

# Effect of Control Frequency on the Performance of Manufacturing Systems with Controllable Production Rates

by

José I. Castañeda Vega

S.M., Mechanical Engineering  
Massachusetts Institute of Technology, 1996

B.Sc., Mechanical Engineering  
Universidad de las Américas – Puebla, 1992

Submitted to the Department of Mechanical Engineering  
In Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY IN MECHANICAL ENGINEERING

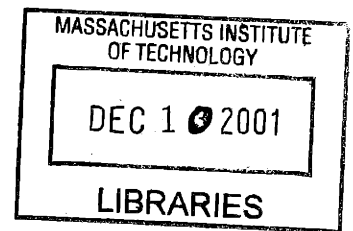
At the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

August 2001

[September 2001]

© 2001 Massachusetts Institute of Technology  
All rights reserved



ARCHIVES

Signature of Author: \_\_\_\_\_

*José I. Castañeda-Vega*

Department of Mechanical Engineering  
August 20, 2001

Certified by: \_\_\_\_\_

*David S. Cochran*

David S. Cochran  
Associate Professor of Mechanical Engineering  
Thesis Supervisor

Accepted by: \_\_\_\_\_

*Ain A. Sonin*

Ain A. Sonin  
Chairman, Committee on Graduate Students



# **Effect of Control Frequency on the Performance of Manufacturing Systems with Controllable Production Rates**

by

José I. Castañeda Vega

Submitted to the Department of Mechanical Engineering  
on August 20, 2001, in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy in Mechanical Engineering

## **Abstract**

Flow-line manufacturing systems represent the most prevalent process structure in industry for the repetitive production of discrete items. Machine breakdowns, however, limit their reliability and efficiency. The control of flow-line manufacturing systems, as a way to compensate for limited reliability, is about the frequent regeneration of production parameters so that the system output conforms to demand requirements in an acceptable manner. The influence of such regeneration frequency –or control frequency– on the system performance as well as its relevance with respect to other control parameters is, nevertheless, currently not well understood. So far, a continuous control action has been widely assumed in the performance analysis of unreliable manufacturing systems. This assumption follows naturally, given that most of the research work on shop floor control has been focused on job shop environments, such as flexible manufacturing systems. The reason for this emphasis on job shop environments could be the lack of flexibility in production rate of traditional repetitive manufacturing systems. This research work, on the other hand, was motivated by the realization that flow-line repetitive manufacturing systems can be designed for short term production rate control. Current implementations of lean manufacturing are cases of such systems.

The intended contribution of this work is a better understanding of the influence of control frequency on the time behavior of flow-line manufacturing systems. The development of pertinent simulation and analytical models for performance assessment is presented. From the simulation results, one effect of control frequency on system behavior motivating further research is evident. An analytical model resembling the simulation one is elaborated. The occurrence of control actions is modeled as homogeneous Markov processes. From system stability considerations –or convergence of the analytic solution– a closed expression involving control frequency is derived. According to simulation results, the analytical model predicts very well the limit of controllability of the system (lowest control frequency required for stable behavior) as well as all the steady-state parameters of interest. Additionally, the closed form of the solution allows a direct study of the complementary effect of control frequency with inventory, capacity and availability on system behavior. These results permit fast assessment of system-wide effects of operational control issues during the design of manufacturing systems.

**Thesis Supervisor: David S. Cochran**

**Title: Associate Professor of Mechanical Engineering**



## Acknowledgements

I would like to thank the numerous people who contributed to the development of this work. First of all, I thank my advisor Professor David Cochran for giving me the opportunity to conduct research in manufacturing system design. His guidance and support significantly contributed to this thesis and my professional growth during the doctorate program at M.I.T. The members of the doctoral committee, Dr. Daniel Whitney and Professor David Hardt, were extremely helpful in defining the scope, approach and depth of the research as well as always very generous with their advice. I also thank Dr. Stanley Gershwin; through his previous work and his friendly advice he strongly influenced this research. I am grateful to all colleagues from the Production System Design Laboratory at M.I.T. for their ideas, their time and specially their friendship. Also, the sponsorship of Ford Motor Company and Visteon Automotive Systems is gratefully acknowledged. The support from and access to both of these companies provided me with motivation and resources to carry out this research. I am thankful to every member of my large family for their continuous and unconditional support and trust. In particular, I am indebted to my beloved parents, María Concepción and José Israel, for everything they have done for me and everything they have given me. My sister Jacqueline provided the support and encouragement I greatly needed to finish this project, my last semester at M.I.T. without her company would have been quite difficult to endure. Finally my sincere gratitude goes to somebody special: Ania Mierzejewska. I dedicate this work to José María Castañeda Vega.



## TABLE OF CONTENTS

<b>1</b>	<b>INTRODUCTION.....</b>	<b>11</b>
1.1	PREVIOUS WORK .....	12
1.1.1	<i>Control of Manufacturing Systems</i> .....	13
1.1.2	<i>Performance Analysis</i> .....	14
1.2	MOTIVATION.....	17
1.3	MODELING APPROACH.....	25
<b>2</b>	<b>THE EVOLUTION OF MANUFACTURING CONTROL PRACTICE AND THE MANUFACTURING SYSTEM DESIGN DECOMPOSITION .....</b>	<b>27</b>
2.1	INTRODUCTION .....	27
2.2	EVOLUTION OF MANUFACTURING SYSTEMS .....	28
2.2.1	<i>First Stage: Pioneering</i> .....	29
2.2.2	<i>Second Stage: Systemization</i> .....	29
2.2.3	<i>Third Stage: Technology and Process</i> .....	30
2.2.4	<i>Fourth Stage: Internal Efficiency</i> .....	31
2.2.5	<i>Fifth Stage: Customer Service</i> .....	32
2.2.6	<i>Sixth Stage: System-Level Reengineering</i> .....	33
2.3	THE MANUFACTURING SYSTEM DESIGN DECOMPOSITION.....	37
2.3.1	<i>Overview</i> .....	37
2.3.2	<i>High-Level Description of the MSDD</i> .....	38
2.4	CHAPTER SUMMARY .....	58
<b>3</b>	<b>ANALYTIC MODEL .....</b>	<b>59</b>
3.1	INTRODUCTION .....	59
3.2	MODEL ASSUMPTIONS AND CONSIDERATIONS.....	60
3.3	ANALYTIC MODEL .....	71
3.3.1	<i>Probability Distribution and Transition Equations</i> .....	74
3.3.2	<i>Solution of Distribution and Transition Equations</i> .....	79
3.3.3	<i>Repair Frequency Equal to Failure Frequency Identity</i> .....	82
<b>4</b>	<b>MODEL VALIDATION AND RESULTS .....</b>	<b>87</b>
4.1	INTRODUCTION .....	87
4.2	MODEL VALIDATION .....	88
4.3	DERIVATION OF PERFORMANCE MEASURES .....	92
4.4	EFFECT OF CONTROL FREQUENCY ON SYSTEM PERFORMANCE .....	94

4.5	GENERALIZATION OF ANALYTICAL MODEL TO MULTI-STAGE MANUFACTURING SYSTEMS .....	106
4.6	DOCUMENTATION OF THE EFFECT OF CONTROL FREQUENCY WITH REAL SYSTEM PARAMETERS .....	109
4.7	CHAPTER SUMMARY .....	112
<b>5</b>	<b>CONCLUSIONS .....</b>	<b>115</b>
	<b>REFERENCES.....</b>	<b>119</b>
	<b>APPENDIX.....</b>	<b>123</b>



## LIST OF FIGURES

<b>FIGURE 1.1</b>	A PRODUCTION PLANNING AND CONTROL HIERARCHY FOR PULL SYSTEMS.....	13
<b>FIGURE 1.2</b>	CASUAL FACTORS TO REALIZE PRODUCTION RATE CONTROL. ....	20
<b>FIGURE 1.3</b>	EFFECT OF CONTROL PERIOD ON STEADY STATE AVERAGE BACKLOG. ....	23
<b>FIGURE 1.4</b>	THE TWO-STAGE, SINGLE-PART TYPE MANUFACTURING SYSTEM.....	24
<b>FIGURE 2.2</b>	HIGH LEVEL ABSTRACT VIEW OF MANUFACTURING SYSTEM BUSINESS OBJECTIVES...	39
<b>FIGURE 2.2</b>	HIGH-LEVEL DECOMPOSITION OF THE RETURN ON INVESTMENT FR.....	42
<b>FIGURE 2.3</b>	HIGH-LEVEL DECOMPOSITION OF QUALITY. ....	48
<b>FIGURE 3.1</b>	LIQUID LEVEL CONTROL SYSTEM AS ANALOGY FOR THE CONTROL .....	62
<b>FIGURE 3.2</b>	BLOCK DIAGRAM REPRESENTATION OF A PRODUCTION RATE CONTROLLER.....	64
<b>FIGURE 3.3</b>	PRODUCTION SURPLUS X OF UNRELIABLE UNIT.....	67
<b>FIGURE 3.4</b>	MARKOV-CHAIN MODEL FOR $X < Z$ OF A GENERAL SINGLE-STAGE .....	73
<b>FIGURE 4.1</b>	COMPLEMENTARY EFFECT OF CONTROL FREQUENCY.....	95
<b>FIGURE 4.2</b>	EFFECT OF INVENTORY ON SYSTEM PERFORMANCE .....	96
<b>FIGURE 4.3</b>	COMPLEMENTARY EFFECT OF CONTROL FREQUENCY.....	97
<b>FIGURE 4.4</b>	EFFECT OF SYSTEM CAPACITY ON SYSTEM PERFORMANCE.....	98
<b>FIGURE 4.5</b>	COMPLEMENTARY EFFECT OF CONTROL FREQUENCY.....	99
<b>FIGURE 4.6</b>	EFFECT OF SYSTEM AVAILABILITY ON SYSTEM PERFORMANCE .....	100
<b>FIGURE 4.7</b>	PROBABILITY DENSITY DISTRIBUTIONS FOR CONTINUOUS CONTROL ACTION. ....	101
<b>FIGURE 4.8</b>	EFFECT OF CONTROL FREQUENCY ON STATE 2 PROBABILITY DENSITY DISTRIBUTION	102
<b>FIGURE 4.9</b>	EFFECT OF CONTROL FREQUENCY ON SYSTEM STATE PROBABILITIES.....	104
<b>FIGURE 4.10</b>	COMPARISON OF ANALYTICAL AND SIMULATION RESULTS. ....	105



# 1 Introduction

This thesis is about the control of the production stages of flow-line manufacturing systems. The operational control of flow line manufacturing systems is intended to compensate for the effects of short term disturbances such as machine breakdowns and consists of a frequent regeneration of production parameters so that the system output conforms to demand requirements in an acceptable manner. A flow-line manufacturing system consists of multiple production stages separated by storage buffers, where parts being processed visit each stage just once and hence, the path of material flow is fixed. This type of manufacturing system is considered as the best-suited process structure for high-volume, limited product-variety applications. It involves shorter throughput times and requires fewer resources for operation and control than other structures.

The main problem that this study addresses is the difficulty in understanding or quantifying the effect of the regeneration frequency –or control frequency– on the time behavior of the system as well as its relevance with respect to other control parameters. The control decision under consideration determines a feasible production rate that once implemented is expected to compensate for the negative effect of limited equipment reliability. For evaluation purposes, the time behavior of the system is measured in terms of its delivery performance. The Manufacturing System Design Decomposition [PSD Lab, 2000] provides a platform to locate the operational control of a flow line system within the context of system design and control.

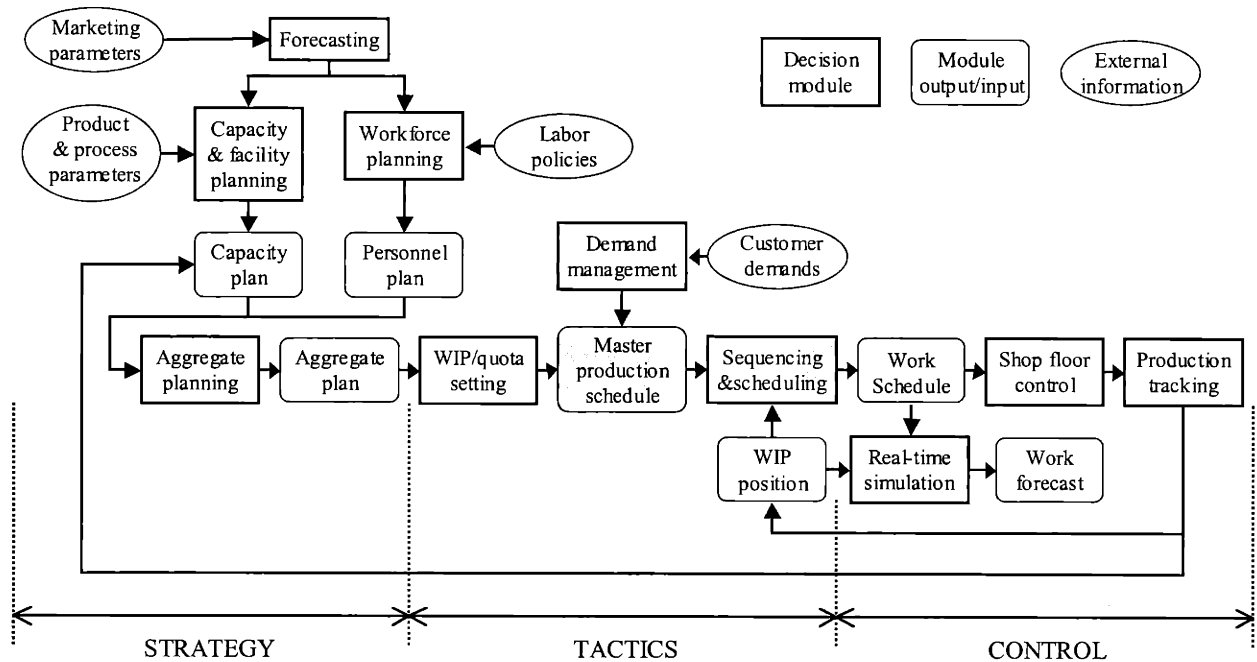
The objective of this work is to contribute to a better understanding of the time behavior of manufacturing systems and its control. While working on this analysis, it became clear that the frequency for control is in fact another system variable that should be included in the specification of a manufacturing system design. As such, control frequency affects and is related to other basic system variables such as capacity, reliability and inventory. The results of this work allow the assessment of the relative effect of control frequency on system behavior with

respect to, or in comparison to the effect of the other basic system variables. In this introductory chapter, previous work in the areas of production control and manufacturing system performance evaluation as well as a simulation study whose results motivated this work are presented.

## **1.1 Previous Work**

Control decisions in manufacturing differ greatly with regard to the length of time over which their consequences persist. For that reason, the definition of different time scales for production planning and control has been a convenient approach in the control of manufacturing systems. Hopp and Spearman [1996], for example, proposed a production planning and control hierarchy for pull systems with strategic, tactical and control time scales corresponding to long, intermediate and short-term actions, as indicated in Figure 1.1. Decisions corresponding to a time scale are taken at frequencies that are orders of magnitude different from the decision frequencies of another time scale. Within any time frame, very few researchers have studied the effect of frequency related issues on system performance [Berkley 1992].

Production control of manufacturing systems is generally classified in the research literature, under either control policies or shop floor control. It has been found, however, that the literature does not make a clear distinction between what should be classified under control policy and under shop floor control, possibly because both are intended to compensate for the many variations in time behavior that are present in manufacturing systems. From the following sections on manufacturing system control and performance analysis it can be said, in general terms, that a considerable amount of research work has been done on understanding the effects of control policies on system performance, and that attention has been given to the study of real-time control and scheduling of unreliable manufacturing systems, within the scope of shop floor control.



**Figure 1.1** A production planning and control hierarchy for pull systems.  
Adapted from Hopp, Spearman, *Factory Physics* 1996.

### 1.1.1 Control of Manufacturing Systems

The term control policy generally refers in manufacturing to a specific approach to the coordination and control of material and information flow in multiple-stage manufacturing systems [Buzacott and Shanthikumar, 1992]. A control policy is specified by the way customer demand information is translated into production authorizations for each unit of the production system. The coordination is achieved by the rules that determine when and how material and information flow through the system, and by the appropriate choice of the parameters of these rules. It can be observed then, that the control action exerted on a specific production unit by the control policy is the release time of production authorizations.

Examples of control policies are MRP (material requirements planning), KANBAN (card system), OPT (optimized production technology), BSS (base stock system), IC (integral control) and CONWIP (constant work-in-progress). They can be classified according to the so-called push and pull control strategies. Spearman [1997] defines a “pure” push system as one that schedules releases considering only due dates, without regard to shop floor status. A “pure” pull system, on the other hand, determines releases entirely by the status of the shop floor while ignoring due dates. Material Requirements Planning, without input/output control modules from Manufacturing Resources Planning (MRP II), is an example of a pure push system. KANBAN and CONWIP working without a production schedule are examples of pure pull systems. Spearman states that in practice there are no pure push or pull systems. All production systems have some sort of feedback regarding either due dates or shop floor status. This is why the control policy specification is a nontrivial part of production system design.

Shop floor control, in contrast to control policies, is focused on the short-term production control of manufacturing systems. Its emphasis is on minimizing the negative effects of system variations such as machine unreliability. Shop floor control includes functions such as status monitoring, work-in-progress and throughput tracking, load forecasting, quality control and capacity feedback. Recently, there has been increasing interest in control theory approaches to the production rate control of manufacturing systems with machines subject to failures and repairs.

### **1.1.2 Performance Analysis**

Kimemia and Gershwin [1983] study a single-stage manufacturing system with an arbitrary number of machine states and product types. Each machine state has a corresponding production rate. The objective of the analysis is to determine an optimum production policy under each machine state, so that a known demand production rate target is met with minimum inventory holding cost and demand backlog cost. By formulating this problem as a dynamic programming problem they show that in general it is extremely difficult to obtain a true optimal control law since it requires solving a complex Hamilton-Jacobi-Bellman equation. However, they show that

the solution for the optimum production rate in each machine state can be reduced to the solution of linear programs provided that a cost function is available.

A very important result of this work is the concept of optimal control structure. Kimemia and Gershwin [1983] and Akella and Kumar [1986] showed that for systems with homogeneous Markov processes, the hedging point policy is optimal. A homogeneous Markov process has constant transition rates, which means that the mean transition times of the process are constant. Under the hedging point control policy, a nonnegative safety stock of parts is maintained at times of excess capacity in order to hedge against future capacity shortages. In other words, for each feasible machine state (in which there is enough capacity to keep up with demand), there is a specific target called hedging point. The optimum production policy will move the inventory level towards the hedging point as quickly as possible, and when at this point, it will keep it there as long as the machine state remains unchanged. This result has been regarded as the first structural result of the optimal production control of failure prone manufacturing systems [Hu and Xiang 1994]. Structural properties of the optimal control are very helpful because they can facilitate the derivation of near-optimal controls and, in some cases, simplify the problem to obtain analytically a true optimal control.

A considerable amount of recent work on hedging point policies covers multiple aspects of manufacturing systems. Akella and Kumar [1986] find an exact solution of the optimal inventory level for the case involving a single-stage single-product and two-machine-state (up and down) system. The analysis is quite complicated even for this simple case. Bielecki and Kumar [1988] use another approach to solve the single-product single-stage two-machine-state problem. Assuming that the hedging point (corresponding to the up state) is known, they calculate the steady-state probability density function of the surplus and the probability mass at the hedging point. The optimum hedging point is then found by minimizing the average cost per time unit over all possible values of the hedging point. This result shows that for an unreliable manufacturing system under a continuous-time hedging point control policy, a zero hedge point, or zero-inventory policy can actually be optimal. This follows from relative effects of machine unreliability and holding and backlog costs.

Sharifnia [1988] describes a method for solving the short-term production control problem of a single part-type single-stage manufacturing system with multiple machine failure states. Each machine failure state corresponds to a failure mode where some machines in the set have failed, resulting in a reduced capacity. He assumes that a hedging point policy is optimal and derives equations describing the steady state probability density function distribution of the surplus and the probability mass of a set of hedging points. Then the average surplus cost is easily calculated in terms of the values of the hedging points. This average cost is then minimized to find the optimum hedging points. Basically, this is the same approach of Bielecki and Kumar [1988]. Glasserman [1995] works on the same problem but assumes that the machine state is governed by a semi-Markov process. His results require that intervals in which demand exceeds production are exponentially distributed.

Hu and Xiang [1994] consider the case where each machine has a failure rate that depends on the age of the machine, as opposed to the more commonly assumed constant failure rates in homogeneous Markov processes. The production system under consideration processes a single part-type to meet a constant demand rate with unreliable machines and with multiple failure modes. They explain that because in practice the older the machine the more likely it is to fail, it makes sense to relate the failure rate to the age of the machine. They show that the closer to the zero capacity state the system is, the larger the hedging point should be.

Other authors contributed in the area of non-homogeneous Markov processes in terms of the dependence of machine failure rates on production rates. Hu, Vakili and Yu [1994] study the necessary and sufficient conditions for the optimality of the hedging point policy for the control of production systems in which the failure rate of the machines depends on the rate of production. They focus on the usual one part-type single-stage system with infinite discounted horizon cost. They claim that the linearity of the failure rate function is both necessary and sufficient for the optimality of the hedging point policy. Liberopoulos and Caramanis [1994] confirm these observations following a different approach.

Some other authors, on the other hand, extended the approach to the area of multiple part type manufacturing systems. Srivatsan [1993] studied a single-stage two-part generalization of



the Bielecki-Kumar problem. However, he was able to obtain a complete solution only when both hedging points were zero. Srivatsan and Dallery [1998] provided heuristic insights to develop controllers for manufacturing systems producing many part-types. Ryzin, Lou and Gershwin [1993] studied the optimal production control of a two-stage, single-product manufacturing system. They suggest that appropriately constructed hedging point policies are close to optimal.

Research on the production control of more general systems has been limited by the complexity of the analysis. In this area most of the work is based on simulation models and suggests algorithms for real-time production control [Bai and Gershwin 1994], [Gershwin 1997].

## 1.2 Motivation

The most general problem motivating the work mentioned above is about the production rate control of *flexible manufacturing systems*, or multi-stage manufacturing systems with multiple product part-types and failure states. According to Hill [1994] flexible manufacturing systems are a combination of standard and numerical control machines, automated materials handling and computer control for the purposes of extending the benefits of numerical control to mid-volume manufacturing situations. According to Gershwin [1994] a flexible manufacturing system is a system whose machines are able to perform operations on any sequence of parts with little or no time or other expenditure for changeover, and with choice of one or more stations for each operation.

In practice, the term flexible manufacturing system refers to systems in which the operations at the workstations and the material handling system are entirely under computer control. The FMS control computer makes decisions such as what parts should be loaded into the system and what workstations each part should visit next. Human intervention is necessary, but only when unusual or unanticipated events take place. Frequent regeneration of production schedules is required to satisfy demand requirements and exercise control over the system so that the output conforms to the schedule. On the other hand, the term job shop refers to a specific type of process structure where relatively small lots of parts are produced with a high variety of

routings through the system [Hopp and Spearman 1996]. In job shops the flow of material through the system is jumbled, setups are common and the environment has more of an atmosphere of project work than pacing. Hence, based on these observations, it can be concluded that a flexible manufacturing system is basically an automated job shop.

The hedging point policy has also been considered in the study of the control of flow line process structures, see Bonvik [1996]. With an increase in volumes and the repeat nature of products, companies select connected and disconnected flow lines as the effective way to meet the requirements involved. Because the products are repeated, companies consider investment at each of the necessary manufacturing steps. This includes engineering the best sequence to make the product, as well as jigs and fixtures and equipment dedicated to products with similar characteristics. The majority of manufacturing systems in industry resemble the disconnected flow-line environment to some extent [Hopp and Spearman 1996]. In disconnected flow lines, individual stages are not paced by a centralized control or material handling system and therefore, inventories can build up between stages. In a connected flow line, on the other hand, products are fabricated or assembled along a rigid routing connected by a paced material handling system. Moving assembly lines and machining transfer lines are examples of connected flow-line processes.

Linked-cell manufacturing systems, having a limited number of identifiable routings dedicated to products of the same family, belong to the disconnected flow line process structure. Linked-cell manufacturing systems, however, differ from more traditional flow-line processes in the sense that they are flexible to operate over a range of production rates [Cochran and Charles 1997] through a method called Shojinka in the Toyota Production System. Shojinka is of significant importance when the number of workers assigned to a production unit is changed to respond to fluctuations in demand. Most implementations of multi-stage flow manufacturing allocate inventory under the assumption of a constant-cycle or a constant-quantity replenishment of material between adjacent processes. Then, when fluctuations in demand become too large with respect to the smoothed, aggregated or average demand, the allocation of inventory is increased. The corresponding approach at Toyota plants, however, is increasing the production rate of the system while keeping inventory levels rather constant.

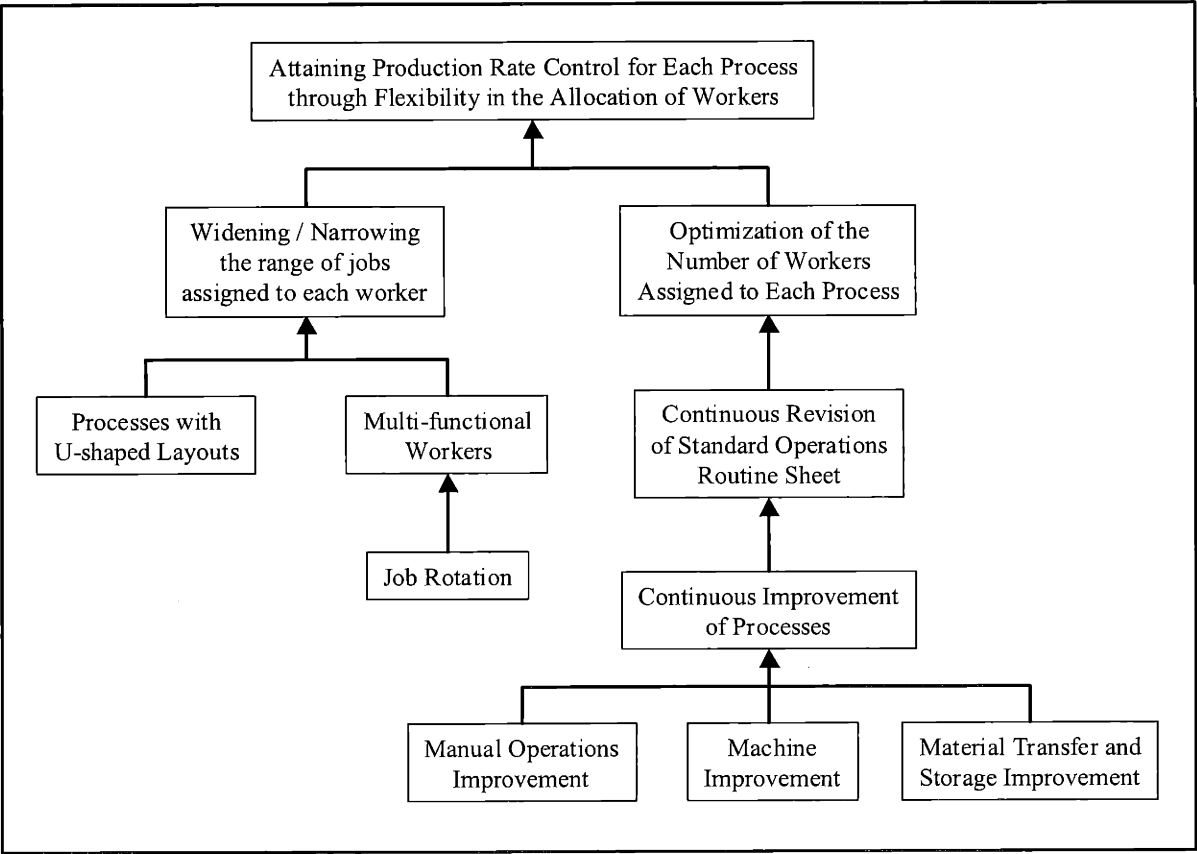
The term Shojinka in the Toyota production system refers to attaining flexibility in the number of workers at the shop floor. Shojinka in the Toyota production system relates to the capability of allocating a flexible, multi-skilled workforce to decrease or increase the production rate of the system [Monden, 1993]. This approach is unique when compared to more traditional inventory systems. Excess inventory is seen as the root cause of multiple kinds of waste, while volume fluctuations in demand are not considered a justification for higher inventory levels. Thus, inventory levels are kept rather constant at Toyota. When the average daily demand increases, the lead-time should decrease. In other words, when the demand increases, the cycle time of standard operations is reduced by changing the allocation of workers in the line, that is, the production rate of the system is modulated by changing the allocation of workers in the line. Production rate control then, allows the system to respond to demand fluctuations and other transitory behavior variations such as those resulting from limited machine reliability and non-perfect quality yield.

Shojinka is seen as a way to increase labor productivity by the adjustment and rescheduling of human resources. Monden [1993] identifies three factors as prerequisites for the implementation of Shojinka:

- Proper system layout design.
- Versatile, multi-skilled and well-trained workforce.
- Continuous evaluation and periodic revision of the standard operations routine.

The ideal process layout for Shojinka is the U-shaped cell. With this layout, the range of jobs for which each worker is responsible can be widened or narrowed and hence the allocation of workers in the cell is flexible. However, this flexibility requires the existence of multi-skilled workers. Multi-skilled workers at Toyota are developed through a multi-level job rotation system that covers the supervisors and the workers within each process or cell and ultimately promotes rotations several times per day. Additionally, revisions of standard operations routines involving improvement of manual and automatic tasks are made continuously. The purpose of such improvements is to reduce the requirements on labor for each production rate or demand level. The relationship among these fundamental prerequisites is shown in Figure 1.2.

The actual allocation of production resources at each process within Toyota factories is not determined rigidly or automatically by specific formulae. The supervisor has the capability to influence the inventory and labor requirements of the system during operation. In fact, each supervisor is expected to continuously reduce the production requirements while keeping the process behavior at the desired level. To do so, the supervisor needs to establish a standardized operations routine and a consistent or *level* production pace to result in a predictable production flow that by its nature displays the problems as they appear and allows the implementation of prompt corrective actions.



**Figure 1.2** Casual factors to realize production rate control.  
Adapted from Monden, *Toyota Production System*, 1993.

According to Rother et al [1998] small, consistent releases of work enable predictable production flows. They recommend, according to their experience in flow manufacturing, to release production instructions corresponding to intervals worth between 5 and 60 minutes. They refer to the consistent increment of work as pitch and is often calculated based on the smallest packing container capacity. For example, if the takt time is 30 seconds and the pack size is 20 pieces, then the pitch is 10 minutes = 30 seconds  $\times$  20 pieces.

The amount of work that is released at regular intervals into the system is related to the frequency of control in the sense that it determines how often the system performance with respect to customer demand is monitored and a corresponding compensatory action implemented. If the production work is released in batches corresponding to a week or so, then it is understood that the system state is monitored once a week or so too, resulting in a situation where it is impossible to produce to takt time. Such situation is not desired because the system could be either getting recurrently to far behind schedule and damaging the system behavior or be making use of unnecessary resources.

There are multiple ways to practice paced release of small, consistent quantities of work. A tool used at some companies to help level both the mix and volume of production is a load-leveling box called Heijunka. A load-leveling box has a column of Kanban slots for each pitch interval, and a row of Kanban slots for each product type. The load-leveling box indicates not only the quantity to be produced, but also how long it takes to produce such quantity (based on takt-time). Kanban cards are loaded into the leveling box in the desired mix sequence by product type. Then the material handler withdraws those cards, one at a time and at each pitch increment, and releases the corresponding production authorizations to the process or cell. Hence the material handler monitors and controls the production unit by releasing production authorizations according to the leveled schedule and if needed reallocating workers to avoid system behavior degradation.

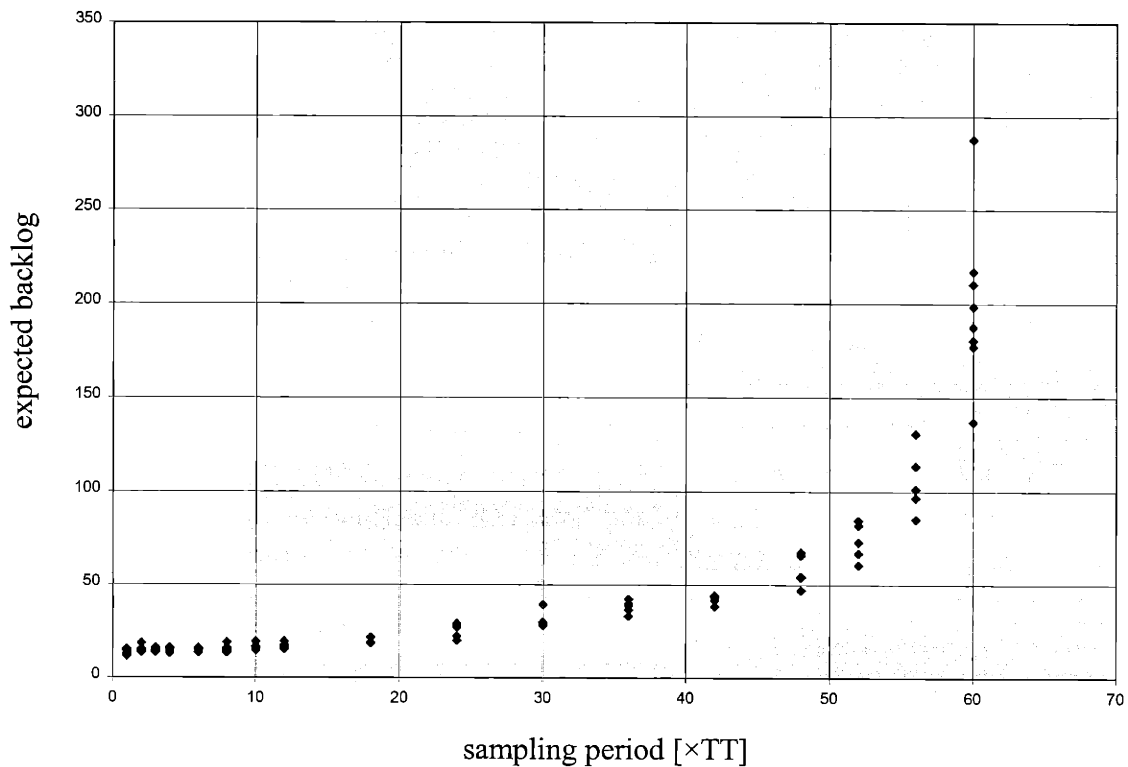
There is a limit, however, to how frequently the production rate can be changed so as to compensate for system output variations. First, analytical results presented in Chapter Four indicate that the improvement in system behavior from more frequent control actions becomes

small after some frequency. This is exactly the same effect observed on inventory allocation. After some inventory level, more inventory allocation has a minimal effect on system behavior. Second, there may be in practice a cost or penalty associated to changing the production rate. In other words, bringing additional labor to the production unit may involve a cost that could result higher than the savings related to the corresponding improvement in customer satisfaction through on-time delivery. Hence, there is an intrinsic optimization problem in system design related to the specification of control frequency. The decision on what control frequency is best for a specific situation is strongly influenced by the way the system performance is measured and the cost is allocated and is beyond the purpose of this work.

So far, a continuous action of the hedging point control policy has been assumed. While this may be true in the case of the control of job shop systems, given that a control decision has to be made every time a part is released into the system or is completed at a station, the control of repetitive manufacturing systems is different. The operation of repetitive manufacturing systems does not require control actions to be made every time a part is completed or released. There is no decision to make concerning routing or sequencing, and the main reason for control is compensation for limited machine reliability. For that reason, the short-term production control of repetitive manufacturing systems can in practice consist of intermittent actions aimed at adjusting the production rate depending on the system status. In implementations of linked-cell systems, for example, a supervisor exercises control intermittently with intervals ranging from minutes or hours to weeks. The realization of this fact motivates a study of the influence of control frequency on the performance of unreliable flow-line manufacturing systems.

Two main issues emerged during an exploratory study with discrete event simulation of the effects of control frequency on manufacturing system performance. There is an upper limit on how long the control period can get for a stable system response, as indicated on Figure 1.3, and inventory and control frequency have a complementary effect on system performance. Figure 1.3 below shows the effect of control frequency on the performance of the two-stage, single-product manufacturing system under constant demand presented in Figure 1.4. Both processes are unreliable, with normal distributions assumed for the time to failure and time to repair. Each process is controlled by a hedging point policy, where the hedge point is equal to

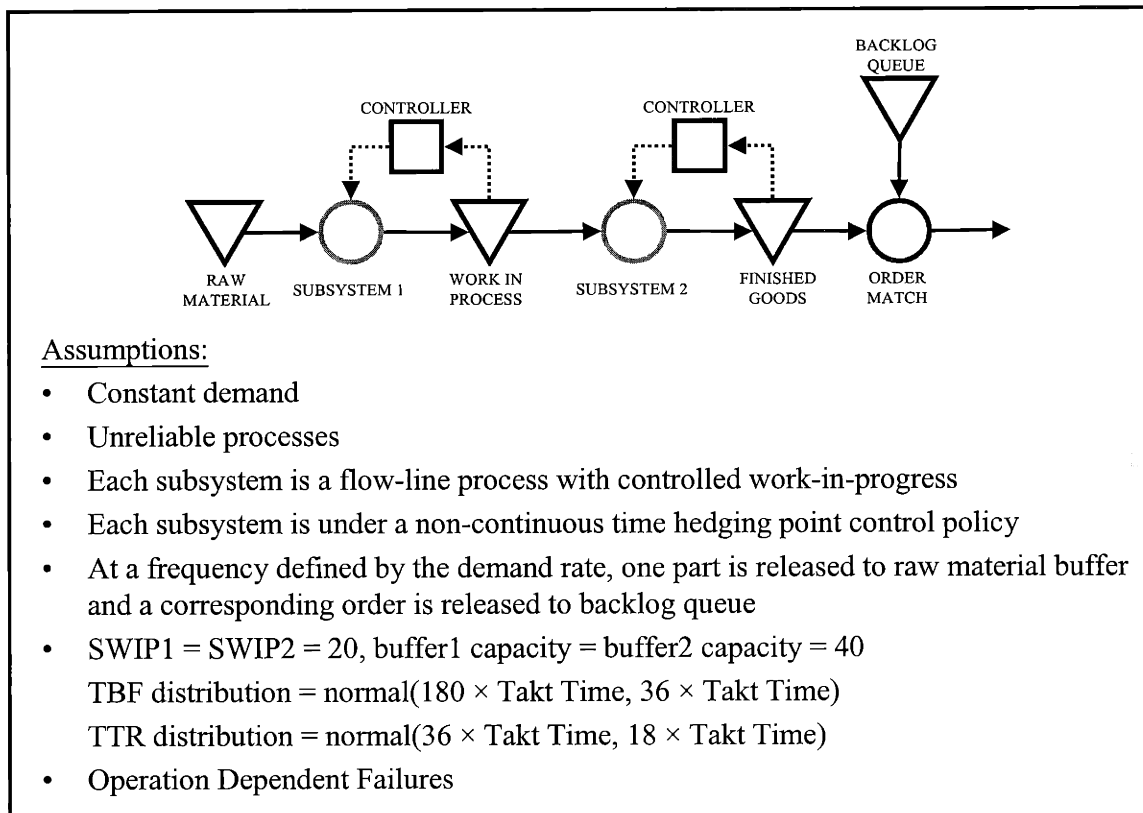
half the size of the corresponding buffer. In fact, each process in itself is a connected flow-line process with constant work-in-progress, just as the individual cells of a linked-cell manufacturing system. The performance of the system is evaluated in terms of the steady-state average backlog. The points in the graph were generated by changing the frequency at which the controller corrects the production rate as dictated by the hedging point policy, while keeping all other parameters unchanged. It was observed that allocating more space for inventory or increasing the control frequency result in a better system performance, and hence deduced that inventory and control frequency have a complementary effect on system behavior.



**Figure 1.3** Effect of control period on steady state average backlog.  
(Control or sampling period in multiples of demand takt-time)

As can be observed on Figure 1.3, the effect of control frequency on system performance can become significant. For the set of parameters describing the system under consideration, (buffer sizes, hedging points, parameters of time distributions for machine unreliability), it was

observed that the longest sampling period for stable behavior was about sixty times the takt time. Furthermore, it can be observed that the control frequency also influences the system performance variability. Steady state values of the average backlog cover increasing ranges as the control frequency decreases.



**Figure 1.4** The two-stage, single-part type manufacturing system studied with discrete event simulation.

With the intention of improving the current understanding of issues related to the control frequency of production rate controls for unreliable systems, it becomes clear that a more analytical approach is needed for evaluation. Previous work has already covered many relevant issues of hedging point controllers. Analytical models are preferred over simulations because they allow the development of an understanding about the structural issues that affect the behavior of a complex system. Simulation models have already proved to be very helpful for



exploratory studies, but they can also be useful for validation purposes. Computation requirements limit the use of simulation models, the generation of the data presented on Figure 1.3, required no less than five full days of computation time.

## 1.3 Modeling Approach

The modeling approach proposed in this study is based around a homogeneous Markov process to represent an unreliable, volume-flexible flow-line production stage. The occurrence of control actions is modeled in the same way failures and repairs are modeled, with the purpose of taking advantage of the simplified mathematics. This implies that the occurrence of control actions is assumed to follow an exponential distribution as opposed to be a deterministic event defined by a constant sampling period. Due to the fact that in almost every manufacturing system a human supervisor is in practice the agent implementing control actions, and hence the occurrence of controls actions takes place at intervals of variable length, this assumption seems to be reasonable.

The solution of the boundary problem resulting from the model has a closed form that allows the derivation of closed expressions for relevant parameters such as steady-state state probabilities and probability density functions. The solution and its results are then validated with discrete event simulation. The closed form of the results allows a better understanding of the relative effect of each parameter involved. Then, the results are extended to the analysis of multi-stage manufacturing systems. A conversion of parameters that groups control and failure transitions into a single transition allows the generalization of the results, by translating the parameters of a discretely controlled system into those of an equivalent continuously controlled system. The generalization is possible because once the parameters have been lumped together, then the analytical tools already available for the evaluation of flow-line manufacturing systems can be applied directly to the evaluation of multi-stage manufacturing systems under discrete control actions.



# **2 The Evolution of Manufacturing Control Practice and the Manufacturing System Design Decomposition**

## **2.1 Introduction**

The design of manufacturing systems is concerned with the allocation, specification and arrangement of production resources. The control of manufacturing systems deals with the decisions affecting the coordination of production resources that determines how much, what and when to produce. Such coordination is required so that the system delivers the desired quantities, with the appropriate quality, at the required time and at competitive costs, meeting the functional or operational requirements of the manufacturing system and the objective of the system design. The control decisions involve a vast spectrum of constraints, objectives and trade-offs, covering the strategic, tactical and operational levels of the enterprise. Hence, the design and the control of manufacturing systems are closely related, and both should be integrated for development and improvement of manufacturing systems.

A specific control, being appropriate for a given set of production resources under a specific situation, can lead to the best achievable performance such manufacturing system can attain. That same control approach, however, can be misused if applied to different resources or under different situations, leading to waste and inefficiencies. An evolution of manufacturing systems has been observed in industry. Such evolution of manufacturing systems seems to follow a dynamic environment of ever-changing economic, social and technological requirements.

This chapter reviews different practices for manufacturing system design and control that have been observed since the end of the nineteenth century, and the industry situation they best

seem to suit. Then the Manufacturing System Design Decomposition (MSDD) as an integral approach to the design and control of manufacturing systems is introduced. This material is intended to serve as context and definition of scope for the study presented in subsequent chapters.

## **2.2 Evolution of Manufacturing Systems**

Manufacturing firms are on an eternal quest for efficiency. Concerned about waste elimination or at least waste reduction, efficiency is broadly regarded as a competitive advantage. Competitive threats and corresponding increasing demands for efficiency seem to take firms along a path of evolution and in some industries along a cyclic pattern of change and adaptation to those demands [Fine, 1998]. Traditionally, manufacturing has been seen as the most difficult link to integrate into a supply chain. Manufacturing has been considered as the root cause of problems with both procurement (for delivering defective or late products) and with sales and marketing (for not meeting customer needs) [Porteus and Whang, 1991]. It could be said, therefore, that manufacturing is commonly considered as inherently inflexible.

Mckay [2001] proposes a path of evolution for manufacturing systems along six identified stages or phases of manufacturing practice: pioneering, systemization, technology and process, internal efficiency, customer service and system-level reengineering. The stages are defined in terms of the specific case of each issue identified as relevant for competitiveness, such as inventory management, internal resource coordination and supply, and distribution and sales. At each stage, the environment, by setting the customer expectations, defines the focus of manufacturing practice for all competitors. The way a firm meets these expectations, and its competitive position, is understood as a direct consequence of its specific approach to the competitive issues mentioned above. The stages of manufacturing practice, along with the corresponding environments, are explained below with the purpose of providing a base for understanding of how the design of a manufacturing system is influenced by its environment.

## **2.2.1 First Stage: Pioneering**

Firms enter this stage after a major change in the product, the processing technology or the environment. A major internal change could be an invention resulting in what is regarded as a “killer technology”, because of its effects in the existing competing technologies. It could be an invention resulting in a much better product or service, or resulting in a better manufacturing process. The advent of the automobile, integrated circuits, ceramics, are cases of such changes. A major external change leading to this stage does not necessarily involve a technological change, but still leads companies to reinvent themselves or disappear. World War Two could be regarded as this type of change.

This stage is characterized by the existence of minimal or no serious competition, limited availability of required technology resulting in high barriers to entry, vertical integration, as well as limited availability of skilled labor. This is a stage of fast growth and expansion, the firm can sell everything it produces. Delivery, quality and cost are not of concern, the focus is on getting into the market by making as much as possible. Inefficiencies in the production system are, therefore, very common. The firm just does not justify improvement efforts because waste and inefficiency are seen as an economical approach or the only way to address the huge demand.

The manufacturing practice would consequently include:

- Maximization of resource utilization at bottlenecks.
- Cloning of factories to reduce start-up problems.
- Design of product for high volume manufacturing.
- Product simplification, single variant or very limited variations offered.
- Large allocations of inventory.

## **2.2.2 Second Stage: Systemization**

After a stage of rapid growth focused on getting products to market, firms eventually realize that they need to change their approach. In a rush to maximize production output, firms usually give

individual departments freedom to do their own equipment purchasing and system design, allocating new equipment wherever there is space available in the existing facilities. The size and complexity of the situation often leads to chaos, few people knowing where something is, where it is going next and when it is going to reach the customer. This confusion leads to the recognition that standard operating procedures are necessary once the focus shifts to overall efficiency and waste reduction.

This shift in focus is generally in response to initial threats from the competition that by then begin to make their presence felt. Usually the situation still allows high profit margins with limited variety of products, with the main concern being the large amounts of most obvious or visible wastes. Because the firm is still selling everything it makes, the firm is still in the position to set the price, and therefore there is still a focus on maximizing production output. Large amounts of inventory are needed to hedge against the many variations and instabilities in and out of the system. While detailed inventory deployment is not the main concern, inventory control has become a major problem carried from the pioneering stage. Hence, the goal of the systemization stage is to bring order to the system and avoid shortages or backlog situations. The manufacturing practices that would change with respect to the previous stage include:

- System view of efficiency and waste elimination.
- Definition of standard operating procedures to avoid confusion in operation and additional system complexity.

### **2.2.3 Third Stage: Technology and Process**

This stage is entered once the firm starts to make what the sales force can sell instead of selling everything that can be made. Increased competition threatens the market share of the firm and a potential for over-capacity exists. Hence, competition shifts focus to product quality and features, aimed at product differentiation. But this is a stage of the so-called “friendly competition” where players aim at larger market shares but nobody’s survival seems to be in danger.

Production technology by now offers better productivity and continues to be seen as the main source of competitive advantage. People, however, start to be seen as a constraining resource. Workers become more specialized but at the same time less flexible as change becomes harder to implement. Standardization and improvement efforts continue. The environment is changing from one with a few, inefficient competitors to that with a number of established players competing with similar processes, lead times and costs. The cost of holding inventory for a firm does not seem, however, to be significant enough to be the deciding factor between business success and failure, in part because profit margins are not sufficiently low yet. Then, no attention is paid to detailed inventory control. Efforts at cost reduction and efficiency improvement at this stage tend to yield limited results as more features are added to the product to make it more attractive to the customer. The increase in product varieties moves the situation from a low mix and high volume towards a low or high volume and high mix situation, resulting in higher complexity and cost. The manufacturing practices of interest at this stage are:

- Development of flexible processes (handle a variety of products with limited resources)
- Reduction of chaos and waste
- Ensure material availability

#### **2.2.4 Fourth Stage: Internal Efficiency**

Once the volumes are not high enough for dedicated resources, the job shop structure emerges naturally with batches, setups and coordination complexity. Increased allocations of inventory are required to avoid production coordination problems, resulting in a loss of leanness. The competition may look less friendly as new competitors enter the market and redefine or challenge the customer perception of product quality. These new competitors usually have a radically different structure, organization, or business approach.

New competitors are quick to identify those few product varieties that account for most of the revenue (the 80/20 rule). Then they address such markets by setting up relatively lean and efficient processes in a continuous process structure as opposed to the established competitors

that have by now evolved into a departmental process structure or even job-shop structures. New competitors focus on developing dedicated lines with a system-wide view of productivity and quality, sometimes even at the expense of limited resource utilization.

Established firms respond by grouping product variations into families or platforms. The goal is to limit the flexibility requirements by grouping similar requirements into product families of sufficiently high volume to justify dedicated lines with continuous flow within the plant. Firms start to look at outsource rationalization, hence abandoning their traditional vertically integrated structure. Suppliers are asked to participate in concurrent engineering, and expected to improve the design of the product or reduce the target cost. This approach is sometimes taken to the limit where the firm pushes its problems outside of its plants and into the supplier's or the rest of the supply chain. It is at this stage when some of the concepts of just-in-time delivery begin to be requested or enforced on the suppliers. The firm is trying everything to reduce costs, but often it just passes the demands directly to the suppliers. The situation inside the plant most likely still displays large amounts of inventory to allow smooth production and support the large number of product varieties. Eventually the firm recognizes that more attention needs to be paid to internal operations.

### **2.2.5 Fifth Stage: Customer Service**

At this stage firms and industries find themselves in a state of excess production capacity, declining prices and profit margins, increased competition, shorter product or model life cycles, and uncertainty and instability in customer demand. Hence, the customer and customer related issues are the focus of the competition in this stage of manufacturing. All competitors have access to the same production technologies and materials and are subject to the same time constraints.

Just-in-time techniques of material delivery are adapted extensively. Manufacturing practice focuses on identification of core competencies, outsourcing, responsiveness, lead-time reduction and hence, there is a return to continuous flow process structures. Within the four walls



of the plant, emphasis is on dock-to-dock measures. Smaller, specialized or focused, more manageable plants are preferred. Inventory requirements are minimized with detailed just-in-time inventory controls. Continuous improvement efforts in every process intend to eliminate every source of waste. It is just a matter of time before all competitors get to very similar levels of competence and efficiency, pretty much everybody is doing the same. It is then when the difference between failure and survival can be defined by the extent to which a company succeeds to focus its resources and remove all and any source of waste from the truly value-adding operations, covering processing, storage and transportation operations.

The work force is included into the efficiency movement. At this stage, it is desired to change the worker culture into a more positive, pro-active force. Workers are engaged into continuous improvement efforts and problem solving initiatives. Worker empowerment and teamwork are encouraged. Training is intended for general skills, multi-functional labor is sought. The worker is expected to develop standard operating procedures and document them. The firm seeks less dependence on specific skills.

### **2.2.6 Sixth Stage: System-Level Reengineering**

By now, firms have addressed all relevant issues of internal processes. Main competitors have very similar fixed and variable costs, service levels and lead times. While efforts are still intended to promote additional savings from continuous improvement programs and reduce costs and flow times, the return from these efforts is relatively insignificant, or seems not enough to address the mounting pressure. Then indirect and overhead costs are challenged next. The organization is re-engineered or re-structured at the plant and corporate levels. Divisions are rationalized, workforces dramatically reduced.

The links to the customer are strengthened and the entire system operation is under pressure to meet well-know industry standards and customer expectations for delivery, cost, quality and variety. The manufacturing practice associated with this stage is mass customization, but process technology limitations prevent its feasibility beyond Assembly-to-Order approaches

in most industries. Strong collaboration throughout the supply chain comes into being. Previously held policies and strategies about outsourcing are abandoned with strategic alliances established with suppliers. Work is subcontracted to specialist firms that are asked to design, develop, manufacture, deliver and even assemble within the firm's plants entire subassemblies or modules.

Some short-term actions, however, could be taken at the expense of future viability and business health with the intention of providing immediate cash flows or cost reductions. Mergers and acquisitions take place along the supply chain as firms attempt to influence where profits are taken. Barriers to entry seem to have become very low and everything is perceived as happening fast with shrinking windows of opportunity as competition in a global scope increases [Fine 98, Goldman et al 95].

It is considered that most manufacturing firms follow at least some of the six eras of manufacturing control presented above as they evolve within an industry. Specific issues drive the evolution of the industry, the most obvious one being a never-ending quest for efficiency in cost and time. Because of the high capital investment required in manufacturing, once an operational base has been established, it usually becomes very difficult to implement changes in search of efficiency improvements. The effort or cost of any change has to be justified by the expected benefits. Hence, until there exist sufficient threat, the firm will not address problems, waste or inefficiencies in the system. Unless it is proven to be necessary in order to meet the objectives of the firm, no effort, time and resources are spent in efficiency improvements.

When a firm approaches the sixth era, however, there are unavoidable risks. Those risks are associated to the speed of change that becomes necessary to stay in business. Radical changes in power, distribution, market, or product and process technology that have traditionally followed the sixth era have allowed just a few of firms to reposition themselves as pioneers in the first stage of the next cycle of evolution. Those changes were triggered by events such as the introduction of interchangeable parts, steam and electrical power, railroad networks for distribution, automobiles and related transportation, the 1921 recession, the 1929 depression, world war two, global market opportunities after world war two, electronics, plastics, synthetics

	I. Pioneering	II. Systemization	III. Technology and Process	IV. International Efficiency	V. Customer Service	VI. System Level Reengineering
<b>Level of competition</b>	Monopolistic (minimal competition).	Not significant competition (few non-established competitors).	Friendly competition for market share (established competitors).	Local Threat (competition for local survival).	Global Threat (competition for survival)	Global Threat (merges and acquisitions)
<b>Barriers to entry</b>	High barriers to entry (limited access to technology and skilled labor). Vertical	Economies of scale of major competitors.	Market saturation.	Declining prices.	Competitors with similar cost base.	Low barriers to entry (dissemination of methods technologies and processes throughout the industry)
<b>Integration</b>	Vertical			Shift to horizontal		Redirection back to vertical.
<b>Constraint</b>	Capacity and materials supply.	Delivery reliability (complex, chaotic output).	Labor flexibility.	Process technology.	Product and process technology.	Speed of change.
<b>Business situation and focus</b>	Demand greater than supply. Price set by firm. Focus on growth, expansion.	Demand greater than supply. Focus on maximizing production output.	Over capacity (supply greater than demand). Focus on product differentiation (quality, variety).	Price set by market. Pressure on reduced prices. Focus on cost reduction.	Shorter product life cycles. Uncertainty and instability on product demand. Renewed focus on product differentiation, quality, and customer service.	Shrinking life cycles and windows of opportunity. Focus on quick incorporation/change to incoming technologies. Focus on indirect/overhead costs.
<b>Manufacturing practice and focus</b>	Maximization of resource utilization at bottlenecks. Design of product for high volume manufacturing. Product simplification, single variant or very limited variations offered. Large allocations of inventory. Low efficiency and general resource utilization, long lead times.	System view of efficiency and waste elimination. Definition of standard operating procedures to avoid confusion, reduce complexity. Focus on delivery, reduction of backlog, waste reduction. Substantial savings resulting from cost reduction efforts.	Larger product variety, reduced volumes resulting in job-shop structures or departmental layouts. Interest on flexible manufacturing systems. Focus on material availability.	Competitive advantage based on improvement of internal efficiency. Product simplification by definition of product families and/or platforms. Detailed inventory control. Concurrent product/process design. Shift to modular product architecture.	Focus on core competencies, outsourcing, and responsiveness. Emphasis on dock-to-dock measures. Continuous flow process structures. Just in time inventory controls, continuous improvement. Focus on overall efficiency and workforce culture. Multi-skill flexible labor.	Goal is mass customization. Major rethinking of the role of supply chain, workforce, organizational structure, manufacturing practices. Strategic alliances, merges and acquisitions. A major change in product or process technology or environment sends industry back to stage 1.

**Table 2.1** Characteristics of Stages of Manufacturing Practice Evolution. Based on [Mckay, 2001].

and composites. These events and inventions had far-reaching impact on competition, technology, efficiency, cost and so on. Firms that are quicker to adapt to or to incorporate change to improve constantly will be the ones to succeed. Table 2.1 shows in a condensed manner the definition of the different stages of evolution of manufacturing practice. The definition is in terms of the associated type or level of competition, barriers to entry, type of integration, main constraint for change, business situation and focus and corresponding manufacturing practice and focus.

Being the main driver of evolution, the quest for efficiency in manufacturing systems has in fact been a trade-off in system design between responsiveness and wasted or reserve resources. The ideal or objective for design is a manufacturing system that is responsive and at the same time free of waste. It has also been observed, however, that absolute efficiency and leanness inherently lead to fragile, rigid systems. Any system that has been left with not much more than the absolute minimum resources required for operation, will not have reserve resources and therefore will operate near full capacity whenever possible. In normal operating conditions such system will be seen as very lean and efficient. Any disruption, however, will result in a crisis.

Flow-line manufacturing systems, because of their leanness and reduced throughput times, are preferred to job-shop structures when efficiency becomes relevant. Job-shop structures inherently have longer throughput times, larger inventory requirements and lower overall efficiencies and utilizations. But even for flow-line manufacturing systems, physical limitations prevent the removal of absolutely all reserve resources, resulting in the efficiency-reliability trade-off mentioned above. Unless flexibility or disturbance rejection or variation compensation is built into the production system, chaotic behavior will result.

Thus, an integrated approach to manufacturing system design and control is required to address the performance requirements of manufacturing systems. The end-customer expectations result in system requirements that cover multiple levels of the organization and span many time frames. Because of the size and complexity involved in designing a manufacturing system, the practice has been for a company to specify a standard method of manufacturing system design.

While those standard procedures have facilitated the design activity, it has become clear that they tend to ignore the context of evolution described above. Some of those approaches may be adaptations of competitor's approaches, or they could be prescriptive tools without a sound understanding of the issues involved. The next section presents the Manufacturing System Design Decomposition as an integrated approach to the design and control of manufacturing systems.

## **2.3 The Manufacturing System Design Decomposition**

### **2.3.1 Overview**

The development of the Manufacturing Design Decomposition (MSDD) was motivated by a confusion observed in industry during the initial concept design phases of manufacturing system design. The relevance and consequence of decisions taken early in the design process is repeatedly misunderstood. The design of manufacturing systems seems to be characterized by a lack of context understanding and formal development process. System objectives are often not stated. Detailed process designs are not related or linked to system-wide objectives. Local optimizations are common. The urge to implement whatever manufacturing practice is in fashion at the moment leads to the implementation of off-the-shelf solutions, without an understanding of the business or industry situation they best suit. Consequently, these implementations repeatedly fail to achieve the expected results. Linck [2001] enumerates several key points related to the practice of manufacturing system design in industry:

- Few approaches cover the design activity from conceptual design to preliminary and detailed design, implementation and operation.
- Each approach by itself provides valuable support for manufacturing system design, but it is difficult or impossible to integrate with other approaches.
- Lean manufacturing in practice is mainly focused on operational issues (not strategic), without formally stating system requirements.
- There is no clear distinction between system requirements and design solutions.

- The practice of manufacturing system design usually does not follow a formal process of evolution and knowledge management. When there is a process for system design in manufacturing, very likely it is a rigid, tool prescription method.

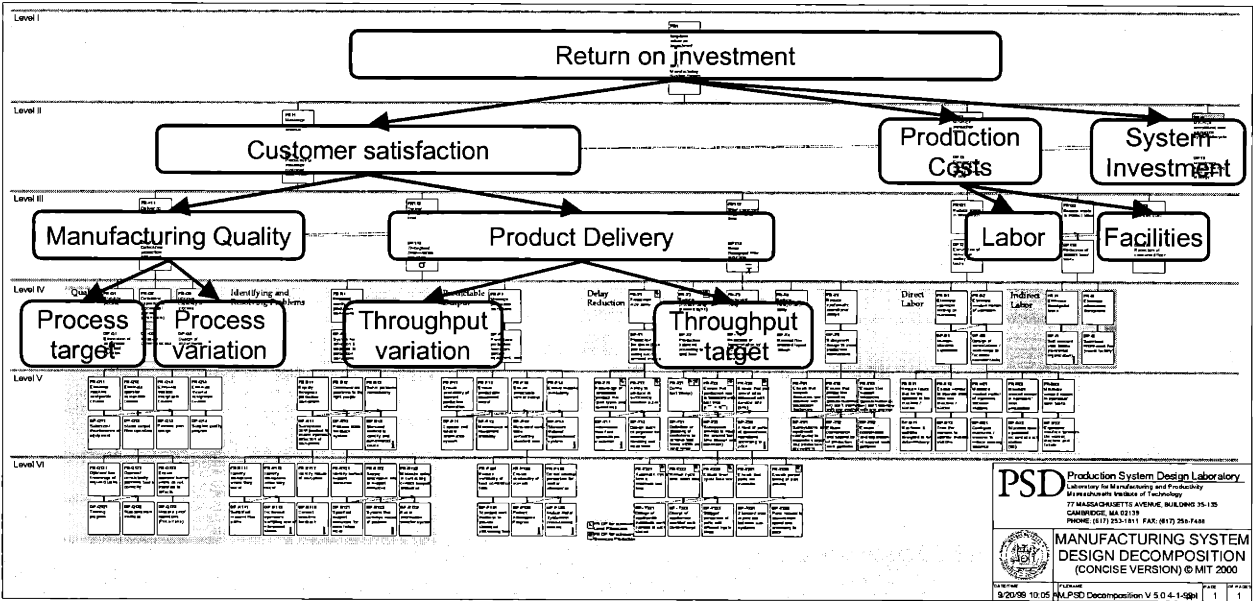
Hence, confusion is manifested by the difficulty of a design team to state clear high-level, system-wide objectives and translate them into detail process specifications. The MSDD is intended to assist designers in understanding the relationships among design objectives and means or design resources. By clearly distinguishing objectives from means, the MSDD is expected to ease the specification of manufacturing systems that best achieve the requirements on manufacturing performance resulting from general business objectives and goals.

The MSDD is envisioned as a comprehensive manufacturing system design methodology that integrates the design of a manufacturing system with its control and operation. It is intended to facilitate the integration of different tools and approaches of system design and manufacturing practice. By promoting a structured step-by-step process it is intended to assure that detail design specifications are related or aligned to high-level system requirements, and more generally, to distinguish between goals and means of manufacturing system design. Thus, the MSDD allows an understanding of operational issues for control in the broad context of manufacturing system design and control.

### **2.3.2 High-Level Description of the MSDD**

The MSDD is a structure or hierarchy representing how manufacturing system requirements and resources relate and affect the entire system design. It is intended to guide the development and design activities by following a top-down methodology, and hence provide a system-wide perspective. By stating first the organization or firm business objectives, it intends to guide the design of the system components. This approach is contrary to bottom-up hierarchies that seem to be better suited for improvement efforts than for system design. Blanchard et al [1998] considers that bottom-up methodologies are based on known elements, whose physical presence is assured, and concludes that bottom up methodologies can not warrantee that high-level system

requirements are met by implementation of specific elements. Hence, detailed frameworks that specify a set of elements of a production system do not necessarily result in an effective manufacturing system, as defined by the highest-level requirement. In other words, the implementation of a specific tool, say Kanban inventory control, is not enough by itself to improve the performance of a system to a level that compares to the best performer that is using Kanban control.



**Figure 2.2** High level abstract view of manufacturing system business objectives  
(PSD Lab, 2000)

The MSDD then, represents objectives and means of a general conceptual-level manufacturing system design. To limit complexity and keep generality, as well as to effectively represent the logic of the relations among the system resources, the MSDD has to use some level of abstraction. The level of detail is limited to six levels, upper levels related to strategic issues, and more operational issues being covered by lower levels. Each Functional Requirement (FR) is coupled to a corresponding Design Parameter (DP) and grouped with other FR-DP pairs according to its subject area. Figure 2.1 shows how the abstraction of the MSDD allows the

representation of the logic of the decomposition by relating lower level areas to the high level objectives they support.

The decomposition begins at the highest level with a single functional requirement designated as the desired return on investment. Return on investment is discussed in the next section and decomposes into branches related to customer satisfaction, production costs and system investment. Customer satisfaction requirements are decomposed into manufacturing quality requirements and product delivery requirements. Production costs are divided into labor cost and facilities costs. These first three levels of the decomposition allow a classification of more detailed key requirements, such as high quality processes, rapid problem resolution, predictable output, throughput time reduction and effective use of labor resources. This structural view serves as a frame to guide the classification of more detailed requirements in each branch of the decomposition. Hence, the MSDD effectively communicates a manufacturing system design process by describing in a top-down manner how every FR is satisfied by its corresponding DP and how the different FR-DP couples relate among themselves.

### **2.3.1.1 Financial Objectives**

The highest level FR in the MSDD corresponds to the maximization of return on investment [Suh et al., 1998]. Return on investment (ROI) in common financial terms is the difference of revenue minus cost divided by investment. The MSDD gives the view that the manufacturing system meets the ROI requirement by decomposing it into customer satisfaction, production cost and system investment requirements.

Care must be taken to avoid evaluating a manufacturing system by ROI alone. ROI performance measures have been criticized for encouraging less participation from upper management by reducing the selection or evaluation process to a single parameter consideration, ignoring many other aspects of a manufacturing system [Hayes 1980, 1982]. In fact, it is proposed that any ROI evaluation has to be placed within a context or a time frame for evaluation, corresponding to the policies and strategy of the enterprise at the moment of



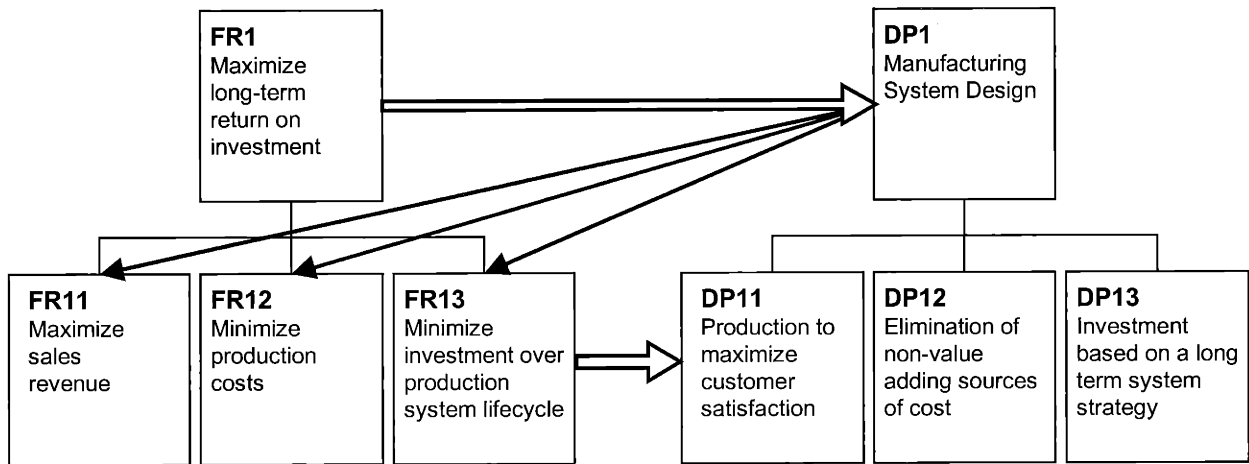
consideration. A long-term ROI will affect the design of the manufacturing system in a very different way than a short-term ROI. Future work will link this high-level of the MSDD to an enterprise design decomposition that will ease the specification of a time frame for ROI evaluation.

ROI evaluation problems are aggravated by the difficulty to measure the impact on ROI of investments resulting in benefits such as manufacturing quality, technology innovation, workplace ergonomics and safety, and so on. Therefore, a single number evaluation, without observation of the related context, will most likely be misleading. A specific example of the risks associated with ROI measures can be seen in outsourcing decisions. A company may decide to improve its short-term ROI by outsourcing processes or entire components to the lowest cost supplier. A lack of vision for knowledge management and competitive advantage development results in a justification for outsourcing based on short-term cost considerations only. The danger of this approach is that companies eventually drive themselves out of business once they have no competitive advantage to survive. Another consequence of outsourcing, commonly ignored by ROI evaluations, is the higher complexity and related operational costs for coordination of a manufacturing system that has many outsource operations as compared to the simpler coordination of in-house operations.

Once the risks of focusing on ROI without a related time frame or context have been understood, ROI is a valid and very useful starting objective for manufacturing system design. Hill [1994] discusses the need for a strategic view of investments and advocates investment evaluation based on how it supports or contributes to the success of a company's business strategy. Kaplan [1984] proposes incorporation of the evaluation of intangible assets into standard financial accounting statements. The MSDD agrees with these views by addressing issues of manufacturing and business strategy not typically considered in usual ROI calculations. For example, issues related to product quality and variety as well as product delivery speed and responsiveness are considered even if they do not have a cost directly associated.

### 2.3.1.2 ROI Decomposition

The decomposition of ROI is shown in Figure 2.2, where Manufacturing System Design is stated as the DP that is expected to meet the requirement related to the highest FR: maximization of the long-term ROI. To avoid the short time horizon bias of ROI measures described above, the functional requirement emphasizes the need for a long-term vision when designing and investing in manufacturing systems. The Manufacturing System Design is decomposed into three functional requirements: maximization of sales revenue, minimization of production costs and minimization of investment over production system lifecycle.



**Figure 2.2** High-Level Decomposition of the Return On Investment FR.

Maximization of sales revenue is affected by sales volume and unit price. The approach to the maximization of these parameters is affected by the stage of evolution of the industry. It is very rare that a company would ever find itself in the position to control both volume and unit cost of sales. More commonly, firms operate under competition conditions resulting in limited market share and fixed unit price set by the market itself. Only through a consistently better manufacturing performance a firm under such conditions can expect to improve its market position in the long-term. A good manufacturing performance has to include product quality, cost and delivery. If the design of the product addresses the required functionality, the production

system has the challenge to meet customer expectations related to quality compliance, cost and delivery. The manufacturing system meets customer expectations by producing the product at the required level of quality and within a required period of time to ensure on-time delivery. Therefore it can be seen that the manufacturing system collaborates in meeting customer satisfaction through unit cost, quality compliance and delivery time.

The second FR of this level is minimization of production costs. System design decisions taken early in the design process affect strongly the final production costs. The MSDD, with its top-down approach and its emphasis on system level objectives, addresses this fact of manufacturing system design and control. On the other hand, the MSDD also focuses on operational costs. Costs are minimized by elimination of all non-value adding operations that are not absolutely necessary. Conceptually, non-value adding operations are understood as any activity that does not increase the value of the product as perceived by the customer.

The third component affecting ROI is investment in manufacturing. A long-term view of investment minimization is in fact a main driver of efficiency and flexibility in system evolution. The objective stated by the FR is to minimize the required investment over the manufacturing system lifecycle. The corresponding DP states that investments should be based on a long system strategy. Hence, the focus on investment needs to be the procurement of manufacturing resources that allow the delivery of finished products that consistently meet customer expectations. This focus means that the manufacturing system must meet operational requirements as the product moves through its lifecycle. In other words, the manufacturing system must be flexible to adapt to changes in business situations resulting in changes of product characteristics, product and process technologies and market size. Furthermore, in agreement with the ROI objective, the functional requirement also stresses the importance of considering long-time horizons for investments. Hence, manufacturing investments should correspond with the manufacturing strategy of the company. For example, manufacturing investments should consider short-term or operational flexibility requirements imposed on the system in terms of product volume and variety, as well as longer-term flexibility requirements resulting from shrinking product lifecycles.

### 2.3.1.3 Customer Satisfaction

For a manufacturing system to maximize customer satisfaction, and hence sales revenue, it must meet operational requirements related to quality compliance and delivery. The level of quality compliance of the finished product and the timelines of product delivery determine whether any customer finds the expectations satisfied better than by the competitor's output. Three requirements are derived from the customer satisfaction branch that as a whole define the effect of the manufacturing system design on sales revenue. These three requirements are the top-level functional requirements of branches of the decomposition that are related to manufacturing quality, mean throughput time and throughput time variation.

A fundamental requirement for customer satisfaction is quality compliance of the product. This requirement refers to how well a manufactured product meets its design specifications, indicating the level of manufacturing quality. On the other hand, the capability of a product to meet customer expectations, or in other words the overall quality of the product, is a result of many activities of manufacturing such as product design, production, distribution, marketing and after-sales services. At this stage of the MSDD development, the decomposition is focused on the production part of the manufacturing system. Therefore, the customer satisfaction branch of the MSDD focuses on what needs to be done in production to assure customer satisfaction. This focus does not disregard concurrent engineering efforts, it just acknowledges that production is only part of the entire manufacturing system and therefore, it is not supposed to compensate for poor performance in other areas of manufacturing.

Therefore, the emphasis of the customer satisfaction branch of the MSDD is to ensure that the output of every manufacturing process yields products with characteristics that comply with the design specifications in terms of dimensions, materials and functions (design specifications). The quality branch of the decomposition begins with the requirement that the manufacturing system entails processes that deliver products on the design target instead of just within tolerance. Manufacturing processes whose output show a minimal variation from the target can meet this requirement if quality improvements intended to move the output mean are clearly distinguished from those intended to reduce the output variance.

The second and third requirements of the customer satisfaction branch, describe the time dimension of the requirements for customer satisfaction, or final product delivery. As a manufacturing enterprise follows the evolution of manufacturing practice described above, the relevance of time for customer satisfaction, and ultimately for survival of the firm, grows. Throughput time is a broadly used evaluation parameter of time performance in manufacturing systems. In a make to order manufacturing system, throughput time is understood as the time elapsed from the moment a customer places an order to the moment the corresponding product is delivered. Similarly, throughput time in a make to stock manufacturing system is understood as the time required to replenish a product that has been removed from the stock of finished goods. The MSDD focuses on the production activities that take up part of the throughput time. The sum of the times related to all required production activities, both value-adding and non value-adding, is often called manufacturing throughput time, or manufacturing cycle time.

Manufacturing throughput time in any real system has many sources of variations. The complexity of manufacturing systems, namely the many variables affecting manufacturing throughput time in the form of variations and disturbances in every input to the system, is in fact the motivation for improvement and even research effort on manufacturing control. The variations inherently related to manufacturing throughput time are in fact the reason for the need for control in manufacturing. Manufacturing throughput time is usually characterized, for representation purposes, by a probability density function with a specific mean, variance and shape.

The mean and the variance of the manufacturing throughput time characterize the time performance of a manufacturing system. The mean of the throughput time indicates how much time is spent in average to produce a complete product. This system parameter, in addition to other parameters such as production volume, and product size, variety, unit cost, and shelf life, affect system design decisions such inventory strategy, control policy, and distribution and sales. At the end, the customer wants to have the desired product available, when he wants to buy it, where he wants to buy it. Therefore, the design of manufacturing system has to be done to ensure availability of the end product. Such availability is easier to accomplish with short manufacturing

throughput times and low variations in throughput time. Therefore, reducing the variation in manufacturing throughput time, allows reduced quoted lead-times and improved on-time delivery. Methods for reducing the variation in throughput time are decomposed into two functional branches: problem identification and resolution, and predictable output of manufacturing resources.

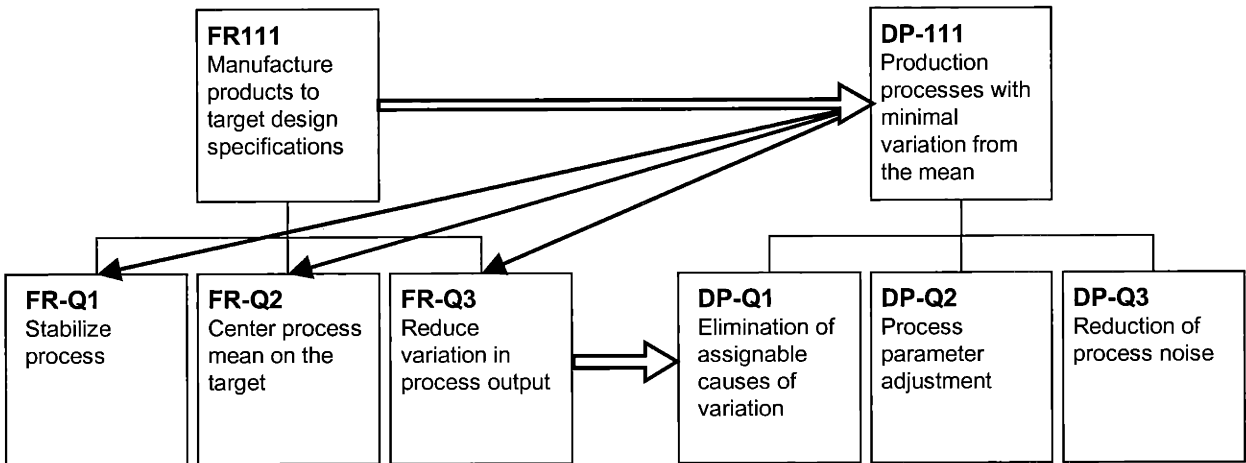
The FR related to meeting the customer expectations on lead-time emphasizes the need for the manufacturing system of being able to meet reliably and confidently quoted lead times. Customer satisfaction in terms of lead-time is maximized when the quoted time equals customer's expectations. The customer expected lead-time should be considered as the upper bound for manufacturing throughput time during the design process. In other words, the manufacturing system should be designed to ensure that its throughput time is less than or equal to what the customer is willing to wait for his product. The design parameter related to the mean throughput time reduction is accordingly decomposed into the different methods leading to reductions of manufacturing throughput time.

It has become clear that a manufacturing system must have a consistent time behavior in order to meet its customer expectations. From the above discussion, the influence of quality compliance, mean throughput time and variation of throughput time on system time behavior for customer satisfaction is clear. A system that meets the requirements of customer satisfaction will have clear manifestation of the dependencies of high quality manufacturing processes and throughput time variation. If the relations among quality and time output are not clearly understood and addressed during the design and operation of the manufacturing system, it will become very unlikely that the system will meet customer expectations. For example, poor process quality will result in lower yields that in turn result in increased variability in system throughput times. If compensation or corrective resources are not built into the system during the design process, quality and time variations develop stronger dependencies resulting in higher difficulty and complexity during the operation of the system. Quality problems become more difficult to resolve when throughput times are lengthy or when problem detection involves long delays because corrective actions have corresponding longer delays, hence the relevance of quick problem detection and correction becomes obvious.

#### **2.3.1.4 Quality**

Manufacturing quality is nowadays expected or assumed since customers no longer tolerate repairing or returning defective products. Products are expected to function as specified from their first use to the end of the life cycle. Total quality management methods [Shiba, 1993] emphasize quality as an expectation of “fitness to use”, whereas more traditional conformance measures of quality relate to the expectation of meeting minimum tolerance specifications from design, referred to as “fitness to standard”. For the purpose of decomposition of the customer satisfaction requirement in the MSDD, quality is understood as the requirement that manufacturing processes yield products that are on the design target, with a minimal variation from such target.

This requirement for quality agrees with the quality loss function proposed by Taguchi [1989] that considers a “loss to society” that is minimized when the performance of the product or the process is at the nominal or target quantity specified by design. The DP: Production processes with minimal variation from the mean is decomposed in terms of the steps that are required to attain high quality output from manufacturing processes: statistical process control and process design and improvement, as shown in Figure 2.3. Then, the focus of manufacturing quality is to ensure that every process output is centered on the target from design and shows minimal variation, as opposed to the traditional approach of just having process outputs within tolerance specifications.



**Figure 2.3** High-level decomposition of quality.

Process stability is identified as a requirement for minimal variation from target. Montgomery [1985] defines a process in control as a process whose variations are not related to assignable causes. Assignable causes are non-random events that can affect the output of a manufacturing process. Examples of non-random events in production are tool wear, and improper adjustments and settings. Hence, in order to achieve process stability it is required to eliminate all sources of assignable causes of variation. This design parameter corresponds to the first major step in statistical process control techniques used to quickly detect the occurrence of assignable causes at the machine. Quick detection of process output variations allows implementation of corrective actions before many or even any non-conforming units are delivered to the next process. A process may be stable and yet the process variation is such that a large amount of parts have variations beyond the specification limits. On the other hand, a process could be unstable but still be producing an acceptable number of parts within the specification limits. Two functional requirements from the decomposition provide the necessary conditions for high quality outputs from stable processes, as described below.



A process can be stable and have a sufficiently small standard deviation and still be producing a large number of rejected or out of tolerance parts. This could happen if the process has a distribution with a mean that is close to a specification limit and therefore many parts fall out of tolerance. Then, many defective parts are produced even if the standard deviation is low. In the decomposition, a functional requirement is specified with regard to the location of the process mean. The corresponding design parameter indicates that to correctly place the process mean at the required design target, adjustment of process requirements is involved. A measure of process variation and mean centering is the process capability ratio  $C_{PK}$  as defined in equation 2.1. The dimensionless ratio  $C_{PK}$  is defined by the distance from the process mean to the closest specification limit divided by the process standard deviation. The closest distance from the process mean to the specification limits is used because that distance corresponds to the most damaging case of process mean shift. The ratio  $C_{PK}$  is broadly used in practice, and as seen in equation 2.1, the process parameter adjustment intended to achieve process centering involves the control of the mean and the standard deviation through operational process adjustment, given that the upper and the lower specification limits are dictated by product design.

$$C_{PK} = \frac{1}{3} \min\left(\frac{\mu - LSL}{\sigma}, \frac{USL - \mu}{\sigma}\right) \quad (2.1)$$

The third high-level requirement of the quality branch is to reduce variation in process output. Variation that is seen in the output of any process, including stable processes, is a consequence of uncontrollable noise factors of such process. Phadke [1989] defines noise factors as variations that cannot be controlled by the designer or the operator and that lead to quality loose. Process noise is then another source of variations in process output. To reduce such variation in process output, it is required to limit noise variations in all process inputs as well as the process sensitivity to those input noise variations, as indicated in the decomposition.

The MSDD indicates that the first step to reduce process variations is to study the process stability. This study should involve, whenever possible, the conversion of previously assumed uncontrollable causes to assignable causes. Statistical process control methods do not indicate how the conversion of uncontrollable to assignable causes can be done, but they acknowledge

that such conversion is necessary to improve process capability. The next step is to develop robust designs that show minimum sensitivity to input noise, and hence ensure an output mean on target. This step requires a deep understanding of sources of noise in the production environment as well as the effect they have on process output, it is to say how noise or environment variations result in process output variations.

### **2.3.1.5 Delivery**

Customer satisfaction requires, in addition to the capability of the end product to meet the functional expectations, the opportune delivery of such product to the customer. Customer satisfaction, as related to delivery, is maximized when the delivery time meets the customer expectation. Delivery time represents the amount of time from the moment the customer places an order to the moment such customer gets the requested product. The customer expectation on delivery time is difficult to determine reliably, but it is strongly influenced by the business environment at the moment. In other words, if the competition is able to deliver similar products in short time, customers naturally will expect that same amount of time from every other company. The time taken to transform raw materials into finished products is called manufacturing throughput time, and includes the time required for processing, inspection, storage and transportation. Manufacturing throughput time is influenced by the way the manufacturing system has been conceived, and more specifically, by the process structure of the manufacturing system involved. For example, it has been explained already that job shop process structures have inherent longer throughput times than flow line process structures. Throughput time is, therefore, affected by the manufacturing system design. Manufacturing throughput time, on the other hand, is only a component of delivery time. Order processing and distribution are also important components of delivery time.

Customer lead-time represents an upper bound on the maximum time the customer is willing to wait for delivery of the product. Therefore, the manufacturing system should be designed so that the sum of all components of delivery time, including manufacturing throughput time, is less than the customer lead-time. As mentioned above, real manufacturing systems have many sources of variations that make of throughput time a random variable. Such randomness

means that real systems deliver finished products at times that for practical reasons are conveniently represented by a mean and a variance parameters. Such parameters and the specification of a customer lead-time can be used to predict the likelihood of the system delivery being able to meet the customer expectations. Concepts such as service level and fill rate are used extensively in practice as measures of delivery performance of manufacturing systems. Both parameters combine the mean and variance of manufacturing throughput time with customer lead-time to assess the performance in time of manufacturing systems.

Service level, according to Hopp and Spearman [1996] is defined as the probability that the throughput time of a make-to-order manufacturing system is less than the customer lead-time. Fill rate on the other hand, is the probability of any demand being satisfied with material readily available from stock in a make-to-stock manufacturing system. The MSDD identifies two high level requirements related to delivery: reduction of mean throughput time and reduction of throughput time variance. Reduced manufacturing throughput time mean and variance results in a better time performance for a given customer lead-time specification. It has been observed in practice, however, that performance improvement efforts tend to be focused on reduction of mean throughput time, overlooking reduction of throughput time variation. It is commonly forgotten or ignored that process time variations are the single most important justification for inventory allocation. While it is generally well understood how inventory allocation affects mean throughput time it is more difficult to relate inventory allocation to throughput time variation. Fill rate and service level analyses, by computing starvation probabilities, show that inventory allocation directly affects throughput time variations. This is the reason why it is proposed that firms should examine the variability of their manufacturing processes first, then understand how process variability translates into throughput time variability and only then approach the reduction of throughput time mean and variance.

#### A. Throughput Time Variation Reduction

Reduction of variation in throughput time is required for on-time delivery of products. Manufacturing throughput time variation is fundamentally a consequence of multiple process variations in the system and the way compensatory resources work to resolve those variations. In the MSDD, a disruption is understood as any problem that leads to system output variability. The

MSDD classifies two types of disruptions: those related to quality and those related to delivery. For customer satisfaction, it is clear that quality and delivery are fundamental, as indicated by the customer satisfaction branch of the MSDD. In fact, quality and delivery are coupled in the sense that a good product quality is a pre-requisite for good delivery. Such coupling is indicated in the MSDD by the relative location of the branches, where an FR-DP couple is located to the right of another couple that influences it. Then the MSDD approaches the design of manufacturing systems beginning with the issues expressed in the quality branch, which is located to the left of all other issues of manufacturing system design.

Decreased variation in throughput time has in fact many consequences on the relative position of a manufacturing firm. A more predictable throughput time allows a company to have better fill rates or service levels, more customers find that the requested products or services are available when needed or when promised and hence the customer satisfaction is better. The customer eventually confirms that delivery is consistent and begins to gain confidence in the ability of the manufacturer to deliver on time every time.

The capability of the manufacturing firm to deliver with predictable times reflects the ability of the system to constrain or delimit process variations and their effects as well as to implement measures to deal with unavoidable sources of process noise. Delivering in a consistent and timely manner requires a deep understanding of how production resources work as well as how those resources should be coordinated so that the overall system output meets the demand requirements in a reliable and predictable manner. It has been observed that in practice or during operation, the parameters that affect the time behavior of a manufacturing system are usually many more than what is considered of relevance. Some of those parameters could be classified as inputs to the system, such as incoming material and related resources availability, quality, and variation of supply; or as control commands to the system, such as production authorizations and operational parameters for rate and product mix. Some other parameters are better classified as internal process parameters such as frequency of machine breakdowns, repair requirements, system availability, flexibility in production rate or product variety, sensitivity to machine parameters and quality yield.

The requirement on rapid response to production disruptions indicated in the MSDD means that there must exist an established procedure for quick detection and resolution of problems. Rapid problem response and resolution involves identification of the nature, location and time of the disruption, determination of and communicating with the required support resources, and hopefully, quick resolution of the problem. During operation of a manufacturing system, many of these steps should actually happen in a natural or automatic way as a consequence of the design and the implementation of the system. Andon boards and visual control systems for example, have been developed as a practical implementation of some of these steps.

Another requirement indicated in the MSDD and related to throughput time variation reduction is about the reduction of production disruptions. Practically any manufacturing resource, such as equipment, people, material or information can contribute to disrupt production. It is important at this point to remember that a reduction of the frequency of occurrence of disruptions by itself does not necessarily improve the delivery of the system. This requirement, as indicated by the MSDD, refers to a continuous improvement activity that seeks to reduce the frequency of disruptions once those disruptions become identified in the operation of a manufacturing system. By reducing the frequency of failures that are handled with a standard procedure, and hence the time it takes to resolve those failures or disruptions is independent of the frequency of occurrence, the actual availability of the system becomes better. On the contrary, if the availability of a system is kept constant while the frequency of failure events is reduced, in fact the system time performance degrades. It is known, that frequent disruptions that are solved quickly is a much better situation to handle than that one with the same availability but corresponding to infrequent disruptions that take long to be solved. This can be understood, at least in part, by its effect on inventory sizing. Smaller amounts of inventory are enough to filter frequent, short duration disruptions, but larger amounts of inventory are required to compensate for infrequent, long disruptions, even if these two cases result in the same overall system availability. The four classifications of manufacturing resources (equipment, people, material, and information) are used to define classifications in lower levels of the decomposition related to availability.

From the above discussion on the negative effect of disruption length on system behavior it follows that a rapid response to production problems is fundamental to achieve a predictable production output. Ideally, production disruptions should be related to sources of process noise or to unpredictable machine failures only. If any disruption that happens during normal operation has an assignable cause, in other words, the disruption is related to a system parameter that can be controlled, then any time such disruption happens its root cause must be identified and a solution implemented so that such failure never happens again. Root cause identification techniques are effective means to improve the reliability of machines whenever they seem to behave below the expected. Total preventive maintenance programs are also very effective to improve system availability by reducing the probability of occurrence of failures during operation.

Manual operations are usually related to large variations. Manual operations result in variations that are not only time dependent but also operator dependent. Ensuring consistent output of manual operations, or predictable behavior from manual processes, is equally important as ensuring reliable output from automatic processes because both contribute to product quality and delivery time. The design of a manufacturing system that involves manual operations should include the development of standard work sequences for each manual task. A standard work sequence is what lets an operator know what should be done when. In fact, the documentation of a standard work sequence is a good tool to improve manual operations. Without a standard sequence, as a first step for performance improvement, it is very difficult to identify motions or tasks that are not needed and therefore can be eliminated. A standard work sequence, in addition to improving the efficiency of manual operations, also helps to reduce variability of those operations as well as to identify overloaded or stressing stations.

## B. Mean Throughput Time Reduction

The MSDD decomposition distinguishes variations in throughput time from components of mean throughput time. Five components of throughput time, or delays that occur during production are commonly identified: lot, process, run, transport and operational delay. The nature of those delays and a practical approach to reduce each of them is presented below.

Lot size delay refers to the time parts have to wait after being processed at a station in order to form a batch that once completed is taken to the next process. The reason for forming batches after processing is generally related to transportation considerations. Transportation batches, also called transfer batches, are sized depending on the transportation requirements and resources. Hence, in order to reduce lot delays, the manufacturing system should require and allow only small lots of parts, because smaller transportation lot sizes result in reduced lot delays. However, physical limitations frequently prevent the implementation of small lot sizes, given the limited capacity of material handling systems and the cost of conveying very small lots to downstream operations. If successive operations can be grouped together in a flow line process structure such as a manufacturing cell, then single piece flow within the cell becomes practical and substantial improvements on throughput time can be attained. With a lot size of one, a part does not have to wait for any more parts to be processed before it can be taken to the next process. Shortly after the process completes the cycle, the part is taken to the next operation.

Process delay is experienced whenever the rate of arrival of parts is higher than the rate of processing. Process delays then, result from unbalanced operations. A queue of parts will form before a stage if the all the operations of the system are not balanced. The average waiting time is used to estimate the amount of process delay, and can be computed according to *Little's law* [Little, 1961], where  $L$  corresponds to average queue length [parts],  $W$  corresponds to the process delay [time] and  $\lambda$  is the processing rate [parts/unit time]:

$$L = \lambda W \tag{2.2}$$

A common practice in manufacturing system design is ensuring that all operations are either balanced or paced by the bottleneck of the system. In practice, however, the stochastic nature of machine failures moves the location of the bottleneck or creates multiple bottlenecks in the system. Machine failures therefore, are a very common root cause of lot delays. Limited machine reliability, by affecting the available processing time, defines transitory slower processes that develop long queues before them, contributing to the total amount of process

delay in the system. The MSDD proposes the design of system configurations that allow the implementation of balanced operations with quick corrective measures that limit the growth of queues in the system. Prompt detection of problems and implementation of corrective measures as well as the definition of restraining mechanisms such as reduced allocations of space for queues of material delimit the amount of process delay in the system.

Run size delay results from the limited flexibility of a process to handle multiple part types, resulting in a justification to process same-type parts in batches of size greater than one. From the point of view of efficiency of the isolated process, it could seem more convenient to define large run sizes in order to limit the number of required changeovers and hence the associated changeover cost. Run size refers to the number of parts of the same type that must be produced before a changeover is done to begin production of a different part type. The smaller the run size, the easier it is for a manufacturing system to satisfy any product mix demanded by the customer. Run size delay occurs whenever the product mix processed does not correspond to the product mix required. If more parts of a given part are produced than needed, the remaining parts will have to wait for the next production cycle to be taken to the next process. The only way to avoid run size delays then, is by producing exactly the product mix demanded. The MSDD proposes the implementation of control policies that determine how customer demand information should be translated into production authorizations in a straightforward way, producing the desired mix and quantity within each production cycle.

The MSDD specifies that production must be based on actual demand. Forecasts should be used for capacity and production planning only. For the system to be able to meet demand, it needs to be driven by actual demand instead of forecasted demand. Variations in actual demand with respect to forecast demand should be handled by or with compensation resources built into the system. The Kanban approach to shop floor control is a proven approach to handle limited amplitude variations in customer demand. The fundamental reason for the implementation of compensation resources in the system is that no forecast can be completely reliable to predict customer demand. Such compensation resources, such as short-term production rate flexibility, allow a system to meet actual demand while keeping inventory requirements low and therefore keeping manufacturing throughput times short.



The demand volume of any specific product usually displays higher variations with time than the aggregate volume of the set of products belonging to the same family or platform. Therefore, it is convenient to have a manufacturing system capable of handling different mix ratios, corresponding to different specific product demand volumes. To produce the correct mix and quantities that are demanded by the customer, the MSDD proposes production in small batch sizes. This requirement in turn calls for a rapid changeover capability and an information system that levels the customer mix over the manufacturing cycle. Additionally, demand requirements travel upstream in the manufacturing system and the supply chain based on actual demand. Because a finite number of production authorization cards is used to transmit the demand information between any two production stages of the manufacturing system, variation amplification and compensation of non-perfect quality yield is achieved as the demand information travels upstream. These requirements are stated in the decomposition for the design of proper shop floor control systems.

Transportation delay relates to the time spent to move parts from one operation to the next. System layouts that locate equipment in close proximity reduce this delay, because short distances facilitate transportation in smaller batch sizes. Transportation delays should be considered when determining paths for material flow within and between processes.

In addition to lot, process, run and transportation delays, systematic operational delays are also commonly present in manufacturing systems. Systematic operational delays are deterministic because they have a known cause that actually presents a frequency of occurrence. Avoiding or reducing the occurrence of systematic delays requires avoidance or minimization of interactions among manufacturing resources that result in interference. For example, the system design must ensure that support resources do not interfere with production resources and that production resources do not interfere among themselves. It has been observed that people and automation resources frequently result in implementations where automation actually interferes with people in the sense that it does not collaborate with the worker but prevents him from improving the operations.

A common flaw of manufacturing system design that results in systematic operational delays is the specification of equipment that prevents continuous production. Routine tasks such as removal of chips or any other by-products and cleaning can be done only if the machine is idle and therefore prevent continuous production. Other disruptions occur when the path of maintenance resources intersects the path of material flow. Some subsystems involve a critical interaction of resources during every cycle. If such interaction is disturbed, for example when an operator path is blocked, immediate operational delays could occur. The design of the system should ensure that material replenishment activities and maintenance activities can be done without disrupting continuous production, for example providing access for maintenance and material replenishment from the rear of the station.

## **2.4 Chapter Summary**

This chapter has presented views of the relevance of manufacturing control in the context of evolution of manufacturing practice and manufacturing system design. Manufacturing systems, while involving a multitude of resources and variables and hence being intrinsically complex, are influenced by their environment and forced to evolve and adapt to dynamic requirements and operating conditions. No system design that involves fixed, static rules or procedures can ensure by itself the satisfaction of the related functional or operational requirements. A good system design includes the definition of reserve or compensatory resources to cope with the effects of disturbances, variations or disruptions. Production rate control, as a way to compensate for the limited reliability of production equipment, is just one of the many compensatory resources that could be used to improve the overall system behavior. Three other relevant compensatory resources have been identified: inventory allocation, production capacity and system availability. With this observation as context, it becomes necessary to understand better the effect of control frequency on the time behavior of manufacturing systems, and more generally, the relevance of production rate control with respect to the other compensatory resources identified.

# 3 Analytic Model

## 3.1 Introduction

This chapter presents the development of an analytic model for a general single-stage, unreliable manufacturing system with controllable production rate. The model is based on a Markov chain process that approximates the flow of material in the system with a continuous variable. The model is intended for performance analysis, as the goal of the study is the development of better understanding of the effects of control frequency on the time behavior of manufacturing systems. Exact analytic solutions to queuing networks are in general not available for any but the simplest systems. For the interested reader, Kleinrock [1975] is a good introduction to queuing theoretic methods, and Buzacott and Shanthikumar [1993] a good review of applications of queuing methods to manufacturing system performance.

Discrete event simulation was used in the exploratory stage of this investigation. Simulation studies required long computation times to produce meaningful evaluations. Some performance measures of interest such as average backlog show a high degree of variation and hence a large number of simulations are required for evaluation. Such computation-intensive simulations are fine for limited research purposes, but are impractical for manufacturing system design, and hence the need for analytical models.

This work is intended to provide approximate yet accessible analysis of control frequency effects for manufacturing system design. A set of mutually independent and collectively exhaustive modeling parameters for a general single-stage, unreliable manufacturing system with controllable production rate is first identified. Those parameters serve as inputs for analysis, namely: mean times before failures, mean time to repair, control frequency, and capacity and

demand. With the intention of understanding better the relevance of production rate control with respect to other system resources for system behavior, the explicit goal of the analysis is to quantify the effect of control frequency on the time behavior of the system. Hence, the compensatory action of production rate control for limited machine reliability and specifically, the effect of control frequency, could be better understood and considered into the models of inventory deployment, production planning, scheduling and control, and capacity planning currently in practice.

The motivation for elaborating analytic models is the development of a better understanding or intuition about the relation between the structure and the time-behavior of the system under consideration, while developing such understanding through simulation studies seems to be impractical. The analytic models considered in this study allow the derivation of probability functions describing the steady state distributions of system states as well as steady state values of relevant performance measures. Hence, the effect of a given parameter on the time behavior of the system can be understood in a more convenient or effective way than through simulation analysis.

This chapter first presents an introduction to the analytic model, the assumptions, generalizations and considerations of most relevance. A Markov-chain model of the general single-stage manufacturing system is introduced next. Then, the rest of the chapter derives probability distribution and transition equations and at the end, the solution of the resulting problem is presented.

## **3.2 Model Assumptions and Considerations**

The probability density functions considered involve the system states  $(x, \alpha)$  and time  $(t)$  as independent variables. The state variable  $x$  is used to indicate a status or condition resulting from the flow of material into and out of a specific point of analysis. If the point of analysis is an inventory buffer downstream of a production unit, the state variable  $x$  represents the production surplus, or the difference between production and requirements of such production unit. This

variable and its use are explained in more detail below. The state variable  $\alpha$  is used to indicate the operational condition of the production unit. Examples of operational conditions are fast, slow, idle, down, starved, blocked and so on. First, an isolated production unit or single-stage manufacturing system is considered. Three states are considered: fast, slow and down. Because of the specific operational conditions assumed some states are not applicable. No coupling effects among multiple stages, infinite supply of raw material and a constant demand result in states such as idle and starved being not applicable. For purposes of modeling, we represent an unreliable flow-line production unit as a single continuous time; continuous and discrete state Markov process. This means that  $x$  and  $\alpha$  can change state in continuous time, where the  $\alpha$  state is discrete and the  $x$  state is continuous.

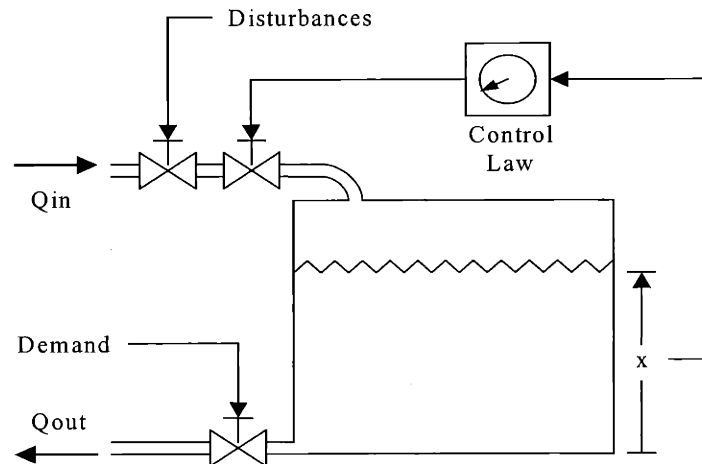
Because  $x$  is a continuous variable, the flow of material into and out of the production unit is not restricted to take place in the form of complete, discrete units. While this is a realistic representation of continuous processes such as those of the chemical industry, it is a modeling simplification in the case of discrete manufacturing processes. Similarly, time is also modeled as a continuous variable and therefore, the system can change state in continuous time. The state variable  $\alpha$  is discrete because the production unit can be in only one operational state at any time and there are discrete choices for the states. Then, to characterize the process we need to derive density functions  $f(x, \alpha, t)$  that vary over discrete values of  $\alpha$  and continuous sets of  $x$  and  $t$  values.

A model representing a general, single-stage flow-line manufacturing system with controllable production rate is developed with the intention of providing a set of generalizations about the influence of control frequency on the time performance of systems that can be represented by such model. Hence, there is a motivation to keep the model general enough so as to allow the development of a general understanding. On the other hand, however, the model needs to be kept simple, not only because it should not represent a single or specific case, but also because we intent to develop an analytical solution for it. In conclusion, we want to isolate a system that is both general and simple, which is manageable to allow analytical modeling but detailed enough to cover the issues of interest.

Specifically, the issues of consideration are:

- Multiple, discrete choices for production rate.
- Non-continuous or discrete-time control actions.
- Limited reliability of production equipment.
- Blockage effects resulting from a limited allocation of space for inventory buffers.

Given that we are interested in systems with multiple choices for the production rate, our model requires a minimum of three system states: fast, slow and down. Three states are needed because there should be at least two choices for the rate in order to have flexibility in production rate. The behavior of the flow-line manufacturing system with controllable production rate under consideration can be better understood through an analogy with a fluidic system, as the one depicted on Figure 3.1.



**Figure 3.1** Liquid level control system as analogy for the control of an unreliable flow-line manufacturing system.

The behavior of an unreliable, volume-flexible production unit can be better understood through an analogy with a random valve controlling the flow rate of a fluid entering a reservoir. The fluid flow represents the flow of material in the manufacturing system. The reservoir

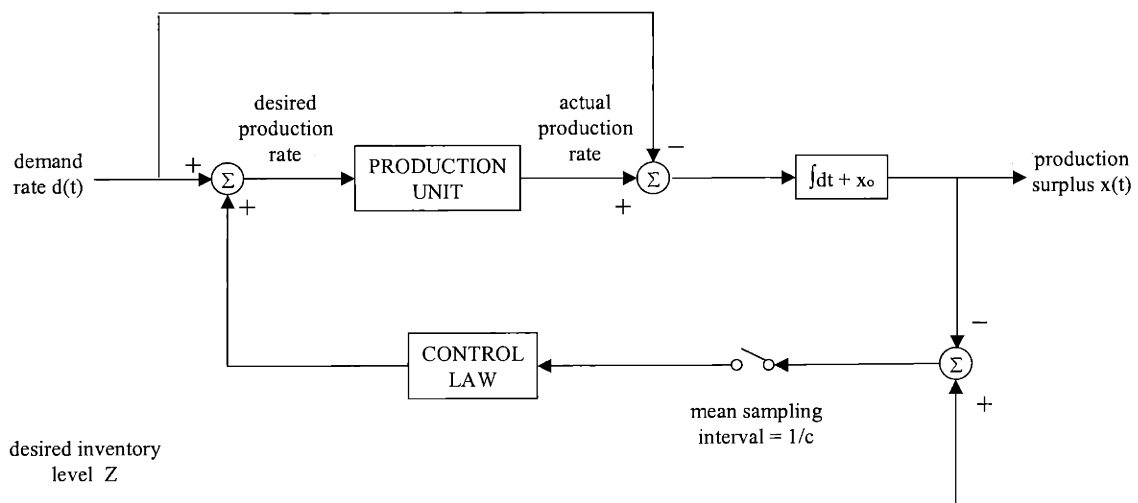
corresponds to an inventory buffer located downstream of the production unit. When the valve is completely open, it delivers material to the reservoir at a rate  $\mu$  per unit time, representing a production unit working at full capacity. When no fluid flows, it represents a condition where the production unit is either broken or commanded to wait for some customer demand. Because the manufacturing system has the capability of processing at different rates, it can be regulated. If it is producing more than a desired amount, we can slow it down. If it is falling behind schedule, we can speed it up in an attempt to recover.

The decision on when and by how much to change the production rate of a real manufacturing system in operation could have, however, potentially disruptive consequences for the downstream production and distribution stages. Actually, a better understanding of such consequences has motivated this work. In many manufacturing systems currently in operation, it seems like there is a set of conditions and complex flows of information that as a whole determine how the material flows within the factory. Conceptually, there exist straightforward principles dictating how manufacturing systems are controlled, that is, how the material flow and the information flow are related. However, because of the complexity of real systems, understanding the flows of material and information is usually very difficult. In some high-volume production systems currently in operation in the automotive industry, it can be quite difficult to understand, even for the operators, how the material flows. Deciding to change the production rate of the manufacturing system, therefore, is not easy to do in such cases.

Some real production systems such as linked-cell systems, on the other hand, offer a much simpler pattern of material and information flow. For that type of system, the analogy of the fluidic system is considered to be enough to represent the dynamics of the material flow related to a single production unit. For this analysis, in particular, the decision on the production rate to implement is taken depending on the state of the system at the moment of the observation. The production surplus  $x$  is taken as the state variable. The production surplus, when positive, corresponds exactly to the inventory level in the buffer. But, contrary to inventory level, the production surplus can become negative to represent backlog. If at a given time the inventory level in the buffer is zero, and the production unit is down and the demand rate is not zero, then the production surplus will keep decreasing (becoming negative) while the inventory level stays

at zero. Then, the inventory level has a lower bound equal to zero and an upper bound defined by the allocation of buffer space. Production surplus, on the other hand, has only an upper bound equal to the allocated buffer space; it approaches  $-\infty$  when the system gets close to its limit of stability.

Figure 3.2 below shows a block diagram representation of the information flows involved in the most general case of the production rate controller being considered. It indicates that two pieces of information are enough to control the operation of the system, a description of the demand rate as a function of time, as well as the desired inventory level. In real systems, these two pieces of information are commonly available. Actually, in many systems in operation, these two parameters are not supposed to change at all during normal operation. Both parameters are more tactical than operational for control. The desired inventory level, for example, is determined as part of an inventory deployment strategy of the production system design. As mentioned in the introduction chapter, this parameter is not supposed to be changed at all once the allocation of inventory along the linked system is determined. Only continuous improvement actions should result in reductions of inventory allocation corresponding to improvements in equipment reliability, throughput times and so on.



**Figure 3.2** Block diagram representation of a production rate controller for an unreliable production system.



For the purposes of analysis of control frequency effects on the behavior of flow-line manufacturing systems, the demand rate indicated in Figure 3.2 corresponds to an aggregated or leveled demand rate over a period of around four weeks. In the Toyota production system, capacity requirements are computed every month or so in terms of aggregate or leveled requirements [Monden, 1993]. Hence, the demand rate  $d(t)$  is influenced by the particular scheduling system in place and is assumed to be constant, or change infrequently with changes that occur at sufficiently low frequencies that allow the system response to reach steady state conditions before the next change in demand occurs.

In our analytical and simulation models, constant values for the demand rate and the desired inventory level are assumed, as well as an unlimited supply of raw material. Given the reasons explained above, these assumptions are considered valid. They are necessary so that the solution is analytically tractable, that is, a closed expression of a meaningful solution can be derived. In cases of multiple-stage manufacturing systems, coupling effects such as blockage, starvation and variation amplification make the assumptions on constant demand and unlimited supply of raw material for each stage unsupported. This fact is the main reason we have to limit the initial analysis to manufacturing systems with a single production unit.

The CONTROL LAW block in Figure 3.2 represents the action of choosing or computing the production rate to be commanded to the system. The signal entering the block, which represents the information needed for such computation, is the difference between the desired inventory level  $Z$  and the production surplus  $x$ . This computation is a case of a *hedging point control law* [Gershwin, 1994].

Then, if:

$u(t)$  indicates the production rate at time  $t$ ,

$\mu$  the capacity rate or maximum value of  $u(t)$  [parts/unit time],

$d$  the demand rate [parts/unit time],

$p$  the failure rate [probability/time interval],

$r$  the repair rate [probability/time interval],

$\alpha$  is the operational state:

$\alpha = 2$  means the unit is operational with a production rate  $u = \mu$

$\alpha = 1$  means the unit is operational with a production rate  $u = d$

$\alpha = 0$  means the unit is down and the production rate  $u = 0$ ;

A necessary condition for stability of the system response is that  $d < \mu$ . And in fact, due to the limited reliability of the system, the initial capacity requirement is

$$d < \frac{\mu r}{r + p}. \quad (3.1)$$

Equation (3.1) states that in order to have a stable response from a manufacturing system under continuous control, the demand rate is required to be lower than the product of the system capacity  $\mu$  and its availability  $\frac{r}{r + p}$ .

The production rate can be controlled. In an ideal case we may want to have the capability to implement any  $u$  in the range

$$0 \leq u \leq \mu.$$

In most manufacturing systems, however, there is a limited number of available production rates. We assume there are two choices for the production rate. Then, if  $x$  represents the production surplus, or the difference between cumulative production and cumulative demand as explained above, it behaves according to

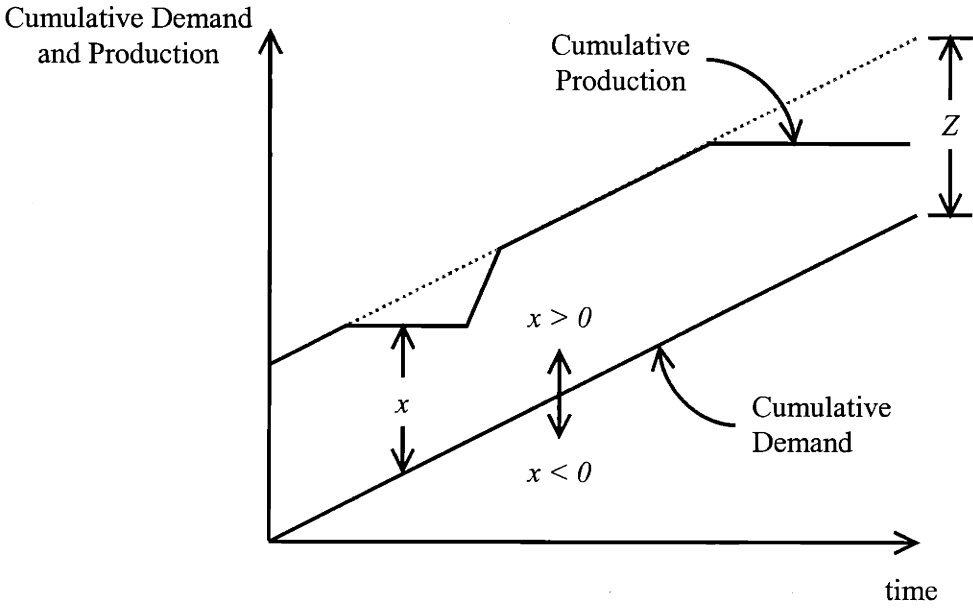
$$\frac{dx}{dt} = u - d. \quad (3.2)$$

Hence, just by checking the value of the state variable  $x$  one can assess how the system is meeting its demand requirements. It is understood that the physical space allocated for storage in the inventory buffer specifies the desired amount of work in process. In other words, only as

much inventory space is allocated as it is needed to accommodate the desired amount of Standard Work In Process (SWIP). Therefore,  $x$  is limited to be less than or equal to  $Z$ ;  $x$  cannot be greater than  $Z$ .

If  $x$  is positive but less than  $Z$ , on the other hand, the system is behind schedule but still able to meet demand without making customers wait and hence without affecting the fill rate of the system. When  $x$  gets to zero the system has just run out of stock to meet demand. If  $x$  is negative then there is backlog, there is a queue of unhappy customers as big as the magnitude of  $x$ , waiting for their products.

As an approach to compensate for whatever moves  $x$  below the desired value  $Z$ , the production rate of the unit can be regulated. This compensation is intended to be a transitory, or short-term, control action. Accordingly, it is meant to compensate for short-term disturbances such as machine failures. Longer-term disturbances such as trends in demand cannot be approached with this type of control actions. Graphically, the dynamic response of a system under these control actions is expected to display a behavior similar to the one presented in Figure 3.3 below.



**Figure 3.3** Production surplus  $x$  of unreliable unit under hedging point law for production rate control.

In Figure 3.3, the desired inventory level  $Z$  is the maximum value of  $x$ , and is called the hedging point or the desired amount of standard work in progress (SWIP). This inventory level is the operational condition the system seeks to attain. Then the control law can be expressed by the decisions:

- if  $x < Z$ ,  $u = \alpha \mu$ ,
- if  $x = Z$ ,  $u = \alpha d$ ,
- if  $x > Z$ ,  $u = 0$ .

The first condition of the hedging point control policy states that if the production surplus does not equal the desired inventory level ( $Z$ ), the system is run at its maximum production rate ( $\mu$  when the system is up, 0 when down). When the system is up, equation 3.2 implies that  $x$  is increasing. The second condition says that if the inventory level is at its desired value and the system is operational, it is run only at the demand rate  $d$ . Therefore,  $x$  remains at  $Z$ , according to equation 3.2. The third condition says that if for any reason  $x$  becomes larger than  $Z$ , the system is commanded to wait idle. It is important to realize that the above conditions as well as blockage effects from the buffer with limited capacity imply that  $x$  can never exceed  $Z$ .

In addition to discrete choices for the production rate, the control actions commanding the production rate also take place at discrete times. Hence, in the case of discrete production rate control with no blockage effects,  $x$  could reach  $Z$  and the system still be commanded to operate at maximum rate. Only after the next time  $x$  is sampled and the production rate corrected,  $x$  will stop increasing if no blockage effects prevent  $x$  from growing beyond  $Z$ . Non-continuous control actions can also result in a situation where, resulting from a machine failure,  $x$  is too far below  $Z$  but the system is still commanded to produce at rate  $d$ , instead of  $\mu$  so that  $x$  approaches  $Z$ . Therefore, if the allocated storage space is much larger than the desired or standard amount of work in progress, then the effect of control frequency is expected to show up not only on the average amount of inventory held but also on the fill rate of the system.

In this initial model, we assume that the space allocated for storage in the buffer is equal to the desired amount of WIP, or  $Z$ . In terms of the analogy of Figure 3.1, our assumption corresponds to a control system intended to keep the reservoir filled at capacity at all times.

When the reservoir gets full, the inflow rate should be set equal to the outflow rate. Even if it is not set to that rate, it will be forced to equal it, because the fluid is assumed to be incompressible. This condition, however, is not desired because the system is making use of resources not needed and because some damage could actually happen to the pump. In the case of manufacturing systems this condition corresponds to a case where we have workers idle most of the time of each cycle, there are more operator work loops than needed and therefore operators may interfere with each other. A frequent control action allows a quick detection of this condition and a corresponding prompt corrective action.

There are many ways to compensate for short-term disturbances affecting the time behavior of a manufacturing system. One approach could be to implement a full capacity production rate  $\mu$  whenever possible. Someone could argue that we could just let the system process at capacity and control it by means of blockage effects. When the downstream buffer gets full and the maximum production rate of the system (rate at capacity) is greater than the demand rate, then the system is forced to work at whatever the demand rate is. When a part is removed from the buffer the system processes a corresponding part, working at capacity rate but just a fraction of the demand cycle time, and waiting idle for the rest of the cycle.

But at this point it is helpful to remember that commonly there are many feasible physical designs or approaches for a corresponding set of requirements. The designer develops what is thought to be the best answer to the problem according to some evaluation criteria, by arranging resources to form a solution. In the case of manufacturing systems, the resources involved are many times reused or designed for general purpose, as a consequence of the high capital investment involved. Those resources however, very often increase the complexity of the system, due to limitations imposed by the processes themselves and the technologies available. Limited machine reliability is a common problem in production equipment. It is just one case of the limitations the system designer has to deal with. Its effects, however, shape the way manufacturing systems are designed.

Limited machine reliability reduces the availability of production equipment and hence, affects strongly the performance of manufacturing systems. Among the multiple ways to compensate for limited equipment reliability we have:

- Allocation of large amounts of inventory.
- Specification of system capacity in excess of demand requirements.
- Duplication of equipment to ensure better overall system reliability.
- Allocation of time buffers (instead of continuous operation).
- Implementation of real-time production rate control.

Depending on the case, some alternatives are more expensive or difficult to achieve than others. There are cases where one alternative is not really an option, due to the requirements of the specific case, or the cost incurred with such option. When the implementation of equipment with controllable production rates seems to be a better option, the reasons justifying such decision in general also define the objective or goal for control in terms of the desired system behavior. In other words, the selection of this approach is based on some criteria that also help define the control law for production rate.

This research has been motivated by a desire to understand better how operational policies affect the behavior of flow-line manufacturing systems. In particular, as mentioned in chapter 1, the frequency at which the manufacturing system state is monitored and some corrective action taken has been found to have a strong effect on the behavior of the system. So far, however, such effect has not been fully understood.

It has been observed during the implementation of manufacturing system designs that very frequently a production system must be developed from equipment already in operation. In such cases, the goal is to improve the behavior or the efficiency of the system by reconfiguring the equipment already available. Issues such as product design, plant location and process technology are already specified and therefore the designer is left with the definition of the path for material flow, the layout, the control policy, the inventory strategy and the operational policies in order to complete the specification of the system. Consequently, most or all design parameters related to equipment specification such as capacity, availability and volume

flexibility are fixed and the designer must integrate those pieces of equipment in the way that best suits the needs of the application. Within this context, this research attempts to develop an understanding base to aid in the specification of operational control policies through a better understanding of the effects of control frequency.

### 3.3 Analytic Model

The proposed analytic model is developed around a Markov process. It has already been mentioned that time and production surplus are modeled as continuous variables while operational state as discrete. This section is intended to cover the specific Markov chain model of our study. For a detail explanation of Markov processes, see [Gershwin, 1994].

First, the nomenclature of the parameters involved in the model is as follows:

$\alpha$  is the operational state:

$\alpha = 2$  indicates the unit is operating at capacity, hence  $u = \mu$

$\alpha = 1$  indicates the unit is operating at the demand rate, hence  $u = d$

$\alpha = 0$  indicates the unit is down and the production rate  $u = 0$ .

$p$  is the failure rate [probability/time interval], such that mean time to failure =  $1/p$ .

**Failure** is any event that results in a transition from state 1 or 2 to state 0.

$q$  is the partial failure rate [probability/time interval], such that mean time to partial failure =  $1/q$ .

**Partial Failure** is any event resulting a partial loss of capacity, modeled as a transition from state 2 to state 1.

After a failure, the unit spends some time in state zero until a repair event takes it back to the same operational state it had immediately before the failure.

$R$  is the repair rate [probability/time interval], such that mean time to repair =  $1/R$ ;

$r$  is the repair rate to state 1, then mean time to repair back to state 1 =  $1/r$ ;

$s$  is the repair rate to state 2, then mean time to repair back to state 2 =  $1/s$ ;

Specification of  $r$  and  $s$  is necessary to ensure that every repair event takes the unit back to the operational state it had right before the failure. The mathematical manipulation that enforces this condition is explained below in the form of an identity. Also,  $R = r + s$ .

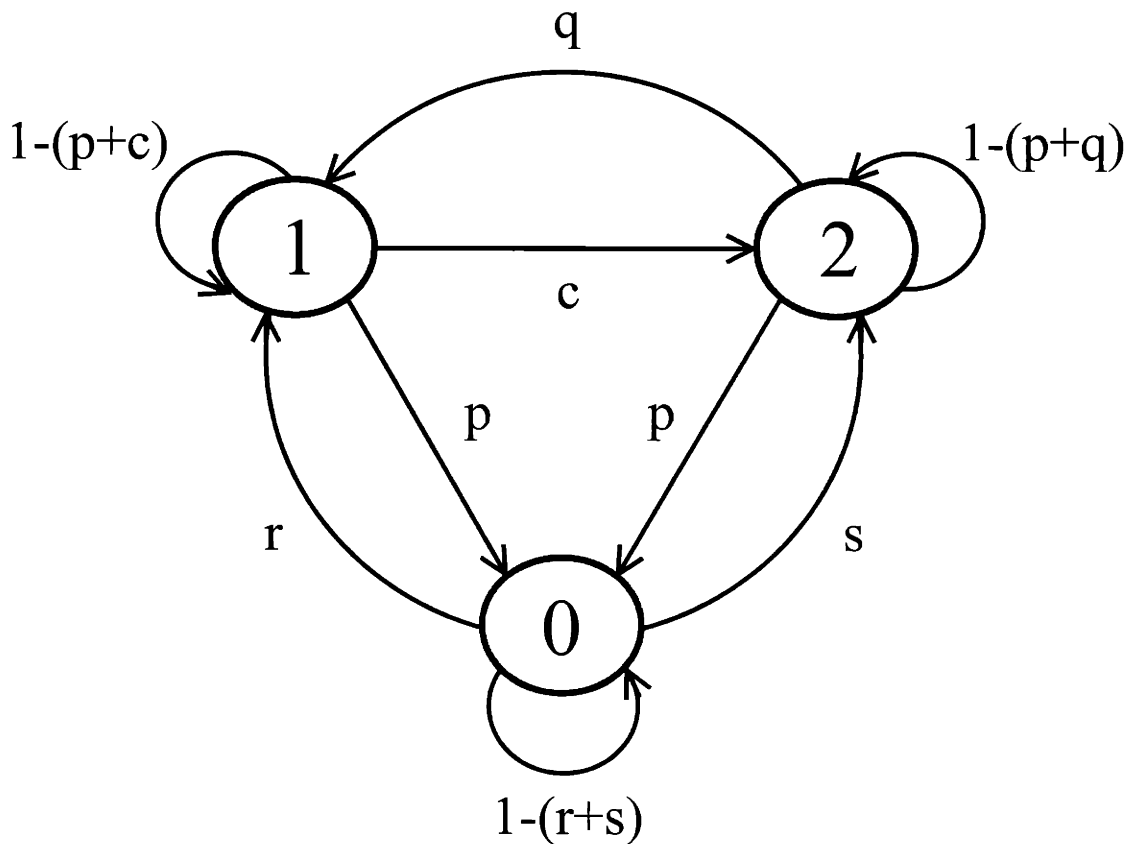
$c$  is the control rate or control frequency. Hence, the system is monitored, a control law is evaluated and a corresponding corrective action implemented in average every  $1/c$  time units.

As indicated in Figure 3.4, it is assumed that the system has three operational states. State zero represents a failure state; the production unit has some problem with at least one process. Independently of the production rate being commanded, the system's production rate is zero. State one corresponds to any situation where the production unit is processing material at the demand rate and therefore the production surplus or backlog is constant. State two is used to represent a state of fast processing, at a rate faster than the demand rate, when the system is working at full capacity. Hence, the control actions resulting from the implementation of the hedging-point control policy mentioned above will consist of transitions being commanded from state 1 to state 2, whenever the system state  $x$  is falling behind the desired amount of work in progress  $Z$ . Blockage effects will result in transitions from state 2 to state 1 whenever the inventory buffer gets full, or in other words, when the system state equals the desired amount of work in progress.

It is assumed that whenever the system is operational, that is whenever the system is in state 1 or 2, the system fails or a transition to state 0 happens in average every  $1/p$  time units, and that the average time the system stays operational follows an exponential distribution (with mean  $1/p$ ). Then, it is said that failures are operation time dependent with a mean time to failure (MTTF) of  $1/p$ . This statement implies that, for the purpose of measuring the time before the next failure, time counts only if the system is working. In other words, the system can fail only if it processing material, that is, it is operational and not blocked nor starved. Once the system gets to state 0, it is assumed that the time required to repair the production unit follows also an exponential distribution with a mean of  $1/r$ , or a mean time to repair (MTTR)  $1/r$ .

Because it is desired to uncouple the effects of control frequency and machine unreliability, it is assumed that after a repair is completed, the system returns to the same operational state it had before the failure. Hence, the frequency of transitions between states 1 and 2 is affected only by the frequency of the control actions and by how often the system variable  $x$  gets to  $Z$  and not by the frequency of the failures.





**Figure 3.4** Markov-Chain model for  $x < Z$  of a general single-stage manufacturing system with controllable production rate.

Numbers on arcs are transition rates.

It is also assumed that when the system is in state 2 there can be an undesired transition to state 1. Such transition represents a partial failure resulting in a loss of some capacity. If the production unit has parallel processing, that is, multiple machines performing a common operation, then a partial failure represents an event where some but not all of those machines fail. If the production unit is a cell with multiple operators, a partial failure represents an event where one or more but not all operators decide to leave the area for some time. As a consequence of any of these events, the production unit loses some of its capacity and produces at a rate lower than

the commanded rate corresponding to state 2. With the intention of keeping the complexity of the model limited, state 1 is also used to represent states resulting from a partial loss of capacity.

It is assumed that the probabilities describing the transitions among states, are constant. Therefore, the system is being modeled as a homogeneous Markov process. From these observations, a set of 5 mutually independent and collectively exhaustive modeling parameters is identified. The set that is required to completely characterize a production unit or stage of a flow line manufacturing system includes:

- $\mu$ : maximum production rate, or production rate at capacity [parts/unit time]
- $d$ : demand production rate [parts/unit time]
- $1/p$ : mean time to failure
- $1/r$ : mean time to repair
- $1/q$ : mean time to partial failure (undesired transition from state 2 to state 1)

Additionally, the model under consideration involves the following fundamental conditions and assumptions:

- The system reliability is not affected by the production rate.
- Upon completion of a repair, the system returns to the same operational state before failure.
- The system has at least two choices for the production rate. The production rate at capacity is greater than the demand rate.
- The time required for switching between production rates is negligible.

### 3.3.1 Probability Distribution and Transition Equations

It is important, at this point, to state that the explicit goal of the analysis is to derive closed expressions for the probability density functions  $f(x, \alpha, t)$  for  $\alpha=0$ ,  $\alpha=1$ , and  $\alpha=2$  and for  $P(Z, \alpha, t)$ . Because the probability  $P(x, \alpha, t)$  can be greater than zero at the specific point  $x=Z$ , then  $P(Z, \alpha, t)$  is a probability mass.

For  $x < Z$  the probability distribution  $f(x, \alpha, t)$  is defined as:

$f(x, \alpha, t) \delta x = \text{prob}(x \leq X(t) \leq x + \delta x \text{ and the machine state is } \alpha \text{ at time } t).$

Where  $X$  is a random variable that at time  $t$ , is expressed as  $X(t)$  and represents a specific value of the state variable  $x$ .

Then, for  $x < Z$ , according to Figure 3.4:

$$f(x, 0, t + \delta t) = p \delta t f(x - (\mu - d) \delta t, 2, t) + p \delta t f(x, 1, t) + [1 - (r + s) \delta t] f(x + d \delta t, 0, t)$$

Three terms are needed because at time  $t + \delta t$  the system could be in state  $(x, 0)$  if at time  $t$  the system was in state  $(x - (\mu - d) \delta t, 2)$ , processing at rate  $\mu$  (because  $\alpha = 2$ ) and a failure occurs in the interval  $\delta t$  (with probability  $p$ ); or in state  $(x, 1)$ , processing at rate  $d$  (because  $\alpha = 1$ ) and a failure occurs in the interval  $\delta t$  (with probability  $p$ ); or in state  $(x + d \delta t, 0)$  in breakdown state and no repair occurs during the interval  $\delta t$  (with probability  $1 - (r + s)$ ).

In steady state, once the  $t$  arguments are removed:

$$f(x, 0) = p \delta t f(x - (\mu - d) \delta t, 2) + p \delta t f(x, 1) + [1 - (r + s) \delta t] f(x + d \delta t, 0) \quad (3.3)$$

This statement is an approximation, and is only accurate to within the order of  $\delta t$ . A more accurate argument for  $(x - (\mu - d) \delta t, 2)$  would be some  $x'$  within  $[x - (\mu - d) \delta t, x + \delta t]$ . If the failure occurs early in the interval, then  $x'$  must take the value  $x + d \delta t$ ; if it occurs close to the end of the interval, then  $x'$  must have been  $x - (\mu - d) \delta t$ . However, the difference between either end of the interval and  $x$  is on the order of  $\delta t$ , and expansion of the terms of  $f(x, 0)$  will result in the term  $\delta t$  being multiplied by the  $p \delta t$  coefficient, and hence it can be neglected. Therefore, we may as well consider  $f(x, 2)$  to be evaluated at  $x$  rather than  $x'$ . As a convention, we will assume failure and repair transitions to happen just before the end of the interval  $\delta t$ .

To analyze (3.3),  $f(x+d\delta t, 0)$  as well as  $f(x-(\mu-d)\delta t, 2)$  must be expanded. The expansions are given by:

$$f(x+d\delta t, 0) = f(x, 0) + \left[ \frac{\delta f}{\delta x}(x, 0) \right] d\delta t + o(\delta t).$$

Where  $o(\delta t)$  is an error of order  $\delta t$ . Similarly, for state 2:

$$f(x-(\mu-d)\delta t, 2) = f(x, 2) + \left[ \frac{\delta f}{\delta x}(x, 2) \right] (\mu-d)\delta t + o(\delta t).$$

Consequently, expansion of (3.3) is

$$f(x, 0) = [1-(r+s)\delta t] \left\{ f(x, 0) + \left[ \frac{\delta f}{\delta x}(x, 0) \right] d\delta t \right\} + p\delta t f(x, 1) + p\delta t \left\{ f(x, 2) + \left[ \frac{\delta f}{\delta x}(x, 2) \right] (\mu-d)\delta t \right\}$$

Discarding terms of second order in  $\delta t$ :

$$0 = pf(x, 2) + pf(x, 1) - (r+s)f(x, 0) + d \frac{\delta f}{\delta x}(x, 0). \quad (3.4)$$

Similarly, for  $x < Z$ :

$$f(x, 1, t+\delta t) = q\delta t f(x-(\mu-d)\delta t, 2, t) + [1-(p+c)\delta t] f(x, 1, t) + r\delta t f(x+d\delta t, 0, t)$$

And at steady state:

$$f(x, 1) = q\delta t f(x-(\mu-d)\delta t, 2) + [1-(p+c)\delta t] f(x, 1) + r\delta t f(x+d\delta t, 0)$$

Expanding the terms around  $f(x, 2)$  and  $f(x, 0)$  leads to:

$$f(x,1) = q\delta t \{f(x,2) + [\frac{\delta f}{\delta x}(x,2)](d-\mu)\delta t\} + [1-(p+c)\delta t]f(x,1) + r\delta t \{f(x,0) + [\frac{\delta f}{\delta x}(x,0)]d\delta t\}$$

or,

$$0 = qf(x,2) - (p+c)f(x,1) + rf(x,0). \quad (3.5)$$

Finally, also for  $x < Z$ :

$$f(x,2,t+\delta t) = [1-(p+q)\delta t]f(x-(\mu-d)\delta t,2,t) + c\delta t f(x,1,t) + s\delta t f(x+d\delta t,0,t)$$

At steady state:

$$f(x,2) = [1-(p+q)\delta t]f(x-(\mu-d)\delta t,2) + c\delta t f(x,1) + s\delta t f(x+d\delta t,0)$$

Expanding leads to:

$$f(x,2) = [1-(p+q)\delta t] \{f(x,2) + [\frac{\delta f}{\delta x}(x,2)](d-\mu)\delta t\} + c\delta t f(x,1) + s\delta t \{f(x,0) + [\frac{\delta f}{\delta x}(x,0)]d\delta t\}$$

or,

$$0 = -(p+q)f(x,2) + (d-\mu)\frac{\delta f}{\delta x}(x,2) + cf(x,1) + sf(x,0). \quad (3.6)$$

Boundary conditions must be satisfied at  $x = Z$ , when the production surplus equals the desired amount of work in process.  $P(Z,1)$  is a probability mass because there is a non-zero probability of finding the system state at  $(x, \alpha)$  when  $x = Z$  and  $\alpha = 1$ . Because the demand rate is assumed to be constant, whenever the production unit is down,  $x$  decreases. Therefore,  $P(Z,0) = 0$ . On the other hand, we also assume that blockage effects force the unit's production rate to equal the demand rate as soon as the buffer gets full, that is when  $x = Z$ . Hence, if the system

gets to  $(Z,2)$  from  $(x,2)$ , then as soon as  $x$  equals  $Z$ ,  $\alpha$  changes from 2 to 1. Therefore,  $P(Z,2) = 0$  also.

Then, to get to  $(Z,1)$  from  $(x,2)$  where  $x < Z$ :

$$P(Z,1) = P(Z,1)(1-p\delta t) + \text{prob}(Z-(\mu-d)\delta t < X < Z, \alpha=2)(1-p\delta t) + o(\delta t) \quad (3.7)$$

That is, the state can be at  $(Z,1)$  at time  $t+\delta t$  if it was at the same state at time  $t$  and no failure occurred in the interval  $[t, t+\delta t]$  (because when  $\alpha = 1$ ,  $x$  does not change). Or the state variable  $x$  could have been just below  $Z$  at time  $t$ , state  $\alpha$  equal to 2 and therefore  $x$  increasing, approaching  $Z$ , and the production unit did not fail in  $[t, t+\delta t]$ . We do not need to consider repair transitions from states of the form  $(x,0)$  in which  $Z-(\mu-d)\delta t \leq x \leq Z$ , because the corresponding terms would be of second order in  $\delta t$ .

Note that, from the definition of the probability density function  $f$ :

$$\text{prob}(Z-(\mu-d)\delta t < X < Z, \alpha=2) = f(Z,2)(\mu-d)\delta t$$

And therefore, equation (3.7) can be rewritten as:

$$P(Z,1) = P(Z,1)(1-p\delta t) + f(Z,2)(\mu-d)\delta t (1-p\delta t) + o(\delta t)$$

Or, to first order,

$$0 = -pP(Z,1) + (\mu-d)f(z,2) \quad (3.8)$$

To get to the interior ( $x < Z$ ) from  $x = Z$ , observe that by definition, where  $\delta x = d\delta t$ ,

$$\text{prob}(Z-\delta x < X < Z, 0) = f(Z,0)\delta x + o(\delta t)$$

But also, due to a failure of the production unit when  $x = Z$ ,

$$\text{prob}(Z - \delta x < X < Z, 0) = P(Z, 1)p\delta + o(\delta)$$

There are no other transitions with probability of first order in  $\delta t$ . Then,

$$f(Z, 0)d = P(Z, 1)p \quad (3.9)$$

To summarize, the following set of equations must be solved:

$$0 = pf(x, 2) + pf(x, 1) - (r+s)f(x, 0) + d \frac{\delta f}{\delta x}(x, 0). \quad (3.10)$$

$$0 = qf(x, 2) - (p+c)f(x, 1) + rf(x, 0). \quad (3.11)$$

$$0 = -(p+q)f(x, 2) + (d-\mu) \frac{\delta f}{\delta x}(x, 2) + cf(x, 1) + sf(x, 0). \quad (3.12)$$

$$0 = -pP(Z, 1) + (\mu-d)f(Z, 2) \quad (3.13)$$

$$df(Z, 0) = pP(Z, 1) \quad (3.14)$$

### 3.3.2 Solution of Distribution and Transition Equations

To solve (3.10) to (3.14) the following reasonable guess is tried:

$$f(x, \alpha) = A(\alpha) e^{bx}$$

Where  $A(0)$ ,  $A(1)$ ,  $A(2)$  and  $b$  must be determined.

Substituting into (3.9) and dividing by  $e^{bx}$ :

$$0 = [(d-\mu)b - (p+q)]A(2) + cA(1) + sA(0) \quad (3.15)$$

From (3.10)

$$0 = qA(2) - (p+c)A(1) + rA(0) \quad (3.16)$$

From (3.11)

$$0 = pA(2) + pA(1) + (db-r-s)A(0) \quad (3.17)$$

Solving for  $P(Z,1)$  in (3.14) and substituting into (3.13)

$$dA(0) = (\mu-d)A(2) \quad (3.18)$$

Resulting in the system of algebraic equations:

$$0 = [(d-\mu)b - (p+q)]A(2) + cA(1) + sA(0) \quad (3.19)$$

$$0 = qA(2) - (p+c)A(1) + rA(0) \quad (3.20)$$

$$0 = pA(2) + pA(1) + (db-r-s)A(0) \quad (3.21)$$

$$dA(0) = (\mu-d)A(2) \quad (3.22)$$

Solving for  $b$  from the set (3.19) – (3.22):

$$b = \frac{cr}{d(p+c)} - \frac{p+q}{\mu-d} + \frac{cq}{(\mu-d)(p+c)} + \frac{s}{d} \quad (3.23)$$



Renaming  $A(0)$  as  $A$ , the solution can be written in the form:

$$f(x,0) = A e^{bx} \quad (3.24)$$

$$f(x,1) = \frac{r(\mu-d)+qd}{(p+c)(\mu-d)} A e^{bx} \quad (3.25)$$

$$f(x,2) = \frac{d}{(\mu-d)} A e^{bx} \quad (3.26)$$

$$P(Z,1) = \frac{d}{p} A e^{bx} \quad (3.27)$$

$A$  is a normalizing constant, chosen so that

$$P(Z,1) + \int_{-\infty}^Z [f(x,0) + f(x,1) + f(x,2)] dx = 1 \quad (3.28)$$

Then,

$$A = \frac{bp(p+c)(\mu-d)}{db(p+c)(\mu-d) + p\mu(p+c+r) + pd(q-r)} e^{-bZ} \quad (3.29)$$

The solution (3.23) – (3.29) involves the repair rates  $r$  and  $s$ , while our set of modeling parameters describing the production unit includes a single repair rate parameter defined by a mean time to repair parameter (MTTR). This follows from the initial assumption where it is considered that the time required to fix the unit after a failure does not depend on the operational state just before failure. For the model to be complete, it is needed to relate  $r$  and  $s$ , the two repair rates of the analytical model (one repair rate related to each operational state), with the single repair rate of our descriptive model.

### 3.3.3 Repair Frequency Equal to Failure Frequency Identity

For a system in steady state, for every failure there is a corresponding repair. This is self-evident if we think of a system that has been observed for a long time. At any time, the system will be either operational or failed. If it is operational, the next transition will be to another operational state or to a failure state, but the next transition cannot be a repair. In order for a repair transition to happen, it is required that the system is down. On the other hand, if the system is down then the following transition, whenever it happens will have to be a repair, and a failure transition is not possible. Therefore, no two failure transitions or two repair transitions can happen together, and hence, the cumulative number of failure transitions changes with the same rate as the cumulative number of repair transitions.

Then, if

$$P(0) = \int_{-\infty}^Z f(x,0)dx \quad (3.30)$$

$$P(1) = P(Z,1) + \int_{-\infty}^Z f(x,1)dx \quad (3.31)$$

$$P(2) = \int_{-\infty}^Z f(x,2)dx \quad (3.32)$$

Then the repair/failure frequency identity can be written as:

$$pP(1) = rP(0) \quad (3.33)$$

and

$$pP(2) = sP(0) \quad (3.34)$$

The parameter  $p$  is present on both equations because we consider the system equally reliable on both operational states. Because the descriptive model has only one parameter to represent the mean time to repair MTTR, or the average time it takes to repair the system back to state 1 or state 2, it can be said as usual:

$$\text{MTTR} = \frac{1}{R}, \text{ where:}$$

$$R = r + s \tag{3.35}$$

Evaluating (3.30) – (3.32):

$$P(0) = \frac{1}{kb} \tag{3.36}$$

$$P(1) = \frac{d}{kp} + \frac{r(\mu - d) + qd}{(p + c)(\mu - d)} \frac{1}{kb} \tag{3.37}$$

$$P(2) = \frac{d}{\mu - d} \frac{1}{kb} \tag{3.38}$$

Where,

$$K = \frac{db(p + c)(\mu - d) + p\mu(p + c + r) + pd(q - r)}{bp(p + c)(\mu - d)} \tag{3.39}$$

Substituting the expressions for  $P(0)$  and  $P(2)$  into the identity (3.36 and 3.38 into 3.34):

$$s = \frac{dp}{\mu - d}$$

Substituting this expression for  $s$  into (3.23) a simpler expression for  $b$  results:

$$b = \frac{cr}{d(p+c)} - \frac{qp}{(\mu-d)(p+c)} \quad (3.40)$$

For convergence of the assumed solution with form  $f(x, \alpha) = A(\alpha) e^{bx}$  and therefore, for a steady state solution and a stable system response to exist, it is required that  $b > 0$  (see integration limits of equations 3.30, 3.31, 3.32). Hence, the control frequency that is required for stable system response is found to be:

$$c > \frac{pqd}{R(\mu-d) - dp} \quad (3.41)$$

In other words, the right side of this inequality is a closed solution for the lowest control frequency  $c$  that is necessary to meet the demand requirements and hence have a stable system response.

And finally, the state solution takes the form:

$$f(x,0) = \frac{bp(p+c)(\mu-d)}{db(p+c)(\mu-d) + p\mu(p+c+r) + pd(q-r)} e^{b(x-Z)} \quad (3.42)$$

$$f(x,1) = \frac{bp[r(\mu-d) + qd]}{db(p+c)(\mu-d) + p\mu(p+c+r) + pd(q-r)} e^{b(x-Z)} \quad (3.43)$$

$$f(x,2) = \frac{dbp(p+c)}{db(p+c)(\mu-d) + p\mu(p+c+r) + pd(q-r)} e^{b(x-Z)} \quad (3.44)$$

$$P(Z,1) = \frac{db(p+c)(\mu-d)}{db(p+c)(\mu-d) + p\mu(p+c+r) + pd(q-r)} \quad (3.45)$$

$$P(Z,0) = P(Z,2) = 0$$

(3.46)



# 4 Model Validation and Results

## 4.1 Introduction

The solution of the analytical model as stated at the end of Chapter 3 is in a form that by itself conveys little information about the time behavior of the manufacturing systems under study. Straightforward manipulations, however, allow the derivation of expressions for relevant performance measures. From the definition of those measures and the expressions for the probability density distributions stated in the model solution, closed expressions for steady-state performance measures such as fill rate, average backlog, average production surplus and system state probabilities can be derived. Since those performance measures are crucial for evaluation of competing manufacturing system designs, the validation of the model solution and the related results is required.

The calculations related to the solution of the analytical model and the results derived from it have been implemented in Matlab. This chapter presents the validation of the analytical model and the derivation of some relevant results. Generality of results is limited in some cases only by presentation requirements. While the results are applicable to any manufacturing system with the same structure as the one described in Chapter 3, generation of graphical results usually requires the specification of every system parameter. Validation results show that the analytical model predicts extremely well the performance of the system. The presentation of the results shows the relevance of control frequency as system parameter. Also, the influence of control frequency on system behavior and its complementary effect with that of other system parameters becomes apparent from the presentation of the results. Towards the end of the chapter, an approach to the generalization of the analysis to the case of multi-stage manufacturing systems is presented.

## 4.2 Model Validation

Since the model and the information derived from its results are crucial for understanding of control frequency issues during the development of manufacturing systems, the accuracy and correctness of the solutions needs to be validated. The validation of the results has been done in different ways, as explained below.

Result derivations were done following standard procedures and assumptions. In fact, this work is fundamentally different from previous works in the way available analytical tools are applied to the analysis of flow line manufacturing systems. Fundamental assumptions that became needed for the solution were tested in several ways and the results compared. Every assumption in the model was tested to result in meaningful outputs. For example, the failure rate-repair rate identity and the probability normalization equation were implemented in different forms and the results were tested for consistency. A very useful test of the correctness of many derivations consisted on making sure that every result that could be included into an expression for the system availability should result in an expression for the system availability that after simplification did not include control frequency terms. This follows from the recognition that system availability is independent of control frequency.

The analytic model results were also validated with discrete event simulations. Analytic results were compared with those of the simulation model of a corresponding manufacturing system, that is, an unreliable, volume-flexible unit involving the same approximations. Initial validation efforts used commercial simulation software. However, due to the level of accuracy required for validation purposes, and the existence of discrepancies between analytical and simulation results with no explainable reason, it was needed to implement the code of both, analytical and simulation models in the same language (Matlab). The need to develop a discrete event simulation program from basic programming commands was not considered during the early exploratory stages of this project. This experience shows the difficulty encountered in this research area to integrate multiple, very different tools and approaches that become necessary for the completion of the analysis.



Tests were run to study the behavior of the model results across any feasible or meaningful case. The sensitivity of the accuracy of the results was checked by changing every parameter of the modeling set, one by one. These tests also allowed a better understanding of the effect of every parameter on the entire system behavior. It was decided to validate the analytical model by testing its accuracy in predicting the limit of stability (equation 3.41), or the slowest control frequency required for stable response, the four system state probabilities as well as more operational measures such as average production surplus and backlog (measures listed on Tables 4.1 to 4.5). The reasons for taking this criterion for validation are the initial motivation of the project (determination of the limit of stability) and the intended application of the results. This procedure is based on the validation standards that are observed in most works documented in the literature. Previous works with good documentation of the validation procedure are [Sharifnia, 1998], [Burman, 1995] and [Bonvik, 1996].

Stable response in this context means that the system under consideration is able, at steady state, to meet the demand. Then a bounded, even if very large, steady-state backlog queue is related to a stable response, even if the fill rate of such system is very poor. Test points were defined by holding all parameters constant except the control frequency. For each test point, fifteen simulation runs or replications were completed, each for  $1 \times 10^9$  time units, with a takt time of 50 time units. For each simulation run the random number generators were initialized with a different seed, so that the set of results can be considered an independent sample of size fifteen. For each measure tested, the sample mean and variance was calculated from the sample of size fifteen. Then a 95% confidence interval was calculated using a Student's t-distribution with fourteen degrees of freedom, according to:

$$x \in \bar{X} \pm t_{14}^{0.025} \frac{S^2}{\sqrt{n}} \quad (4.1)$$

Where  $x$  represents some performance measure,  $\bar{X}$  is the sample mean of that performance measure,  $S^2$  is the sample variance and  $n$  is the sample size (15 for this case) [DeVor, et al, 1992]. This approach assumes that each performance measure from the replications has a distribution that can be approximated by a normal function. Equation 4.1 is rewritten as:

$$x \in \bar{X} \pm 0.554S^2 \quad (4.2)$$

For each performance measure, the validation test begins with the hypothesis that the analytical model output correctly predicts the corresponding performance measure. This hypothesis is rejected at the 5% confidence level if the analytical value falls outside the 95% confidence interval of the corresponding simulated performance measure, as determined by equation 4.2. In other words, 5% confidence level means that on average, correct hypotheses will be rejected in 5% of the cases. If the analytical result falls within the confidence interval, the hypothesis cannot be rejected. The results of the validation tests are presented in the following tables.

c	u	d	p	q	R
1/25	1/35	1/50	1/5000	1/2000	1/1800

Performance Measure	Confidence Interval	Analytic Result	Validation
$P_0$	$0.2645 \pm 0.0004$	0.2647	Pass
$P_Z$	$0.1100 \pm 0.0013$	0.1094	Pass
$P_{1x}$	$0.0083 \pm 0.0000$	0.0083	Pass
$P_{2x}$	$0.6172 \pm 0.0009$	0.6176	Pass
Average surplus	$286.4116 \pm 2.8191$	284.4638	Pass
Average backlog	$26.9023 \pm 1.5727$	27.3050	Pass

**Table 4.1** Control Period =  $0.5 \times$  Takt Time.

c	u	d	p	q	R
1/200	1/35	1/50	1/5000	1/2000	1/1800

Performance Measure	Confidence Interval	Analytic Result	Validation
$P_0$	$0.2647 \pm 0.0004$	0.2647	Pass
$P_Z$	$0.0540 \pm 0.0013$	0.0538	Pass
$P_{1x}$	$0.0639 \pm 0.0001$	0.0639	Pass
$P_{2x}$	$0.6174 \pm 0.0009$	0.6176	Pass
Average surplus	$48.5854 \pm 19.0638$	33.8386	Pass
Average backlog	$159.3344 \pm 16.0679$	168.9455	Pass

**Table 4.2** Control Period =  $4.0 \times$  Takt Time.

c	u	d	p	q	R
1/350	1/35	1/50	1/5000	1/2000	1/1800

Performance Measure	Confidence Interval	Analytic Result	Validation
$P_0$	$0.2651 \pm 0.0002$	0.2647	Fail
$P_Z$	$0.0079 \pm 0.0007$	0.0089	Fail
$P_{1x}$	$0.1088 \pm 0.0001$	0.1087	Pass
$P_{2x}$	$0.6182 \pm 0.0005$	0.6176	Fail
Average surplus	$-2340.9 \pm 436.5566$	-2436.7	Pass
Average backlog	$2381.4 \pm 433.8352$	2480.7	Pass

**Table 4.3** Control Period =  $7.0 \times$  Takt Time.

c	u	d	p	q	R
1/365	1/35	1/50	1/5000	1/2000	1/1800

Performance Measure	Confidence Interval	Analytic Result	Validation
$P_0$	$0.2647 \pm 0.0003$	0.2647	Pass
$P_Z$	$0.0054 \pm 0.0010$	0.0046	Pass
$P_{1x}$	$0.1130 \pm 0.0001$	0.1131	Pass
$P_{2x}$	$0.6170 \pm 0.0007$	0.6176	Pass
Average surplus	$-3865.8 \pm 873.2849$	-5239.5	Fail
Average backlog	$3895.5 \pm 869.5876$	5262.5	Fail

**Table 4.4** Control Period =  $7.3 \times$  Takt Time.

c	u	d	p	q	R
1/375	1/35	1/50	1/5000	1/2000	1/1800

Performance Measure	Confidence Interval	Analytic Result	Validation
$P_0$	$0.2650 \pm 0.0003$	0.2647	Pass
$P_Z$	$0.0022 \pm 0.0006$	0.0017	Pass
$P_{1x}$	$0.1159 \pm 0.0001$	0.1159	Pass
$P_{2x}$	$0.6170 \pm 0.0004$	0.6176	Fail
Average surplus	$-9078.1 \pm 2926.7$	-14960.0	Fail
Average backlog	$9089.7 \pm 2926.0$	14969.0	Fail

**Table 4.5** Control Period =  $7.5 \times$  Takt Time.

It can be seen that the validations seem to fail more frequently for smaller values of  $c$ , corresponding to lower control frequencies. For the specific set of parameters considered, the limiting control frequency is around  $1/380$ , or a control period =  $7.6 \times$  Takt Time. It is however, interesting to note that according to the confidence interval defined, it can happen that the analytical model predicts the state probabilities correctly yet fail to predict correctly the steady state average production surplus and production backlog, as seen in the case for  $c=1/375$  in table 4.5. The opposite seems to happen for the case where  $c=1/350$  presented in table 4.3, where the steady-state value of the average production surplus and production backlog are predicted correctly even if three state probability tests failed. This behavior seems to be counterintuitive, given that the expressions for average production surplus and backlog are derived from terms related to the probability density functions that also determine the state probabilities, as indicated with the derivations of performance measures in the following section.

It was observed that by redoing the experiments the validation results change, it is to say, performance measures that initially failed the validation eventually pass while occasional failures take place somewhere else. The variance of the samples affects the confidence interval and specifically the mean value for the state probability  $P_0$  as predicted by the simulation model changes as it can be seen in the tables, while in reality  $P_0$  is a parameter that is expected to be constant. It is concluded that these variations are spurious failures from the statistical simulation procedure. Therefore, the analytic model is taken as valid and its results are considered completely reliable as long as the system is not operating very close to the stability limit. And even close to the limit of stability, the model can be used for quick performance evaluation if it is considered that in any system operating close to capacity the effect of variations becomes stronger.

### 4.3 Derivation of Performance Measures

From the relation between state probability and probability distribution function:

$$P_0 = \int_{-\infty}^z f(x,0)dx \quad (4.3)$$

Where  $f(x,0)$  is given in the analytical model solution, equation 3.42. Similarly, for the other two system states:

$$P_1 = \int_{-\infty}^{\infty} f(x,1)dx \quad (4.4)$$

$$P_2 = \int_{-\infty}^{\infty} f(x,2)dx \quad (4.5)$$

Where  $P_1$  in equation 4.4 does not include the probability mass  $P(Z,1)$ . Substitution of the expressions for the density distributions, equations 3.42, 3.43 and 3.44 into 4.3, 4.4 and 4.5 results in expressions for the state probabilities:

$$P_0 = \frac{p(p+c)(\mu-d)}{db(p+c)(\mu-d) + p\mu(p+c+r) + pd(q-r)} \quad (4.6)$$

$$P_1 = \frac{p[r(\mu-d) + qd]}{db(p+c)(\mu-d) + p\mu(p+c+r) + pd(q-r)} \quad (4.7)$$

$$P_2 = \frac{dp(p+c)}{db(p+c)(\mu-d) + p\mu(p+c+r) + pd(q-r)} \quad (4.8)$$

From the definition of production surplus, the average production surplus is computed from:

$$\bar{x} = \int_{-\infty}^{\infty} x[f(x,0) + f(x,1) + f(x,2)]dx \quad (4.9)$$

Similarly, from the definition of backlog, the average production backlog is computed from:

$$\bar{B} = - \int_{-\infty}^{\infty} x[f(x,0) + f(x,1) + f(x,2)]dx \quad (4.10)$$

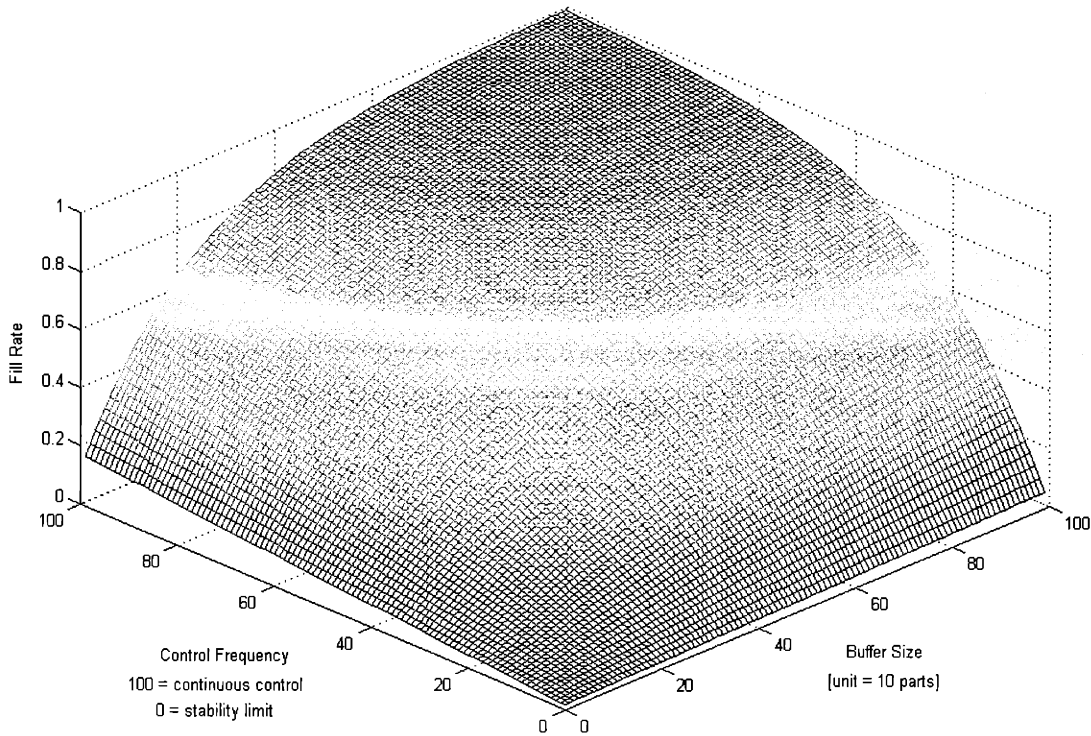
Substitution of equations 3.42, 3.43 and 3.44 into 4.9 and 4.10 yields expressions for the average production surplus and backlog in terms of known system parameters. Because of space limitations and the complexity of the resulting expressions, the explicit form of the results is not presented here. Instead, the results are presented in graphical form in the following section.

## 4.4 Effect of Control Frequency on System Performance

This section presents the results derived from the analytical model solution. The results are intended to show how control frequency affects the time behavior of a single stage manufacturing system, in other words, how the parameter describing control frequency relates to the other system parameters in terms of its effect on system behavior. By providing an analytical tool for quick assessment of system behavior, it is expected that the manufacturing system designer could better specify supervisory actions for the system under development. Once the effect of control frequency on a single stage manufacturing system has been explained, an approach for analysis of multi-stage manufacturing systems is presented. This approach is based on a conversion of parameters, where the effect of discrete control actions is included into the reliability parameters. Then, the converted parameters can be used with the already available tools in the literature for performance evaluation of multi-stage manufacturing systems.

Figure 4.1 shows the effects of control frequency and buffer size on the fill rate of the system. The system parameters are  $u=1/35$ ,  $d=1/50$ ,  $p=1/5000$ ,  $q=1/2000$  and  $R=1/1800$ . The buffer size varies from 0 to 1000, each unit along the axis representing 10 parts. The control frequency varies from the minimum frequency required for stable response at the origin of the axis to a very high frequency corresponding to a continuous control action at the end of the axis. It is seen that the fill rate of the system increases monotonically with control frequency and buffer size. As indicated in the expression for the limiting control frequency, equation 3.41, the amount of inventory allocated does not affect the limit of stability. This is indicated in figure 4.1 by the fact the fill rate function approaches the value of zero along an edge that is parallel to the buffer size axis. It is noted that sufficiently large allocations of buffer space and sufficiently frequent control actions can assure any fill rate desired.

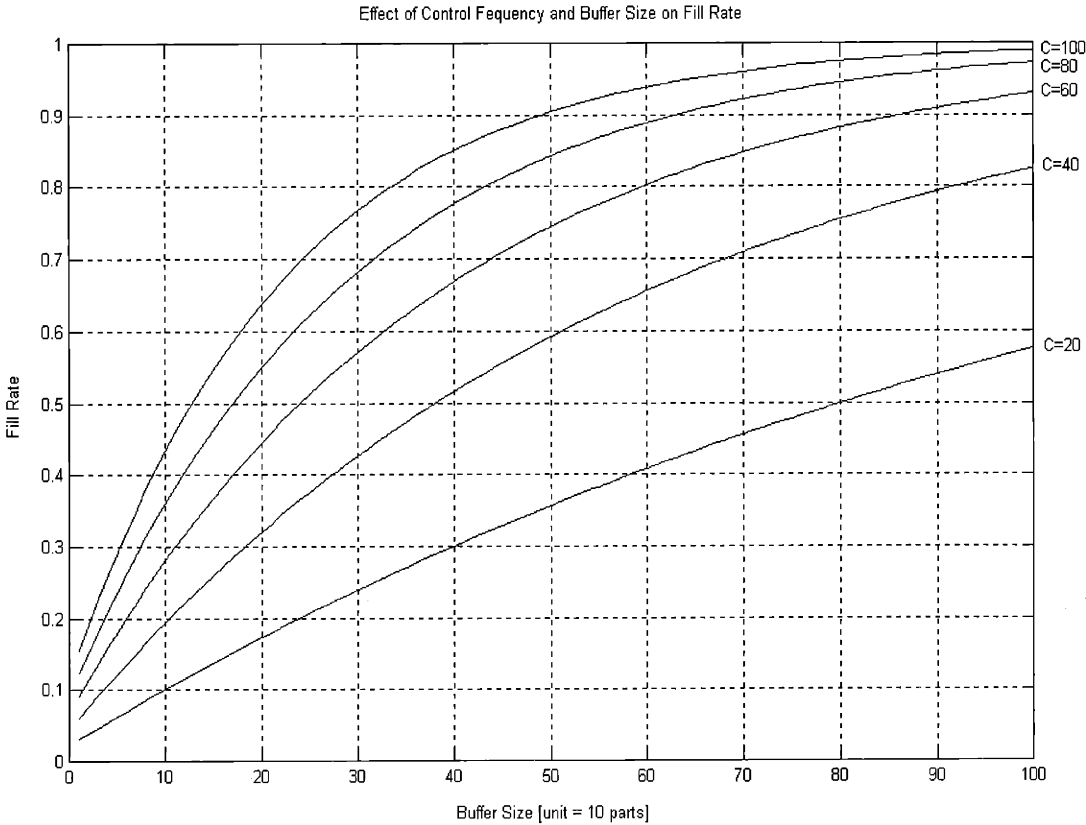
Complementary Effect of Control Frequency and Inventory on System Performance



**Figure 4.1** Complementary effect of control frequency and inventory allocation on system performance.

The relative effect of control frequency and inventory for the system under consideration can be better understood on figure 4.2. The fill rate is presented as a function of buffer size for different values of the control frequency. The parameter  $c$  in figure 4.2 corresponds to the control frequency parameter indicated in figure 4.1. Hence,  $c = 100$  corresponds to a continuous control action and 0 (not included) corresponding to the limit of stability, and all values in between represent locations along a linear scale between the two limits. This figure intends to show that different approaches can be taken for design. For example, if it is desired to ensure a 90% fill rate, the designer could have multiple choices. A continuous control action and a buffer size of 50 could be specified, as indicated in the figure 4.2. Or a sampling period of 150 seconds (corresponding to the control frequency indicated by 60, that is 60% of the range from limit of

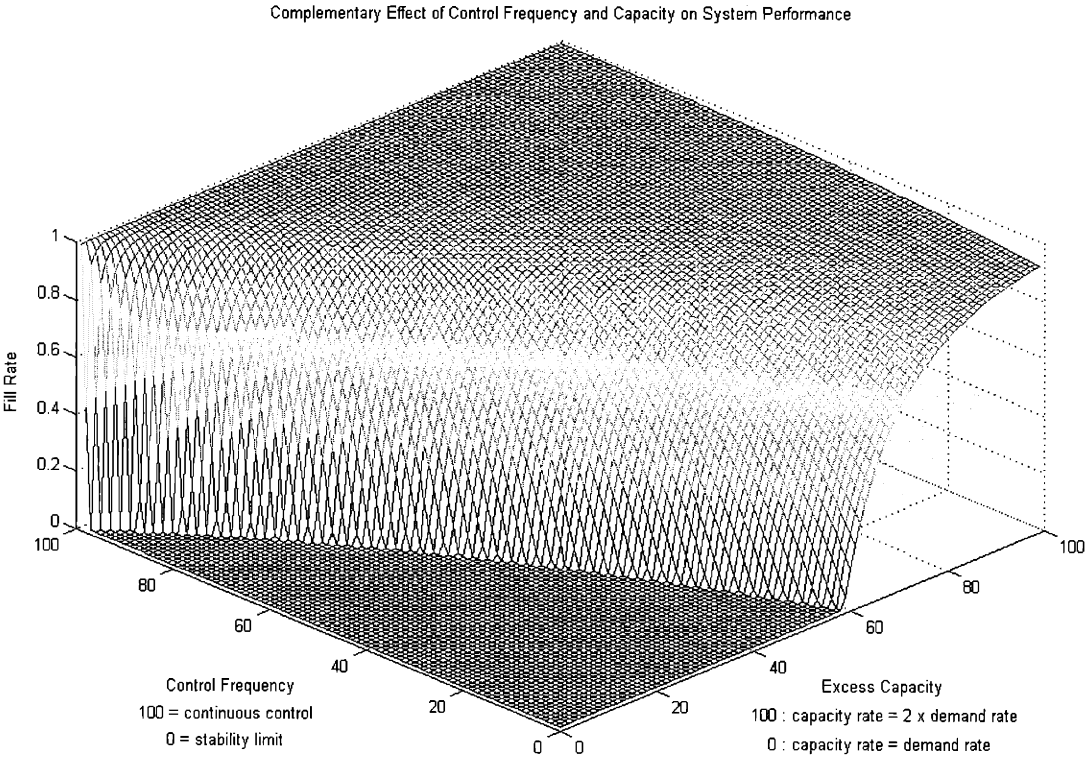
stability to continuous control), and a buffer size of 85 could also be specified to meet the requirement. Given that for the system under consideration the takt time is 50 seconds, the two scenarios considered above correspond to a permanent supervisory action or discrete control actions every three parts. Some other considerations enter into the decision. The designer will have to assess the feasibility and the cost associated with each option, and more importantly, evaluate the system-level effects of each option. A system may incur in high costs in order to implement different production rates frequently, or competitive issues may demand quick and reliable delivery of products, resulting in savings in throughput times and inventory that offset the costs associated with the implementation of real time production rate control.



**Figure 4.2** Effect of inventory on system performance for different control frequencies.

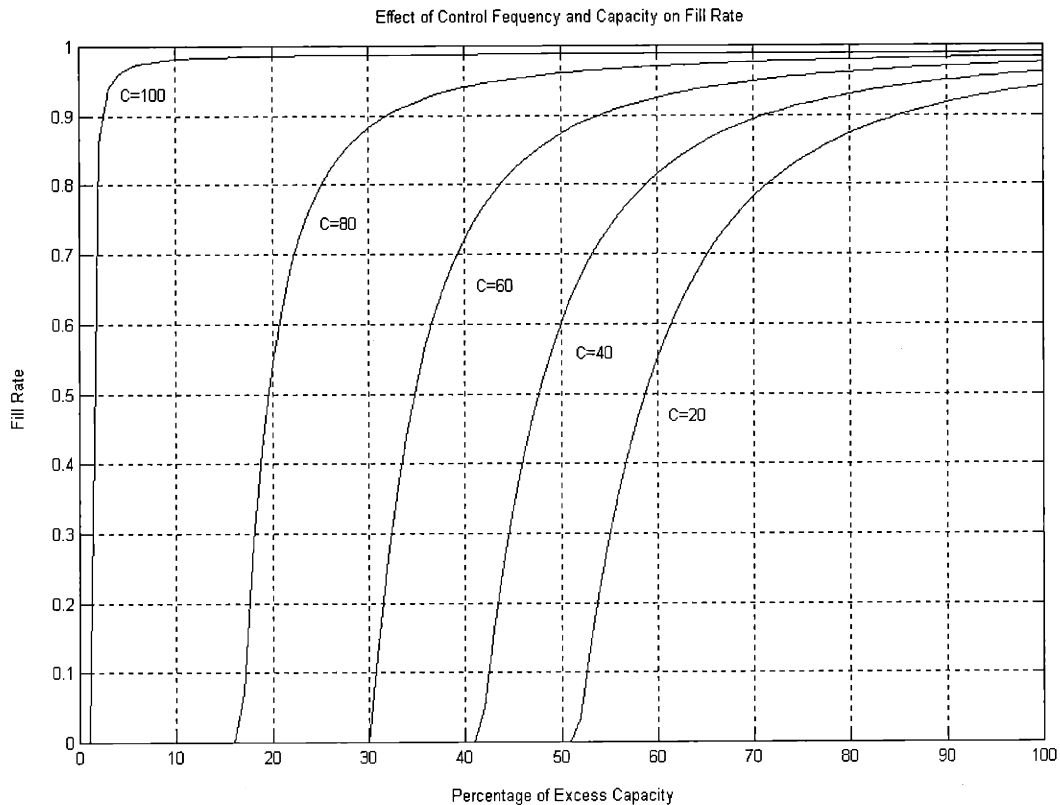


Figure 4.3 demonstrates the effect of control frequency and system capacity on the fill rate of the system. The system parameters are the same as in figures 4.1 and 4.2. The range of values for the control frequency are also the same, it is to say from the limiting frequency corresponding to a control period of 380 seconds to a continuous control action. The other independent variable, system capacity, is plotted along the axis in terms of excess capacity as percentage of the demand requirements. Thus, zero excess capacity corresponds to a system capacity that equals the demand rate. An excess capacity of 100 corresponds to a system capacity twice as large as the demand requirements.



**Figure 4.3** Complementary effect of control frequency and capacity allocation on system performance.

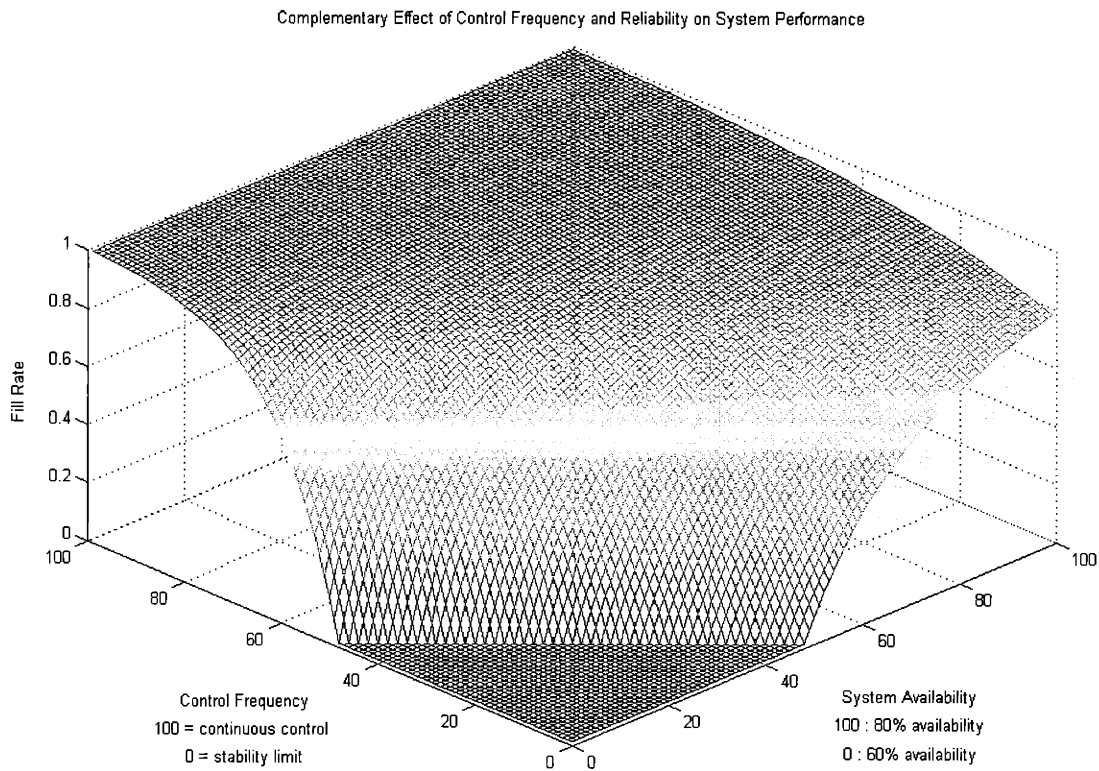
It can be seen in figure 4.3 that there is a region where the fill rate is zero. The curve defining the boundary of such region is in fact a representation of the expression for the limiting frequency, equation 3.41. Actually, only along such curve in the horizontal plane, the fill rate is zero. Everywhere else in the region of the horizontal plane bounded by the limiting function 3.41, the system response is unstable and the production backlog grows without bound.



**Figure 4.4** Effect of system capacity on system performance for different control frequencies.

Figure 4.4 displays the same information as figure 4.3, for five specific values of the control frequency. Again, each curve is designated as in the surface representation, figure 4.3. The main message is intended to be that under some conditions, the effect of control frequency on system behavior could define whether a system can handle the demand requirements with an

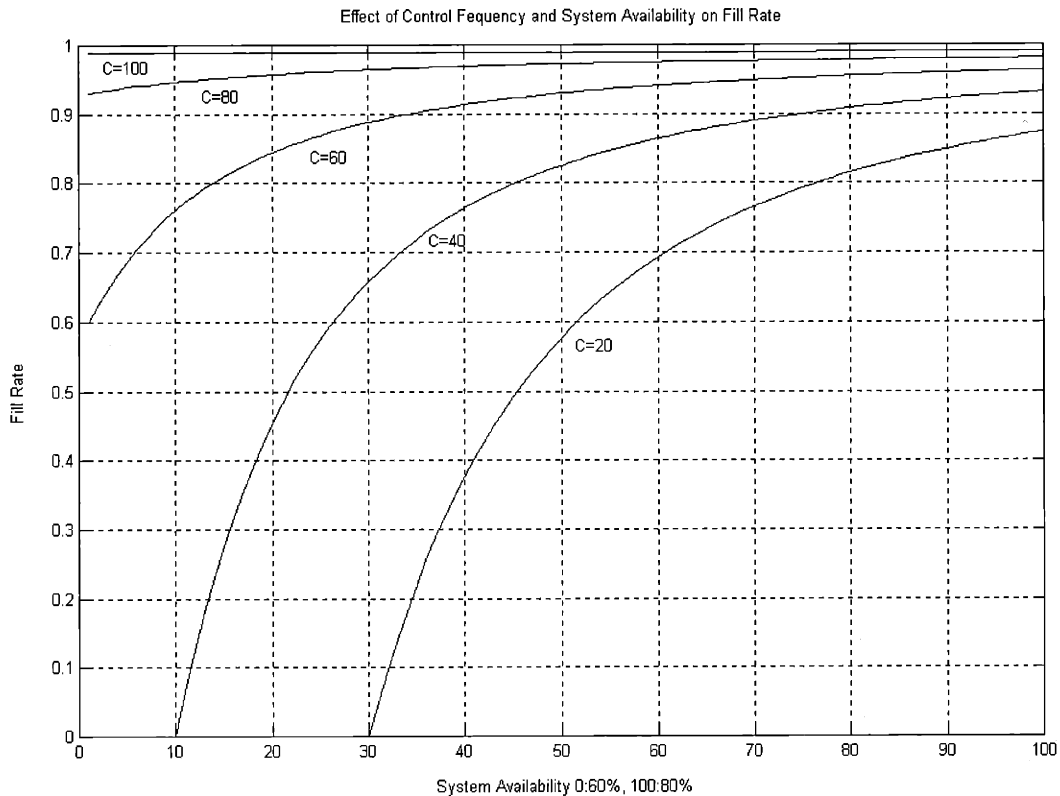
acceptable time behavior. Every system parameter considered for analysis:  $p$ ,  $q$ ,  $R$ ,  $u$ , and  $d$  affects the fill rate function by changing the pronunciation or attenuation of the function but still keeping the same basic shape. On the other hand, the operating conditions and constraints will define the range of feasible values for the system parameters. Then, the region of interest for evaluation of the fill rate function is defined by the system parameter values and the ranges of analysis for the design variables of interest such as: control frequency, system capacity, system availability and inventory allocation.



**Figure 4.5** Complementary effect of control frequency and availability on system performance.

Figure 4.5 indicates the effect of control frequency and system availability on fill rate. The system parameters are the same as in the previous figures. The range of control frequencies considered is the same again. The range of system availability considered is from 60% to 80% availability indicated by a linear scale from 0 to 100. The fill rate increases monotonically with

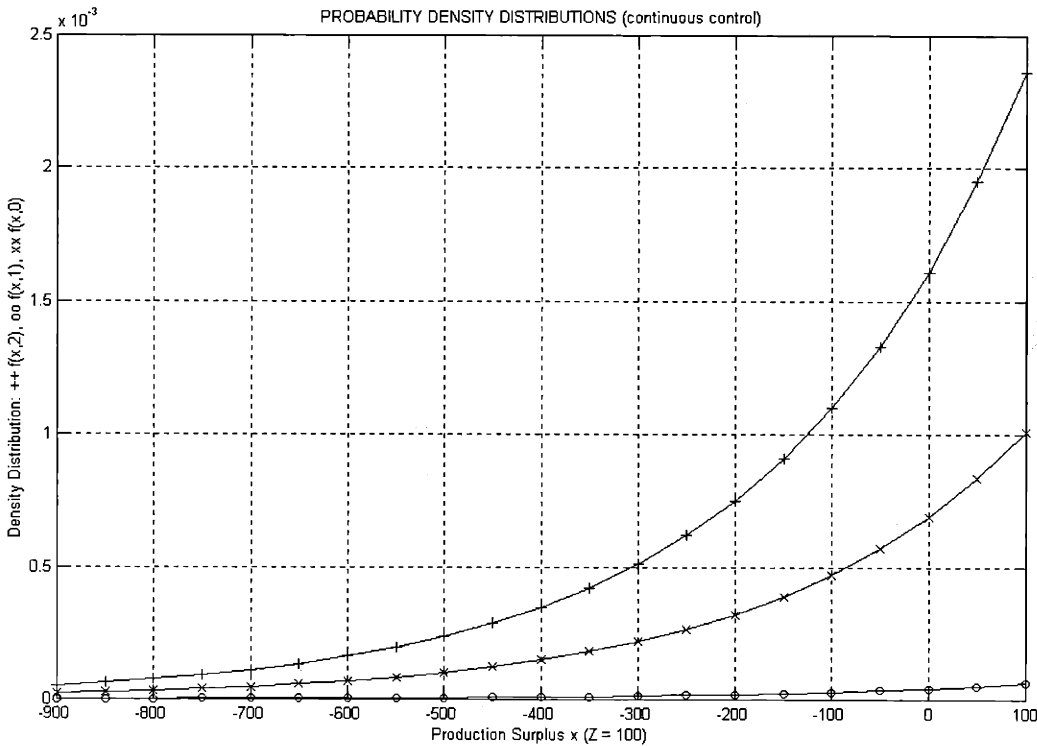
control frequency and system availability. In theory, it is possible to specify a control frequency and a system availability to meet any requirement on fill rate. Again, operating conditions and constraints will define ranges of feasible values for the design variables and the focus of the practice will define a cost allocation that will determine the option of choice. Figure 4.5 indicates that control frequency can determine whether a system is capable of meeting some operation requirements. Again, the area covered on the horizontal plane corresponds to a fill rate of zero, and if the condition specified by equation 3.41 is not met, the response is unstable. This same function is presented in figure 4.6 for specific values of the control frequency.



**Figure 4.6** Effect of system availability on system performance for different control frequencies.

Figure 4.7 indicates the probability density distributions for each of the three system states. The system parameters are the same as in every figure described so far. The buffer size, or space allocated for inventory is 100. Figure 4.7 shows the probability density functions

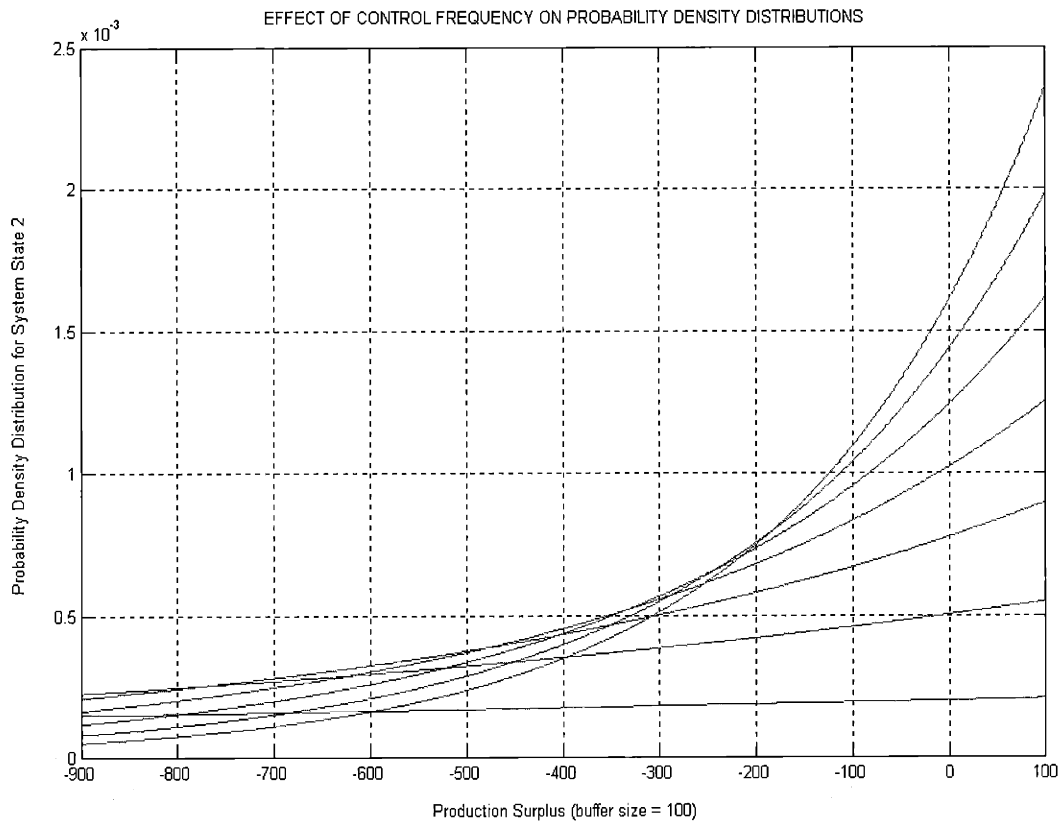
corresponding to a continuous control case. Ideally it is desired to have large values for the probability distributions in the neighborhood of the maximum value for the production surplus, which is the buffer size. This effect is desired so that the probability of finding the system with any operational state and a large negative reduction surplus, corresponding to a large backlog, is small. In fact, the control frequency also has an effect on the system state probabilities that is presented below. In any case, the ideal situation is such that the state probability  $P(Z,1)$  is maximized.



**Figure 4.7** Probability density distributions for continuous control action.

The effect of control frequency on the probability distributions can be seen in figure 4.8. Only the distribution for state two is presented, that is  $f(x,2)$ . The effect on the other probability distributions is similar. As the control frequency is reduced towards the limit of stability, the exponential shape of the distribution related to a continuous control action observed in Figure 4.7

flattens towards a uniform distribution that corresponds to an unstable system response. The higher the control frequency, the more likely it is for the production surplus  $x$  to be found closer to the buffer capacity  $Z$ . This analytical result agrees with our intuition about the effect of control frequency on the dynamic behavior of manufacturing systems, and confirms the reliability of the analytical results.

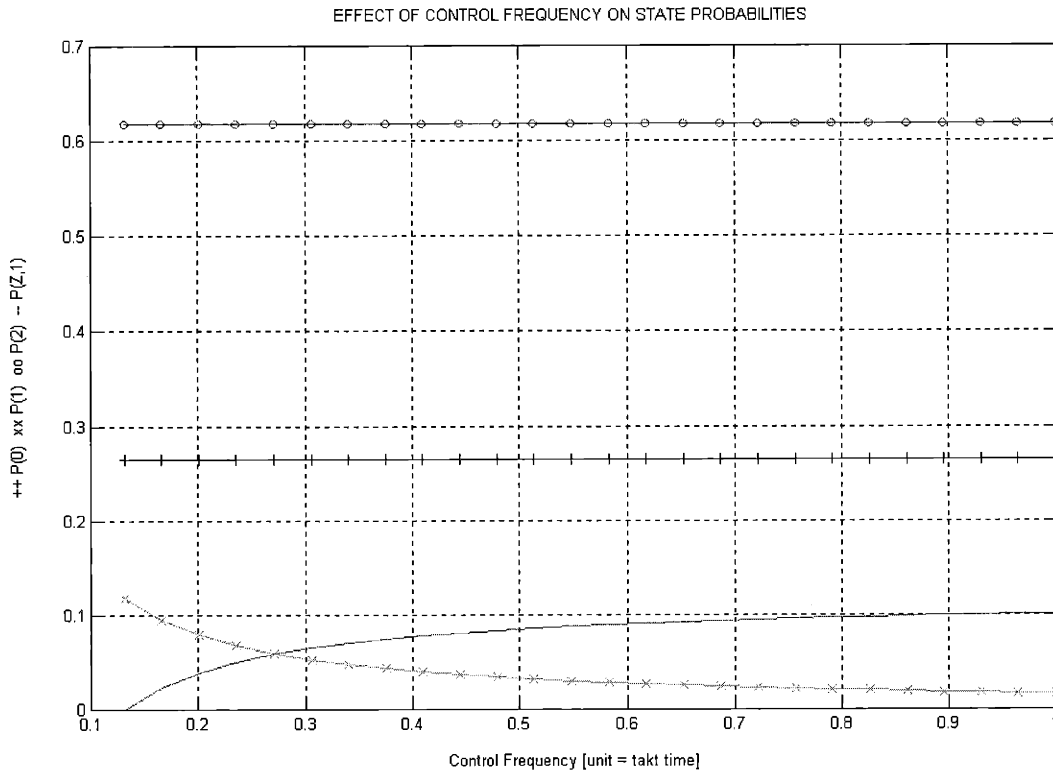


**Figure 4.8** Effect of control frequency on state 2 probability density distribution

Figure 4.9 shows how control frequency influences the state probabilities. This analytical result shows that the state probability  $P(0)$  is not affected by the control frequency. This result is expected, given the assumption that the control frequency does not affect the availability of the

system. The mean time to failure and mean time to repair of any machine or line or cell are commonly regarded as parameters that are independent of the control frequency. This observation helps to confirm the validation of the analytical results.

Figure 4.9 shows that as the control period approaches the limiting condition for stability defined by equation 3.41, the state probability  $P(Z,1)$  approaches zero.  $P(Z,1)$  is maximized when the control action is implemented continuously.  $P(Z,1)$  represents the probability at any time of finding the system operating at the demand rate and with the production surplus at the desired inventory level.  $P(Z,1)$  then, represents the desired operating condition. Figure 4.9 shows also a very interesting phenomenon: the fact that  $P(2)$  is also independent of control frequency. This result indicates that at steady state, the system is operated at capacity (state 2) a fraction of the time that is not influenced by the control frequency. In other words, the system could be controlled at higher frequencies, and hence have better fill rate or on-time delivery, and make use of its full capacity for the same amount of time as when the system is operating poorly under low control frequencies. Then at steady state, the system makes use of the same total amount of capacity independently of the control frequency. By having more frequent control actions, the system is better suited to use its resources when they are needed, and therefore the state probabilities  $P(x,1)$  and  $P(Z,1)$  behave as complementary functions of the control frequency. A confirmation of this behavior results by computing the average production rate of the system. In order to have a stable system response, the steady-state average production rate of the system must equal the demand rate. Given that  $P(0)$  has been proven to be independent of the control frequency, the fraction of time that the system has to be in state 2 has to be also independent of the control frequency and equal to whatever is necessary to compensate for the system down time so that the actual production rate equal the demand rate. This result, once again, agrees with the intuition and assumptions of the model and therefore confirms the validation of the model.

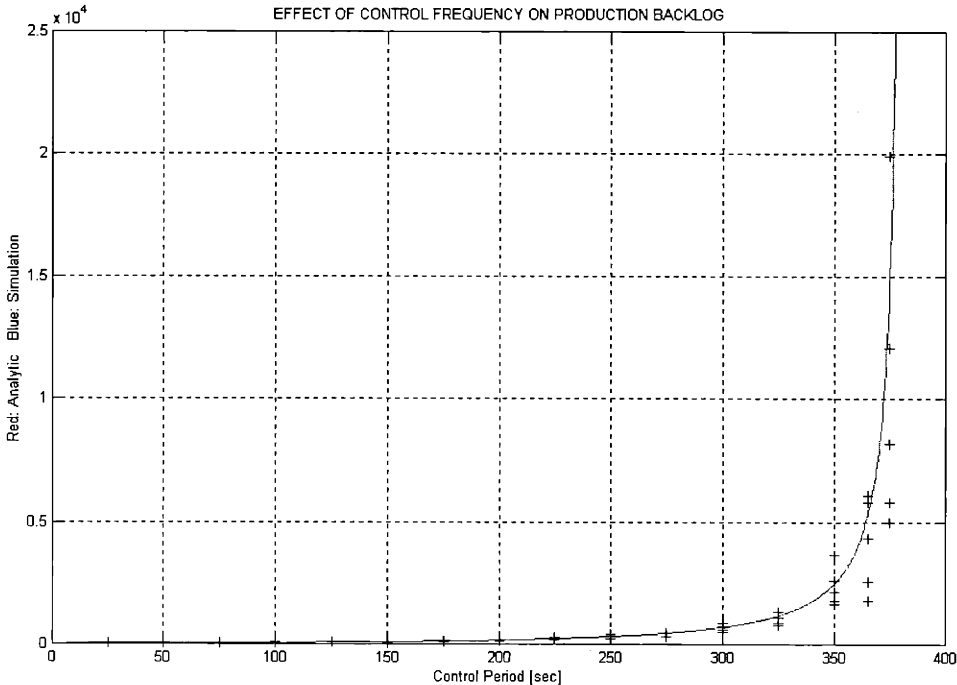


**Figure 4.9** Effect of control frequency on system state probabilities.

Finally, Figure 4.10 shows a comparison of results for the system backlog, a more operational parameter. The average system backlog is presented as predicted by the simulation and the analytical models. This picture shows information whose behavior is basically the same as that of the validation results presented in tables 4.1 to 4.5. It can be seen in Figure 4.10 that the analytic model predicts very well the limit of stability and the production backlog. It is clear that when the system operates close to capacity the effect of variations on its performance becomes stronger. The variation in the predicted average backlog gets larger as the control frequency is reduced. It is concluded that the analytical model predicts exactly the behavior of the system as long as the system is not being operated close to its capacity or stability limit (as determined by equation 3.41). And when close to capacity, the analytical model seems to be still valid, but the large amount of variation in system performance may result in a real system



behavior observation behaving quite different from the prediction. A large amount of observations are necessary in order to be able to draw some conclusion about the behavior of the real system, and then compare against the analytical model predictions. In any case, from the validation tests of the simulation results it can be said with confidence that the model is valid and can be used reliably.



**Figure 4.10** Comparison of analytical and simulation results. Average backlog of single-stage manufacturing system for increasing sampling period.

## 4.5 Generalization of Analytical Model to Multi-Stage Manufacturing Systems

The generalization of the analytical model to the study of control frequency effects on multi-stage manufacturing systems is based on a conversion of discrete control effects into reliability effects.

From equation 3.40

$$b = \frac{cr}{d(p+c)} - \frac{qp}{(\mu-d)(p+c)}, \quad (4.11)$$

Also, from equation 3.35

$$r = R - \frac{dp}{\mu-d} \quad (4.12)$$

Substituting (4.12) into (4.11) results in:

$$b = \frac{cR}{d(p+c)} - \frac{cp+qp}{(p+c)(\mu-d)} \quad (4.13)$$

Equation 4.13 can be rewritten as:

$$b = \frac{c}{c+p} \left[ \frac{R}{d} - \frac{p + \frac{qp}{c}}{\mu-d} \right] \quad (4.14)$$

For convergence requirements of the steady-state solution,  $b$  must be positive. Then:

$$\frac{c}{c+p} \left[ \frac{R}{d} - \frac{p + \frac{qp}{c}}{\mu-d} \right] > 0 \quad (4.15)$$

From physical considerations, the parameters  $c$  and  $p$  are strictly positive, and therefore this condition can be rewritten as:

$$\frac{R}{d} - \frac{p + \frac{qp}{c}}{\mu - d} > 0 \quad (4.16)$$

The expression for the corresponding parameter  $b$  in the case of the continuous control of a similar system, an unreliable, single-stage manufacturing system under a hedging-point control policy [Gershwin, 1994] is:

$$b = \frac{R}{d} - \frac{p}{\mu - d} \quad (4.17)$$

Then it can be seen that if  $c \rightarrow \infty$ , that is, the control period is reduced to zero and the control action is implemented continuously, equation (4.14) in the limit converges to equation (4.17) as expected. Equation (4.14) can be rewritten in the form:

$$b = \frac{R}{d} - \frac{p}{\mu - d} \left[ \frac{d(c + q) + R(\mu - d)}{d(c + p)} \right] \quad (4.18)$$

By comparison of equations (4.17) for the continuous control case and (4.18) for the discrete control case, it can be seen that the last coefficient in equation (4.18) is all that is different between them. It is important to understand the behavior of the magnitude of this coefficient because it can be used to translate the parameters of a discretely controlled system into those of an equivalent continuously controlled system. If a system under discrete control can be translated into an equivalent system under continuous control then the study of control frequency effects covered so far can be extended to the time performance evaluation of multi-stage manufacturing systems. This generalization is possible because the analytical tools already available for the performance evaluation of multi-stage manufacturing systems under continuous control can be used for the evaluation of the time behavior of multi-stage manufacturing systems under discrete control actions. Those tools have been developed with assumptions regarding

demand, supply, starvation and blockage effects for each stage that correspond to the assumptions of the analytical model presented in Chapter 3.

Hence, the equivalent model can be defined by converting the modeling parameters of a system that is being controlled under continuous actions to include the effect of discrete control actions. The coefficient in equation (4.18) that defines the difference between corresponding parameters is proposed as a conversion factor for the failure rate parameter. If the magnitude of such coefficient is greater than one, then it could be concluded that the effect of discrete control actions, for the set of parameters under consideration, is an increment in the failure rate of the equivalent continuously controlled system. Then, the proposed conversion is:

$$p_e = p \left[ \frac{d(c+q) + R(\mu-d)}{d(c+p)} \right] \quad (4.19)$$

Where  $p_e$  is the failure rate of the equivalent system that is controlled under continuous control and whose time performance is the same as that of the original system under discrete control actions and described by the parameters on the right side of equation (4.19). If the hypothesis  $p_e > p$  is established, in accordance to the intuition that the effect of discrete control actions is to increase the failure rate of the equivalent system, then:

$$\left[ \frac{d(c+q) + R(\mu-d)}{d(c+p)} \right] > 0$$

Which can be rewritten as:

$$\frac{R}{d} > \frac{p-q}{\mu-d} \quad (4.20)$$

But from requirements for stability of the continuous control system output, or  $b > 0$  in (4.17), it is already known that:

$$\frac{R}{d} > \frac{p}{\mu-d}$$

Given that  $q > 0$  from physical considerations, inequality (4.20) is therefore true and the hypothesis  $p_e > p$  holds as long as the system parameters ensure that the continuously controlled system is capable of meeting the demand requirements.

## **4.6 Documentation of the Effect of Control Frequency with Real System Parameters**

Parameters from real manufacturing systems have been used as inputs of the analytical model with the purpose of studying the applicability of the research results. Information on system specifications was obtained from the Variable Swash-Plate Compressor Program of the Climate Control Division of Visteon Automotive Systems. Such program involved the development of a system for the yearly production of 1.2 million compressor units for climate control.

The aggregate volume of 1.2 million units per year involved three product families differentiated by the displacement of the compressor unit and about four variations within each family. Variations of the compressors belonging to the same family were defined by different mountings and connecting pipes, due to the different engine and different customer they were intended to meet. Hence, different alternatives were considered involving different number of flow-line systems dedicated to each product family, meeting the demand on product variety with an Assemble to Order (ATO) approach, given that all the internal components of all compressors belonging to the same family are the same.

The takt time of each subsystem depended on the number of parallel lines that was possible to implement, as dictated by the constraints on machine performance and investment limitations. Three parameters, however, were specified as required for any subsystem to be implemented, as follow:

Overall Line Availability  $\geq 95\%$

Quality Rate  $\geq 97\%$

Performance Efficiency  $\geq 92\%$

A common practice for performance evaluation is to take the limiting values and specify that the analysis results refer to a corresponding limiting condition. Hence, if the availability and quality rate parameters are lumped together into the availability parameter of the analytic model, from the definition of availability:

$$\text{Availability} = \frac{R}{R + p} = 0.95 \times 0.97 = 0.9215 \quad (4.21)$$

Performance efficiency is specified for a process in isolation and hence no starvation or blockage effects are considered. A process in isolation represents a single-stage manufacturing system operating with unlimited supply of raw material and unlimited downstream storage. Performance efficiency then represents that fraction of the time a process in isolation is commanded to be in state 2 that such process actually stays in state 2. Inefficiency results from undesired losses of partial capacity that are represented in the analytic model by transitions from state 2 to state 1 at rate  $q$  in Figure 3.4. It is important to realize, however, that efficiency is not the same as  $P(2)$  in the analytic model which is the steady state probability of the system being in state 2. Performance efficiency is specified by Visteon Process Engineering assuming a continuous supervisory action. Then, if  $\frac{1}{q}$  represents the mean time before a partial loss of capacity happens when the system is in state 2, and  $\frac{1}{TT}$  the mean time it takes for the system to be taken back to state 2 after a partial loss of capacity:

$$\text{Efficiency} = \frac{\frac{1}{q}}{\frac{1}{q} + \frac{1}{TT}} = 0.92 \quad (4.22)$$

For equation 4.21 no explicit information is available on Mean Time To Failure ( $\frac{1}{p}$ ) nor Mean Time To Repair ( $\frac{1}{R}$ ) and in any case such information is machine or line specific and dependent instead of a general specification for design (such as the case of availability, quality yield and efficiency).

Then if a MTTF of 5000 seconds is assumed for a takt time of 50 seconds, the modeling parameters result to be:

$$q = \frac{1}{575.0}$$

$$p = \frac{1}{5000.0}$$

$$R = \frac{1}{425.936}$$

The amount of excess capacity needs also to be assumed. From personal communications with the chief process engineer of the project [Jackson, 2001] a typical value for the excess capacity is 30%. Hence, all the required information for the analytical model is available. By substituting into equation 3.41 for the required control frequency for a stable system response:

$$c > \frac{pqd}{R(\mu - d) - dp} \quad (3.41)$$

A maximum control period of  $46.35 \times$  takt time, or 38.62 minutes is obtained. As a specific application of the results of this analysis it is concluded then, that the container capacity should be less than 46 parts in case the material handler samples the state of the system only when material is delivered. For the case when the supervisory and the material delivery tasks are independent, the results indicate that the supervisor should check the system status at intervals of less than 38.62 minutes to ensure that the system response is stable, i.e., the system manages to eventually meet all the demand. As explained above, if cost allocation and poor performance penalty are available, then an optimization problem can be formulated to find the most convenient control frequency for specification.

## 4.7 Chapter Summary

The purpose of this chapter is to gain insight into the influence of control frequency on system time behavior. From the results presented so far it can be seen that the control frequency should be regarded as a system design variable. System capacity, inventory allocation and system availability are commonly taken as system design variables. It is proposed that control frequency, because of its potentially strong effect on system time behavior, should be considered as a basic system design variable and its effect on system performance studied during the design of a manufacturing system.

The results presented in this chapter relate to the detailed analysis of an unreliable, volume flexible, single-stage manufacturing system. The analytic results have been proved to be valid for the evaluation of any manufacturing system that can be represented by the model considered, as indicated in Chapter 3. The main limitation of the current approach is considered to be the restriction on two discrete choices for the production rate. The more choices for the production rate that are available in a real manufacturing system, the more system states that need to be considered in the corresponding analytical model, that is, the more states that need to be included into the Markov chain model. This research intended to cover more than two choices for the production rate, but the difficulty in finding a closed form of the solution for the resulting set of equations of the boundary problem prevented the generation of a meaningful analytical result. While some closed solutions were found, none of them had an appropriate physical interpretation.

This chapter has covered the effect of control frequency on operational parameters such as system backlog and production surplus, the relative effect of control frequency with respect to other system parameters such as capacity, inventory and availability, and the effect of control frequency on system state probabilities and probability distributions. Furthermore, an approach has been proposed to extend these results to the study of multi-stage manufacturing systems. In practice, when production rate is specified as a variable for control of the manufacturing system, only limited, or discrete choices for the production rate are commonly specified. The results presented, then, can be very useful for the evaluation of manufacturing systems that are limited



to two options for the production rate. Other systems, such as assembly lines with manual operations could have a production rate defined by the speed of the conveyor, and at least in theory the production rate can behave as a continuous variable. For such cases, the results of this work are applicable to specific choices for the production rate only, and more points could be analyzed by interpolation of results.



## 5 Conclusions

This dissertation studies the influence of control frequency on the time performance of flow-line manufacturing systems, or manufacturing systems intended for repetitive production of discrete items. The term *control frequency* refers to the rate of occurrence of control events such as system state observation, control law evaluation and compensatory action implementation. During the exploratory stage of the research an unambiguous lower bound on the control frequency that is required for stable response of volume-flexible, unreliable, flow-line manufacturing systems was identified. The intended characterization of such limit of stability motivated the development of an analytical model. An approach based on Markov processes to modeling and evaluation of operational control policies that accurately predicts the steady state of the time performance of single-stage systems is presented. A method for the study of control frequency effects on multi-stage manufacturing systems is proposed.

Not fully resolved items throughout the thesis offer opportunities for further research. These items are listed in this section and some promising research strategies are described. The study has been validated and its results presented covering the case of single-stage manufacturing systems only. Most implementations involve multiple stages of flow manufacturing and hence much more understanding could be developed and the results could have more direct applicability if the study is extended to cover more than one stage. An approach to analyze such systems has been proposed but it could not be tested with simulation. A continuation of this work should begin with the development of a simulation program for multiple stages based on the code listed in the Appendix. Then a specific analytic approach for the study of multi-stage manufacturing systems should be chosen from the multiple options available in the literature and finally the results tested with simulation.

The present work has covered the study of the effect of the frequency of control actions that relate to transitions from a slow state to a fast state. This consideration involves two very

important assumptions that limit the applicability of the results. First, only two production rates have been considered (slow and fast). Second, blockage effects take care of transitions from fast to slow. While it seems very reasonable to limit the choices for the production rate to two options in the specification of a policy for the operation of a production system, it may be of interest to consider more than two options for the production rate. The problem becomes more specific, however, in the sense that decision rules need to be stated on what rate to choose as a function of the system-state, as opposed to the simple and well-known hedging point control policy. The analysis of such case involves more states than those presented in Figure 3.4 and the resulting problem becomes more difficult to solve. Further research could take this direction.

Removing the assumption of control by blockage, on the other hand, would allow the study of the influence of control frequency on average inventory levels as well as inventory size requirements. Removing this assumption, however, requires a very different research strategy. The development of an analytical model involving a token-based control may be the best approach to take.

During the development of the analytic model the strong effect of the operational policy under consideration became evident. The system that has been modeled in this work is one that after the completion of a repair, is set back to the same state it had right before the failure. The analytic model considers in the steady state that same condition. If this single assumption is changed, however, all the results of the model change. For example, if an operational policy involving a transition to state 2 immediately after the completion of every repair is in place then there is in fact a continuous control action for the system under consideration and therefore the results should agree with those in the literature for such case. It is not as simple, however, as to take the results of the model considered and substitute  $r = 0$ , which would result in a contradictory conditions, given that those results are based on the assumption that  $r$  is greater than zero. In any case, different policies can be studied by substituting an according value for the corresponding parameter depending on what the condition is. In any case, however, the policy studied is considered to be the most commonly found in real implementation of production rate control, i.e., systems where the control and the repair actions are independent.

Finally, it is concluded that the effect of control frequency on the behavior of the system has been found to be of a significance that compares to that of much more common design variables as system capacity, inventory allocation and system availability. Control frequency, however, is frequently not considered as part of the specifications of a final system design. The results of this research show that control frequency should be regarded as a design variable for production system design. These results readily allow production system designers to specify better system resources and controls.



# References

- Akella, R. and Kumar, P. R. (1986). Optimal Control of Production Rate in a Failure Prone Manufacturing System. *IEEE Transactions on Automatic Control, AC-31* (2) 116-126.
- Bai, S. X. and Gershwin, S. B. (1994). Scheduling Manufacturing Systems with Work-In-Process Inventory Control: Multiple-Part-Type Systems. *International Journal of Production Research* 32 (2) 365-385.
- Berkley, B. J. (1992). A Review of the Kanban Production Control Research Literature. *Production and Operations Management* 1 (4) 393-411.
- Bielecki, T. and Kumar, P. R. (1988). Optimality of Zero-Inventory Policies for Unreliable Manufacturing Systems. *Operations Research* 36 (4) 532-541.
- Blanchard, B. S., Fabrycky, W. J. (1998). System Engineering and Analysis, Upper Saddle River, N. J. Prentice Hall, Third Edition
- Bonvik, A. M. (1996). *Performance Analysis of Manufacturing Systems Under Hybrid Control Policies*. Ph.D. Thesis, Massachusetts Institute of Technology.
- Burman, M. H. (1995). *New Results in Flow Line Analysis*. Ph.D. Thesis, Massachusetts Institute of Technology.
- Buzacott, J. A., and Shantikumar J. G. (1992). A general approach for coordinating production in multiple-cell manufacturing systems. *Production and Operations Management* 1 (1), 34-52.
- Buzacott, J. A., and Shantikumar J. G. (1993). *Stochastic Models of Manufacturing Systems*. Englewood Cliffs, New Jersey: Prentice Hall.
- Cochran, D. S. and Charles, M. (1997). Design of Manufacturing Systems to Support Volume Flexibility. *Proceedings of the International Conference on Engineering Design ICED 97*.
- DeVor, R. E., Chang, T., Sutherland, J. W. (1992). Statistical Quality Design and Control, Contemporary Concepts and Methods. Prentice Hall.
- Fine, C. H. (1998). *Clokspeed*, Perseus.
- Gershwin, S. B. (1994). *Manufacturing Systems Engineering*. PTR Prentice Hall.
- Gershwin, S. B. (1997). Design and Operation of Manufacturing Systems – Control and System-Theoretical Models and Issues. *Proceedings of the American Control Conference, v 3 IEEE*. ISSN 0743-1619.

Glasserman, P. (1995). Hedging-Point Production Control with Multiple Failure Modes. *IEEE Transactions on Automatic Control* 40 (4) 707-712.

Goldman S. L., Nagel R. H. and Preiss K. (1995). *Agile Competitors and Virtual Organizations*. Van Norstrand Reinhold.

Hayes, R. H., Abernathy, W. J. (1980). Managing Our Way to Economic Decline. *Harvard Business Review*, July-August.

Hayes, R. H., Garvin, D. A. (1982). Managing as if Tomorrow Mattered. *Harvard Business Review*, May-June.

Hill, T. (1994). *Manufacturing Strategy, Text and Cases*. Second Edition. Irwin McGraw-Hill.

Hopp, W. J., and Spearman, M. L. (1996). *Factory Physics: Foundations of Manufacturing Management*, Irwin McGraw-Hill.

Hu, J. Q., Vakili, P. and Yu, G. X. (1994). Optimality of Hedging Point Policies in the Production Control of Failure Prone Manufacturing Systems. *IEEE Transactions on Automatic Control* 39 (9) 1875-1880.

Hu, J. Q. and Xiang, D. (1994). Structural Properties of Optimal Production Controllers in Failure-Prone Manufacturing Systems. *IEEE Transactions on Automatic Control* 39 (3) 640-643.

Jackson, K. (2001) Personal communication. Compressor SBU, Climate Control Systems – Visteon Automotive Systems.

Kaplan, R. S. (1984). Yesterday's Accounting Undermines Production. *Harvard Business Review*, July-August.

Kimemia, J. G. and Gershwin, S. B. (1993). An Algorithm for the Computer Control of Production in a Flexible Manufacturing System. *IIE Transactions* 15 (4) 353-362.

Kleinrock, L. (1975). *Queuing Systems, Volume I: Theory*. New York, John Wiley and Sons.

Liberopoulos, G. and Caramanis, M. (1994). Production Control of Manufacturing Systems with Production Rate-Dependent Failure Rates. *IEEE Transactions on Automatic Control* 39 (4) 889-895.

Linck, J. (2001). *A Decomposition Based Approach for Manufacturing System Design*. Ph.D. Thesis, Massachusetts Institute of Technology.

Little, J. D. C. (1961). A Proof for the Queuing Formula:  $L = \lambda W$ . *Operations Research, Volume 9*. Number 3, 383-387.



- McKay, K. N. (2001). The evolution of Manufacturing Control – What has Been, What Will Be. Working Paper 03-2001.
- Monden, Y. (1993). *Toyota Production System, An Integrated Approach to Just-In-Time*. Second Edition. Engineering and Management Press.
- Montgomery, D.C. (1985). *Introduction to Statistical Quality Control*. John Wiley and Sons.
- Phadke, M. S. (1989). *Quality Engineering Using Robust Design*, Prentice Hall.
- PSD Laboratory (2000). *The Manufacturing System Design Decomposition*, Internal Document, Version 5.1.
- Rother, M., Shook, J. (1998) *Learning to See*. The Lean Enterprise Institute, Brookline, MA.
- Ryzin, G. V., Lou, S. X. C. and Gershwin, S. B. (1993). Production Control for a Tandem Two-Machine System. *IIE Transactions* 25 (5) 5-20.
- Sharifnia, A. (1988). Production Control of a Manufacturing System with Multiple Machine States. *IEEE Transactions on Automatic Control* 33 (7) 620-625.
- Shiba, S., Graham, A., Walden, D. (1993). *A New American TQM: Four Practical Revolutions in Management*. Productivity Press.
- Spearman, M. L. (1997). On the theory of constraints and the goal system. *Production and Operations Management* 6 (1), 28-33.
- Srivatsan, N. (1993). *Synthesis of Optimal Policies for Stochastic Manufacturing Systems*. Ph.D. Thesis, Massachusetts Institute of Technology.
- Srivatsan, N. and Dallery, Y. (1998). Partial Characterization of Optimal Hedging Point Policies in Unreliable Two-Part-Type Manufacturing Systems. *Operations Research* 46 (1) 36-45.
- Suh, N. P., Cochran, D. S., Lima, P. L. (1998). Manufacturing System Design. 48<sup>th</sup> General Assembly of CIRP, *Annals of the CIRP, Vol. 47/2*, 627-639.
- Taguchi, G. (1989). *Quality Engineering in Production Systems*, McGraw-Hill.



# Appendix

## Simulation Code

```
clear;

u=1.0/35.0;           % TIME UNIT IS SECONDS
d=1.0/50.0;           % capacity [parts/time]
Z=500.0;              % demand rate [parts/time]
time=0.0;             % inventory allocation [parts]
simtime=1000000000;  % current time
transition=1;         % simulation length
                    % transition identificator

cummsurplus=0.0;     % integral of surplus over time
cummbacklog=0.0;     % integral of backlog over time
actsurplus=Z;        % current surplus level
timeP0=0.0;          % cummulative time on state (x,0)
timePZ=0.0;          % cummulative time on state (Z,1)
timeP1x=0.0;         % cummulative time on state (x,1)
timeP2x=0.0;         % cummulative time on state (x,2)
timep=g05fbf(5000,1); % failure rate
timeq=g05fbf(2000,1); % partial failure rate
timeR=g05fbf(1800,1); % repair rate
timec=g05fbf(365,1); % control rate

while time < simtime,
    switch transition
    case 1
        cummsurplus=cummsurplus+(Z*timep);
        timePZ=timePZ+timep;
        time=time+timep;
        timeR=g05fbf(1800,1);
        transition=2;
    case 2
        newactsurplus=actsurplus-(d*timeR);

cummsurplus=cummsurplus+(0.5*(actsurplus+newactsurplus)*timeR);
    if newactsurplus < 0
        if actsurplus < 0

cummbacklog=cummbacklog+(0.5*(actsurplus+newactsurplus)*timeR);
        else
            timepos=(actsurplus/d);
```

```

        cummbacklog=cummbacklog+(0.5*newactsurplus*(timeR-
timepos));
    end
    end
    actsurplus=newactsurplus;
    timeP0=timeP0+timeR;
    time=time+timeR;
    timep=g05fbf(5000,1);
    timec=g05fbf(365,1);
    if timep < timec
        transition=3;
    else
        transition=4;
    end
    end
case 3
    cummsurplus=cummsurplus+(actsurplus*timep);
    if actsurplus < 0
        cummbacklog=cummbacklog+(actsurplus*timep);
    end
    end
    timeP1x=timeP1x+timep;
    time=time+timep;
    timeR=g05fbf(1800,1);
    transition=2;
case 4
    cummsurplus=cummsurplus+(actsurplus*timec);
    if actsurplus < 0
        cummbacklog=cummbacklog+(actsurplus*timec);
    end
    end
    timeP1x=timeP1x+timec;
    time=time+timec;
    timep=timep-timec;
    timeq=g05fbf(2000,1);
    timeZ=(Z-actsurplus)/(u-d);
    if timeZ < timeq
        if timeZ < timep
            transition=7;
        else
            transition=6;
        end
    else
        if timeq < timep
            transition=5;
        else
            transition=6;
        end
    end
    end
case 5

```

```

newactsurplus=actsurplus+((u-d)*timeq);
cummsurplus=cummsurplus+(0.5*(actsurplus+newactsurplus)*timeq);
  if actsurplus < 0
    if newactsurplus < 0
cummbacklog=cummbacklog+(0.5*(actsurplus+newactsurplus)*timeq);
      else
        timeneg=-1.0*actsurplus/(u-d);
        cummbacklog=cummbacklog+(0.5*actsurplus*timeneg);
      end
    end
    actsurplus=newactsurplus;
    timeP2x=timeP2x+timeq;
    time=time+timeq;
    timep=timep-timeq;
    timec=g05fbf(365,1);
    if timep < timec
      transition=3;
    else
      transition=4;
    end
  case 6
    newactsurplus=actsurplus+((u-d)*timep);
cummsurplus=cummsurplus+(0.5*(actsurplus+newactsurplus)*timep);
    if actsurplus < 0
      if newactsurplus < 0
cummbacklog=cummbacklog+(0.5*(actsurplus+newactsurplus)*timep);
        else
          timeneg=-1.0*actsurplus/(u-d);
          cummbacklog=cummbacklog+(0.5*actsurplus*timeneg);
        end
      end
      actsurplus=newactsurplus;
      timeP2x=timeP2x+timep;
      time=time+timep;
      timeR=g05fbf(1800,1);
      transition=8;
    case 7
      cummsurplus=cummsurplus+(0.5*(actsurplus+Z)*timeZ);
      if actsurplus < 0
        timeneg=-1.0*actsurplus/(u-d);
        cummbacklog=cummbacklog+(0.5*actsurplus*timeneg);
      end
      actsurplus=Z;

```

```

    timeP2x=timeP2x+timeZ;
    time=time+timeZ;
    timep=timep-timeZ;
    transition=1;
case 8
    newactsurplus=actsurplus-(d*timeR);

cummsurplus=cummsurplus+(0.5*(actsurplus+newactsurplus)*timeR);
    if newactsurplus < 0
        if actsurplus < 0

cummbacklog=cummbacklog+(0.5*(actsurplus+newactsurplus)*timeR);
        else
            timepos=(actsurplus/d);
            cummbacklog=cummbacklog+(0.5*newactsurplus*(timeR-
timepos));
        end
    end
    actsurplus=newactsurplus;
    timeP0=timeP0+timeR;
    time=time+timeR;
    timep=g05fbf(5000,1);
    timeq=g05fbf(2000,1);
    timeZ=(Z-actsurplus)/(u-d);
    if timeZ < timeq
        if timeZ < timep
            transition=7;
        else
            transition=6;
        end
    else
        if timeq < timep
            transition=5;
        else
            transition=6;
        end
    end
end
end
end

P0=timeP0/simtime
PZ=timePZ/simtime
P1x=timeP1x/simtime
P2x=timeP2x/simtime
Ptot=P0+PZ+P1x+P2x;
avesurplus=cummsurplus/simtime
avebacklog=-1.0*cummbacklog/simtime

```