling areas are regarded by machinists, executives, Production Planning and demand schedulers as the bottlenecks of most manufacturing processes within the shop. Furthermore, turning and milling comprise initial operations for the vast majority of SKUs Waters produces in its Milford plant, so these delays in the upstream processes propagate down the entire line to other processes and areas of the facility.

Having established an area to focus on, the next step for the team was to choose part types to follow through this area. To begin with, the initial list of over 1500 Waters components was narrowed down to two critical component categories recommended by project supervisors: pump heads and flow cells. The company was interested in these parts due to their high demand into the assembly area, and the consistency of demand for parts in these categories made them appealing parts to consider; parts with more consistent demand could produce more ensured, impactful change. Moreover, parts in these categories all face initial production steps in turning and subsequent operations in milling before going downstream, so the team could monitor the same flow through the chosen departments for all parts.

From these categories, the team narrowed selection by part families: that is, groups of part types whose setups were similar enough that producing them back-to-back instead of separately would result in a significant reduction in overall setup time. Finally, part types were narrowed down by which machines in turning and milling were dedicated for their production: for ease of data collection, only part types produced at two distinct turning work centers (M1302-2 and M1309) and two distinct milling work centers (M1210 and M1211) were selected.

33

Note that not all of the part types selected fit the profile of high-demand, low-volatility, high-setup-time parts which could benefit in the short term from optimized lot sizing; four of the flow cells, in particular, have annual demand of under 400 parts. However, as the project involved solutions to three separate problems (lot sizing, scheduling and data collection), and as Waters desired minimal disruption during the project's implementation, the part types selected for changes had to be ones which met the most criteria for all proposed solutions.

This narrowed potential candidates for changes to ten total SKUs: four pump heads and six flow cells.

Table 2.4: List of part types selected for project. Part types are color-coded by family.

Part #	Type	Turning Work Ctr	Milling Work Ctr
405015367	Pump head	M1309	M1211
405013109	Pump head	M1309	M1211
405012748	Pump head	M1309	M1211
405008452	Pump head	M1309	M1211
289003308	Flow cell	M1302-2	M1210
wat081193	Flow cell	M1302-2	M1210
289007447	Flow cell	M1302-2	M1210
405001315	Flow cell	M1302-2	M1210
405007811	Flow cell	M1302-2	M1210
405008774	Flow cell	M1302-2	M1210

3   Literature Review

This chapter describes concepts of manufacturing systems, including basic principles of factory physics, production approaches and push-pull systems, and inventory review policies. These principles inform the lot sizing and supermarket establishment methodologies detailed in Chapters 4 and 5. Factory physics principles and batching laws are described by Hopp and Spearman [5]. Inventory policies are described by Simchi-Levi, Kaminsky, and Simchi-Levi [6].

3.1   Overview of Principles of Factory Physics

We can first define the utilization u of a workstation as the fraction of time that the station is up and operating (not failing and not idle due to lack of input of materials). Various manufacturing textbooks and manufacturing environments have their own definitions for workstation utilization [5]. However, for the purposes of the project, the utilization of a machine will be defined as the percentage of time that the machine is drawing power and its spindle is running, since this can serve as a standardized representation of the fraction of time that parts are actually being produced. This can be written as

$$u = \frac{t_{on/up}}{t_{on/up} + t_{on/down} + t_{off/down}} \tag{3.1}$$

where $t_{on/up}$, $t_{on/down}$, and $t_{off/down}$ are times corresponding to the three possible states of the workstation. Spindle uptime $t_{on/up}$ refers to the time period when the machine draws the most power, which is while it performs an operation. The low power state $t_{on/down}$ usually corresponds to time during which the machine is being set up or maintained. The machine will be idle or under repair for the remainder of its available time, which is $t_{off/down}$ [1].

Over time, inventory will accumulate in buffers within a system. Inventory of parts in buffers, which still require operations and are not yet finished, will hereafter be referred to as work-in-progress inventory (WIP).

We can define the bottleneck of a manufacturing system as the workstation with the lowest effective production rate [5] in a given process, and therefore the machine which limits the production rate of the system as a whole.

The fundamental factory relationship relating inventory levels to throughput and lead time of a part is Little's Law [5]:

$$L = \lambda W \tag{3.2}$$

where L represents inventory, λ is system throughput and W represents the time a part spends in a system.

Little's Law can be applied at all levels of a manufacturing system, from individual machines and buffers to an entire production line. With individual machines, L and W will constitute WIP and time spent both waiting in a buffer and being operated on, so the Law can be used to estimate queue lengths and waiting times. With an overall line, L and W will be overall inventory and the lead time of a part, respectively, where lead time is defined as the average amount of time a part will spend in the system before being completed and released, as per the definition Waters uses for its machining center. Therefore, the Law can be used to determine what level of WIP is necessary to handle the maximum throughput of the bottleneck at a given average lead time specification [5].

Typically in manufacturing systems, parts are produced in sets following a single setup, also known as batches or lots; large batch sizes are typical of the push systems which will be described in Chapter 3.2. Little's Law tells us that producing parts in larger batches, and therefore increasing WIP in front of a bottleneck, will increase waiting and therefore lead time of parts for a given throughput, which will drive up production costs and require production to begin earlier to reach a steady state by the time customers demand finished goods. Typically, to keep lead time down, batches are limited to the size leading to inventory which can support the production rate of the bottleneck.

However, producing parts in large batches can also potentially lower costs for two reasons. The first is that producing parts with high demand in small batches will require repeated setups for machines, which will not only cost time and money, but require more manual work from operators and could potentially lead to more mistakes and higher scrap. Producing in larger batches instead can increase product quality [7].

The second is that increasing batch size will reduce the variability of part processing times through variability pooling. Let us define t_0 and σ_0 as the mean and standard deviation of the cycle time at a given workstation for a given part. Let us define the coefficient of variation (CV) as

$$CV_0 = \frac{\sigma_0}{t_0} \tag{3.3}$$

When parts are produced in lots of size n, their aggregate process time will have a mean equal to the sum of the means of individual part process times. However, the variability of the aggregate part processing time will only increase by a factor of the square root of n:

$$CV_{lot} = \frac{\sqrt{n}\sigma_0}{nt_0} = \frac{CV_0}{\sqrt{n}} \tag{3.4}$$

Managing the tradeoff between reducing setups and process time variability and reducing the costs of WIP and high lead times is a major component of deciding batch sizes in manufacturing systems [5].

With relation to the project at Waters, these principles demonstrate that reducing setups in favor of increased production is an argument in favor of larger lot sizes, while controlling the costs of inventory constrains lot size. In addition, Little's Law shows that while reducing setups may increase available capacity, they will increase overall lead times if implemented from the point of order and carried through the entire production process of a part.

37

3.2   Push Production Systems

Push production systems are characterized by having production orders arrive at the beginning of a production line. Batch sizes and order quantities are the result of a centralized decision process, usually coming from a master production scheduler. Decisions on batch/lot sizes and production levels are made irrespective of current demand (which is made at the end of a line by the customer) and are instead based upon a demand forecast, which is "pushed" through the system [8].

Push production systems became the most common production method in the United States for most of the 20th century, at least for systems with few part types. An example of a push system is the assembly line introduced by Henry Ford for the Model T in the 1910s and 1920s. These systems produce parts in large batches, which reduce process time variability and setup times for systems with few distinct parts and an expectation of high demand.

Push systems have drawbacks, however. Firstly, they are not very responsive to changing customer needs, which means their effectiveness is dependent on the quality of the demand forecast which precipitates production. This in turn means that parts in a push system are susceptible to the "bullwhip effect," in which variability of orders increases more and more moving further upstream in a supply chain in response to small changes in customer demands. Due to this high variability in production needs at the highest levels of the production line, companies are susceptible to either surpluses of inventory or backlogs for popular products [4].

3.3   Pull Production Systems and Push-Pull Hybrids

In contrast to push systems, pull systems have production orders arrive at the end of a line. These orders are then "pulled" out of the system, with upstream production compensating for parts being pulled out of WIP downstream [5]. Thus, parts are only produced upstream after orders have come in downstream, either from the customer or someone downstream in an internal supply chain.

The key benefit of a pull system over a pure push system is that it establishes a WIP cap at the various buffers in a production line, preventing the production of WIP that fails to add to throughput [5]. Since less WIP is produced, the average time a part spends in a manufacturing system will also go down, as parts will spend less overall time in buffers. Moreover, as HP showed in a test demonstration in 1983, pull systems can in fact result in higher production rates due to their higher responsiveness to downstream line problems [9]. Thus, pull systems reduce the costs of holding WIP and can increase overall throughput for lines susceptible to failures or demand changes.

Push systems are, however, preferable in cases where demand forecasting is relatively accurate. Producing parts in larger lots with fewer changes in setups can be cheaper overall for a manufacturing system, provided they do not result in explosions in WIP, or provided the cost of holding WIP is relatively low (as it is with high-volume parts at Waters). Finally, producing the same number of parts with less overall setup time means more overall time for production, signifying a higher production rate.

To manage the tradeoff between wanting to produce to a pre-specified forecast, wanting to reduce WIP and wanting to be responsive upstream to downstream problems, most companies use some combination of both strategies, known as a push-pull hybrid system. An example of a push-pull hybrid at work is that a line can have a Kanban authorize the release and flow of parts within a production line, but overall inputs to the line, and therefore production overall, can be delayed by the master scheduler due to a low demand forecast [10].

3.3.1 Kanbans

One way to establish a WIP cap for pull systems is through the use of Kanbans. Kanban systems use cards as "demand signals" at various points in a production line, usually at machines that would otherwise build excessive inventory (such as the supermarket described in Chapter 5). When a finished part or inventory downstream in a line is depleted, or a customer requests a

build-to-order part, a Kanban card is sent to an upstream machine, signaling it to produce additional WIP to replenish the depleted product [5].

Production kanban movement — **Withdrawal kanban movement** — **Material movement**
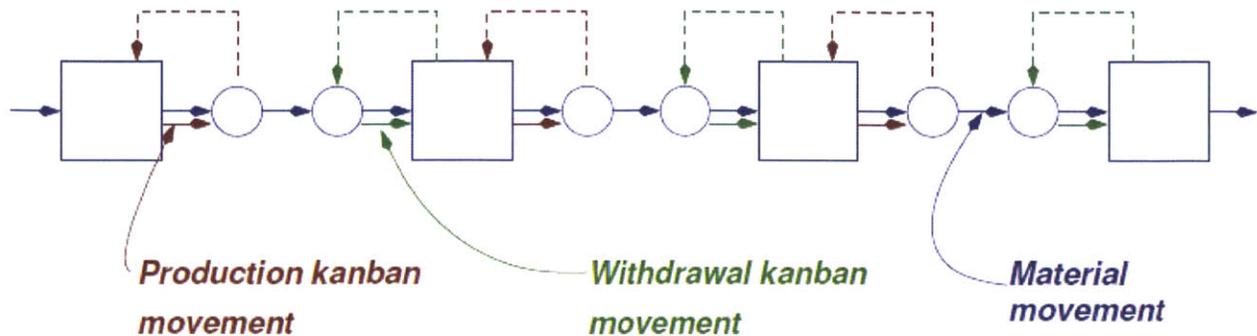
Figure 3.1: Movement of Kanban cards. Dotted lines reflect movement of Kanbans back upstream, signaling a machine to produce more or a buffer to replenish its WIP[14].

Kanban systems use buffers of finite sizes after each machine in the production line, with buffer sizes related to the magnitude of disruptions caused by upstream machines going down. These finite buffer sizes, as well as information flow upstream indicating when production needs to occur, will limit WIP in the system [5].

3.3.2   Supermarkets

A way to manage the tradeoff between the need to limit inventory and the need to produce high-volume parts at a high rate is through the use of supermarkets. Supermarkets are essentially decoupling buffers, storage areas placed at locations within a manufacturing system to serve as intermediate sources of inventory [12]. Upstream of the supermarket, production can be triggered by a Kanban pull system, telling the start of the line when supermarket inventory is sufficiently low. Downstream production can then operate, decoupled from upstream inventory, by simply taking its inputs from the supermarket rather than waiting for upstream production to finish. An example of a supermarket placement within the Waters manufacturing system is shown in Figure 3.2.

---

[14] Lecture notes, "Multi-Stage Control and Scheduling," ©2006 Stanley B. Gershwin. Based upon Monden [11].
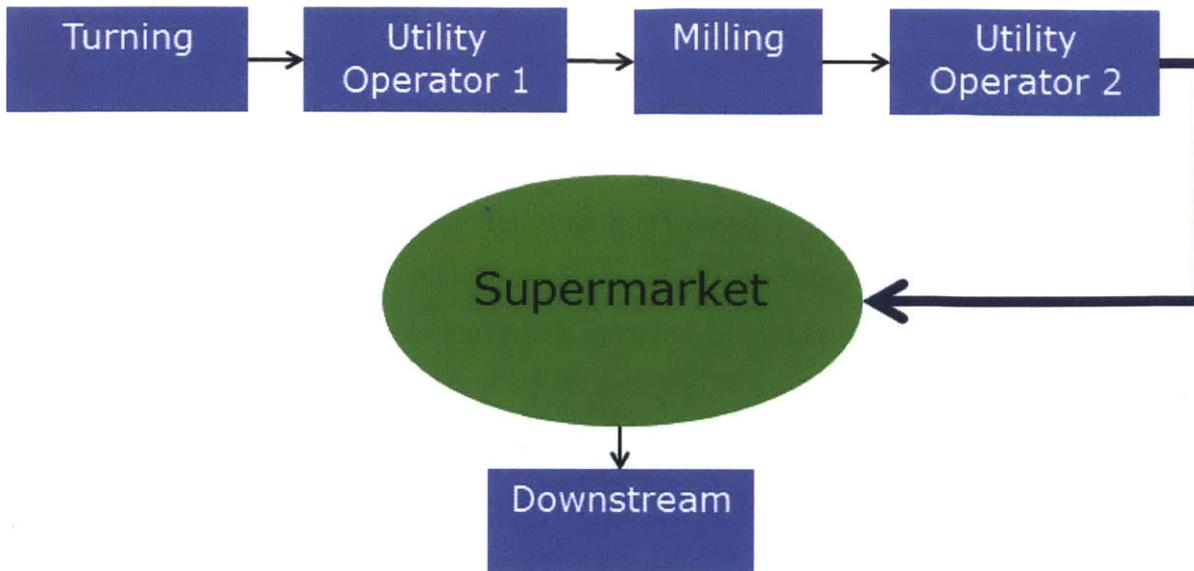
Figure 3.2: Supermarket location.

A supermarket system can provide improvements over a pure push system because it can achieve lower WIP upstream than a push system would typically allow [9]. It also allows for an increase in on-time delivery from order placement in cases where upstream processes can bottleneck downstream production, as turning and milling do at Waters, and can allow different production lot sizes to exist upstream and downstream due to its decoupling effect [12].

3.4   Inventory Review Policies

The two major inventory policies in use in modern manufacturing environments are continuous review and periodic review, described in detail by Simchi-Levi et al [6]. The major underlying concepts behind these policies are outlined in this section. Another policy used commonly in industry, and present at several points in the Waters system, is that of vendor-managed or "breadman" inventory [13], but since the project does not consider interactions with external vendors, this policy is not described here.

To characterize inventory review, it is important to first define inventory position. Inventory position is the sum of inventory on hand and units which have been ordered but have not yet arrived.

3.4.1 Continuous Review

Under a continuous or perpetual review policy, inventory levels are monitored in real time. Whoever is responsible for managing inventory will place an order for additional production of a part once inventory drops below a threshold known as the reorder point [6]. Continuous review results in a responsive inventory management system; however, it also results in high costs from constant inventory monitoring and time taken away from other company activities [14].

In practice, the continuous review model typically follows the (Q, R) approach: an order quantity of Q units is triggered when inventory position reaches the reorder point R.
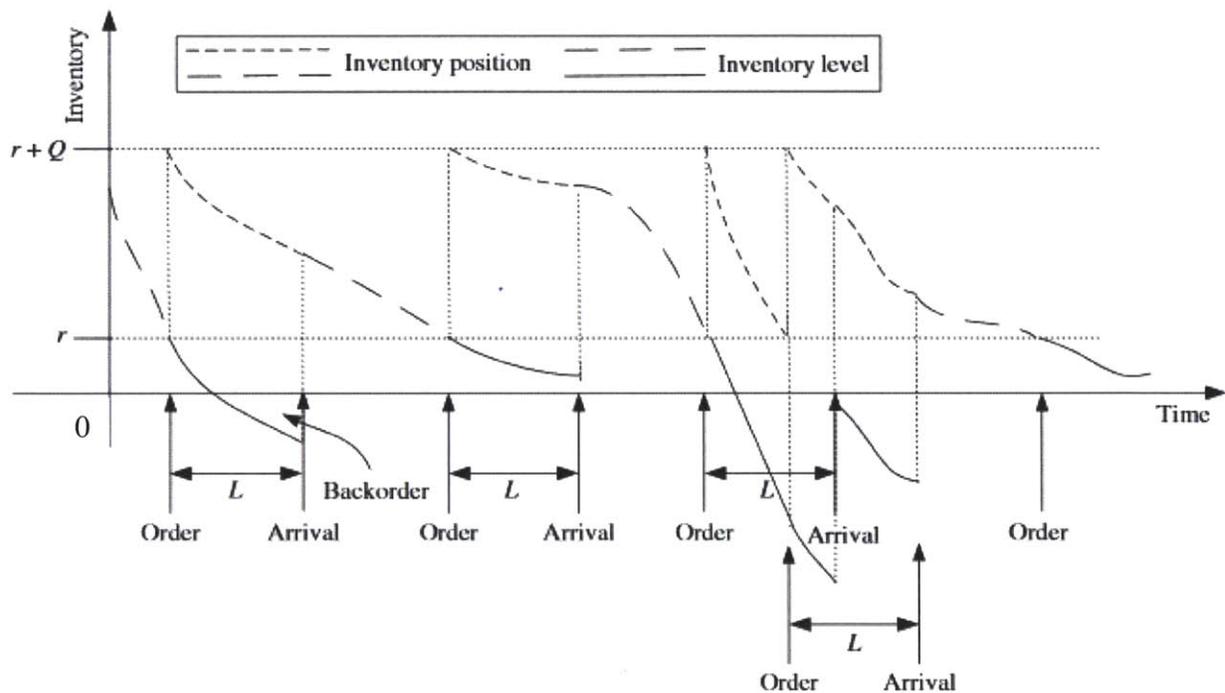
Figure 3.3: Illustration of the (Q, R) policy for continuous inventory review. L is the lead time for an order of quantity Q. Note that orders are placed when inventory position reaches R, even if prior orders have not yet arrived [15].

In most factory settings, the reorder point R consists of two components: forecasted average inventory consumption over the lead time of the product and safety stock to cover

42

demand variation to an acceptable service level, where service level is defined as the probability of meeting all demand over a given time [10].

$$R = Average\ demand\ over\ lead\ time + safety\ stock \qquad (3.5)$$

The value of the safety stock depends upon the demand distribution and the desired service level. The two most common demand distributions considered in systems literature are a basic normal distribution with mean μ and variance σ^2, or, as is more often used when the demand coefficient of variation is very high (distribution is volatile), a lognormal distribution with mean μ^* and variance σ^{*2} [16]. The value of the reorder point R when demand is normal will be

$$R = \mu + z\sigma \qquad (3.6),$$

and if demand is lognormal,

$$R = e^{\mu^* + z\sigma^*} \qquad (3.7),$$

where z in both cases is the standard normal test statistic corresponding to the desired service level,

$$z = \Phi^{-1}(service\ level) \qquad (3.8)$$

The value of the order quantity Q is based upon the optimal preferences for the inventory manager and product distributor [16]. In the context of production rather than a supply chain, Q can be referred to as the "lot size," or the quantity of parts produced in a single run under a single machine setup. Section 3.5 describes methodologies for sizing Q.

43

3.4.2 Periodic Review

The periodic review policy differs from continuous review in that inventory levels are recorded at regular, fixed intervals rather than in real time. At each review, an order of appropriate quantity is placed. [6]. Periodic review is popular due to its being relatively accurate and cost-effective; studies such as that conducted by Sentyaningsih et al [14] show that for most companies and cases, use of the policy not only reduces stockouts and backlogs but is cost-efficient.

Figure 3.4 illustrates the implementation of one form of periodic review. Under this model, a single factor – the base stock level B – characterizes the review. A distributor or warehouse will determine a target base stock level and review period r. It will then review the inventory position at intervals of length r and place orders to replenish inventory back to base stock.
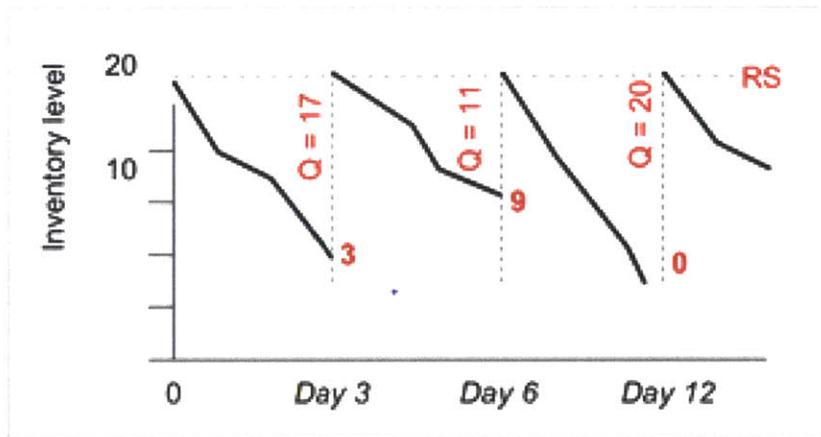
Figure 3.4: Illustration of base stock policy for periodic inventory review. In this example, the base stock level B is 20, the review period r is 3 days, and replenishment lead time is immediate (0 days) [17].

An alternative periodic review model will utilize a safety stock level and constant order quantity, as in the continuous review model.

The base stock level should be set up to ensure against stockouts during the lead time for delivery of additional inventory. When an order becomes necessary, it will arrive at intervals of r + L days, where L is the product lead time. The base stock level must be able to cover average

lead time demand with a certain safety factor, taking into consideration the uncertainties in demand forecast. This is given by [6]:

$$B = Average\ demand\ over\ (r + L)days + Safety\ stock \qquad (3.9)$$

3.5 Lot Sizing

Two common approaches have been taken in the past to determine order quantities: the Economic Order Quantity (EOQ) model and the newsvendor model. However, both of these approaches assume that parts are independent of each other and fail to take into account that multiple part types with different demand rates and setups can go through a machine in a given period. Very little literature exists on lot sizing and scheduling on machines with multiple, dependent part types, but a recent attempt to model such a system came from Pinedo et. al in 2005 [18]. For a description of the newsvendor model and its use in inventory policy, see the original paper by Arrow, Harris and Marshak [19]; however, the main models of interest to this thesis are the EOQ model used as a basis for Waters' current lot sizing and the Pinedo model used for the author's systems improvement methodology.

The EOQ model originally developed by Ford Harris in 1913 [20], computes lot size Q^* to minimize cost for one part type, taking into account inventory holding cost and ordering cost (analogous to setup cost for production):

$$Q^* = \sqrt{\frac{2Dk}{h}} \qquad (3.10),$$

where k is the ordering cost (setup cost), h is inventory holding cost and D is demand rate.

The Pinedo model, which he describes as the "economic lot scheduling problem," is more comprehensive than the EOQ model in describing machines with multiple part types and with setups, as it factors in the setup time, cycle time, demand rate and inventory holding cost for all parts going through a machine [18]:

45

$$Q_i = D_i \sqrt{\left(\sum_{j=1}^{n} \frac{h_j D_j (G_j - D_j)}{2G_j} \right)^{-1} \sum_{j=1}^{n} c_j}$$

(3.11),

where n is the number of parts on a machine, c is the setup cost for each part j, and G is the production rate of each part j.

It should be noted that the Pinedo model is not perfect either; it sizes lots assuming all parts have independent setups and assuming a single-machine system. On top of that, it does not consider other factors which come into play in a complex system such as the Waters factory, such as usage of raw materials or the lot sizes demanded by outside vendors. However, it takes into account the dependence of all parts with a limited capacity on a machine and sizes lots to reduce costs while preventing inventory backlogs.

3.6 Relating Literature to the Waters Lot Sizing Problem

As described in Chapter 2, one of the problems facing the Milford facility is lot sizing which fails to account for limited machine capacity or the interrelatedness of demand for all parts which go through a machine. Little's Law accentuates the need to control inventory in order to prevent costs of WIP or costs of long lead times. However, the benefits of fewer setups and variability reductions from larger lot sizes can outweigh these costs in cases of high-demand, high-volume parts. As a result, an appropriate lot sizing methodology, such as that used by Pinedo, must consider the costs of both setups and inventory as well as the demand of all parts on a machine. Moreover, to mitigate the costs of longer lead times caused by larger lot sizes, as well as to mitigate any other hidden costs of changing lot sizes, production at bottlenecks, such as turning and milling at Waters, should be decoupled from production driven by customer demand by mechanisms such as supermarkets. These concepts, therefore, guided the improvement methodology and modeling described in the chapters that follow.

4 Methodology- Optimized Lot Sizing

Demonstration of improvements to the manufacturing system involved four major steps: selection of part types and machining departments to improve, resizing of the lots in which they are produced, creation of a supermarket for those parts, and reordering of the production schedule to make those parts more efficiently. The proposition for reordering of the schedule guided the sizing of the supermarket. The selection of target areas and part types is discussed in Chapter 2.5. This chapter describes the lot sizing methodology for those part types, based primarily upon the Pinedo multi-part, single-machine model. The third step will be discussed in Chapter 5, and the fourth step is described by Puszko [2].

4.1 Optimized Lot Sizing for Turning and Milling

Recall that the original project aim was to realize an improvement in productivity to help the Milford facility increase its annual output. After discussions with multiple people in the company about how production lot sizes were determined, it became apparent that although the Economic Order Quantity (EOQ) model serves as a baseline for some lot sizes, there is no systematic methodology in place at Waters to determine an optimal size. Instead, the natural restrictions of the manufacturing processes determined logical lot sizes, such as how many parts could fit in one fixture and be worked on simultaneously, or how many parts could be made from a standard length of raw material stock. Our hypothesis was that production lots for some of the selected part types could be increased in order to reduce the total number of setups needed to make the same amount of parts over a particular time period; the time recovered from the reduced setups could be spent making other parts, directly increasing the productivity of the machines chosen.

The challenge of lot sizing and scheduling is balancing production between all the part types that have to go through a particular machine, so that all part types can be made according to their demand and without falling into backorder. While the Economic Order Quantity model is a popular methodology for determining production sizes or order quantities based on setup

costs, demand, and inventory holding·cost, it assumes that parts are ordered or made in a vacuum, independent of the production or procurement of any other part type, and that those orders can be placed at any time. In the case of this project, if one were to apply an EOQ model for all the part types that have to go through one machine, the resulting lot sizes would be too large for the production system, as EOQ does not include the capacity constraints of a multiple-part .one-machine system. This would cause issues with backorders, as the production of certain part types would have to wait for the machine to complete other production runs before being set up; Waters experiences this problem already due to its infinite capacity assumption, although, as discussed in Chapter 2, this problem stems from the other extreme of reducing lot sizes to produce redundant setups.

As mentioned in the literature review, a methodology for optimal lot sizing in the multiple-part single-machine case has been previously developed by Pinedo [18]. The Pinedo model factors in the setup time, cycle time, demand rate and inventory holding cost for all parts going through a machine in order to determine the proper production lot size for each part type that needs to be formed on that machine. As previously stated, the Pinedo model is not perfect, as it does not take into account raw material usage or the need for multiple machines in a production line for a part. The Pinedo model also assumes deterministic, constant demand rates for each of the part types input into it, meaning that high demand volatility can result in backlog when demand is unexpectedly high or inventory surpluses when demand is low. Moreover, high demand volatility for low-volume part types can mean that the optimal method of inventory control is not through consistent production with small lot sizes, as the constant-demand Pinedo model would suggest, but rather by production of low-volume part types at larger lot sizes at less frequent intervals.

Further, if an optimal lot size is determined that can reduce production and holding costs, while considering all parts that need to go through a particular machine, it is not known that these new lot sizes would be larger or smaller than the ones currently set by Waters and stored in SAP. If a methodology determines that optimal lot sizes are in fact smaller than the ones Waters is currently using (due to low demand), then this would lead to more frequent setup changes on

the machines, and thus changing the production lot sizes would not be a solution to the problem of high setup times.

However, the Pinedo model does solve the critical problem of a standard EOQ model, that of part independence, and so it formed the baseline of the lot sizing methodology for this thesis. The equation for the optimal lot size Q of part type i in an n-part, single-machine case is given in Equation 3.11. By gathering this information for all part types produced on the machines under investigation in the milling and turning areas, the optimal lot sizes for the parts in question could be determined for each machine that they went through.

For this project, since all selected part types are monitored through first turning and then milling, lot sizes had to be the same in both departments. However, since the Pinedo model is a single-machine model, and since setup and production rates are different for all part types in turning and milling operations, optimal lot sizes calculated were different between departments. Thus, a choice had to be made as to how lot sizes could be optimized given that constraint.

The solution used for implementation was to simply follow the lot size calculation from the turning department for the ten selected part types. This was based upon several factors: for one, since the overall project goal was to increase production, lot sizes needed to tend towards being higher for the same amount of setup time, and setup times in turning were higher for all selected part types by any data collection metrics as well as the word of project supervisors. Secondly, as mentioned in Chapter 2.2.1, milling suffers from utilization rates consistently around 50%, while turning is closer to 70%. The team conjectured, therefore, that milling had enough productive capacity to handle producing parts above the optimal level of the Pinedo model, given that only ten total SKUs were to be modified. Finally, there were no other discernible reasons, from an external supplier or planning standpoint, why producing larger-than-optimal lot sizes would adversely affect milling.

Further modifications, to reach lot sizes which could be applied on the shop floor, were applied heuristically and considered use of raw bar stock; production planners desired

minimizing waste of bar stock by limiting lot sizes to use increments of half-length bars. These modifications depended on the part types in question and are detailed numerically in Chapter 4.2.

Note, however, that this modification of milling lot sizes is not an advisable rule of thumb in system design, as these lot sizes do not truly take the capacity of the mills into account. Also note that failing to apply the lot sizing methodology to every part type made on a given machine is less than optimal, since capacity is again not fully accounted for. These problems will need to be addressed in future work (see Chapter 8.3). However, within the constraints of the project, the methodology described comes as close as possible to optimizing lot sizes for the selected part types.

4.2   Lot Sizes for Selected SKUs

The final lot sizes for the selected part types are displayed below:

Table 4.1: Final lot sizes for selected part types. Modified lot sizes are highlighted.

Lot Sizes	Final	Original
405015367	276	92
405013109	72	50
405012748	50	50
405008452	76	60
289003308	165	110
wat081193	276	250
289007447	30	30
405001315	30	30
405007811	25	25
405008774	40	40

Lot sizes for the five highest-volume part types need to be increased, with two SKUs in particular needing increases of more than 50%. The optimization methodology estimated smaller lot sizes than present values for the other SKUs on the list, which are all lower-volume; these lot sizes were kept the same during implementation so as not to starve their downstream production.

In order to obtain these lot sizes, the author began by considering the Pinedo model. From Equation 3.11, the following inputs from the turning department were necessary:

1. Demand rates for all part types going through the same machines as selected SKUs, M1302-2 and M1309.

2. Setup costs for all part types going through M1302-2 and M1309.

3. Production rates for all part types going through M1302-2 and M1309.

4. Inventory holding costs for all part types going through M1302-2 and M1309.

4.2.1 Computing Demand Rates

In order to compute the first input, the author had to choose between two separate measures of part demand, both available within Waters' SAP system. The first was the "annual usage" statistic each department reports in SAP, a measure of how many units of the SKU the department has produced in the past year.

The second measure was one the project team termed "consumption" of parts. Within SAP, Waters refers to machines, buffers or any other part locations as "stock locations." One way in which demand for each part type could be measured was by looking at part histories in SAP and determining how many pieces of the part type were removed from machining stock locations and transferred to downstream ones, as well as when each of these transfers occurred. Starting one year before the time of investigation (June of 2013), the team charted 52 weeks' worth of dates and removal amounts for every instance of part transfer away from the machine shop for the ten selected part types. These removal amounts were then grouped by week and compiled into weekly consumption statistics. An example of this procedure for one of the high-volume selected part types, 405015367, is displayed in Chapter 5.

Consumption statistics are more accurate than annual usage in displaying how produced parts are used downstream of the Machining Center, and more accurately display the week-to-week fluctuations in demand. However, we nevertheless chose to use annual usage, divided by 52 weeks, as the measure of weekly demand for all parts for the lot sizing methodology. This decision was driven by two main reasons. The first was that since lot sizing is driven by demand

for production rather than demand for consumption (which are not necessarily the same measure in a complex manufacturing environment), usage statistics pertaining to production would better inform the methodology. The second reason was that since the Pinedo model assumes deterministic demand in any case, using a single estimation of demand rather than a demand distribution simplified the calculation of lot sizes.

4.2.2 Computing Setup Costs

Two factors drive setup costs in manufacturing environments: the time and the raw material needed to perform a setup.

$$cost_{setup} = time_{setup} * \frac{cost}{unit\ time} + cost_{material} \tag{4.1}$$

Investigations revealed that Waters does not track the cost of raw material used in setups, and interviewees suggested that for the selected SKUs, it was a reasonable assumption that material costs could be absorbed into the time costs of the machine shop. As a result, we set material costs to 0 for this methodology.

The cost per unit time for a setup was assumed to be the amount the Machining Center charges to Waters in internal accounting credits for time worked in the turning department, which is \$67 per hour[15]. Spread over a machine's weekly capacity of 135 hours (22.5 hours per day, 6 days per week), this equates to \$9045 per week per machine.

This just left the calculation for the setup time. SAP has standard values for setup times for all SKUs stored from the previous cost roll; unfortunately, the cost roll is two years old, and data for the cost roll, as mentioned in Chapter 2, was not actually measured for every SKU. The automated data collection project [1] was not yet in place during the evolution of the lot sizing methodology and so could not be used to estimate setup times either. Instead, then, the most accurate measure available was data from recent production runs recorded by machinists into a Filemaker program for the turning supervisor. We therefore acquired setup times for the

---

[15] Personal communication, Supervisor of Valve Cell, Waters Corporation, March 21st, 2014.

methodology by averaging run data from Filemaker from January to May of 2014 for all part types going through the machines of interest in turning, M1302-2 and M1309. These setup times were given in hours; in order to express them in relation to the capacity of each machine, and therefore size lots according to that capacity, setup times were converted to weeks by dividing by the machine's weekly hourly capacity of 135 hours.

4.2.3 Computing Production Rates

The formula for maximum production rate of a single part type on a single machine [5], not counting setups, maintenance, or other non-productive steps, is

$$rate = \frac{1}{\tau} * \frac{MTTF}{MTTF+MTTR} \tag{4.2}$$

where τ is the average processing time or cycle time of the part type in hours, MTTF is the mean time to failure of the machine, and MTTR is the mean time to repair a failed machine. The second fraction in Equation 4.2 can also be thought of as the efficiency of the machine.

Interviews with department supervisors and section leaders suggested that machine failures are uncommon and, if they occur, tend to be catastrophic rather than repairable. There was no further up-to-date data relating to the mean time to failure or mean time to repair of any machine; potential failure modes are instead accounted for by being absorbed into the reduced capacity of the machines (even though supervisors consider capacity for all machines to be 135 hours per week, the shop actually runs for 144 hours per week)[16]. We assumed that all efficiency losses subtracting from production rate were accounted for by reduced capacity.

Therefore, production rate calculations came out to

$$rate \left(\frac{parts}{week}\right) = \frac{1}{\tau} * 135 \; hours \; of \; capacity \; per \; week \tag{4.3}$$

---

[16] Personal communication, Head of Machine Shop, Waters Corporation, June 24th, 2014.

As with setup times, processing times were found with machinist-collected data imported from Filemaker for all parts on M1302-2 and M1309.

4.2.4 Computing Inventory Holding Cost

Two factors drive inventory holding costs: the space needed to hold inventory and the depreciation in value of the product while it is held.

$$h = space * \frac{cost}{unit\ space} + Depreciation * original\ value\ of\ WIP \qquad (4.4)$$

Although the demand rates of each of the selected SKUs vary, the part types were filtered based upon their low obsolescence risk. All of the part types are components used in Waters' newest liquid chromatograph model, and Waters still sells components from 25-year-old models at original prices. Therefore, for the purposes of the methodology, we assumed no cost of depreciation from holding WIP.

As for the cost of space, the Machining Center charges Waters about $648 per square foot of space per year for its production[17]. This equates to about $12 per square foot per week. Bins containing parts occupy an area of about 2 square feet, which equates to $24 per bin per week. Depending on the part type, for the selected SKUs, bins can contain either 24 or 30 parts. Since Waters does not have a present method of tracking inventory holding cost, we wished to make the estimate for the methodology conservatively high; thus, for the purposes of calculating holding cost, all bins were assumed to contain 24 parts.

This resulted in a final estimation for inventory holding costs for all selected parts of $1 per part per week.

The reader should note that this estimate overstates the true cost of holding inventory. Although depreciation cost is in reality slightly higher than 0, the cost of space used for the

---

[17] Personal communication, supervisor of CNC turning and milling, Waters Corporation, June 13th, 2014.

estimation includes charges for space used by productive machines. Space for inventory, rather than production, is worth much less in reality than the $648 per square foot used above; some sources at Waters believe this may overstate inventory holding cost by a factor of 10 or more[18].

This means that true optimized lot sizes may be higher for all selected part types than those indicated in this thesis. However, without accepted company standards for measuring inventory holding cost, we could not assume that Waters could handle more than a conservative estimate of inventory. Inflation of inventory costs and deflation of lot sizes given the lack of data is an acknowledgement of the constraint inventory has on production and the reason lean manufacturing principles seek to impose WIP caps in production environments. Having an accepted measure of holding cost to improve methodology projections will be necessary future work (see Chapter 8.3).

4.2.5 Final Lot Sizes

The above inputs were calculated and compiled together for all part types on the two lathes in question. Using the Pinedo model with these values, optimized lot sizes were obtained. The lot sizes displayed in Table 4.2 only include the ten selected SKUs, and not others on the investigated machines.

---

[18] Personal communication, Supervisor of Valve Cell, Waters Corporation, June 17th, 2014.

Table 4.2: Optimized lot sizes based upon Pinedo model.

Part material No.	Demand rate (parts/week)	Processing time (hours)	Production rate (parts/week)	Setup time (weeks)	Work center cost ($/week)	Material cost	Setup cost ($/week)	Inventory holding cost ($/part/week)	Lot size	EOQ	Current Lot Size
289003308	29.62	0.13	1023.71	0.04	9045	0	316.73	1	184	172	110
wat081193	48.42	0.11	1285.61	0.03	9045	0	247.06	1	300	225	250
405001315	1.15	0.18	758.57	0.04	9045	0	385.25	1	7	25	30
405007811	2.50	0.62	218.92	0.02	9045	0	195.42	1	14	41	25
405008774	7.12	0.39	342.90	0.03	9045	0	268.00	1	41	90	40
289007447	2.42	0.29	470.93	0.03	9045	0	284.75	1	14	37	30
405015367	70.90	0.24	558.78	0.01	9045	0	120.04	1	262	220	92
405013109	19.85	0.42	322.03	0.01	9045	0	125.63	1	73	102	50
405012748	11.46	0.21	654.32	0.02	9045	0	167.50	1	42	84	50
405008452	20.77	0.37	366.98	0.02	9045	0	195.85	1	77	110	60

Data Source	SAP/Existing
Color Coding	Filemaker
Legend:	Calculated

56

Two modifications were made to the above table in order to arrive at final lot sizes. First, those SKUs whose lot sizes were reduced by the Pinedo methodology were reset to their original lot size. This was done because reducing lot size would lower productivity by adding setup time, and because production of those parts downstream of turning and milling would be starved if fewer parts traveled downstream. The estimates for low-volume part types display one of the limitations of the Pinedo methodology: its single-machine assumption can cause reduced utilization downstream if lot sizes become smaller upstream, and its constant-demand assumption suggests that low-volume part types should be consistently produced in small amounts rather than in larger lots all at once. These estimates also display the limitations of any systems improvement methodology: their dependence on accurate input data. Inventory holding cost, if properly tracked, would likely have been lower than assumed, indicating that high inventories of each low-volume part type could be held and resulting in larger lot sizes than those displayed in Table 4.2.

The second modification to the table was to make adjustments to the remaining lot sizes based upon raw material inputs. Due to a desire to reduce wasted bar stock, Planning requested that lot sizes be such that orders consumed only full bars or half-length bars of raw material. This required major adjustments to the lot sizes for three parts: 405015367, WAT081193 and 289003308. Each of these sizes had to be adjusted to be a multiple of half the original lot size (WAT081193 had an original lot size of 92, which was changed to 250 for unrelated reasons just prior to the start of the project). The end result of that change was that lot sizes were not quite optimized with respect to the Pinedo model. However, this represents an example of changes presented in a real-world manufacturing environment relative to literature: Waters found the bar stock usage with the finalized lot sizes more manageable.

Thus, the methodology provided a template for adjustment of lot sizes. Five part types needed an increase in lot size, displaying the dangers of unnecessarily reducing inventory by adding setups and reducing capacity. Five others were held at existing lot sizes, displaying the limitations of applying any methodology in all cases without considering demand volatility, inflated inventory holding costs or the potential for inaccurate data inputs to the system.

5   Methodology- Establishment of a Part Supermarket

The lot sizes described in Chapter 4 were used only for the turning and milling areas, with lots in downstream production decoupled by establishment of a supermarket. This supermarket is designed to be managed through a constant-order-quantity, continuous inventory review policy with reorder points meant to maintain a 95% service level. The reorder points are calculated assuming one week of manufacturing lead time to proceed through turning and milling for each of the ten selected part types. This chapter describes the rationale behind the supermarket and its inventory review policy, as well as the computation of the reorder points.

5.1   Part Supermarket

One potential issue with the lot sizing methodology, mentioned briefly in discussing the difference between optimal lot sizes in turning and milling, is that this investigation only looked at the production of parts in those upstream areas. The optimized lot sizes developed for these areas may be sub-optimal for processes down the line. A large batch in a quicker upstream process can wreak havoc on a downstream process with longer cycle times than in turning or milling, causing orders of other parts that need to be processed on the same machine to fall into backorder. Moreover, further down in the production processes for the selected SKUs, as well as other SKUs the methodology could be applied to in the future, the influence of external vendors or customers and their lot size restrictions becomes more prevalent[19]. As the machining area is set up as a job shop with a variety of different processes that pull from and feed each other in a variety of combinations, it would be difficult to generate and implement optimized lot sizes for every part type and every process, especially since the Pinedo model is multi-part, single-machine rather than multi-part, multi-machine. If the optimized lot sizes for the ten selected SKUs in turning and milling were to be implemented, then, there needed to be a way to decouple the production of milling and turning from the rest of the system, so that production in those areas could be optimized without disruptions downstream.

---

[19] Personal communication, Production Planning group, Waters Corporation, July 7th, 2014.

The implemented solution to this issue was a part supermarket at the end of milling (after processing by the milling utility operator) for all SKUs under investigation. The part supermarket is an inventory holding area for parts that are completed in the milling area and stored before moving on to the next process. In this situation, production in turning and milling operate in order to maintain the inventory of the part supermarket, and downstream processes can pull from that inventory as orders come in. With a part supermarket in place, production in turning and milling can occur at their optimized lot sizes, while not affecting the production of the downstream processes; the next step of the production process can be completed at any lot size by pulling any number of parts from the inventory, based on the order quantity (see Figure 3.2).

In this new line, whenever an order for one of these SKUs is placed in the system by SAP, the "production" of the SKU will begin by pulling the desired number of parts from the supermarket after milling, instead of starting at the beginning of turning. Since these parts had previously been completed in turning and milling before being held at the supermarket, they are immediately ready for the next step of the process. The inventory levels of the part types at the supermarket will be monitored, and when the levels fall under a certain threshold, production will be triggered at the beginning of turning to refill the inventory at the supermarket. This production will occur at the optimized lot sizes detailed in Section 4.2 for each SKU.

An additional benefit to the part supermarket will be to decrease the lead time of parts from point of order to arrival at assembly, which will increase their on-time delivery. Although, under Little's Law, increasing lot sizes will increase a part's time in the system, decoupling lot sizes at the supermarket imposes separate lead times for forecast- or customer-generated orders downstream and inventory-based orders through turning and milling. Thus, when orders get scheduled to go downstream in SAP, the lead time through turning and milling will be removed from the order lead time. Currently, turning and milling consist of about 50% of the total lead time for production across all parts in the machining center, so removing those processes will significantly decrease order lead time. If the quoted time of arrival remains the same, then the on-time delivery percentage is expected to increase dramatically. While the focus of this work is on the productivity and utilization improvements possible on a multi-part machine due to

optimized lot sizing, investigations of improved order lead time and on-time delivery from the supermarket are possibilities for future work (see Chapter 8.3).

It should be noted that of the ten selected SKUs, only five needed an adjustment to existing lot sizes based upon the methodology detailed in Chapter 4. Therefore, the supermarket only needed to be established for five part types to satisfy the need for decoupling upstream and downstream production. However, as part of the data collection [1] and scheduling [2] projects, the remaining SKUs were still used as test part types for implementation. As a result, they were produced outside of the usual Waters production schedule and were not necessarily ready to proceed downstream, so the supermarket included them as well.

5.2   Triggering Production: QPR Inventory Control Policy

To trigger production of the selected part types in turning and milling, the team needed to select a baseline inventory review policy at the supermarket.

One consideration was to use a base-stock (S, s) policy as detailed in Section 3.4.2. This would have established an inventory cap at the supermarket, preventing overuse of space or excess inventory costs. However, the base-stock idea was dismissed when the team opted for lot size optimization; since (S, s) only allows production up to a certain inventory level over specified review periods, using it for the supermarket would have resulted in a necessity for different lot sizes every time production was necessary. With parts having long setups in turning and milling irrespective of lot size, producing in different amount for different orders was undesirable.

Moreover, discussions with executives, finance and machine shop supervisors revealed no company-wide method to track inventory holding costs, and since all selected parts have low obsolescence risk, the cost of holding them in a supermarket is predominantly derived from the space they take up. With a rack available for the supermarket, the team conjectured that the cost of WIP was small compared to the cost of stocking out of one of the selected parts, and that having a WIP floor to meet demand was preferable to maintaining a WIP ceiling. Such an

inventory review model would inherently have a WIP ceiling anyway, equivalent to the WIP floor plus the order quantity.

As a result, the production trigger was one which produced in constant order quantities according to a WIP floor. A modified (Q, R) policy was proposed to manage the inventory levels in the part supermarket and trigger production of the selected SKUs in turning and milling. This modified policy is designed to take advantage of the reduced total production times that stem from producing part types within the same family back-to-back, where a part family is defined as a set of part types where the setup time and costs of going from one part to another and vice versa is much less than the same transformation between any other parts in the system. Thus, in order to reduce total production time in a multi-part system, it is optimal to produce parts in a family back-to-back, in order to reduce total setup times. This inventory policy has been coined "QPR" and is discussed in further detail by Puszko [2]. The values of Q and R, however, are calculated in the same way as in a conventional (Q, R) policy, and are relevant to this thesis.

To review, a (Q, R) continuous review inventory policy, as detailed in Section 3.4.1, triggers the ordering or production of a part type whenever the inventory level reaches below a certain reorder point R. At that time, a fixed lot size of Q parts is ordered/ produced. The reorder point R is set so that the inventory left in the system can cover the demand during the manufacturing lead time, so that stockouts are minimized.

For implementation, Q was determined by the optimized lot sizing model shown in Section 4.2. The reorder point R for each part type was determined based upon an assumption that each part followed a lognormal demand distribution, using Equations 3.7 and 3.8. The target service level was set at 95% based upon discussions with supervisors in various departments.

The methodology assumed a lognormal distribution rather than a normal one because several parts featured a high coefficient of variation (well above 0.5), meaning that demand fluctuations could result in high potential for negative demand if using a normal distribution, an impossibility which could skew calculations.

61

The calculation of R also assumes demand follows a distribution over a specified, constant lead time. Since SAP could be manipulated to break demand into weekly buckets, and since actual production times in turning and milling were much shorter than one week, the lead time through turning and milling for the purpose of calculating R was set to be one week. Thus, R would, in theory, cover one week worth's of downstream demand at a 95% service level until production of Q replenished the supermarket. Note that the specified lead time for the calculation can be easily adjusted for application in other production systems.

The inputs for demand and subsequent calculations for reorder points in the supermarket are presented numerically in the next section.

5.3   Supermarket Reorder Points for Selected SKUs

Calculation of the reorder points for the supermarket, based upon Equations 3.7 and 3.8, required inputs for weekly mean and standard deviation of demand. Chapter 4.2.1 introduced the project team's concept of "consumption," a measurement of parts being removed from Machining Center stock locations and transferred to downstream operations. Although this method of computing demand was not used for lot sizing, it was more appropriate than annual usage for the calculation of supermarket reorder points. The reasoning for this was that, unlike production lot sizes, supermarket inventory will vary based upon downstream consumer demand rather than production demand. In addition, calculating reorder points requires consideration of demand as a stochastic rather than deterministic process. Consumption measurement more accurately assesses fluctuations in downstream consumer demand than production demand, which tends to fluctuate only between no demand and demand for the production lot size.

The project team compiled consumption results over the 52-week period starting in June of 2013 and broke it down into both weekly and monthly buckets. An example of monthly data for a high-volume pump head, 405015367, is displayed in Table 5.1. Sample statistics based upon both weekly and monthly data are shown in Table 5.2.

Table 5.1: Consumption data for removal of 404015367 from machine shop , displayed in four-week buckets.

4-wk	Consumption
5/3/14-5/30/14	161
4/5/14-5/2/14	156
3/8/14-4/4/14	299
2/8/14-3/7/14	226
1/11/14-2/7/14	134
12/14/13-1/10/14	111
11/16/13-12/13/13	298
10/19/13-11/15/13	369
9/21/13-10/18/13	521
8/24/13-9/20/13	388
7/27/13-8/23/13	335
6/29/13-7/26/13	257
6/1/13-6/28/13	236

Table 5.2: Statistics for consumption data for 405015367.

Stats				
Weekly			4-wk	
Mean	67.13		Mean	268.54
Stdev	39.33		Stdev	117.07
CV	0.59		CV	0.44
Median	73.50		Median	257.00

Since calculation of R covered one week of lead time, weekly consumption statistics were applied to calculate reorder points.

Equation 3.7 requires inputs for μ^*, the lognormal mean parameter, and σ^*, the lognormal volatility parameter. These can be computed from the mean and standard deviation by manipulating the formulas [16]

$$\sigma^{*2} = \log(CV^2 + 1) \tag{5.1}$$

$$\mu^* = \log(mean) - \frac{\sigma^{*2}}{2} \tag{5.2}$$

The Normal test statistic z in Equation 3.8 is 1.644 for a target service level of 95%.

The values of these parameters for each of the selected part types, and the resulting values of the supermarket reorder points R, are shown below in Table 5.3:

Table 5.3: Supermarket inventory review data and reorder points.

Part No.	Mean	Stdev.	CV	$\mu\_^*$	$\sigma\_^*$	z (95%)	R
405015367	67.13	39.33	0.59	4.06	0.54	1.64	142
405013109	14.32	13.11	0.92	2.36	0.78	1.64	38
405012748	13.63	12.20	0.90	2.32	0.77	1.64	36
405008452	13.40	10.88	0.81	2.34	0.71	1.64	34
289003308	36.02	50.58	1.40	3.04	1.04	1.64	116
wat081193	34.19	53.04	1.55	2.92	1.11	1.64	114
289007447	7.21	10.90	1.51	1.38	1.09	1.64	24
405001315	1.85	6.34	3.43	-0.66	1.60	1.64	7
405007811	6.10	11.28	1.85	1.06	1.22	1.64	22
405008774	4.75	6.54	1.38	1.03	1.03	1.64	15

The inputs to this methodology can be easily changed. R can be based upon an expected lead time of two weeks or four weeks rather than the one week assumed above. Moreover, R can be dynamic, as consumption demand inputs can be changed to always include only the previous 52 weeks. For the purposes of implementation, however, the values in Table 5.3 were the reorder points used.

6   System Implementation

The lot sizing and supermarket methodologies were implemented over a four-week span. Implementing the proposed and modeled system within the Machining Center involved three main steps:

1. Instructing operators to produce parts at the new lot sizes.
2. Establishing a location for the supermarket and instructing the milling utility operator to place excess parts within it.
3. Establishing a method to monitor supermarket inventory.

Implementation required clearance of several logistical hurdles, both with documenting part flow for company records and with establishment of space for inventory and bin locations. However, not only was implementing optimized lot sizing and a supermarket feasible without a major disruption to the shop, but a framework is in place to continue the implementation beyond the time constraints of the project.

6.1   Implementation of Optimized Lot Sizing

Over the course of four weeks of implementation, the project team worked with Production Planning as well as with turning and milling supervisors to place orders for the selected SKUs using optimized lot sizes. For the purposes of the project, orders for the parts were placed into the SAP system and made part of the production schedule, with the optimized lot size values set as the order quantities. When orders for part types were already within the production schedule, those orders themselves were modified to specify desired production lot sizes.

This practice did not hold for all orders, however. All five of the part types chosen for modification were behind Waters' production schedule entering our implementation period. Two in particular, WAT081193 and 405015367, had multiple orders scheduled back-to-back in order

65

to make up for lateness of production. In those cases, we allowed Planning to schedule multiple successive orders in SAP rather than a single order at a larger lot size.

These orders were followed through the turning and milling departments, at which point the milling utility operator was tasked with inputting "partial" orders into SAP. Any order quantities which were originally a part of the production plan continued downstream, while excess parts resulting from modified lot sizes were held back in the supermarket.

In the course of implementation, the actual lot sizes produced did not always equal those specified in the methodology. For a variety of reasons, such as the presence of a few additional inches of bar stock, extra space in the bins, or a desire to avoid new setups preceding shift changes, operators produced either several units below or units above the indicated lot sizes.

In addition, allowing Planning to schedule 405015367 and WAT081193 as successive small-lot orders rather than single large-lot orders resulted in the order schedule being changed without notice. Other orders, based purely upon delivery due dates rather than setup-dependent production sequences, were pushed ahead of orders of 405015367 while it was in the midst of production runs. Thus, although six orders of 405015367 were placed during implementation at a lot size of 92, only one set of three orders were actually carried out back-to-back-to-back to form the designed optimal lot size of 276. WAT081193 was produced at an aggregated lot size of 526 in turning rather than the designed size of 276; although this did not pose any obvious problems during the implementation period, the fact that the schedule changed without the team's approval was an evident barrier to the success of the experiment and limited the number of data points the team was able to acquire. These deviations from the designed lot sizes affected results, although the final conclusions were not very sensitive to these differences (see Chapter 7.1).

6.2   Establishment of a Supermarket Location

In order to establish a supermarket that could persist beyond the time constraints of the project, a modification had to be made in SAP to register its existence within the Waters system. Waters currently keeps records for inventory storage locations embedded in its data management

system in SAP, with part inventory areas referred to as "stock locations." In order to establish the supermarket within SAP, it was necessary to create a new stock location corresponding to it.[20] This way, when the milling utility operator places parts into the supermarket, all he or she needs to do is note in SAP that a certain number of parts have been transferred into that stock location.

The supermarket stock location was created during the course of implementation and took about one week to be ready to use. Logging the transfer of parts between stock locations is already a part of the utility operator's job, so supervisors, machinists and IT staff foresaw no future hurdles with establishing SAP records for the supermarket location or other potential supermarket locations.

As far as the supermarket itself, a cart was set aside to hold bins for the parts under investigation. This cart was placed in the milling area, adjacent to the utility operator incoming rack, so that parts could be transferred easily into it after completing milling operations.

Having a rack in an existing department did pose occasional problems during the implementation period, as milling operators placed partially completed parts in supermarket rather than the milling utility rack; parts were supposed to enter the supermarket only after clearing the milling utility operator. A dedicated inventory area needs to be set up for the supermarket at a convenient location in the shop should implementation continue.

6.3   Monitoring Supermarket Inventory

All of the selected SKUs began the implementation period in backlog due to unforeseen delays. As a result, most produced parts immediately went downstream rather than being held in the supermarket, and the implementation period was not long enough to populate the supermarket and test the effect of the inventory control policy. However, the project team still produced a method for checking supermarket inventory outside the time constraints of the project. To monitor inventory in the supermarket and trigger production starting in turning, the

---

[20] Personal communication, Supervisor of Valve Cell, Waters Corporation, July 18th, 2014.

team produced Kanban cards. Whenever people remove parts from the supermarket for downstream production, they will pull a blank Kanban card from a designated bin and fill it out with the date, time, and number of parts removed before placing the card in another designated bin for the supermarket supervisor. The supervisor can use these cards to determine remaining inventory in the supermarket, subtract from prior inventory values, and trigger production of new orders when inventory for a part drops below its reorder point.

Figure 6.1: Designated bins on supermarket rack for blank and completed Kanban cards.

Figure 6.2: Example of a completed Kanban card.

Inventory monitoring was also established through SAP. The data management system is capable of tracking inventories of all parts in its stock locations, provided that operators record a transaction whenever they remove parts from the supermarket. As a result, the creation of the SAP stock location when establishing the supermarket also provided a method of inventory monitoring and control.

Going forward, in the short term, the department supervisor will monitor the Kanban cards in order to determine when parts need to be produced. Long-term, operators will, when they remove parts from the supermarket, record movement from the supermarket stock location to a new stock location in SAP. Use of the Kanban cards can continue at that time, as a secondary method of determining inventory accuracy.[21]

There are, of course, challenges to the implementation of physical Kanban cards. Their reliability is contingent upon operators' discipline in filling the cards out when they remove parts from the supermarket. They also need to be disciplined in recording part transfers in SAP. A lack of discipline in recording data and communicating part movement between departments of the Machining Center can undermine any scheduling driven by inventory rather than by a forecast. Even if operators are conscientious about filling out Kanban cards and reporting transfers, these acts of communication cost the Machining Center operator time and money.

The potential for human error with the implementation of physical Kanban cards is a reason to advocate for a passive data collection system to monitor supermarket inventory long-term. Possibilities for alternate inventory monitoring methods, such as the RFID system tested by Perez [1], are mentioned in Chapter 8.3.

6.4   Implementation Challenges

A framework exists for both adjusting lot sizes and establishing supermarkets in company documentation. However, beyond the challenges described in the previous sections, three hurdles

---

[21] Personal communication, Supervisor of Valve Cell, Waters Corporation, July 23rd, 2014.

persist in terms of implementation throughout the Waters shop, and can be applied to other manufacturing environments as well. One, as alluded to in Chapters 4 and 5 as well as the previous section, is that there is uncertainty over whether inputs to the lot sizing methodology are correct. There is also uncertainty over whether an active method of data collection, such as physical Kanban cards for monitoring supermarket inventory, will be consistently accurate. While incorrect input data is not a barrier to implementation, it is a barrier to good performance in any environment.

Another challenge surrounds the practice of splitting up lots when they reach the supermarket. During implementation, the team provided the utility operator with instructions to hold partial orders at the supermarket rather than letting them move downstream, and tracked those orders personally. Long-term, however, production systems implementing supermarkets need to consider part orders separately upstream and downstream of the supermarket, since it decouples those production processes.

The final, most difficult barrier to implementation is internal resistance to change. This presented itself multiple times during the implementation period, particularly when dealing with Production Planning. Even when a proposal has potential to improve production, companies with tenured employees and established practices often have the perspective that even minor changes are not worth the time needed for implementation. This perspective exists at Waters, as it does at any company which has seen success doing things a certain way for a long time. In order to implement any system improvements long-term at any factory, the willingness to embrace change needs to be impressed on everyone affected by them. In addition, someone must take over the lot sizing and supermarket methodology following this project and assume the burden of advocating for these changes to continue. The implementation will gain permanence only with a strong advocate and an internal willingness to adapt.

7   Results

This chapter provides the results, in terms of setup hours and productive hours, of producing the selected parts with the optimized lot sizing methodology. It also provides a simulation of productivity with modified lot sizes over a yearlong span.

The project results do not demonstrate the value of supermarkets to on-time delivery or inventory control, as the implementation period was not long enough to populate the supermarket. Exploration of these benefits will be a topic of future research (see Chapter 8.3).

The results presented in this chapter also do not consider the benefits of Puszko's scheduling methodology [2]; they attempt to isolate the effect of lot sizing from the effect of scheduling. Perez's data collection system [1] was not used to obtain the results presented here, as its implementation was in an experimental rather than a practical context.

This chapter also provides scrap rates for the selected parts during the implementation period. As mentioned in Chapter 3.1, a benefit of reducing setups is that operator error becomes less prevalent, which can result in reduced scrap over time. However, the implementation period did not result in enough production runs to verify this hypothesis or produce enough data to simulate it. Tracking scrap rates over more runs will also be a topic of future research.

7.1   Results of System Implementation

Tables 7.1 and 7.2 provide data, obtained through Filemaker, from each production run of each of the five part types with modified lot sizes over the implementation period, including their run times, parts produced and estimated cycle times, and scrap rates through turning and milling. Note that net cycle time refers to cycle time excluding scrap parts.

Table 7.1: Summary of implementation results in CNC turning.

Part material No.	Previous Lot Size	Designed Lot Size	Order Lot Size	Quantity Produced	Good Parts	Scrap Parts	Scrap Rate	Setup Time (Hours)	Processing Time (Hours)	Order Cycle Time (Hours)	Net Cycle Time (Hours)
289003308 (Order 1)	110	165	165	171	165	6	3.51%	3	16	0.09	0.10
289003308 (Order 2)	110	165	165	167	160	7	4.19%	8.5	16	0.10	0.10
WAT081193	250	276	526	524	522	2	0.38%	3	38.5	0.07	0.07
405015367	92	276	276	287	285	2	0.70%	1.5	62.5	0.22	0.22
405013109	50	72	72	72	68	4	5.56%	0.5	26.5	0.37	0.39
405008452	60	76	76	94	85	9	9.57%	2	26	0.28	0.31

Table 7.2: Summary of implementation results in CNC milling.

Part material No.	Previous Lot Size	Designed Lot Size	Order Lot Size	Quantity Produced	Good Parts	Scrap Parts	Scrap Rate	Setup Time (Hours)	Processing Time (Hours)	Order Cycle Time (Hours)	Net Cycle Time (Hours)
289003308 (Order 1)	110	165	165	164	158	6	3.66%	12	40	0.24	0.25
289003308 (Order 2)	110	165	160	161	152	9	5.59%	0	37.5	0.23	0.25
WAT081193	250	276	522	517	516	1	0.19%	5	45.5	0.09	0.09
405015367	92	276	276	284	276	8	2.82%	6.5	51	0.18	0.18
405013109	50	72	72	63	63	0	0.00%	0.5	27.5	0.44	0.44
405008452	60	76	76	82	82	0	0.00%	0.5	20.25	0.25	0.25

These tables do not show setup times; those are displayed in Tables 7.3 and 7.4. It is important to note that, due to the small number of production runs, the implementation of improved scheduling concurrently with optimized lot sizing [2], and failures incurred during some setups, the setup times obtained for these orders deviate significantly form the sequence-independent standards presented in SAP and from the Filemaker data used for the lot-sizing methodology.

However, the purpose of this section is to demonstrate the effect of lot sizing over the course of the implementation period. Therefore, in order to determine how much the increased lot sizes added to available production hours, one must hold every other sequence of events constant and consider the effect lot sizes have on setup time. So, in calculating the hours saved by lot sizing, assume that everything else (part order in the production schedule, operator errors during setup, etc.) would have remained the same under the original lot sizes, except that Planning would have needed to cut additional orders for parts.

Tables 7.3 and 7.4 display the setup times for the orders of the five part types, as well as the projected number of orders whose setups were skipped due to increased lot sizes:

Table 7.3: Setup times for orders completed during implementation in CNC turning.

Part material No.	Previous Lot Size	Turning Setup Time (Hours)	Quantity Produced	Est. Setups Skipped	Turning Processing Time (Hours)
289003308 (Order 1)	110	3	171		16
289003308 (Order 2)	110	8.5	167	1	16
WAT081193	250	3	524	1	38.5
405015367	92	1.5	287	2	62.5
405013109	50	0.5	72	1	26.5
405008452	60	2	94	0	26

Table 7.4: Setup times for orders completed during implementation in CNC milling.

Part material No.	Previous Lot Size	Milling Setup Time (Hours)	Quantity Produced	Est. Setups Skipped	Milling Processing Time (Hours)
289003308 (Order 1)	110	12	164		40
289003308 (Order 2)	110	0	161	1	37.5
WAT081193	250	5	517	1	45.5
405015367	92	6.5	284	2	51
405013109	50	0.5	63	1	27.5
405008452	60	0.5	82	0	20.25

Table 7.5 provides an estimate of the setup hours needed to produce at least as many parts under the original lot sizes as were produced during the implementation period, and assesses the impact of the lot sizing methodology by directly estimating setup hours saved.

Table 7.5: Estimation of benefits from adjustment of lot sizes.

Part material No.	Est. Setups Skipped	Avg. Turning Setup Time (Hours)	Avg. Milling Setup Time (Hours)	Est. Setup Hours Saved	Turning Processing Time (Hours)	Milling Processing Time (Hours)
289003308	1	5.75	6	11.75	32	77.5
WAT081193	1	3	5	8	38.5	45.5
405015367	2	1.5	6.5	16	62.5	51
405013109	1	0.5	0.5	1	26.5	27.5
405008452	0	2	0.5	0.0	26	20.25
Total	5	12.75	18.5	36.75	185.50	221.75

As the tables show, using the larger lot sizes over four weeks of production on five parts and four machines resulted in an estimated aggregated savings of 36.75 hours of setup time over 407.25 productive hours of processing time. Thus, the lot sizing methodology caused as much as a 9% reduction in hours needed to produce these parts.

It should be noted, however, that standard setup times for all of these parts are shorter in milling than they are in turning. However, during the implementation period, setup times were higher in milling due to tool failures or other operator-reported problems. If this methodology were to be carried out over a longer period of time, we would expect that savings in setup hours would skew more heavily towards the turning department compared to milling than they do in our implementation results.

7.2   Simulation of One Year of Optimized Lot Sizing

Although the implementation demonstrated reduction in setup time, it did not show the long-term benefit of increasing lot sizes in the Waters manufacturing environment. It also did not demonstrate the value derived from having increased production while still meeting the demand of all parts within the capacity constraints of a given machine, or estimate the cost of additional inventory accrued due to higher production rates.

The following simulation takes setup and cycle times for each part going through one of the four machines investigated during the project, M1309. The production over time of each part on the machine, relative to its demand, is plotted hourly over the course of one year (52 weeks). The simulation is run with two different choices for the lot sizes of 405015367, 405008452 and 405013109: the original ones and the ones chosen in Chapter 4. Its purpose is to show the estimated difference in hours of productive output over a year, as well as to show the difference in ability to meet demand for all parts given the capacity of the machines.

The simulation was not run for the other three machines investigated due to their preponderance of part types. While M1309 makes only ten SKUs with an annual usage of over 300 parts, M1302-2 has 40 such parts, M1210 has 56, and M1211 has 28. Since the simulation is

meant as a simple demonstration of the value of lot sizing, it was unnecessary to run it with inputs for such high part mixes. If one were to run it for another machine, similar results could be expected for the other lathe, M1302-2, since the lot sizing methodology was based upon demand through turning. Lot sizes are not optimized with respect to the mills, so the results may not be as powerful should the simulation be run for M1210 or M1211.

The simulation is based on several assumptions. The first is that demand for each part is constant and equivalent to the annual usage value used in Chapter 4 divided into hourly rates. This is not realistic, as true demand is stochastic, but the conclusions reached by the simulation are not very sensitive to demand fluctuations.

Another assumption is that although the lot sizes calculated in Chapter 4 are based upon Filemaker data, the setup and cycle times in the simulation will equal the standard times present in Waters' SAP system. This assumption introduces data which is known to be obsolete; however, Filemaker data is not readily accessible for all parts. The SAP data will at least be accurate enough for an informed estimate of savings; as more data gets collected, both by machinists and by a potential automated system [1], the input parameters to the simulation can be changed.

This assumption also introduces the problem of using different data sources for different parameters, since setup and process time inputs are based on SAP and lot sizes are based on Filemaker. However, the simulation is meant to demonstrate the effect of the newly imposed lot sizes to Waters, so the lot sizes used for it must be the ones actually used for implementation rather than those computed by the Pinedo methodology with SAP data.

Setups are also assumed to be sequence-independent for each part type. In reality, Puszko improves scheduling by accounting for sequence-dependent setups [2], but our goal here to isolate the effects of lot sizing from those of scheduling.

The simulation does not consider production outside of turning, so for its purposes, assume the supermarket exists at the end of turning. In this scenario, it serves only as an inventory location.

Finally, the simulation does not include the effects of the continuous review inventory policy. Since demand is assumed to be constant, incorporating continuous review would be redundant, as parts would be produced in regular intervals just as in a push system.

The outputs of the simulation are displayed below; Matlab code for it is available in the Appendix.

7.2.1 Increase in Productive Hours

Table 7.6: Simulation part types and lot sizes.

Part material No.	Standard Setup (hours)	Standard Processing (hours)	Annual Usage	Original Lot Size	Adjusted Lot Size
405001659	4	0.12	873	125	125
405003366	2	0.137	2,880.00	320	320
405005119	2	0.16	9,828.00	351	351
405005223	2	0.06	922	131	131
405008452	4	0.38	1,080.00	60	76
405012748	4	0.19	596	50	72
405013109	4	0.38	1,032.00	50	50
405013611	2.5	0.15	300	50	50
405015367	4	0.25	3,687.00	92	276
WAT057209	5	0.17	455	100	100

Table 7.7: Aggregate output for M1309 from simulation.

	Average Inventory	Time of Simulation (hours)	Setup hours	Productive hours	Setup/Productive
Original	539.5	7034.3	526	4108.3	0.13
Optimized	616.4	7020.5	382	4188.5	0.09
Difference with Optimized	76.9	-13.8	-144	80.2	-0.04

Table 7.7 shows an aggregate increase of 80 productive hours due to the new lot sizes, as well as 144 hours gained through reduction of setups. Those 144 hours can be converted into productive hours by adding more SKUs to the machine, and account for about 9% of the machine's annual output. Considering the internal accounting credits imposed by each department to Waters ($67 per hour for turning[22]), this increase will result in additional internal accounting credits of about $15,000.

It should be remembered, as well, that this is a conservative estimate. The simulation results in nearly 2500 hours of idle time (the simulation covers 7000 hours, but setup and production times only account for about 4500 hours), so if Waters can achieve machine utilization higher than 65%, The increase in accounting credits should be even higher than estimated here.

Note also that this simulation only covers one machine and three adjusted parts. Depending on part mix and setup and run times, consistency of demand and other unaccounted-for factors, the results for other machines will vary from those for M1309. However, if we take M1309 as an average representative of machines in the turning and milling departments, and estimate that similar effects can other with the lot sizing methodology on all fourteen machines across the two departments, the estimate increases to $210, 000 annually in internal credits. If more parts from each machine get added to the methodology, that figure can climb even higher.

7.2.2 Increased Inventory in the Supermarket

The side effect of increasing lot sizes is increasing inventory. Cumulative production less cumulative demand equals inventory in the simulation. Table 7.7 also provided a cumulative average inventory estimate for both original and adjusted lot sizes for M1309.

As the table shows, the cumulative average inventory following M1309 over one year is estimated to increase by about 77 units. Using the conservative estimate for inventory holding

---

[22] Personal communication, Supervisor of Valve Cell, Waters Corporation, March 21st, 2014.

cost established in Chapter 4 ($1 per part per week), the holding cost over one year will be about $4,000, nearly four times less than the savings in added production hours estimated in Chapter 7.2.1. Using an estimate that Waters machine shop supervisors consider closer to the true cost[23], about 10% of the lot sizing estimate, that inventory holding cost will only be about $400.

Again, the simulation used obsolete data inputs and only serves as a demonstration of how increased lot sizes for high-volume parts can provide value to Waters. However, it does show that the benefits in added available production hours from optimized lot sizing far outweigh the costs of adding inventory.

---

[23] Personal communication, Supervisor of Valve Cell, Waters Corporation, June 17th, 2014.

8   Conclusions and Future Work

8.1   Conclusion

The team was given a project goal of realizing an improvement in productivity to help the Milford facility increase its annual output. We opted to increase productivity by reducing overall setup time through adjustments to lot sizes, allowing Waters to produce to demand within the capacity constraints of their turning and milling departments. Using a methodology to optimize production lot sizes, considering a multi-part system with high setup times and low inventory holding costs, this project achieved a total estimated reduction of 36.75 setup hours over 407.25 productive hours on five selected part types in the Waters turning and milling departments. These savings can directly be converted into production hours for other part types.

A simulation for these effects using SAP and Filemaker data inputs predicts that operating with modified lot sizes for the selected part types can result in 224 additional productive hours over the course of a year, or nearly 10% of annual output gained.

Finally, implementation of a supermarket has allowed production of optimized lot sizes in turning and milling without disrupting downstream production, and having a supermarket in place will control inventory at levels which justify conservative estimates for increased inventory holding costs.

8.2   Recommendations

The recommendation going forward is that Waters should maintain the specified lot sizes and the supermarket for the part types selected for this project, and should further apply the lot sizing and supermarket methodologies to other high-volume parts on all machines in turning and milling. Any parts which see high-volume production and are currently produced with repeated setups are good candidates for the optimization methodology. Waters should ensure that standard lot sizes are modified in SAP to reflect these changes for Production Planning and finance.

In addition, an intermediary supermarket should be established between turning and milling, separate from the one implemented after milling in this project. It was mentioned in Chapter 4 that applying the same lot sizes to turning and milling did not actually optimize production in milling. The lot sizes were only manageable because milling sees low utilization anyway. In order to decouple productive output and provide more accurate predictions for each department, turning and milling will need separate supermarkets.

Finally, it will be necessary to institute a part number break in SAP for part types in the supermarket[24]. Managing lot sizes which are consistently different upstream and downstream of the supermarket will require breaking orders off and sending partial orders downstream again and again. To institute a permanent, systematic way of decoupling lot sizes, a part number break will be the best solution, based upon discussions with Production Planning. Under a part number break, a component will be considered as two separate parts upstream and downstream of the supermarket, and orders for those two parts will be made separately of each other.

8.3   Future Work

Future work can take place with regard to the lot sizing and supermarket methodologies described in this thesis. Investigations during the project revealed other avenues for future work as well, most prominently the use of the utility operator and implementation of efficient data collection and production scheduling.

8.3.1 Lot Sizing and Supermarkets

With respect to optimized lot sizing and supermarkets, Waters can perform additional research elsewhere in the Machining Center. The lot sizing methodology can be applied in downstream departments such as the valve cell or column cell; it may be that adjustments to lot sizes are necessary in those departments as well.

---

[24] Personal communication, production planning group, Waters Corporation, July 8th, 2014.

Another point of interest to Waters is the long-term effect of larger lot sizes on scrap rates for selected part types. The change in scrap over an extended period of time (at least six months) with new lot sizes will need to be tracked to obtain useful statistics to answer this question.

Future work will also involve collecting data on the effects of supermarkets. Waters can monitor parts, either in the supermarket established in this project or elsewhere, and determine the impact the supermarket has on downstream order lead time and on-time delivery.

Waters can also investigate alternate methods for managing supermarket inventory than the one proposed in Chapter 5. Due to the high setup and processing times associated with the turning and milling departments, a fixed order quantity model would be preferable to a base stock policy. However, the continuous review model can be modified so that reorder points cover different lead times than the one week specified in Chapter 5. The supermarket may be more effective at handling part withdrawal if it holds enough inventory for two or four weeks of lead time instead, given the perception that the lot sizing methodology in Chapter 4 underestimated inventory holding cost. Another alternative is that the supermarket can serve as a push-pull boundary, with production in turning and milling operating according to a forecast and downstream production pulling from the supermarket. This provides less inventory control; however, if inventory costs are sufficiently low, this system can still work for high-volume parts.

8.3.2 Part Queueing Times at Utility Operator Stations

Regardless of whether a part in turning or milling requires secondary operations or not, Waters policy dictates that all parts going through those departments must be signed off by the department utility operator before moving downstream. The utility operator is exclusively responsible for verifying completion and scrap rates and logging those numbers into SAP.

The problem with this arrangement is that only one utility operator is employed in each department, and that operator only works an eight-hour shift. This means that for 16 hours a day, parts may be completed and ready to move on to their next operation, but must sit idle until the utility operator comes in the next day and moves them. In this way, the utility operator can be

81

thought of as an inefficient, bottlenecking machine in the production line (see Figure 8.1). Unfortunately, this phenomenon is not considered in SAP scheduling, causing longer lead times than predicted by SAP, reducing on-time delivery and necessitating daily schedule adjustments.

Figure 8.1: Model of production line including the utility operator as a separate machine, a bottleneck between upstream and downstream.

As a part of its future work, Waters needs to reconsider the role of the utility operator in the turning and milling departments. One suggestion as to how the position can more efficiently be used is requiring machinists to transfer completed parts downstream when they do not need secondary operations, leaving the utility operator to focus on specialized secondary machining. Another suggestion is that two additional utility operators can be hired for each department, ensuring the station is staffed at all hours and enabling part flow during the night shift.

8.3.3 Data Collection and Scheduling

This thesis repeatedly mentioned that a limitation to the lot sizing methodology, as well as the calculation of supermarket reorder points, was uncertainty over whether data inputs to the equations were correct. Waters needs more accurate, more consistent, more standardized data on setup and process times as well as demand, to ensure inputs to the lot sizing methodology are correct, to track supermarket inventory and to achieve benefits outside the scope of this thesis, such as tool life and mean times to machine failure or repair. There are a number of methods detailed in Chapter 5 of Perez [1] for collecting this data, including automated methods such as power and auto RFID monitoring, computer vision software and barcode scanning.

From a scheduling standpoint, Puszko [2] discusses how sequence-dependent scheduling can cause a reduction in setup time. Waters will need to fill out a setup matrix to get setup times and distributions for all part types in a given family, and to adjust their scheduling procedures to take advantage of the increased utilization and capacity [18] this can bring.

References

[1] Perez, A. A., 2014, "The Value of Accurate Automated Data Collection to Manufacturing," M.Eng. Thesis, Massachusetts Institute of Technology, Cambridge, MA.

[2] Puszko, G., 2014, "Efficient Scheduling to Reduce Setup Times and Increase Utilization in a Multiple-Part Manufacturing System," M.Eng. Thesis, Massachusetts Institute of Technology, Cambridge, MA.

[3] Józefowska, J., 2007, *Just-in-Time Scheduling: Models and Algorithms for Computer and Manufacturing Systems*, Springer, New York, Chap. 1.

[4] Wangphanich, P., Kara, S., and Kayis, B., 2010, "Analysis of the Bullwhip Effect in Multi-Product, Multi-Stage Supply Chain Systems – a Simulation Approach," International Journal of Production Research, 48(15), pp. 4501-4517.

[5] Hopp, W. J., and Spearman, M. L., 2011, *Factory Physics*, 3rd Ed., Waveland Press, Long Grove, IL, 720 p.

[6] Simchi-Levi, D., Kaminsky, P., and Simchi-Levi, E., 2008, *Designing and Managing the Supply Chain: Concepts, Strategies, and Case Studies*, 3rd Ed., McGraw-Hill/Irwin, Boston, MA, 498 p.

[7] Feng, W., Zheng, L., and Li, J., 2012, "Scheduling Policies in Multi-Product Manufacturing Systems with Sequence-Dependent Setup Times and Finite Buffers," International Journal of Production Research, 50(24), pp. 7479-7492.

[8] Gershwin, S. B., 1994, *Manufacturing Systems Engineering*, Prentice Hall, Englewood Cliffs, NJ, xix, 501 p.

[9] paulmaguirehome, 2011, "Stockless Production Part 1" [Online], YouTube, Available: http://www.youtube.com/watch?v=yZDRB_z51MQ [2014, March 21].

[10] Gupta, S., 2013, "System Improvements in Valve Manufacturing Cell at Waters Corporation," M.Eng. Thesis, Massachusetts Institute of Technology, Cambridge, MA.

[11] Monden, Y., 2011, *Toyota Production System: An Integrated Approach to Just-in-Time,* 4th Ed., CRC Press, Boca Raton, FL, 566 p.

[12] Faccio, M., Gamberi, M., and Persona, A., 2013, "Kanban Number Optimisation in a Supermarket Warehouse Feeding a Mixed-Model Assembly System," International Journal of Production Research, 51(10), pp. 2997-3017.

[13] Jackson, J., 1998, "Effectively Implementing Vendor Managed Inventory," Pioneer Standard Electronics [Online], Available: http://www.ism.ws/files/Pubs/Proceedings/JacksonEE.pdf [2014, June 29]..

[14] Setyaningsih, S., and Basri, M. H., 2013, "Comparison Continuous and Periodic Review Policy Inventory Management System Formula and Enteral Food Supply in Public Hospital Bandung," International Journal of Innovation, Management and Technology, 4(2), pp. 253-258.

[15] Zhao, X., Fan, F., Liu, X., and Jinxing, X., 2007, "Storage-Space Capacitated Inventory System with (r,Q) Policies," Operations Research (Informs), 55(5), p. 855.

[16] De Treville, S., Bicer, I., Chavez-Demoulin, V., Hagspiel, V., Schurhoff, N., Tasserit, C., and Wager, S., 2014, "Valuing Lead Time," SSRN, New York, SSRN-id2431698.

[17] Bozarth, C., 2011, "Periodic Review System: Inventory Management Models: A Tutorial," NC State University [Online], Available: http://scm.ncsu.edu/scm-articles/article/periodic-review-system-inventory-management-models-a-tutorial [2014, July 30].

[18] Pinedo, M. L., 2005, Planning and Scheduling in Manufacturing and Services, 2nd Ed., Springer, New York, pp. 144-151, Chap. 7.

[19] Arrow, K. J., Harris, T., and Marshak, J., 1951, "Optimal Inventory Policy," Econometrica, 19(3), pp. 250-272.

[20] Harris, F. W., 1990 [Reprint from 1913], "How Many Parts to Make at Once," Operations Research (Informs), 38 (6), pp. 947–950.

Appendix

Matlab Code for Simulation

```
    % Lot sizing simulation for M1309

clear;
clc;

% There are ten SKUs going through M1309 with an annual usage of 300 parts
% or higher. This simulation will include only those parts.

% INPUTS: Note that since setup and machine cycle times are given in hours,
% demand rates will also be in hours.

%   Demand of each SKU in parts per hour, divided from annual usage assuming
%   constant demand. Note that machine capacity is 135 hours per week.
D5119 = 9828/52/135;
D5367 = 3687/52/135;
D3366 = 2880/52/135;
D8452 = 1080/52/135;
D3109 = 1032/52/135;
D5223 = 922/52/135;
D1659 = 873/52/135;
D2748 = 596/52/135;
D7209 = 455/52/135;
D3611 = 300/52/135;

% Demand vector to loop through. Note that the list is ordered so that
% parts with the highest demand come first and will have priority in order
% scheduling.
demand = [D5119 D5367 D3366 D8452 D3109 D5223 D1659 D2748 D7209 D3611];


% Order quantity (lot size) of all SKUs. Note that 5367, 8452 and 3109
% have increased lot sizes in the thesis methodology, so their order
% quantities will vary depending on the simulation run.
Q5119 = 351;
Q5367 = 92 or 276;
Q3366 = 320;
Q8452 = 60 or 76;
Q3109 = 50 or 72;
Q5223 = 131;
Q1659 = 125;
Q2748 = 50;
Q7209 = 100;
Q3611 = 50;

% Lot size vector to loop through, ordered the same as the demand vector.
lot_size = [Q5119 Q5367 Q3366 Q8452 Q3109 Q5223 Q1659 Q2748 Q7209 Q3611];
```

```
% Cycle time in hours of all SKUs, from SAP.
C5119 = .16;
C5367 = .25;
C3366 = .137;
C8452 = .38;
C3109 = .38;
C5223 = .06;
C1659 = .12;
C2748 = .19;
C7209 = .17;
C3611 = .15;

% Cycle time vector to loop through, ordered the same as the demand vector.
cycle_time = [C5119 C5367 C3366 C8452 C3109 C5223 C1659 C2748 C7209 C3611];

% Setup time in hours of all SKUs, from SAP.
S5119 = 2;
S5367 = 4;
S3366 = 2;
S8452 = 4;
S3109 = 4;
S5223 = 2;
S1659 = 4;
S2748 = 4;
S7209 = 5;
S3611 = 2.5;

% Setup time vector to loop through, ordered the same as the demand vector.
setup_time = [S5119 S5367 S3366 S8452 S3109 S5223 S1659 S2748 S7209 S3611];


% INITIALIZED VALUES

% Let us specify the tracking period to be one year, or 135*52 hours.
tracktime = 135*52;

% Cumulative production and demand values at any given time for all SKUs.
cum_prod = zeros(1, length(demand));
cum_demand = zeros(1, length(demand));
%cum_demand = [9828 3687 2880 1080 1032 922 873 596 455 300];

% Cumulative production vectors of all SKUs, measured each hour.
CP = zeros(length(demand), tracktime);

% Cumulative demand vectors of all SKUs, measured each hour.
CD = zeros(length(demand), tracktime);

% Let us also keep an hourly tab of the time passed, starting with hour 0.
time = zeros(1, tracktime);

% Initialize time, setup hours and productive hours. Also initialize vector
% for sum of cumulative production less cumulative demand for all part
% types, measured at each time instance.
t = 0;
setup_hours = 0;
```

```matlab
run_hours = 0;
cum_inv = zeros(1, tracktime);

% Run production of parts in a loop until a year has passed.
while(t<tracktime)

    % Search through each SKU to see if it needs to be produced to meet
    % demand. The search will begin with the highest-demand SKU at the
    % beginning of each instance of the while loop, but will search each of
    % the ten SKUs before restarting the loop.
    for i=1:length(demand)

        % Produce if cumulative production has not met cumulative demand.
        if cum_prod(i)<=cum_demand(i)
            % Save the most recent moment in time to decrement demand.
            t_old = t;

            % Increment time by time for setup of the SKU.
            t_setup = t_old + setup_time(i);

            if t_setup > tracktime
                break
            end

            setup_hours = setup_hours + setup_time(i);

            % Increment demand for all SKUs by the demand over SKU i's
            % setup time.
            for j=1:length(demand)
                cum_demand(j) = cum_demand(j) + demand(j)*(t_setup-t_old);
            end

            % Increment the time, production and demand lists over the
            % setup time for the SKU.
            for j=ceil(t_old):ceil(t_setup)
                time(j+2) = time(j+1) + 1;
                cum_inv(j+2) = sum(cum_prod-cum_demand);

                for k=1:length(demand)
                    CP(k,j+2) = CP(k,j+1);
                    CD(k,j+2) = CD(k,j+1)+demand(k);
                end
            end


            % Now we consider the production of the SKU, separately from
            % its setup. Increment time by the run time of the SKU
            t = t_setup + lot_size(i)*cycle_time(i);

            if t > tracktime
                break
            end

            % Increment demand for all SKUs by the demand over SKU i's
```

```matlab
        % run time.
        for j=1:length(demand)
            cum_demand(j) = cum_demand(j) + demand(j)*(t-t_setup);
        end

        % Increment production of SKU i by its lot size.
        cum_prod(i) = cum_prod(i) + lot_size(i);

        run_hours = run_hours + lot_size(i)*cycle_time(i);
        % Increment the time, production and demand lists over the
        % run time for the SKU.
        for j=ceil(t_setup):ceil(t)
            time(j+2) = time(j+1) + 1;
            cum_inv(j+2) = sum(cum_prod-cum_demand);

            for k=1:length(demand)

                % Increment cumulative production hourly for the SKU
                % being made, and hold other production constant.
                if k==i
                    CP(k,j+2) = CP(k,j+1)+lot_size(i)/(ceil(t)-
ceil(t_setup));
                else
                    CP(k,j+2) = CP(k,j+1);
                end

                CD(k,j+2) = CD(k,j+1)+demand(k);
            end
        end

    % If a part does not have demand for production, make nothing and
    % add an hour of demand to the system.
    else
        t = t + 1;
        for j=1:length(demand)
            cum_demand(j) = cum_demand(j) + demand(j);
        end

        for j=ceil(t-1):ceil(t)
            time(j+2) = time(j+1) + 1;
            cum_inv(j+2) = sum(cum_prod-cum_demand);

            for k=1:length(demand)
                CP(k,j+2) = CP(k,j+1);
                CD(k,j+2) = CD(k,j+1)+demand(k);
            end
        end
    end
        end
    end
end

setup_hours
run_hours
A = sum(cum_prod - cum_demand)
t
```

```matlab
% Eliminate trailing zeros from the inventory output.
time = time(1:find(time,1,'last'));
cum_inv = cum_inv(1:find(cum_inv,1,'last'));

mean(cum_inv)


plot(time,cum_inv)
title('Cumulative Inventory over Time, M1309, Original Lot Sizes')
xlabel('time (hours)')
ylabel('production-demand (parts)')
```