"Islands of Simulation": Using Simulation as a Decisions Support Tool for Long-term Layout Planning

by

Caryl Burnette Brown

B.S. Material Science and Engineering, University of Florida (1992)
Minor, Business Administration, University of Florida (1992)

Submitted to the MIT Department of Material Science and Engineering
and the Alfred P. Sloan School of Management at MIT
in Partial Fulfillment of the Requirements for the Degrees of

Master of Science in Materials Engineering

and

Master of Science in Management

In conjunction with the
Leaders for Manufacturing Program
at the
Massachusetts Institute of Technology
June 1995

Signature of Author ______________________________
Department of Material Science and Engineering
MIT Sloan School of Management
May 12, 1995

Certified by ______________________________
Anantaram Balakrishnan, Associate Professor
Thesis Supervisor

Certified by ______________________________
Stuart Brown, Associate Professor
Thesis Supervisor

Accepted by ______________________________
Carl V. Thompson II
Professor of Electronic Materials
Chair, Departmental Committee on Graduate Students

Accepted by ______________________________
Jeffrey A. Barks
Associate Dean, Master's and Bachelor's Programs
MIT Sloan School of Management

"Islands of Simulation": Using Simulation as a Decisions Support Tool for Long-term Layout Planning

by

Caryl Burnette Brown

Submitted to the Department of Material Science and Engineering and
the Sloan School of Management on May 12, 1995 in partial
fulfillment of the requirements for the Degrees of

Master of Science in Materials Engineering
and
Master of Science in Management

ABSTRACT

A plant producing aluminum extrusions and casting parts for an all-aluminum spaceframe car was facing the need to increase capacity of its plant. Having decided that the machining area was the bottleneck, the plant decided to add more CNC high-speed milling machines (HSMs) and an automated material handling system using automated overhead robotic cranes. Queueing and simulation models were developed and used to evaluate the feasibility of a potential layout of the new machining area and the material handling system proposed for the area. The models assisted with both the hard-system configuration decision and also some soft-system configuration decisions regarding dispatching and scheduling rules of the automated material handling system. Both the queueing model and the various versions of the base simulation model judged the proposed layout and material handling system as feasible. Two job dispatching heuristics for the robotic cranes which we call proactive and reactive were studied, with the proactive heuristic delivering higher service levels to the HSMs than the reactive. The tradeoff between dynamic and static part assignment was also investigated and found that increases in part assignment flexibility did not improve significantly the amount of machining capacity lost due to waiting on the material handling system. The system was found to be relatively robust to partial failures experienced by the material handling system. Some generic learnings comparing the approach of using localized "islands of simulation" vis-à-vis plant-wide, global models are discussed, along with some insight gained regarding some of the people-related issues of gaining confidence in simulation models as a decision aide for layout planning.

Thesis Advisors:
Anantaram Balakrishnan, Associate Professor, MIT Sloan School of Management
Stuart Brown, Associate Professor, MIT Department of Material Science & Engineering

Table of Contents

Table of Figures

Acknowledgments

I wish to gratefully acknowledge the support and resources made available to me for this work through the Leaders for Manufacturing Program, a partnership between the Massachusetts Institute of Technology and major United States manufacturing companies.

I would like to express my tremendous gratitude to Mohammad Zaidi, Michael Clifford, Rod Heiple, Michael Brandt, Earl Lockwood, Benno Schlatmann, Jürgen Weiß, and the rest of the P&T group for the wonderful assistance and support I received during the course of my internship at their plant.

Separate thanks go out to Sven, Johann, Benny, Roger, Axel, Georgio, and the Lippstadt Baseball Manschaft for many hours of good laughs, good pizza (Didomino's at 12:30 a.m.), good beer, and, most importantly, great friendship while I was in Germany. I will never forget you guys!

A special thank you goes to my thesis advisors, Anant Balakrishnan and Stu Brown. Your input and feedback were always valuable, helpful, and appreciated. The extra mile you both went (no pun intended) given my internship was an ocean away went above and beyond the call of duty. I have learned much from both of you.

A bit of thanks goes to the technical support staff at Imagine That!, especially Dave Krahl and J. Steven Lamperti. Their key insights with the software package I was using assisted me greatly in my modeling efforts. I would also like to thank Rhonda Reynolds, George Lanphere, and David O'Keefe at The Boeing Company for their assistance in understanding better the machining of aluminum.

I want to thank my fellow Fellows who have been a great source of support and who have helped me keep things in perspective during these somewhat stressful two years. I have learned a great deal beyond what is taught in the courses from all of my friends in the LFM Program.

On a more personal note, a bit of thanks goes to my two roommates Anand Mehta and Eric Roche for putting up with my frequent "dèplace de l'air" and for allowing me the use of Anand's computer for typing this thesis into the wee hours of the late evening.

Finally, I wish to extend an eternal debt of gratitude to my parents Gary and Pat and my best friend and lovely girlfriend, Susan Ipri. Mom and Dad, I would never have had the opportunity to achieve the accomplishments I have without your continued love, guidance, and support for all these years. Susan, you have been the greatest blessing for

which a guy could ever hope. Thanks for putting up with me -- then, now, and in the distant future. I can only hope I have been as good to you as you have to me. I love you.

Dedication

I would like to dedicate this thesis to the most

important influences in my life:

My phenomenal family

and

my most marvelous girlfriend and bestest buddy, Susan Lynn Ipri.

Thank you both for making my life as excellent as it has been.

Introduction and Overview

"What we experience of nature is in models, and all of nature's models are so beautiful."

-R. Buckminster Fuller (Hoffman, 1994)

This thesis presents a detailed example of applying queueing and simulation models to evaluate the feasibility of a proposed machining cell layout with its associated material handling system. An aluminum parts manufacturer knew it had to add additional in-house machining capacity to its capital base. Plant personnel created a proposed layout for these new machines and included with it an automated material handling system involving an overhead crane system. The plant had difficulties, however, in assessing how well the proposed system would perform. Specifically, plant personnel were concerned that machining capacity would be lost due to machines waiting to be served by the material handling system. Estimating this factor incorrectly could cost the plant millions of dollars in idle capacity or could cause the plant to starve its sole customer.

To deal with these concerns, a queueing model and multiple simulation models were developed to help predict the amount of machining capacity that would be lost due to machines waiting on the material handling system. The thesis describes why the machining process could not simply "go faster," and then focuses on the development process used for the models along with their results. The thesis addresses issues associated with assumptions involved in the models and the process used to verify and validate the models. Plant-specific findings and more generic insights gained are presented.

This thesis serves as a good example of how queueing and simulation models can together be a valuable decision support tool when dealing with plant layout issues. Many of the insights gained from this work are applicable to a variety of situations and circumstances where modeling might be found useful. As manufacturing systems become more complex and automated, the techniques and tools used in this thesis will become more and more necessary to aid intelligent layout and system design decision making.

Chapter 1 describes the products and process flow involved in the plant where the work for this thesis occurred. The chapter highlights some of the details of each process step. It also discusses briefly some of the challenges or limitations faced by each process step.

In Chapter 2, the problem faced by the plant is described. This problem is that the plant must add machining capacity but does not know how to determine the expected performance level of a particular layout and material handling system. Material-based and machine-based limitations to machining speeds of aluminum are also presented in this chapter, with the true limiting factors being machine-based.

The literature review is presented in Chapter 3. The review is conducted in two segments, the first of which presents a basic approach to simulation along with different classification schemes for simulation models. The second section briefly reviews a selection of previous research completed in the areas of material handling modeling and crane scheduling.

The primary chapter of this thesis is Chapter 4. In this chapter, the specifics of the proposed machining area and its material handling system are described. A queueing theory model is then developed and its many assumptions and limitations discussed. Following the presentation of the queueing model's results, the simulation model and its results are presented. First, the structure and underlying assumptions of the base case model are presented. The method used for data collection is outlined, followed by a discussion of different versions of the base case model that test various elements of the system being modeled. These elements include the flexibility of part assignments to machines, system robustness, and crane dispatching heuristics. The results of these different models are presented along with steps taken to verify and validate the model.

Chapter 5 describes insights and inferences gained from the work done for this thesis at two levels. The first level at which insights are presented is the plant-specific, project-specific level. The second, more important level is the generic, more widely applicable level. Some of these insights are related to using "islands of simulation" vis-à-vis a single, plant-wide simulation. Other points raised deal with gaining end-user confidence in the model's results. In the final conclusion, simulation was found to require some degree up-front investment in learning and becoming proficient with a simulation package, but reaped significant rewards as a decision support tool once that investment is made.

Chapter 1: Products and Processes

1.1 Introduction

This chapter will briefly introduce the products and processes involved at the plant where the problem studied in this thesis persisted. This background is useful in understanding the situation that drove the project.

This project involves a start-up automotive supplier. The facility currently being used has been in operation for less than one year. The primary customer for this plant was steadily increasing its daily production of automobiles.

1.2 Products and Process Flow

This auto supplier plant produces primarily two types of products, both of which require the use of newly developed process technologies. The first type of product is relatively large vacuum-die-cast aluminum castings. The second type of product is thin-walled aluminum extrusions with complex geometries and tight tolerances. Together, these parts are joined to form an all-aluminum automobile chassis utilizing a spaceframe approach as opposed to the traditional monocoque approach. To ensure high product quality, tight dimensional tolerances must be met, and key physical and mechanical properties must be maintained. The customer for these parts is a high-end automobile original equipment manufacturer (O.E.M.). The O.E.M. uses the spaceframe produced with the parts as the basis for a high-end yet light-weight luxury sedan.

Thirty-five types of die cast parts are cast on site or off-site and finished on-site. For these parts, the plant receives billets, or metal bars, of a special alloy. The billets are melted and then cast. Following a trimming operation, the parts are deburred by passing through an abrasive medium on a vibrating table. The parts are placed into baskets that travel on a monorail system throughout the plant. The baskets are grouped into groups of 6 (called 6-packs) and sent through a solutionizing heat treatment and then handstraightened. Following the straightening operation, the parts are aged and then receive a surface treatment by being submerged in degreasing and etching baths. The parts are then inspected, both mechanically and visually, and shipped. This process flow is summarized in Figure 1.1.

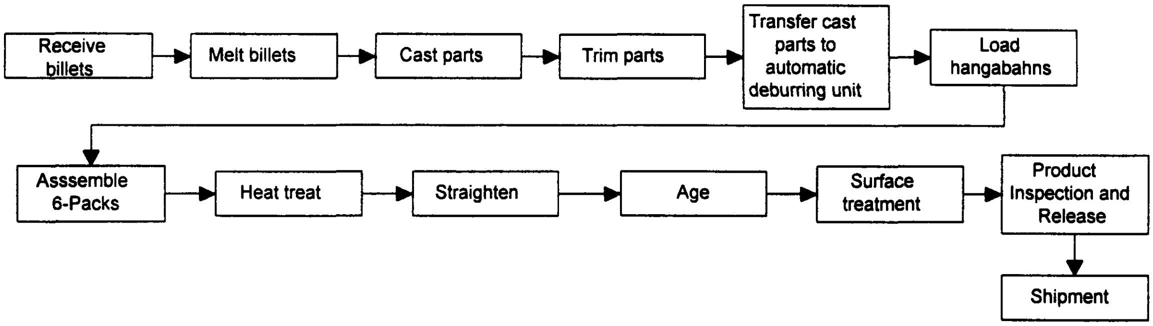


Figure 1.1: Vacuum die cast parts' process flow.

The 43 different extruded parts have a separate process flow. Straight extrusions are received from an external supplier and immediately marked with a unique serial number for traceability purposes. After marking, some parts go directly to a machining step (sometimes off-site), while others are first bent and possibly sawed before processing at the 5-axis high speed milling machines. (The machining process will be described in greater detail later in this chapter.) Following machining, parts are manually deburred and then cleaned with a parts washer. The cleaned parts are aged and then receive a surface treatment similar in nature to the die cast parts. The final step is product inspection and shipment. Figure 1.2 summarizes these steps.

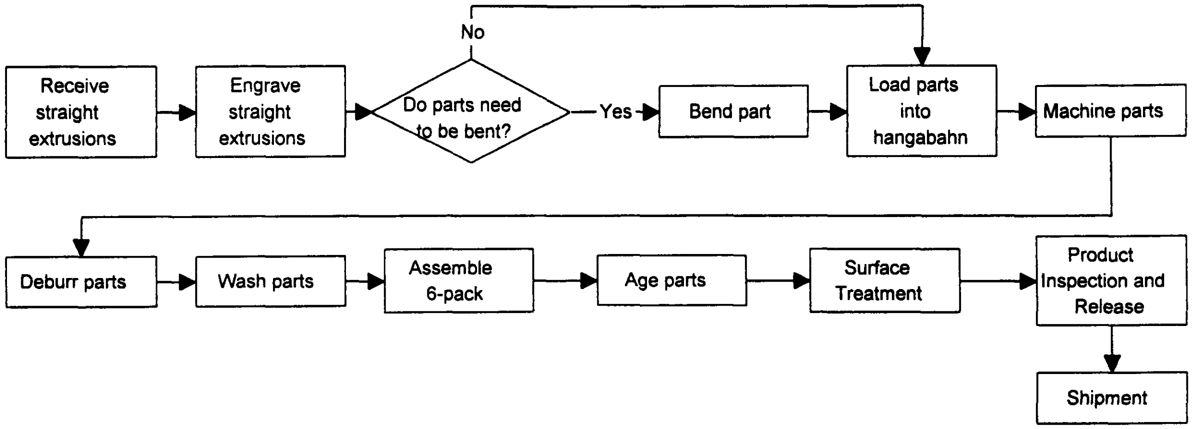


Figure 1.2: Extruded parts' process flow.

1.3 Processes and their Challenges

As introduced in Section 1.2, an array of metal forming and treating processes is used to make the parts produced by this plant. This section summarizes some of the key elements of each process. Also presented are some of the more pressing challenges associated with some of these processes.

1.3.1 Vacuum die casting of aluminum

The die casting process used by the automotive supplier to produce these castings was only recently developed. This process is used to produce high-quality thin-walled castings with improved elongation and ductility relative to more traditional casting methods. These castings were also more uniform with regard to metallurgical consistency compared to other casting technologies, mainly due to the removal of air from the mold prior to molten metal entering the mold. Trapped air tends to create porosity that tends to make the metal brittle.

Initially, a billet of a special ingot called C119 (from the 6XXX series of alloys) is preheated in a crucible heater and then placed into a caster's source furnace. When ready, the caster first evacuates all air from the die, earning this process the trade name the "Vacural" process. After the die is evacuated, the metal is injected into the die and is completely filled in less than 0.25 seconds. After cooling, the upper and lower dies part and the casting is removed, checked automatically for dimensional accuracy, and quench-cooled in a water bath.

The process and alloys being used have been under development for many years by the auto supplier. The evacuation of gas from the die prior to metal injection was seen as crucial for manufacturing castings with superior elongation and impact energy absorption. One of the most difficult challenges of this process is keeping the machines running. The casters themselves are complex pieces of machinery that are difficult to maintain. The casting alloy is very low in iron. The dies, however, are steel. Because of the difference between the partial molar free energies of iron (i.e., iron's chemical potential) in the aluminum alloy and the steel dies, unprotected iron from the steel die would dissolve into the molten Al-alloy. Thus, the steel dies must be continually nitrided and renitrided in a salt solution to provide an inert, protective layer so as to not slowly dissolve.

1.3.2. Extrusion of aluminum

The extrusion process consists of initially heating a billet of C210 (also a 6XXX series alloy) to approximately 70-80% of the melting temperature of the alloy. This heated billet is then placed in an extrusion press where the softened metal is pushed through a special steel die by large hydraulic presses. The extruded metal is cooled quickly in a very controlled manner as it is rolled out on a 100 yd.+-long cooling table of rollers. The extrusions age at room temperature for a week, and then are shipped to the customer.

The process as used by the auto supplier incorporated new extrusion alloys, innovative die design, and enhanced control of the cooling process. Many challenges are associated with this process. One of the most significant challenges is to minimize bow and twist in the straight extrusion, for these forms of dimensional variation cause problems in later downstream forming operations. Uniform mechanical properties are also very important to achieve throughout the extruded length and the extrusion profile.

1.3.3 Bending of aluminum

Two types of extrusion bending technology are used at the plant. One of these is called stretch bending. A straight extrusion is placed in the machine in such a way that one mandrel on each end of the extrusion grips it. The mandrels form an air tight seal around the hollow extrusion, and then fill it with a gas. Once brought to pressure, the extrusion is simultaneously stretched with tension applied by the two mandrels and pulled against a curved die. Once fully pulled against the die, the gas pressure is released, and the mandrels return to their original positions. The use of tension during the bending process helps to reduce springback once the part is released from the mandrels.

The second forming process is called roll bending. In this process, a straight extrusion is inserted over a mandrel. An outer "sweeper" die then clamps around one half of the extrusion and presses it against an inner forming die. The sweeper die then rotates around, pulling and bending the extrusion around the forming die. The resulting bend is not as smooth as that obtained by stretch forming, but bends of greater radii are obtainable.

The greatest challenge faced by both of the bending processes is making parts that meet specification. Variations in the incoming material's bow, twist, and physical properties were subtle but great enough to cause problems in bending parts to very tight specifications. Closed-loop process modeling was one approach being explored to help improve robustness of the bending processes. Efforts were also underway to reduce the variation of the incoming stock of straight extrusions.

1.3.4 Machining of aluminum

This process step and the machining cells used to machine the extruded parts are the main focus of this thesis. 5-axis CNC high-speed milling machines (i.e., HSMs) are used to machine both the extrusions and a few types of castings. This process step was clearly the bottleneck operation in the extrusion part flow and needed to have its capacity increased. Some of the reasons why this process step was the bottleneck are discussed in

Chapter 2, and the modeling work used to find the best way to add capacity to this area is presented in Chapter 4.

Two types of HSMs were in use within the facility. The first type, Type A, was two-table machines equipped with an automatic table changing mechanism. All five axes of movement (horizontal (X and Y axes), vertical (Z axis), and two rotational axes (A and C)) were all performed by the robotic arm suspended tracks that ran along the interior ceiling of the machine. The arm could pick from any one of eight different tools with which to machine the parts within its machining envelope. Because the arm could move along all five axes, the table remained stationary during the entire machining cycle.

The second type of HSMs used by the plant, Type B, was a one-table system. This type of HSM could also perform 5-axis machining, but accomplished it in a different manner than the Type A machines. The robotic arm only moved in the X, Y, and Z axes. The table, however, moved along the X and Y axes during machining. By carefully synchronizing the 3-axis motion of the robotic arm with the 2-axis motion of the table, 5-axis machining was accomplished. This machine could store up to 22 different types of tools at one time. This almost 3-fold increase in tool carrying capacity reduced relative to the Type A HSMs the need to exchange tools within the machine whenever the machine was set up for a different part type. The Type B HSMs suffered less from vibration-induced tolerance problems than the Type As, as only three axes were associated with the B's suspended robotic arm vis-à-vis all five axes with the Type As.

The tools used by both types of HSMs were fairly standard carbide-tipped machining tools. Saws of various radii were common. Drill tips of different diameters were frequent used. Milling tools were also quite common. The machines were also equipped with a coordinate measuring machine (CMM) end effectors. This "tool" permitted the HSMs to be set up carefully relative to the parts and their associated fixtures in order to facilitate machining the parts to very tight tolerances. Fine-tuning of the CNC program was facilitated by the use of the CMM end effector.

This process step faced many challenges. The machining operation was capacity constrained. Machine downtime was high. Set-up times were high. Minor variations in the incoming parts were difficult for the CNC machines to handle. Great strides were being made by the plant in addressing each of these problems at the time this thesis research was being conducted.

1.3.5 Heat Treatment, straightening, and aging

Aluminum alloys can be grouped into two broad categories: heat-treatable and non-heat-treatable. The difference between these two groups lies in how they obtain their strength. Non-heat-treatable alloys gain their strength from strain hardening and by solid solution. Heat-treatable alloys have their alloying elements dissolved in the aluminum at high temperatures and are maintained in solid solution by quenching from this high temperature. Following this solution heat treatment, the strength can be further increased by precipitating a portion of the alloying elements still in solution by aging the alloy at moderately elevated temperatures (Weisman, 1981).

Both the extrusion and casting alloys used for the products made by this plant are heat treatable. Both types of products are subjected to controlled heating and cooling to convert them from the quenched and natural aged condition (called T4) to that where the peak combination of strength and impact absorption are reached (called T6). The machining and bending processes are conducted on parts while they are in the weaker T4 condition as to require less energy to deform them or remove material from them. After all machining and bending processes are completed, the thermal treatments convert the parts from the weaker T4 condition to the stronger T6 condition.

The cast products are subject to two types of thermal treatments: heat treatment and aging. The cast products are first subjected to a heat treatment. This heat treatment exposes the parts to temperatures between 400 and 500 °C for a time between one-half and one hour. Following this heating, the parts are quenched with water that is heated to temperatures between 60 and 90 °C. Because the casting have very thin walls, they deform slightly under their own weight while they are softer at the elevated temperatures and also experience some thermal distortion due to quenching. A straightening process is used following the heat treatment to bring the dimensions of the parts back into specification. Following this straightening process, the parts are exposed to a thermal age treatment. This treatment consists of exposure to temperatures between 190 and 230 °C for a time between 1.5 and 2.5 hours.

The extruded part types are only heated once during an aging process. They do not require a separate heat treatment to achieve the desired physical properties. The extrusions are aged at a temperature between 200 and 250 °C for a time between 2 and 3 hours. Because the alloy is not excessively soft at this aging temperature, they do not deform under their own weight during this step. Consequently, no straightening process is required for these parts.

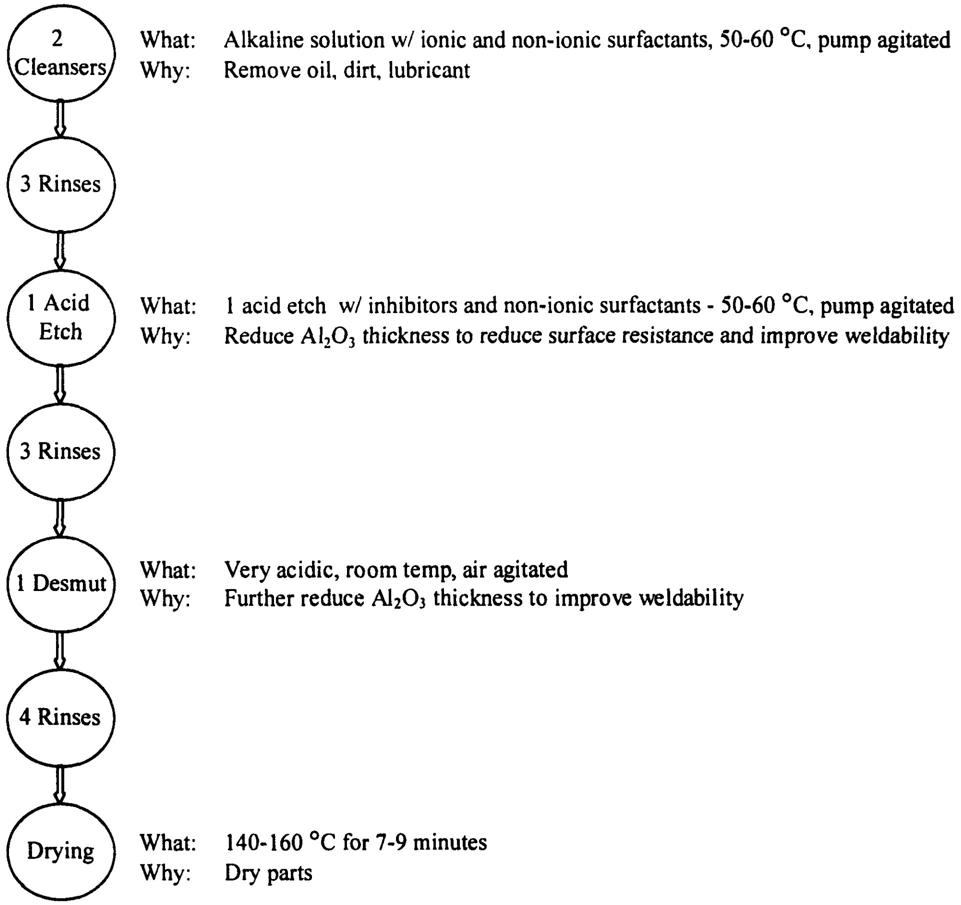
The primary challenge for this process step is to minimize or eliminate the need for straightening of the cast products. New alloys are being explored that will not require the heat treatment step to achieve the desired physical properties. Other approaches are being tried to reduce or eliminate the need for the straightening step.

1.3.6. Surface treatment

The final step (besides inspection, packing, and shipping) is surface treatment. This step has two primary functions:

- cleansing the parts of machining lubricant, dirt, and aluminum shavings, and
- improving the weldability of the extruded and cast parts by chemically modifying their surfaces

The parts are soaked in a total of 14 different baths. The series of baths used for both the extrusions and castings is given in Figure 1.3. A quick discussion as to the purpose of each type of bath is also given.



What: Alkaline solution w/ ionic and non-ionic surfactants, 50-60 °C, pump agitated
Why: Remove oil, dirt, lubricant

What: 1 acid etch w/ inhibitors and non-ionic surfactants - 50-60 °C, pump agitated
Why: Reduce Al_2O_3 thickness to reduce surface resistance and improve weldability

What: Very acidic, room temp, air agitated
Why: Further reduce Al_2O_3 thickness to improve weldability

What: 140-160 °C for 7-9 minutes
Why: Dry parts

Figure 1.3: Surface treatment steps for both extrusions and castings.

This process step has associated with it a few challenges. Minimizing cross-bath contamination is essential and accomplished in part by having at least three plain water baths between each chemical bath. Even then, the technique used for stacking the parts into the bath baskets and the geometry of the parts themselves can lead to cross-bath contamination. Maintaining the baths' composition, pH, and temperature is also challenging.

Given the basic process steps and some of the material science behind the parts made in the plant, the rest of this thesis will focus on one significant challenge the plant was facing during the time the research for this thesis was taking place. It will then discuss briefly some of the relevant literature on this subject, and then present the modeling work conducted to assist the plant in meeting this challenge.

Chapter 2: Capacity Planning Problem

2.1 Introduction

This chapter describes one of the major problems that was faced by the plant where the work for this thesis was completed: too little in-house machining capacity. It then goes on to discuss the factors that limit the rate at which aluminum can be machined. A trade-off between spindle speed and horsepower seems to be the most limiting factor with regard to aluminum removal rates.

2.2 Problem Faced by Plant

This plant has one primary customer for its parts. This customer was gradually increasing throughput and did not want to be starved by one of its most important suppliers. The plant was able to meet or exceed the demand at the customer's current throughput rate, but felt it could not continue to do so if the customer's throughput rate significantly increased. In order to avoid finished part shortages, the plant desired to always have capacity greater than or equal to the customer's throughput rate.

The plant began to examine where it needed to increase capacity most. Plant management knew that the bottleneck operation for the plant was the machining area (i.e., the HSMs). These HSMs were expensive, general-purpose CNC machines that cost roughly $600,000 each. The HSM area ran the maximum number of hours allowed by law and still had to contract out some work to outside machine shops. Thus, the plant's major challenge was to decide how to increase machining capacity. Much work was done to decrease cycle times of parts on the HSMs through optimization of the machines' programs and tool use. However, even assuming "optimal" cycle times, the total in-house machining capacity would be insufficient to meet the expected eventual demand of the customer. Some basic cutting operations were off-loaded to an earlier process step, thus, increasing the throughput of the machining area by decreasing even further the average cycle time per part. However, due to the complex end cuts required by the parts' designs, the amount of machining off-loadable to less expensive, non-5-axis machining technologies was very limited. Accordingly, the increased throughput realized by off-loading selected machining steps to cheaper machining technologies was minimal. The plant concluded it must increase the number of CNC 5-axis milling machines it had in-house. Chapter 4 of this thesis describes modeling work done to help in the decision process, particularly dealing with machine layout and material handling.

One might think that the machines could just "machine faster." In other words, why can't these machines just drill, saw, and cut faster? With regard to drilling, one must separate the process into two components: cutting speed and feed. Cutting speed is the speed at which the drill rotates. Feed is the distance that the drill is fed into the workpiece with each revolution. Together, these two factors help determine the time required to produce a hole. The revolutions per minute optimal for drilling a hole is a function of the metal, it's thermal and work-hardening history, and the diameter of the drill bit. Trying to drill too fast for a given set of parameters can result in chipped cutting edges, drill breakage, and the drill heating (Walke, 1973).

As far as cutting and milling go, other factors come into play. The choice of cutting tool is a function of the quality of finish among other things. The number of teeth per inch on a given saw blade or cutter impacts significantly the speed at which material removal can occur and the quality of the resulting finish (Walke, 1973). These factors alone, however, do not explain fully what limits the rate (in^3/min.) at which aluminum can be machined. The following sections address this question in a manner rooted in material and mechanical engineering.

2.3 Material-based Limitations to Machining Speeds

During the machining process, much heat is generated. This generation of heat is caused by the friction between the material being machined and the cutting tool itself. The dissipation of heat generated by the material removal process must be considered as a potential limiting factor. A majority of this heat is contained in the chips produced by the cutting tool. Therefore, one might think that faster removal of chips from the cutting area would permit faster cutting speeds. Also, for a given amount of aluminum removed, decreasing the average chip size increases the amount of heat stored in the chip vis-à-vis the parent metal, helping to minimize heat-related distortion in the workpiece.

McDonnell Douglas/ Lockheed is currently working on improving material removal rates with a new "ultra high speed" machining process. This process produces very fine chips and uses jets of high velocity air to quickly remove the chips and their associated heat (Koszarek, 1995). The results of their work are not yet conclusive as to the magnitude of the gains they will make with this approach for increasing machining speeds.

One might think the rate of plastic dislocation movement within the aluminum might be the rate-limiting step in the material removal process. According to an industry expert (O'Keefe, 1995), this is definitely not the case. The speed and ease of plastic dislocation movement through a lattice increases as the temperature of the lattice increases. This tendency is demonstrated by the observation that in most metals, yield strength decreases as temperature increases. As stated before, the friction between the cutting tool and the workpiece creates heat in the workpiece. Thus, in the heated aluminum face-centered cubic (FCC) matrix, the dislocations can move quickly enough not to be the rate limiting step in the machining process.

Some studies show that as one machines faster, the task of machining actually becomes easier, not harder, due to material-related phenomena. The faster a metal is machined, the greater is the heat generated due to friction. As the amount of heat generated increases, so does the temperature of the meal. As the temperature of the metal becomes higher, its yield and shear strengths become lower. As these two strengths become lower, so too does the force required to machine and remove material from the metal. One can thus see that as the speed of the cutting tool increases, the ease of material removal also increases.

Fast feed rates also provide another advantage. The tool/ metal interface can be modeled for thermal considerations as a point heat source. If the feed rate is slow and thus the point source slow moving, the thermal gradient lines extend far into the parent metal and ahead of the point source. If the heat point source moves more quickly, it gradually begins to catch up with and then pass the thermal gradient lies. Also, these lines collapse and compress against the cutting interface. This phenomenon causes a larger percentage of heat to be stored in the chip and thus removed when the chip is removed. It also helps to minimize thermal deformation that can occur if heat from the cutting process is transferred deep into the parent metal. Thus, for thermal management reasons, fast metal-tool interface movement is quite desired, and the degree of thermal damage done can be decreased if the feed rate is increased.

If material-based limitations do not currently limit the rate at which aluminum can be machined, the machine tool itself must be the limiting step. The following section introduces some of these limitations.

2.4 Machine-based Limitations to Machining Speeds

According to industry and academic experts, machine tool related factors limit the rate at which aluminum can be machined. Specifically, spindle speeds are the most dominate limiting element of the entire system (Hardt, 1995). One might think that with today's modern technologies, higher spindle speeds could be easily maintained. In actuality, today's typical commercial spindle speeds and feed rates for aluminum machining are currently limited to 15,000-20,000 revolutions per minute (RPM) and between 500-10000 inches/ minute respectively. The reason that faster speeds and rates are not easily obtainable lies in a complicated trade-off between RPM and the horsepower (where 1 HP = 743 W) of the spindle.

As explained before, fast spindle speeds (high RPM) and high HPs are desired to maximize material removal rates. In fact, increasing the RPM decreases to some degree the required HP/ in^3/ min. (King, 1985). However, as one increases the speed of the spindle, the diameter of the spindle's bearings must decrease due to centripetal force considerations. However, smaller bearings cannot carry as large side and axial loads as larger bearings. These side and axial loads increase with increasing HP. Thus, as the HPs of a spindle increase, so should also the diameter of the its bearings (O'Keefe, 1995).

Here is the essence of the trade-off. Faster spindles require smaller bearings, while spindles with greater HP require larger bearings. Some newer spindle technologies are attempting to work around this inherent trade-off. Hydrostatic and magnetic bearings are being tried, but so far, these approaches are unproven, and their crash reliability is not has high as more traditional mechanical bearings.

One might think that the cutting tools themselves might be limiting the rate at which aluminum can be machined. For machining aluminum, this is not the case. When machining aluminum, the temperature of the metal/ cutting tool interface cannot exceed 660.37 °C, the melting point of unalloyed aluminum. The carbide and cutting tool steels used as materials for the cutting tools are not affected by exposure to this temperature, so tool life is not shorted by these temperatures. (Machining steel, however, is quite a different story.) Thus, one can see that the mechanism that currently limits the rate at which aluminum can be removed to 3.5-4.0 in^3/ min./ HP is rooted in the spindle RPM/ HP trade-off.

Given that simply "increasing machining speeds" is not a viable solution for meeting the increasing effective in-house machining capacity needed by the plant, the plant investigated other potential solutions to increasing the effective in-house 5-axis machining capacity. As mentioned before, off-loading machining steps did not gain much in terms of reduced average cycle times as the complexity of the parts' designs mandated 5-axis abilities. The customer of the parts was not willing to redesign the parts to a great enough degree to substantially reduce the average cycle time on the HSM. Technologies were being investigated to reduce the average set-up time between part types, but such investigations had not proven significantly successful to date. Improvements in the operation from modifying the sequencing of parts were also being investigated, but were still far from being implementable. Thus, given cheaper options to increase the throughput of the machining area in-house in the near future, the plant decided to add more HSMs to the asset base. The plant was not sure of the best way to layout the new machines it would be buying. Material handling systems were also of important concern for the new HSMs. Chapter 4 describes queueing and simulation models developed to assist the plant with these decisions. Chapter 3 discusses briefly literature related to simulation and specifically, modeling of the type of system being considered by the plant.

Chapter 3: Literature Review

3.1 Introduction

This section summarizes relevant information and prior work done associated with the modeling and simulation of systems similar to those modeled in this thesis. Specifically, the literature review first presents some general approaches to simulation modeling. Following this presentation, previous work completed specifically relevant to material handling systems and machining cells will be summarized.

3.2 Basic Simulation Concepts

Simulation models can be classified according to their purposes or missions. The models developed in this thesis have two purposes: to assist with a hard-system configuration decision and also to assist with some soft-system configuration systems regarding dispatching and scheduling an automated material handling system. In the end, a "go/ no-go"--type decision dealing with the layout of eight machining centers and its material handling system would have to be made based on the results of this modeling work.

Simulation has been used as a powerful and useful tool in many more areas that those dealt with in this thesis (Law, 1992). Simulations have been used to evaluate new military weapons or tactics. They have been used to aid with the design and operation of transportation facilities like freeways, airports, ports, and subways. Simulations have also been used to analyze financial and economic systems. The discipline of "systems thinking" depends heavily on using system dynamic simulation models to determine where high leverage points for change exist in a given system's structure (Sterman, 1994).

Pegden (1990) presents a concise series of steps one should follow when executing a significant simulation modeling project. These steps are summarized below:

1. *Problem Definition* -- Clearly defining the goals of the study so that the purpose is known.
2. *Project Planning* -- Being sure adequate resources (i.e., hardware, personnel, software, and time) are available.
3. *System Definition* -- Determining the boundaries used in defining the system.
4. *Conceptual Model Formulation* -- Developing a preliminary model graphically or with pseudo-code.
5. *Preliminary Experimental Design* -- Selecting the measures of effectiveness.
6. *Input Data Preparation* -- Identifying and collecting the input data needed by the model.
7. *Model Translation* -- Formulating the model in an appropriate simulation language.
8. *Verification and Validation* -- Confirming the model operates the way intended and the output is believable by the customers of the model.
9. *Final Experimental Design* -- Selecting the final types and number of experiments that will be conducted.
10. *Experimentation* -- Executing the simulation to generate the desired data and sensitivity analysis.
11. *Analysis and Interpretation* -- Drawing inferences from the data generated by the simulation.
12. *Implementation and Documentation* -- Putting the results to work and documenting the model and its use.

The simulation work described in this thesis roughly followed the above steps. "Problem definition" and "System definition" were presented in Chapter 2. The company where the study took place provided all the necessary resources. Specifically, a

Macintosh-based simulation program called Extend produced by Imagine That!, Inc., was purchased by the plant to conduct simulation studies. The work was conducted on a Quadra 650 computer. The time for the study was predetermined by the fixed length of the internship period during which this work was to occur. Plant personnel were very helpful in providing necessary data on cycle times and other elements necessary to the model. The work completed on the remaining steps (4-12) will be discussed in Chapter 4.

Within simulation theory, certain types of model classifications have developed. One of these is "continuous versus discrete." A continuous model treats change like a continuously occurring phenomenon, while a discrete model describes changes in the status of the system as occurring only at isolated points in time (Pegden, 1990). The simulation models created as part of this thesis were discrete simulation models.

A second classification of simulation models is whether a particular model is static or dynamic. A static model portrays the behavior of a system at a single point in time (e.g., end of year profits), whereas a dynamic model describes the behavior of a system throughout time. The modeling work done for this project contained some static calculations and some dynamic ones too. Special steps were taken to calculate certain key system performance measures only at the end of the model's run in order to speed up the execution of the model run. Other calculations were done at each discrete event step the model took. Using this approach, one could monitor how some aspects of the model were running *in situ*, but the model's execution speed was not slowed down by having it continually calculate metrics that would by truly valid and comparable only until a given model run completed its execution.

3.3 Previous Related Work

Simulation modeling has been used in a wide array of settings, from simulating natural gas discoveries (Jewkes, 1992) to improving the utilization of Red Cross bloodmobiles (Brennan, 1992). Work done to simulate and model automated material handling systems associated with machining cells has increased much recently as technological advances have enabled the creation of flexible manufacturing systems (FMSs). The recent increased use of automated guided vehicles (AGVs) and automatic storage/ retrieval systems (AS/ RSs) has also prompted greater interest in the field to develop analytical optimization and simulation models for these kinds of systems. Problems of this type are sometimes classified "Crane Scheduling" or "Machine Layout" problems and are becoming quite common in the literature. The overall goal of many of the models developed is to minimize risk and increase throughput in automated material handling operations (Trunk, 1989). (A good introduction to the basic structures of FMSs, their material handling systems, and some of the broad range of problems faces by their material handling systems is found in (Kusiak, 1985)).

Matsuo et al. (1988) tackled both of these problems of machine layout and crane scheduling. In their paper, a well-established quadratic assignment program was presented that dealt with heuristically finding a near-optimal machine layout. The goal of the layout was to locate the machines so as to minimize the total distance a material handling crane would have to travel. They then present a cyclic scheduling method called the Equalizing Interval Heuristic that performed better than many common dispatching rules. The near-optimal schedule of parts on machines was found superior to typical dispatching rules like "first come-first serve," "least work remaining," "shortest processing time," and "shortest travel time." Their work modeled a single stacker crane, a system very similar in operation to that being explored in this thesis.

Tüchelmann (1986) addressed some of the issues mentioned above in a more qualitative manner. He describes and recognizes an increasing importance of rationalizing and automating material handling within a factory as flexible manufacturing cells and systems become more commonplace. He also found many advantages to automated, computer controlled overhead traveling cranes like the kind analyzed in this thesis over suspension monorails and AGVs.

Mahadevan and Narendran (1990) have contributed much to this field of study. They address key issues involved with the design and operation of AGV-based material handling systems. They found that single loop configurations with sequential dispatch rules fared better than other rules studied in terms of service levels and other metrics measured. These "other rules" included the "least vehicle utilized" and "farthest idle vehicle" heuristics. Their analytical method developed achieved results very close to those found by simulation efforts they conducted. They later proposed a two-stage approach to designing AGV-based material handling systems for flexible manufacturing systems (FMSs) (Mahadevan and Narendran, 1994). The first stage consisted of using an analytical approach for calculating the number of vehicles required in a particular system. The second stage involves using simulation to study the effects of AGV failures and dispatching rules on the performance of the system being studied.

A system similar to that investigated in this thesis is the linear induction motor tool delivery system studied by Hahn and Sanders (1994). Their work dealt with modeling the delivery and removal of machine tools to HSMs arranged in two parallel straight lines. The tool transporter traveled in a uni-directional loop. This work shared many assumptions made to create the models presented in this thesis including modeling the HSMs with no breakdown time, using dynamic, flexible part assignments, and modeling the speed of the transporter as a constant average speed that took into account acceleration and deceleration. One significant difference between their work and work presented in this thesis is that they assumed the HSMs are never starved for material but can wait for tool delivery. The models developed in this thesis assume tool delivery is included in the cycle time for each fixture but that the HSMs definitely can be starved for material. Their work found that increasing the speed of the transporter increased throughput for the cell until a machine-related bottleneck was hit. They also found that increasing the cycle times of the parts increased machine utilization while it decreased the transporter utilization. These findings are very similar to some of the finding presented in this thesis.

Unlike the study mentioned above, bi-directional transporters have also been investigated (Egbelu and Tanchoco, 1986). These studies used simulation to conclude that having a bi-directional flow of AGVs increases throughput and productivity vis-à-vis unidirectional flow systems. This tendency was found to be especially true in systems that required only a few vehicles. The same authors also used simulation to study various dispatching rules in systems with large material flow volumes (Egbelu and Tanchoco, 1984). This work found several drawbacks with vehicle dispatching rules that are based on distance measures. They explain the shop locking phenomenon and offer a clear classification system for vehicle initiated task assignment rules and work center initiated task assignment rules.

The review of some of the previous studies related to using simulation and analytical models to predict the performance of AGV-based and crane-based systems sets the stage for the next chapter. In this chapter, the queueing model and simulation models developed in this thesis will be described and detailed. Their assumptions will be explicitly stated, and their results discussed.

Chapter 4: Modeling the HSM Area

4.1 Introduction

In this chapter, a variety of approaches to modeling the HSM area will be presented. Initially, the system being modeled will be described. The first model presented of this system is a queueing model that attempts to make a rough approximation of the service level of the robotic material handling system. The remaining models presented in this chapter are simulation models of the HSM area that experiment with different scenarios possible within the area. These scenarios include static or dynamic part assignment, different decision rules used to pick the next machine served, and "what-if" analysis of the material handling system's robustness if one of the two robots, or "servers," goes off-line.

4.2 System being Modeled

The system being modeled is a set of 8 high-speed 5-axis CNC milling machines (HSMs) that have parts brought to them in fixtures by a robotic material handling system composed of a total of four robots. A schematic of the layout is shown in Figure 4.1.

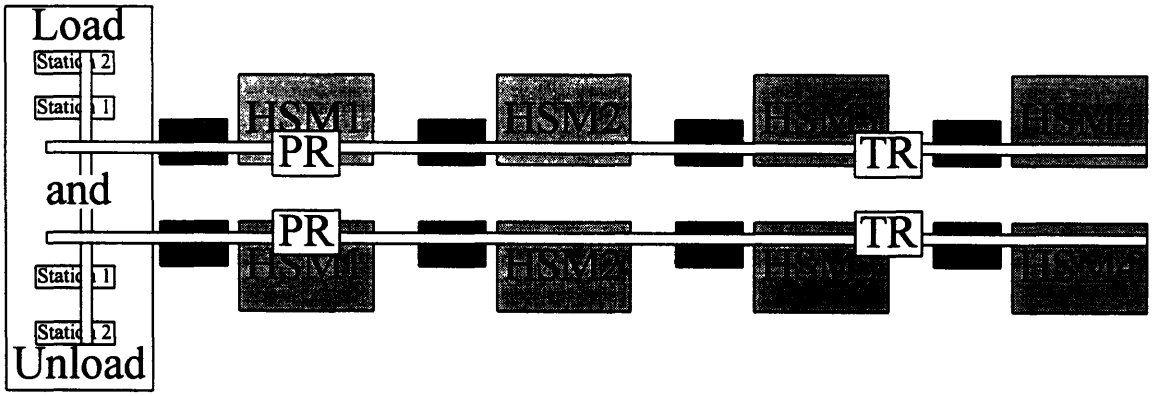


Figure 4.1: Schematic of system being modeled.

As seen above, the 8 HSMs are arranged in two identical parallel lines of four each. Associated with each HSM is a table that holds the next fixture of parts waiting to be machined by the given HSM. (These stand-by tables are indicated by "ST" in the diagram above.) Each line of 4 HSMs is served by two robots that share one overhead track. These robots cannot cross over each other during normal operation. For reference purposes, the robots on the left side closer to the load/unload area will be called the "primary" robots ("PR" in Figure 4.1) and those on the right the "transfer" robots ("TR" in Figure 4.1).

In the real system, a total of 24 different fixtures exists. A "fixture" is a metal structure that holds one or more specific parts while these parts are being machined or transported to a different place within the machining cell. Because some fixtures hold more than one part type (i.e., left and right side pairs of similar parts), 24 fixtures were capable of holding all 32 of the extruded part types machined in-house. (The machining of parts requiring only simple cuts or drilling was outsourced to an external vendor equipped with less expensive yet less technologically capable machining capacity.) The range of cycle times for machining all of the parts in a given fixture ranged from a low of 3.8 minutes to a high of 96 minutes. The number of parts per fixture ranged from a low of 1 part per fixture to a high of 12 parts per fixture.

The normal material handling sequence consists of the following events:

1) An HSM, say HSM3, finishes work on all of the parts in the fixture currently on its table.

2) This HSM signals the primary robot to pick up its finished fixture.

3) The primary robot picks up the fixture of machined parts and carries it to the "Load/unload" area. At the Load/Unload area, the fixture is dropped off to one of two Load/Unload stations (in Figure 4.1, Station 1 or Station 2). Machined parts are unloaded from the fixture and parts needing machining are then loaded onto the fixture.

4) The transfer robot transfers the fixture awaiting machining located at HSM3's stand-by table to HSM3's processing table.

5) The primary robot picks up from the load/unload area a fixture of recently loaded parts in need of machining and delivers them to a vacant stand-by table.

(Note that this order of steps is a stylized example. The service discipline used for the robots queue is one of the variables experimented with in the thesis. The method of part assignment to machine is also an element of the model that is changed in later versions. What these different queueing disciplines and part assignments are and how they are modeled are explained in Sections 4.5 to 4.7.)

One of the primary concerns for the plant was that HSMs would have to wait "in line" occasionally to be served by the robotic material handling system. In other words, occasionally, more than one HSM might have machined parts ready to be removed and could not machine more parts until serviced by the primary robot. A machine waiting to have its fixture picked up cannot receive and begin machining a new fixture of parts and is thus blocked. Also, a HSM cannot start machining parts in a fixture until the transfer robot delivers the fixture from the HSM's stand-by table.

The important metric for the plant was the average percent capacity lost waiting for the material handling system to deliver fixtures to or to pick up fixtures from the HSMs (i.e., capacity lost due to HSMs waiting in the robots' service queue). In terms of the system, "waiting in the queue" implies a machine with a fixture of machined parts waiting for the fixture to be picked up and taken away by the primary robot. "Waiting in the queue" also includes time spent by an HSM waiting to be serviced by the transfer robot. Time spent waiting in the queue was capacity lost on the bottleneck operation, and as Theory of Constraints states, "An hour lost on the bottleneck is an hour lost for the whole plant." (Goldratt, 1992)

4.3 Queueing Theory Model

For the given machining area, a queueing theory model was developed to get a rough approximation of the expected service levels for the robots. For this task, a standard "bank teller with customers" model was used. The HSMs were viewed as customers, and the robotic material handling system was viewed as a bank teller. Given the dimensions of the HSMs, their stand-by tables, and their layout, the distances the primary and transfer robots must travel to complete an average "trip" were calculated. Given these distances and the robots' average traveling speed, average travel times and their variances for each robot were calculated. The time required to pick up and drop off fixtures was equivalent for the primary and the transfer robots. Because the average trip for the primary robot was so much longer than the average trip for the transfer robot and the time required by both to pick up and drop off fixtures was the same, the assumption was made that the transfer robot's operation would not be a bottleneck, and this robot's operation was not included in the model. (In reality, the transfer robot might occasionally "interfere" with the primary robot's travel and visa-versa. This interference would cause the machines to wait longer to be served by the material handling system than predicted by the model. However, for the sake of the queueing model, the possibility of interference was neglected.) With these facts and assumptions in place, the model of the system was established as one with one server (the primary robot) and up to four customers (the HSMs) in the system at one time.

Other assumptions were also made to assist with the development of the queueing model. One such assumption was that only one fixture of each fixture type existed (i.e., no duplicate fixtures) in the whole system, a valid assumption given the plant in reality had only one fixture of each fixture type in the plant in order to minimize required capital expenditure and fixture storage space. This assumption would thus not cause the queueing model to underestimate or overestimate the capacity lost relative to real system.

Another simplifying assumption was that no machine downtime would occur. In the real system, the HSMs will experience planned and unplanned downtime. When a HSM is down, it cannot place a call to the primary robot. Consequently, this "no downtime" assumption causes the queueing model to overestimate the arrival rate of calls relative to the real system. The inflated call arrival rate causes the predicted utilization of the primary robot to be overestimated, and according, causes the expected HSM capacity lost due to waiting on the robot to be overestimated.

Because the lines were identical, the queueing model was made for one line of four machines, and the number of minutes of capacity lost due to all eight machines waiting would be twice those results found for the single line. On the other hand, other performance metrics like robot utilization would not scale by a factor of two and would instead be the same for both the 4 HSM system and the 8 HSM system due to the addition of a second primary robot that comes with the second 4 HSMs. The model was developed based on a 24-hour shift with (24 hours * 60 minutes/hour =) 1440 minutes of available machine time per HSM so that plant management could easily relate the model's inputs, outputs, and parameters to a unit with which it was familiar: one 24-hour day of production.

The next task was to pick distributions representative of the customer arrival pattern and service times. For this system, a "customer arrived" for the primary robot to service whenever an HSM finished machining all of the parts on a given fixture or when a fixture was ready to be transported from the load/unload area to a vacant stand-by table. The distribution that most closely fit this customer arrival process was the Poisson distribution. Using this distribution, however, is far from perfect for the system being modeled. Using a Poisson distribution implies that the probability of an arrival occurring during next differential (i.e., very small) sized time interval is equal to the product of the expected arrival rate λ customers per time unit and the differential time unit, as seen in the following functional form (Hall, 1991):

$$\Pr\left\{ \left[N(t+dt)-N(t)\right] \begin{Bmatrix} =0 \\ =1 \\ >1 \end{Bmatrix} \begin{matrix} =1-\lambda dt \\ =\lambda dt \\ =0 \end{matrix} \right\}$$

Using a Poisson process implies the numbers of arrivals in any pair of disjoint time intervals are statistically independent (i.e., arrival process has independent increments). However, in the real system being modeled, actual arrivals are not independent since a robot cannot call again once it is waiting in queue. Even given this flaw, the memory-less customer arrival pattern seemed accurate enough to get a rough approximation of the system's performance. The reasons for believing this arrival process was at least "semiindependent" was that the finishing of one fixture's parts by one HSM (and needing this fixture to be picked up by the primary robot) was seen as independent of when a different HSM machining parts with different cycle times would enter the queue for service from the primary robot. This flaw is addressed directly later in this section when queueing models for a finite calling population are discussed.

Service times, on the other hand, were seen to more closely follow a general distribution. The "service time" for this system was the average time it takes primary robot to complete one cycle of service (i.e., steps 1-6 listed below). Using a general, or Gaussian, distribution in this model still assumes the arrival process is independent. As discussed in the previous paragraph, this assumption of independence in the arrival process is not perfectly accurate, but it is good enough for the purpose of the model. The average service time for the primary robot was assumed to be the time required to complete the following steps:

1) Arriving to the next machine in the queue in need of having a fixture removed from some average distance away,
2) Picking up the fixture from the HSM,
3) Delivering the fixture to the load/unload area,
4) Picking up the next available fixture of parts needing machining from either Station 1 or Station 2 in the load/unload area,
5) Traveling an average distance to the proper stand-by table, and
6) Dropping off the fixture of pre-machined parts onto the stand-by table.

Outlining the steps as shown above points out an additional weakness of the queueing model as formulated. In the theoretical queueing model, the customer does not leave the system until the entire service time has expired. In the real system being modeled, however, a machine can start working on parts as soon as the transfer robot delivers a fixture from the machine's neighboring stand-by table. Thus, the machine must not wait for the primary robot to complete steps 1-6 before "leaving the queue" (i.e., starting machining), but only for the primary robot to accomplish steps 1-3 and for the transfer robot to complete the transfer. This difference results in the HSMs reentering the population of potential callers earlier in reality than is captured in the queueing model. Thus, because the HSMs reenter the potential calling population sooner in reality than in the queueing model, the calling rate (or arrival rate) experienced should be higher in reality than calculated by the queueing model. In other words, the net effect of this discrepancy is that the queueing model might underestimate the frequency of arrivals vis-à-vis reality. Therefore, given an underestimated arrival rate, the queueing model will also underestimate primary robot utilization which leads to an underestimation of the total amount of HSM capacity lost due to waiting relative to the real system. This underestimation effect, however, is at least partially counteracted by two overestimation effects described later in this section, those effects stemming from the incorrect assumption that the queue can reach infinite size and that the population from which the callers come is of an infinite size.

With this framework established and weaknesses noted, a Poisson distribution for arrival times and a general distribution for the service time were still seen as good enough approximations to arrive at a rough guess of the service level of the system. In queueing theory notation, this combination of Poisson arrival process and a normally distributed service time with one server is called a M/G/1 system. The arrival rate was modeled as a function of demand, while the service time was modeled as independent of demand. The approaches used to calculate the expected values for inter-arrival times and service times are presented in the following two paragraphs.

Based on estimates from plant personnel, an average expected time between arrivals (i.e., inter-arrival time) of 2.57 minutes was calculated. Taking the inverse of this value results in the expected call arrival rate, λ, of 0.388 call/minute. The arrival rate was calculated in the following manner:

1) Calculate the total number of fixture changes required in a day of production. Calculating this number requires knowing the production volume per day by part type and the number of parts of each part type in each fixture. Increasing

production demand increases the number of fixture changes per day and thus also increases the arrival rate.

2) Divide this number by two so that the number of fixture changes required by one line of four HSMs will be known.

3) Divide the number of fixture changes required for one line of HSMs by 1440, the number of minutes the machines will be running given a 24-hour schedule. This quotient of 0.388 is λ, the expected arrival rate of HSMs calling on the primary robot in units of calls arriving per minute.

An average service time per call of 1.26 minutes/service was used. This duration was the time estimated by plant personnel for the primary robot to accomplish the six tasks outlined previously. The inverse of this value, μ, is the number of machines served per minute (i.e., service rate) and is equal to 0.79. A variation (σ^2) in service times of 0.0049 minutes was calculated based on the differences in distances between the HSMs, their stand-by tables, and the load/unload area. An expected robot utilization, ρ, of the primary robot was calculated by dividing the arrival rate by the service rate ($\rho=\lambda/\mu=0.39$ arrivals per minute/0.79 customers served per minute) to get 48.8%.

Queueing theory develops metrics for evaluating the service levels of a M/G/1 queueing system given the parameters λ, μ, (and therefore ρ) and σ (Nahmias, 1993). These metrics include the expected number of HSMs in the queue in steady state (L_q), the expected number of HSMs in the system in steady state (L), the expected waiting time in the queue in steady state (W_q), and the expected waiting time of a HSM in the steady state (W). Of these metrics, L_q is the most significant and important metric to the plant's evaluation of the feasibility of the described cell, since L_q represents the "# of HSMs' capacity lost due to waiting" number the plant was interested in knowing. This performance metric is calculated below for the model presented using the Pollaczek-Khintchine formula (Hall, 1991):

$$L_q = \frac{\lambda^2\sigma^2 + \rho^2}{2(1-\rho)} = \frac{0.39^2 * 0.0049^2 + 0.49^2}{2(1-0.49)} = 0.233 \text{ HSM}$$

One must recall that this value represents the amount of capacity lost due to waiting for the primary robot for only one line of four HSMs, while the real system consists of two lines of eight HSMs. Therefore, to calculate the total equivalent capacity lost due to waiting for the whole system, the above result must be multiplied by 2, resulting in a total capacity lost for the system of 0.46667 HSM.

As mentioned previously, the above approach makes several assumptions. Among these assumptions was that the arrival process was Poisson, whereas in reality it was not truly Poisson. Second, the buffer size was assumed infinite, whereas in the real system the maximum buffer size is four. This flaw is related to another flaw in the model: It assumes an infinite calling population, while in reality the calling population consists of only four machines. As the upper limit on the buffer size decreases, robot utilization decreases. Also, as the size of the finite population decreases, the robot utilization decreases (Hall, 1991). Recall that as the robot utilization decreases, the amount of capacity lost due to HSMs waiting also decreases. Therefore, both of these errors cause the model to over-estimate the total capacity lost due to waiting, helping to cause the model to give a worse case upper bound estimate on the total capacity lost due to HSMs waiting. Recall also that these overestimation tendencies are countered at least in part by the underestimation tendency described previously related to the leaving of the customer from the system before the total service time transpires (i.e., in reality, the HSMs need not wait for the primary robot to complete all six steps of one “service cycle” before they can start machining due to the presence of the transfer robot). Thus, overall, the 0.46667 HSM predicted to be lost by the queueing model was thought to be an overestimation relative to the real system, but determining the magnitude of the overestimation was not possible.

Given the limitations concerning assumptions made, potential robot interference problems, and the dependence in arrival and service times, the plant desired a "second opinion" to confirm the results of the queueing study. The plant felt the queueing model approach was too static and did not capture the dynamic interactions that would occur in the real dynamic, integrated manufacturing system. This approach used broad averages for parameters with large standard deviations like actual fixture cycle times and robot travel times. Also, results from elaborate formulas "pulled out of a book" not familiar to the plant were difficult for plant personnel to believe and trust. The author also questioned the validity of the assumption that the arrival rate would be memory-less and thus modelable with a Poisson distribution along with the validity of other assumptions. For these reasons, further analysis was required.

The use of spreadsheet models was seen as an option for further analysis. Spreadsheets are good for static analysis of hard numbers. Numbers like tooling costs, operating costs, and machine costs can be inputted, and a spreadsheet can perform a wide array of calculations on those numbers. Theoretical capacity calculations can be accomplished with the use of a spreadsheet. Some spreadsheets can do basic simulation work and even linear programs. However, dealing with interactions between systems and modeling multiple sources of variability in a spreadsheet can be difficult. Capturing the robustness of a machine or process relative to a competing alternative can be difficult with a spreadsheet. Technical risk and environmental impact are not easily represented. The merits of flexibility in process or machine can be challenging to evaluate with a spreadsheet. Capturing flow of material and systemic effects of process changes can be difficult. However, the spreadsheet can be quite a powerful tool and simple to use and develop.

Given the limitations inherent in spreadsheet analysis, simulation was viewed as a welcomed source of a second opinion. Simulation models have been found to be useful by some long range capacity planners. Simulation models can be very helpful for capturing the dynamic of a particular system or process flow. A wide range of outputs is available from typical simulation software. Machine utilization, buffer sizes, flow times, starvation times, and total throughput can be ascertained. What occurs between process steps can be examined, along with the timing of when events happen. Simulation models can be built to represent different levels of aggregation, modeling the movement through a plant of individual parts all the way up to the movement of a generic, typical order. Simulations can be very visual in nature, a feature that can aid with management buy-in of the results, especially if the members of management are more qualitative in their thinking and analysis than quantitative.

Simulation models do have their drawbacks. Lead time for developing a large, complex simulation can be over 2-3 months or more. Some packages require much time to learn before useful models can be developed. Many times, due to the pace of change in plans for a plant or plant addition and the time required to create a model, the model (when finally completed) does not answer the questions that the planners want to ask. In other words, what was important to capacity planners two months before a simulation is created may not be as important two months later when the simulation is finished. A model can be difficult to update and maintain if the plant is undergoing a fast pace of change. Much modeling work is usually done by people not at the site being modeled or at the site where the planning of a new plant is taking place. This distance can make the creation of a useful simulation less probable and maintenance of the model, once developed, very improbable. Given the on-site presence of the modeler and the guarantee to conduct extensive knowledge transfer on the use of the models developed to the plant before leaving, these drawbacks were minimized, and simulation modeling was chosen as the approach for further analysis.

4.4 General Simulation Model for HSM Area

Many different versions of a base simulation model were used to help assess how well the proposed machining cell would work and how could it be made to work better. However, before discussing the different versions, the overall modeling framework and structure will be discussed. Assumptions common to all the model's versions will be presented. Data collection will also be presented.

4.4.1 Structure of general simulation model

As briefly mentioned in Section 3.2, a Macintosh-based simulation program called Extend produced by Imagine That!, Inc., was purchased by the plant to conduct simulation studies. The work was conducted on a Quadra 650 computer with a math coprocessor and floating point unit (FPU). To create a typical model using Extend, the modeler must only click and drag the appropriate icons into the model and link them together in a structure representative of the system being modeled. Typical blocks included with the program include first-in-first-out (FIFO) queues, buffers, machines, conveyors, and work stations. Double-clicking on any single icon permits the user to change operating parameters specific to that icon (e.g., maximum buffer size of a buffer icon, cycle times of a machine icon, etc.). Behind each icon invisible to the typical user is up to 15 pages of computer code written in a computer language based on C called "ModL." Because of the uniqueness of the system presented in this thesis, the author was required to break open and extensively reprogram some of the icons in order that their behavior would more closely mimic that of the real system. Some totally new blocks, or icons, were created from scratch by the author also.

A copy of the model showing the general structure of the simulation model is shown on the next page in Figure 4.2.

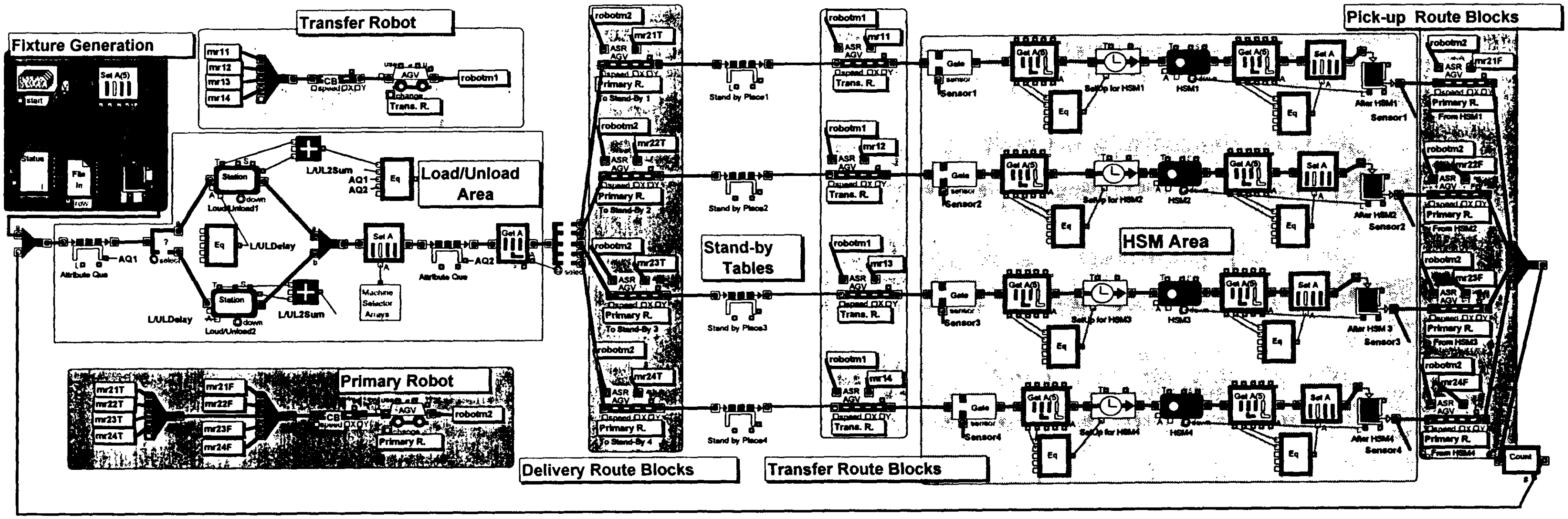


Figure 4.2: Basic layout of general simulation model for HSM area.

The model contains five main sections: fixture generator, load/unload area, primary robot, transfer robot, and the HSMs. Within the "fixture generation" section, eight fixtures are created at the very beginning of the model's run. These fixtures are assigned cycle times and other parameters depending on which version of the model is being run. (The different versions of the base model are described later in this chapter.) These eight fixtures enter a queue to enter the load/unload area.

In the load/unload area, fixtures are accepted and routed to one of two load/unload stations in an alternating fashion. When the model is first initializing, the delay experienced in the load/unload area is 0.0 minutes. In other words, when a model run is just beginning, the fixtures are sent without delay to the HSMs via the two robots. This lack of delay is used to minimize start-up effects. However, during the course of normal operation, the fixture is delayed in the load/unload area for 2.5 minutes, the expected time to unload machined parts from a fixture and load the same fixture with parts that need to be machined. Once a fixture is delayed for 2.5 minutes, it is sent onward to be picked up by the primary robot. On the way to the robot, the fixture is tagged with an indicator telling the robot to which HSM it should deliver the fixture. The heuristic used to decide to which HSM a fixture should be delivered varied between different model versions and is thus discussed in detail within the descriptions of the different versions later in this chapter.

The primary and transfer robots are modeled using what the software calls "AGV" (Automated Guided Vehicles) blocks and "Route" blocks. The AGV blocks (refer to Figure 4.2) serve as sources for robots, one robot in each of the two AGV blocks. When not in use, each robot "waits" in its respective AGV block. The route blocks call the robots from their AGV blocks when a given robot's services are needed. Each route block contains information about the speed the robot travels on that route and the length of that particular route. In the base model, each route block can only call either the primary or the transfer robot, but not both. This limitation is used to model the fact that the robots cannot overlap each other in the single-track system being modeled. (A later version of the base model that tests the performance of the system when one robot must complete the tasks of both the primary and transfer robots removes this restriction.)

Some intricacies of the real fixture handling system were not captured in the models developed. One of these intricacies is the possibility of interference between the primary and transfer robot. One might think that the primary robot and transfer robot will end up having to stop for the other during the course of normal operations due to the fact that they cannot pass each other on the single track. Figure 4.3 below captures what some people might think could occur.

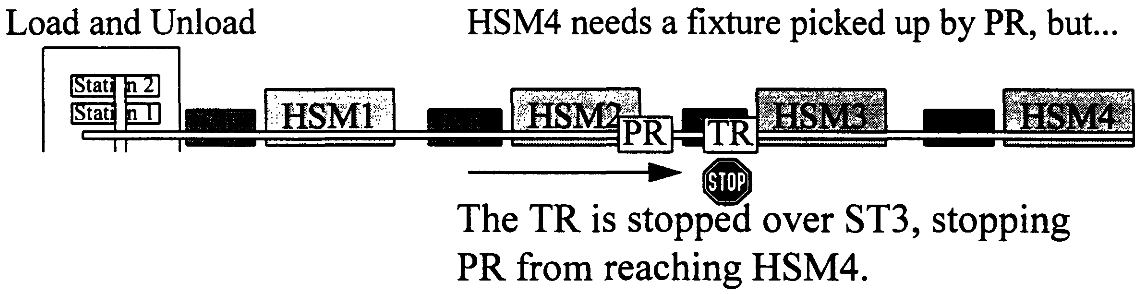


Figure 4.3: Schematic of potential blocking problem related to the fact that the two robots share the same track and therefore cannot pass each other.

A careful thought experiment proves that the above issue is not potentially a problem. Both robots require the same amount of time to pick up and lower fixtures. Both robots travel at the same velocity. For each complete cycle, both robots must pick up a fixture and lower that same fixture. Thus, the time required from both robots to accomplish these tasks is equal and cannot cause one robot to have to wait for the other. The only factor left that could cause one robot to wait on the other is differences in required lengths of travel. One will note that the distances between stand-by tables and their associated HSMs are small relative to the distances between the HSM/stand-by table pairs and the load/unload area. Thus, the distances the transfer robot must travel to complete its part of a complete fixture change cycle are much smaller than the distances the primary robot must travel to complete its part of a complete fixture change cycle. Also recall that the transfer robot cannot accomplish its function till the primary robot removes a given finished fixture. Thus, one can see that the transfer robot will always be waiting on the primary robot and that the transfer robot will not block or interfere with the primary robot.

Another potential concern with the manner by which the material handling system involves using an average "home" position for both robots. An alternative approach would be to keep track of the exact position of each robot at all times and calculate on-line exact travel distances (and therefore times) for each move either robot makes. A relational matrix containing how far any one machine or stand-by table was from another or the load/unload table would be needed for such an approach. Unfortunately, such a relational matrix was not available in the simulation software package used. Therefore, the less accurate approach of having the robots return to some average "home" spot on the overhead track was seen to be the best model given the limitations inherent in the software package.

The last area of concern in the base model is the HSMs themselves. This area is represented as four parallel lines of one HSM each. Once receiving a fixture, the HSM determines if this fixture is the first of a new lot of a different part type, thus, requiring a set-up. In all versions, the time required for a set-up was 0.0 minutes, thus reflecting that any part could be machined on any HSM without an operator adjusting the HSM prior to machining. However, the model versions were created to easily take into account set-ups if the model user desired to do so. The models accomplish this task by "shutting down" a particular HSM for a specified amount of time whenever a different part type is machined on that HSM in order that it can be adjusted for this new particular part type.

A fundamental difference exists between how the simulation models modeled the HSM area and how the queueing theory model represented the same area. Recall that in reality, a HSM can start working on a fixture of parts as soon as the fixture is delivered from the stand-by table to the HSM by the transfer robot. The simulation model captures this dynamic, whereas the queueing model did not. In other words, recall that in the queueing model, the "customer" was not allowed to leave the system until the entire service time had expired. In the simulation, the customer can "leave the system" as soon as the transfer robot delivers a fixture of parts from a stand-by table and must not wait for the primary robot to complete the six steps previously outlined in Section 4.3. The ability to accurately capture this subtly is one reason why plant management had greater confidence in the simulation models' results vis-à-vis the queueing model's results alone.

With this introductory information in hand, one can trace the movement of a given fixture as it cycles through the model. When the model run first initializes, eight fixtures with different cycle times are created and immediately transported to the correct HSM's stand-by table by the primary robot. The transfer robot immediately transfers fixtures from the stand-by tables to their corresponding HSMs. These "immediate deliveries" are accomplished by using accelerated speeds in the route blocks during the first seconds of model simulation time and are used to minimize start-up phenomena. (Start-up phenomena could have also been minimized by starting data collection after the model had had time to start-up or to run the model long enough to wash out end effects.) The HSMs begin machining the parts on the fixture as soon as they arrive. Once all the parts on a fixture are machined, a HSM calls the primary robot, which is actually summoned by the appropriate route block located to the right, or downstream, of the HSM. The route block enters a FIFO queue for service by the robot. When the primary robot is in its AGV block, it looks at the next route block in its queue and services it. If the route block is to the right of an HSM, it picks up the fixture from the HSM and transports it back to the load/unload area where it is delayed for 2.5 minutes to account for unloading and loading of machined and pre-machined parts, respectively. The primary robot returns to its AGV block, traveling a route that approximates it returning to the middle of the track. Thus, the eight fixtures recycle through the closed system once the initialization of the model is complete until one day of production time has occurred, at which time the simulation ends and calculates the desired performance metrics.

4.4.2 Assumptions common to all simulation model versions

During the creation of the general simulation model, many assumptions were made. These assumptions are detailed below:

- The lack of cross-over ability does not significantly impact the transfer robot's service level and arrived to the next stand-by table because the transfer robot will always be able to closely track the primary robot as it returns the fixture to the load/unload area.
- No variation exists in a given fixture's "contact cycle time."[1]
- The time required by the transfer robot to complete its part of one complete cycle of a fixture change is a constant 0.25 minutes plus travel time back to its "home position."
- The average time required to unload a fixture of machined parts and reload the fixture with unmachined parts is a constant 2.5 minutes. This estimation is based on the expectations of engineering staff and operators at the plant.
- The maximum number of fixtures at any one time in the load/unload area is two.
- Some results of simulating one-half of the system (i.e., one line of four machines with two robots) multiplied by two are equivalent to the results of simulating the entire system (i.e., two identical lines of four machines apiece). These results include total machine waiting time and capacity lost due to that waiting time. Other results (e.g., average machine utilization, primary robot utilization, and transfer robot utilization) do not scale and in fact are the same for the single-line system of four HSMs as the double-line system of eight HSMs.
- In all of the models developed, both robots and all four HSMs experience no downtime. The only time a HSM is not running is when it is blocked or starved. The only time a robot is not running is when it is waiting in its AGV block and its queue is empty of any serviceable route blocks.
- The primary robot cannot pick up a fixture of machined parts from a HSM until at least one of the two load/unload spots can accept the fixture from the robot.

---

[1] A fixture's contact cycle time was calculated to be the summation of the contact cycle times of all the parts in the fixture plus the time required by the HSM to move the fixture in and out of the enclosed machining envelop.

- Only one fixture of each type exists in the system at one time.
- A machine can process only one fixture at a time.
- Jobs cannot be preempted once they have begun processing.
- The transfer robot can only carry fixtures to an HSM from its own stand-by table. In other words, HSM2 cannot receive a fixture from the stand-by table dedicated to HSM1.
- Both robots travel as a constant speed. Decelerating to stops and accelerating afterwards were taken into account when estimating the average speed to be used in the simulations.
- Calls placed to the primary robot to pick up fixtures will be serviced in a FIFO discipline as long as at least one vacant spot exists at the load/unload area to receive the fixture. If both load/unload stations are filled, all calls from HSMs to the primary robot to pick up fixtures will be stored in the queue in the order received but not serviced until the primary robot has delivered one fixture from the load/unload area to a stand-by table.
- Calls placed to the transfer robot will be serviced in a FIFO discipline.

4.4.3 Method for data collection

Because this model is of a system that has yet to be built, most data used for this model were "best guesses" from plant and contract personnel. The robot vendor supplied travel speeds and the times required to pick up and set down fixtures. The engineer in charge of the HSM area provided estimates of fixture cycle times that were basically current cycle times with some minor improvements expected from HSM program optimization efforts underway in-house. The travel distances were based on the known distances between the HSM, accounting for the presence of stand-by tables.

As one may recall, the number of different extrusion part types produced in the plant exceeds 8, the number of different fixtures that circulate through the system.

Although more than one part type can be machined in the same fixture (e.g., left and right side versions of the same shape of part have different part numbers, but are collocated on the same fixture), the number of different fixtures required to machine all the parts still exceeds eight, the maximum number of fixtures allowable in the cell at one time. Therefore, representative part types were picked so that the average of the eight fixtures' contact cycle times used in the model was very close to the average of all of the plant's 24 fixtures' contact cycle times. The standard deviation of the cycle times of all of the plant's fixtures and the standard deviation of the cycle times of those fixtures used in the model were very close in magnitude.

With the assumptions and data used for the general model described, the various versions of the model created will be explained. The Figure 4.4 shows what types of models were developed. Following the diagram, each model variation will be explained. After all variations are explained, the results of each type of model will be presented.

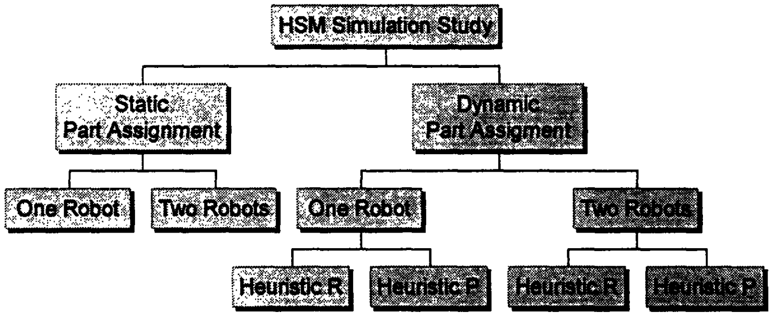


Figure 4.4: Breakdown of different versions of base model created.

4.5 Static versus Dynamic Part Assignment

One important decision for the plant when designing this system was the degree of flexibility desired in part assignment. Basically, the question was what would be gained from having the option to drop a fixture of parts needing machining on any HSM's stand-by table (if empty) relative to having to drop off a fixture on the stand-by table of the HSM from which this fixture originated. In other words, what is gained from having flexible (dynamic) part assignments relative to fixed (static) part assignments.

Dynamic part assignment implies all machines are equally capable and is only feasible if set-ups between different part types require zero (or nearly zero) time. Any fixture can be run on any machine by simply reading a bar-code unique to each fixture, loading in the fixture's CNC program, and having the full array of machining tools needed for all part types available full-time to each machine. Because of the tight tolerances required of the parts being machined, extreme uniformity among fixtures and parts would have to be achieved so that tweaking of the CNC program by the operator would not be needed every time a new fixture was being machined.

The plant wanted to know if working towards gaining dynamic part assignment was worth acquiring both the above mentioned technologies (e.g., bar code readers, common CNC programs, etc.) and the necessary process and product variation reductions. Thus, a set of models was made using dynamic part assignment and another set with static part assignment. The heuristics used by the model to decide where to send each part in the dynamic part assignment models are presented later in this chapter.

In the static models, the fixtures always went to the same, preset HSMs. The same set of fixtures with their associated cycle times used for the dynamic models was used for the static models. The initial assignment of fixtures to HSMs was based on grouping high and low cycle time fixtures together in order to roughly balance the number of fixture changes experienced by each HSM. Thus, the fastest and slowest fixtures were both assigned to the same HSM. The second fastest and second slowest fixtures were paired together and assigned to the next HSM, and so on until all eight fixtures were assigned. The method for pairing fixtures' assignments to HSMs was used to achieve a balanced total throughput of all part types by the end of one simulated day of production. Although this method may have not generated an optimal set of part assignments, the author felt this approximation was good enough so as to not impact significantly the magnitude of the important wait time and lost capacity metrics.

In order to make accurate comparisons between the dynamic and static cases, no set-up times were used in either sets of models. The dynamic systems would only be truly feasible if zero (or near-zero) setup times were required when a HSM was loaded with a new part type. In reality, the static system would not require this assumption, for machines would change part types only after an entire production lot had been completed. An HSM would lose capacity when a set-up was incurred in the static system, but these set-up times were kept at zero to permit an "apples with apples" comparison between the two sets of situations and so that only time lost to HSMs being blocked or starved by the material handling system would be counted.

4.6 'What If?" Analysis and Robustness of System

One concern with the cell as designed by the plant involved the potential consequences of one of the two robots "going down." The plant desired to know how much HSM capacity would be lost due to machines being blocked or starved by the fixture handling system if one of the two robots had to perform both roles of being the "primary" robot and the "transfer" robot. In other words, how robust would the system as a whole be to periodic downtime experienced by the fixture handling system.

To explore this question, one set of models was developed that had both robots in service (i.e., the primary and the transfer robots). Another set of models was developed where one robot with the standard speed of travel was required to perform all fixture movements by itself. The increased HSM capacity lost due to longer waiting in the second set of models would serve as some upper bound of maximum capacity lost because the likelihood of a robot being down is not 100%. The magnitude of the difference between the capacity lost in the one and two-robot versions would be disproportional to the general robustness of the system. In other words, the more robust the system is, the closer the one and two robots versions would be in terms of total HSM capacity lost due to waiting.

4.7 Two different Decision Heuristics

In the dynamic part assignment models developed, some type of decision heuristic had to be developed to determine which HSM the primary robot must serve next. One method programmed into the simulation for choosing the next machine's stand-by table to receive a fixture of parts was reactive in nature ("Heuristic R"). The other heuristic was more proactive in nature and will be referred to as "Heuristic P."

Heuristic R was the reactive decision rule. In these versions of the base simulation, the primary robot sent the next available fixture to the stand-by table of the machine that finished a fixture the most amount of time ago relative to the other three HSMs. The stand-by table for the HSM selected to receive the fixture also had to be unoccupied. When this decision rule was in operation, the HSM would be asked to be placed in the primary robot's service queue as soon as it finished a fixture. The primary robot serviced its queue in a first-in-first-out (FIFO) manner as long as the stand-by table of the first HSM in the queue was unoccupied and ready to accept of fixture of parts needing machining.

Heuristic Proactive P was the proactive decision rule. When following this heuristic, the primary robot would deliver the next available fixture of parts needing machining to the stand-by table of the HSM closest to finishing the fixture it was currently working on as long as it had an available, unoccupied stand-by table. Thus, the simulation was programmed to look ahead in time when deciding which HSM's stand-by table should receive the next fixture as opposed to following some queue of service that was more retrospective in nature.

4.8 Results

The primary purpose of the studies conducted was to find which combinations of part assignment heuristics lead to the least amount of lost HSM capacity due to being starved or blocked by the material handling system. The result of primary importance to the plant was calculation of the equivalent capacity lost due strictly to the HSMs being blocked or starved by the material handling system. (Recall that "blocking" implies the HSM is waiting for a fixture to be picked up, and "starving" implies that the HSM is waiting for the transfer robot to deliver a fixture of parts that need machining.) Another important result for the plant was the average HSM utilization related to lost time in each scenario. Lastly, the plant was interested in the utilization of the primary robot.

The results are presented in Figure 4.5. The first number (1) given in each group is in units of "# of HSMs" lost due to waiting on one or both of the robots. This number was calculated by adding the total time blocked and starved for all four HSMs, multiplying it by two to account for the other line of four HSMs, and dividing this product by 1440, the number of minutes a perfectly run HSM could provide of machining time in one day. In functional form, this value can be represented by the following equation:

$$EquivalentCapacityLost = \frac{\sum_{i=1}^{4}(\text{minutes starved}_i + \text{minutes blocked}_i)}{24 * 60} * 2$$

The second number (2) listed in each group is the average HSM utilization. This utilization figure is the percent of time the HSM was not blocked or starved during the course of the simulation run. It is related to the EquivalentCapacityLost value by the following equation:

$$AverageHSMUtilization = 1 - \frac{EquivalentCapacityLost}{4 * 2}$$

The last number (3) in each group is the utilization of the primary robot. This utilization figure is the percent of time the primary robot was in motion during the course of the simulation run (i.e., one day, or 1440 minutes, of simulated time).

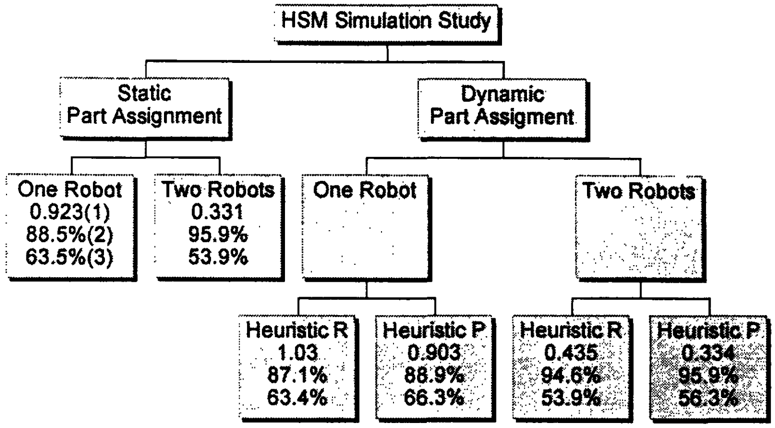


Figure 4.5: Results of various simulation versions.[2,3,4]

Some interesting observations can be made of the results derived:

- The proactive heuristic is better than the reactive heuristic in both the "one robot" (i.e., one of the two robots is shut down) and "two robots" (i.e., both robots are functioning properly) cases.
- Dynamic part assignment with both robots following the proactive assignment heuristic gains you nothing relative to static part assignment with two robots running except slightly higher robot utilization.
- Dynamic part assignment with one robot following the proactive assignment heuristic gains you only 0.02 HSM of capacity per day relative to the static part assignment case.
- The reactive part assignment heuristic in both the one robot and two robot cases performed the worst in terms of having the largest amount of capacity lost.

---

[2] (1) = Equivalent HSM capacity lost per 24-hour day of production

[3] (2) = Average HSM utilization

[4] (3) = Utilization of primary robot

4.9 Verification and Validation

As discussed in the previous chapter, verification and validation are very important steps in any simulation project. One common approach to verification and validation is showing that a model can replicate reality fairly closely. However, for the models presented in this thesis, reality does not exist yet. The system being modeled has not been built yet, so reality does not exist and thus cannot be replicated. Therefore, to complete the validation and verification steps, other methods must be used.

Initially, the results of the model should be checked to see how realistic they are. (This step was taken in section 4.8.) One can also compare the results to those gained from other approaches to modeling. Another approach is to run the simulation at extreme conditions and then to determine if the model behaves as expected.

Given this, one may wish to recall and compare the results of the queueing theory analysis to the results of the different versions of the base case simulation model. The M/G/1 model of the "one server and at most four customers" system predicted that the primary robot utilization would be 48.8% and the total capacity lost per day for a system of 8 HSMs would be 0.4667 HSM per day.

Comparing the results can be accomplished two ways. One way is to calculate the percent difference between the simulations' estimates of the capacity lost and the queueing model's estimation of the same metric. Using this method to compare results, the comparison does not look too accurate. For example, when comparing the two-robot Heuristic P result for "#HSM lost" with that of the queueing model, one gets a ((0.467-0.334)/0.334=) 40% error.

On the other hand, one should ask if the two models in the end make the same recommendation to management: Should or should not this system be built? The answer from the plant was the same whether waiting caused 0.467 HSM to be lost or if it caused only 0.334 HSM to be lost. Along these lines, one can calculate the difference between the two estimates and divide by the total number of HSMs in the system to get an estimate of the error. Using this approach, one calculates an "error" of ((0.467-

0.334)/8=) 2%! The author posits that this margin of error better captures the magnitude of the difference between the two approaches' results. A "2% error" helps to validate the simulation models vis-à-vis the queueing model.

When comparing the results of the queueing model with the simulation models, one must remember one key differentiating factor between the two approaches. In the queueing model, an arrival rate of calls made to the primary robot was calculated based on the required daily production in the plant. In other words, λ was a function of demand. As one might expect based on the Pollaczek-Khintchine formula (Section 4.3), as demand increases so too would the number of fixture changes required per time unit. Accordingly, the arrival rate would increase, increasing the primary robot utilization and the expected total HSM capacity lost due to waiting on the busier robot for service.

In the simulation model, λ was not calculated ahead of time. The demand was not incorporated into the model. Instead, representative fixture cycle times were inputted into the model, and the model was run for one 24-day of production. Therefore, given the simulation model's basic construct, it cannot be validated by increasing demand and seeing if it responds in a manner similar to the queueing model, for demand is not included as one of the simulation model's inputs.

Given varying demand could not be used to validate and verify the simulation model, one version of the model was run at some sets of extreme operating conditions (not including demand) to see if it behaved as expected. The version selected to be run using extreme conditions was the two-robot Heuristic P model. The first extreme condition analyzed was using a 10-fold increase in the speed of both robots and the cycle times for each fixture. One would expect that as cycle times increase and robot speeds increase, robot utilization would decrease, waiting times would decrease, utilization of the HSMs would increase, and total capacity lost due to waiting would decrease. These expectations were fulfilled by the simulation model as Figure 4.6 displays:

	Base Case	High Robot Speeds Large Cycle Times
Equivalent HSM Capacity Lost	0.334	0.0028
Average HSM Utilization	95.9%	99.98%
Primary Robot Utilization	56.3%	0.55%

Figure 4.6: Results of high-speed robots and large cycle times compared to base case.

A similar type of experiment was run with a robot speed at one-tenth the speed of the base case and with the 10-fold increase in cycle times. Compared with the high-speed, large cycle time version, one would expect that this latest combination of extreme conditions would have higher robot utilization and lower HSM utilization due to greater waiting times caused by having the slower robot. These two scenarios are compared in Figure 4.7:

	Low Robot Speeds Large Cycle Times	High Robot Speeds Large Cycle Times
Equivalent HSM Capacity Lost	0.400	0.0028
Average HSM Utilization	95.6%	99.98%
Primary Robot Utilization	53.1%	0.55%

Figure 4.7: Results of high-speed robots and large cycle times compared to low-speed robots and large cycle times.

Thus, with the given evidence, the author feels the base model and its variations are valid, comprehensive, complete, and reliable. With such reliable results, conclusions and insights will be drawn based on both these system specific results and also the entire modeling effort and presented in the next chapter.

Chapter 5: Insights and Inferences

5.1 Introduction

The work presented in this thesis leads to two different types of "findings" or insights. One type is very specific to the plant area layout and the automated material handling system in question. The other type involves more generic learnings, issues raised, and insights developed resulting from working on the specific project studied. This final chapter looks at both types of insights, giving more attention to the more generic learning in order to increase the usefulness of this work for a wider audience.

5.2 Project-specific Insights

A few key project-specific insights were gained from the research presented in this thesis. The insights are prefaced by the fact that they apply directly only to the specific system modeled in this thesis. The author does not posit that these findings apply to a much wider array of layout and material handling scenarios.

- The plant should be sure to design the material handling system in such a way that if one robot shuts down, the other can perform as both the primary and the transfer robot. System performance was slightly degraded in the one robot case, but not nearly as much as if one robot shutting down shut down in turn one whole line of four HSMs.

- For the current part types, dynamic part assignment did not gain much in terms of reducing waiting time relative to static part assignments. Therefore, the plant should allocate resources towards reducing set-up times, not necessarily towards acquiring dynamic part assignment capability. (This finding might not be true as part type parameters change (e.g., cycle times, volume, etc.))

- If the plant acquires dynamic part assignment capability, using a proactive dispatching heuristic for the material handling system as opposed to a reactive one to minimize total machine waiting time results in less machine capacity lost to waiting.

5.3 Generic Insights based on Thesis Work

The work in this thesis is representative of one "island of simulation" within the context of a large manufacturing plant. By the term "islands of simulation," we refer to many smaller models for many different areas of the plant. Many other decisions dealing with long-term layout issues were also assisted by simulation modeling of operational "islands" within the plant. A thorough review of these models is outside the scope of this thesis, but below is a partial listing of islands that either were simulated or planned to be simulated at some time:

- Estimating the size of a single, centralized, all-encompassing work-in-process (WIP) buffer planned to replace the many scattered buffers that currently exist in the plant.
- Forecasting the impact on operations of replacing the manual casting straightening operation with an automated one and how to most efficiently operate the new automated process in harmony with the other existing operations.
- Laying out a new set of welding cells that will be installed soon in the plant.

This thesis raises many issues associated with advantages and disadvantages associated with using "islands of simulation" for capacity planning vis-à-vis all-encompassing, plant-wide global models for long-term layout planning. One advantage of the "islands" approach is the fact that each island can be easily tailored and customized to meet the simulation needs of each layout planning project. Smaller, island-like models on average are simpler to create than one large, all-encompassing plant-wide global model and therefore should require less time and resources to develop. Smaller models run faster than large ones, favoring more "what-if" and sensitivity analysis. Islands lend themselves more easily to rapid evolution of alternative configurations. Simulation islands are more easy to verify and validate using back-of-the-envelope queueing approximations than are more large scale, plant-wide models.

Disadvantages of the "island" approach vis-à-vis the plant wide approach for helping plant layout decisions also exist. When following the island methodology, the modeler must make assumptions as to the inflows into the model and what happens with the outflows (e.g., transfer lot sizes, timing of the inflow and outflows, buffer capacities, potential for blocking by a downstream operation, potential for starvation by an upstream operation, etc.). If many islands are made, so must many inflow-outflow assumptions. On the other hand, in a plant-wide model, only one pair of inflow-outflow assumptions must be made. All hand-offs of parts and material within the plant's various process steps or cells are actually modeled, thus reducing the number of assumptions one must make. This reduction in assumptions might improve the accuracy of any results obtained and capture more precisely the possible interactions between upstream and downstream operations and a particular area, or island, of interest within a plant. Along this theme, in most situations, selecting appropriate boundaries for a plant-wide model is easier than for an island. In general, defining what is in a particular island and what is not is more difficult than defining what is in a plant level model and what is not, for a plant's boundaries are more obvious than any given sub-plant unit's boundaries.

Comparing the island (i.e., local) approach to the plant-wide (i.e., global) approach to using simulation to support layout decisions leads to an obvious question: Why not reap the advantages of using islands, and then connect all the islands to then reap the advantages of a plant-wide model? Ideally, this approach sounds like one that leverages all the advantages while minimize the disadvantages associated with both approaches. However, linking islands into one global model forces the model creators to create the islands with this eventual global linkage as a known eventual goal. If the island makers are different people, coordinating the development of the islands by the different people in such a manner to permit the islands' eventual linking increases the coordination required between the various modelers. If the underlying assumptions embedded in each island are not similar, the validity of the results from a global model composed of these different islands would be in serious question.

An example based on the local models made by the author helps illustrate some of issues. As mentioned before, one of the islands made was of a proposed consolidated buffer for all of the plant's WIP. Part of this model included a representation of the HSM area. However, the level of aggregation between the two models was quite different. The buffer model moved whole orders between process steps and the large buffer, whereas the HSM model moved fixtures of parts between machines. The buffer model was designed to simulate one month's production, while the HSM model was designed to model one day of production. The customers of each model had very different desires for metrics and outputs, one desiring information on the volume of total parts in WIP and the other desiring information on capacity lost due to waiting on the material handling system. One can thus see how linking islands to make a plant-wide model might not be as easy (or even possible) as initially perceived.

Given the advantages and disadvantages of using local models vis-à-vis global models for assisting layout decisions, some further observations can be made. In general, local models seem more useful in helping layout an area within an established plant. The plant-wide approach appears to be more helpful for assisting with the layout of greenfield sites or site where the whole floor is being totally redone. Once running, any plant-wide simulation model used to help layout the plant might be useful as the basis for a plant-wide operations-related model. The global model could be modified to serve as a capacity planning model for daily operational planning. Sequencing, due dates, and scheduling at an aggregate level could be assisted more by such a global model than using many individual island, or local, models.

Another issue raised by this work is the use of queueing models with simulation models together to make predictions on how well a proposed system not yet built will perform. Although the queueing model had embedded in it many simplifying assumptions, it did help validate and verify at least in part the results from the simulation analysis. Developing the queueing model did not require much work beyond the data collection, which presumably must be done for the simulation itself anyway. However, it served as a good check-and-balance to the simulation effort.

One final insight gained from this experience deals with some of the more "soft" issues of using simulation programs to aid in capacity planning. Using simulations is still new for many people, and people tend to fear and not trust the unknown. Given this, how should one go about getting buy-in from plant personnel and confidence in the model's output? How does one "sell" a model's usefulness and results to plant personnel who may not be familiar with simulation modeling? Based on his personal experiences, the author recommends the following advice:

- *Start small and then build up.* Present and sell a skeleton model, one that is easily understood. Following this, move on to more complex models. Starting small first helps demystify the whole idea of simulation for those people unfamiliar with it.
- *Keep the model's customer in the loop throughout the model creation process.* Start getting buy-in from the customer early in the model formation process, not during the final model presentation after it has been totally developed.
- *Clearly state all assumptions and aggregations* used to create the model.
- Time permitting, *develop a simulation of a system (preferably a simple one) that is already in place and has been operating.* Ideally, this system is one that is experiencing some sort of problem that a simple simulation would help solve. Show that the simulation can replicate past behavior and use it to help improve the system. Doing this "for fun" model will greatly enhance the plant's trust in later models of systems that do not yet exist (i.e., layout decision support).
- *Use user-friendly graphical interfaces and outputs* as much as possible when sharing the model as long as they do not slow down the model's execution significantly.
- If possible, *involve the customer in the execution phase of the model.* Transfer sufficient knowledge to enable the customer to run the model. Knowledge has to be transferred and used between people and parts of an organization where it can be

beneficial; doing so can aid with the creation of a learning organization (Senge, 1990).

- If using simulation to assist in predicting how well a proposed layout or system will work, validation and verification are more difficult, for no previous system behavior exists that the modeler can initially replicate in order to gain confidence in the model. In this instance, *using back-of-the-envelope calculations and queueing models* can be helpful in validating the model and increasing eventual model users' confidence in its results. Another method for validating recommendations is to *push the model's parameters to extreme points* and see if the model behaves as one would expect it to behave.

5.4 Concluding Remarks

This thesis has shown how a queueing model and simulation models can be used in tandem to assess and evaluate the feasibility of a proposed machine layout with its associated material handling system. A simulation island was used to conduct scenario analysis and evaluate different robotic material handling system dispatching heuristics. Project-specific and generic insights gained from this work were presented.

In general, the author also learned that if one wants to use simulation as a decision support tool, one must invest time up front to learn about simulation, learn a particular simulation package, and create a "base model." After these investments are made, however, the returns are many. Sensitivity analysis is quick, and different versions of the base model can be generated relatively rapidly. Simulation has the ability to be much less deterministic than more traditional forms of capacity planning. However, as with any modeling work, one must always keep in mind the underlying assumptions imbedded in any model developed and attempt to enumerate them as explicitly as possible. Also, creating confidence in a model's results requires much attention be paid to a wide array of people-related issues, which, in the end, are most likely the most important element of any modeling project.

Bibliography

Boyer, Howard E. and Gall, Timothy L., Ed. Metals Handbook, Desk Edition, Metals Park, Ohio: American Society for Metals, 1989.

Bratley, Paul et al. A Guide to Simulation, 2nd Ed., New York: Springer-Verlag, 1987.

Brennan, J.F. et al. "Go with the flow: Improving Red Cross bloodmobiles using simulation analysis," Interfaces, Vol. 22, 1992, September/October, pp. 1-13.

Brooks, Charlie R. Heat Treatment, Structure, and Properties of Nonferrous Alloys, Metals Park, Ohio: American Society for Metals, 1986.

Egbelu, P.J. and Tanchoco, J.M.A., "Characterization of automated guided vehicle dispatching rules," International Journal of Production Research, 1984, Vol. 22, pp. 359-374.

Egbelu, P.J. and Tanchoco, J.M.A., "Potentials for bi-directional guide paths for automated guided vehicle systems," International Journal of Production Research, 1986, Vol. 24, pp. 1075-1098.

Goldratt, Eliyahu M. and Cox, Jeff. The Goal, Croton-on-Hudson, New York: North River Press, 1992, p. 306.

Groover, Mikell P. Automation, Production Systems, and Computer Integrated Manufacturing. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1987.

Hahn, H.S. and Sanders, J.L. "Performance analysis of a LIM-based high-speed tool delivery system," International Journal of Production Research, 1994, Vol. 32, January, pp. 179-207.

Hall, Randolph W. Queueing Methods for Services and Manufacturing. Englewood Cliffs, New Jersey: Prentice Hall:, 1991.

Hardt, David. Professor, Massachusetts Institute of Technology, personal communication, 1995.

Hoffman, Paul. Extend Software Manual. San Jose, California: Imagine That, Inc., 1994.

Jewkes, E. and Power, M. "Simulating natural gas discovers," Interfaces, Vol. 22, 1992, March/April, pp. 38-51.

King, Robert I. Handbook of High-speed Machining Technology. New York: Chapman and Hall 1985, p.212.

Koszarek, Albert E. LFM '95 Fellow and Employee of The Boeing Company, personal interview, 4/28/95.

Kulwiec, Raymond A., Ed. Materials Handling Handbook, 2nd Ed., New York: John Willey & sons, 1985.

Kusiak, Andrew. "Working Paper # 06/85 - Material Handling in Flexible Manufacturing Systems," Technical University of Nova Scotia Department of Industrial Engineering, Halifax, Nova Scotia.

Law, Averill, and Kelton, David. Simulation Modeling & Analysis, 2nd ed., New York: McGraw-Hill, Inc., 1991.

Lindkvist, R.G.T.. Handbook of Materials Handling. West Sussex, England: Ellis Harwood Limited, 1985.

Mahadevan, B. and Narendran, T.T. "Design of an automated guided vehicle-based material handling system for an flexible manufacturing system," International Journal of Production Research, 1990, Vol. 28, pp. 1611-1622.

Mahadevan, B. and Narendran, T.T. "A hybrid modeling approach to the design of an automated guided vehicle-based material handing system for an flexible manufacturing system," International Journal of Production Research, 1994, Vol. 32, September, pp. 2015-2030.

Matsuo, Hirofumi et al. "Crane Scheduling and Machine Layout Problem in a Computer-Integrated Manufacturing Environment," Working Paper # 88-1-1, University of Texas at Austin, 1988.

Nahmias, Steven. Production and Operations Analysis, 2nd. Ed., Homeword, Illinois: Richard D. Irwin, Inc., 1993, pp. 449-450.

O'Keefe, David. Engineer, The Boeing Company, phone interview, 5/4/95.

Pegden, C. Dennis et al. Introduction to Simulation using SIMAN, Hightstown, New Jersey: McGraw-Hill, Inc., 1990, pp. 6, 12-13.

Senge, Peter. The Fifth Discipline: The Art and Practice of the Learning Organization, New York: Bantam Doubleday Dell Publishing Group, Inc., 1990.

Sterman, John D. Professor, Massachusetts Institute of Technology Sloan School of Management, 15.874 System Dynamics For Business Policy, class lecture notes, 1994.

Trunk, C., "Simulation for success in the automated factory," Material Handling Engineering, Vol. 44, 1989, pp. 64-76.

Tüchelmann, York. "Material flow components for automated flexible manufacturing," International Journal of Production Research, 1986, Vol. 24, pp. 787-797.

Walke, John R., Modern Metalworking, South Holland, Illinois: Goodhert-Wilcox Company, Inc., 1973, pp. 34-10-34-11, 38-17.

Warrick, Edward. "High Speed Machining Clinic and Additional High Speed Data," summary of notes from High Speed Machining Clinic in Oakbrook, Illinois (4/20/93-4/21/93), Seattle, Washington: The Boeing Company, 1993.

Weisman, Charlotte, Ed., Welding Handbook, Seventh Edition, Volume 1, Fundamentals of Welding, Miami, Florida: American Welding Society, 1981, pp. 119-120.

Biographical Note

Caryl Burnette Brown was born in Winter Park, Florida, and raised in St. Petersburg, Florida. After graduating Valedictorian from Boca Ciega Senior High School, he attended the University of Florida from Fall 1987 to Spring 1992 and participated in its Honors Program. While there, he was awarded the Florida College Student of the Year Award in 1990 and the Outstanding Male Graduate Award in 1992. He conducted internships with Honeywell and Exxon during the summer months and graduated in 1992 with his Bachelors Degree in Material Science and Engineering and a Minor in Business Administration.

Caryl then attended MIT for three years from Fall 1992 to Spring 1995. For the last two of those years, he was a Leaders for Manufacturing (LFM) Fellow. While at MIT, he served as the president of the Graduate Student Council for one year and was awarded the William L. Stewart, Jr. Award for his contributions to the MIT community. He conducted a 6-month internship in Germany with an aluminum company as part of the LFM Program that served as the basis for the work presented in this thesis. He will graduate Spring 1995 with a Masters of Science in Materials Science and Engineering and a Masters of Science in Management. He will then begin employment with ITT Automotive in Auburn Hills, Michigan, as a Manufacturing Operations Specialist in the Body Hardware and Structural Stampings Group.