STOCHASTIC MODELING AND CONTROL OF SOME PROBLEMS IN MANUFACTURING

by

Charles James Conrad

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Computer, Information and Control Engineering)
in The University of Michigan
1986

Doctoral Committee:

Professor N. Harris McClamroch, Chairman
Professor Yoram Koren
Professor Arch W. Naylor
Assistant Professor Demosthenis Teneketzis
Associate Professor A. Galip Ulsoy

ABSTRACT

STOCHASTIC MODELING AND CONTROL OF SOME PROBLEMS IN MANUFACTURING

by

Charles James Conrad

Chairman: N. Harris McClamroch

A control theoretic approach to some problems in manufacturing is examined. This approach is motivated by the recognition of the potential for enhancing system performance through the use of information feedback and on-line decision making. The problem of applying the control theoretic approach to manufacturing is discussed.

A class of stochastic processes, called diffusion-threshold processes, is developed. Diffusion-threshold processes can be used to describe discrete event phenomena generated by underlying continuous processes.

Four manufacturing problems are considered in which diffusion-threshold models are featured. The first problem considers the control of a machining process, specifically a simple drilling operation, with respect to economic criteria. Policies for feedrate (feed speed) selection and tool replacement are to be determined. The problem addresses real issues that are often ignored, such as uncertainty, tool failure, scrap production and the discrete nature of parts. A diffusion-threshold model for tool wear/tool life is utilized. Two types of policies are considered: age replacement policies in which the feedrate is constant and tool replacement is based upon age; and feedback policies in which tool wear measurements are occasionally available, the feedrate is allowed to vary, and tool replacement is based upon the wear measurements. Potential performance improvement (economic and otherwise) resulting from the additional information and control freedom present in the feedback policies is investigated. Cost per time and cost per part measures are developed for the age replacement case. A class of cost functionals, called one step costs, is introduced for the feedback case. The optimal policies for both cases are described. Comparative results are presented for a special case of interest, where tool life is assumed to obey a Taylor formula in the mean.

The results from the first problem are used in the development of two additional problems. The second problem considers a machine with multiple tools. Age replacement and feedback policies are analyzed. The third problem considers decentralized control in serial transfer lines. Simulation results for a two machine example are presented.

The last problem uses a diffusion-threshold process to model work progress on a job. A supervisor is allowed to make occasional inspections of the job and re-assign resources. The scheduling of the inspections and the resource allocation are to be determined. An optimal control formulation of the problem results in a quasi-variational inequality.

The thesis offers new viewpoints of some contemporary manufacturing problems and demonstrates how new mathematical tools can be used to approach these problems.

For family,
friends,
teachers,
co-workers;

Each played no small role.

ACKNOWLEDGEMENTS

One cannot hope to summarize in a few words the contributions made by so many over the years that mark the evolution from student to graduate. To those I neglect, please accept my apologies.

My sincere thanks go to my advisor, Prof. Harris McClamroch, for the guidance he has given me as I wandered along the murky path that a student must take. Many thanks go to the members of my committee for the suggestions, criticisms, and insights they have offered. I am also appreciative of the assistance provided by the Robot Systems Division of the Center for Research on Integrated Manufacturing, and by the faculty and staff of CICE.

I am forever grateful to Mictron, Inc. and Lamb Technicon, Inc. for the support they have given me that has made this possible. A very special thank you goes to Fran Leo for his unwavering encouragement and confidence in me.

Much credit is deserved by my fellow students and especially my officemates for the contributions made through countless enlightening discussions. To each of them I say thank you.

Finally, many thanks go to my family and friends for their support and understanding, and in particular to my wife, Wendy, who had to endure it all.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

CHAPTER 1

INTRODUCTION

A greater awareness of the importance of manufacturing, brought on by increasing global competition and the impact of a deteriorating manufacturing base, has resulted in a surge of activity in manufacturing research and development. Those new to the field are discovering what veteran workers and researchers have long known: manufacturing crosses many discipline boundaries and is rich in challenging problems.

It might be expected that in any area as diverse as manufacturing there will be pockets of intense research activity and great unexplored problem frontiers. This is the case in manufacturing today. Although incremental improvements in the existing manufacturing knowledge base are important, the future probably lies in new approaches and methods.

The intent of this work is to present one approach to some problems in manufacturing. The problems themselves are minuscule within the context of manufacturing. The approach, though, is believed to be far-reaching and thus of more general value to manufacturing. This approach is called control theoretic, since it draws heavily from control theory. More precisely, the approach recognizes the importance of information feedback in decision making. The concept of the control theoretic approach is to borrow from control theory the framework for approaching the problem of utilizing information feedback in on-line control, and applying it to problems in manufacturing.

The motivation for wanting to utilize information feedback and on-line decision making in manufacturing comes from the success of control theory in other areas. That is, the effective application of the ideas can potentially lead to improved performance and/or more desirable system characteristics. However, a concomitant caution must also be understood, because deterioration of performance is also possible. The effective use of information in decision making is a difficult problem.

The control theoretic approach is beneficial in another way. The explicit allowance of information feedback and on-line control induces new views of old problems. This is particularly evident in the problems considered in this work. New viewpoints do not always lead to improved performance, but they usually contribute to increased understanding of the problems. This by itself is beneficial and useful. It is hoped that this work makes some contribution to the understanding of a class of manufacturing problems.

1.1. Overview of the Research

The presentation continues in Chapter 2 with an expository chapter on manufacturing and the control theoretic approach. The concept of intelligent manufacturing is introduced. In an intelligent manufacturing system, the capability to acquire and process data in a timely way is assumed. This leads into the main theme of the work: How can information feedback be effectively used to improve the performance of manufacturing systems? This question is very general, but provides the motivation for most of this work. The question is explored through the examination of several example manufacturing problems.

The control theoretic approach is one methodology that can be used to address the main theme. The approach is not without difficulty, however. A discussion of problem issues confronting the control theoretic approach is presented. One particular source of difficulty is the modeling of manufacturing systems; more specifically, the development of models of manufacturing systems that capture the important features and maintain a compatibility with the control theoretic framework is difficult.

Since modeling plays a key role in the control theoretic approach, a summary of the important features and characteristics of manufacturing systems is presented. The list is not exhaustive, but does include features not always captured in manufacturing models. These features include complexity, uncertainty, and available control mechanisms. The presence of both discrete event and continuous time phenomena in manufacturing systems is also discussed.

A very brief survey of analysis methods available for manufacturing system study is given, accompanied by a comparison and critique of these methods.

The chapter concludes with a discussion of an important sub-theme that transcends manufacturing: discrete events that arise from underlying continuous time processes. This view of discrete event systems is important because it allows the introduction of information feedback and control concepts into discrete event systems in a very natural way.

Chapter 3 introduces four sample manufacturing problems. The first problem is called the drilling problem and is the most deeply studied problem in the thesis. Fundamentally, it is a machining economics problem, but placed in an intelligent manufacturing context. The drilling problem captures many important manufacturing features in a conceptually simple problem. The objective of the drilling problem is to determine machining parameters and make tool replacement decisions so as to achieve good system performance.

The second problem is an extension to the drilling problem where multiple tools are considered. The third problem poses the question of how to control networks of machines where each machine is similar to that considered in the drilling problem. This problem is extremely challenging. Only a very simple example is considered in this work, resulting in what might best be called speculation. The fourth problem is called the supervisor's problem and is a radical departure from the first three problems. The supervisor's problem is a type of scheduling problem, but viewed in a very different way. In particular, an attempt is made to more accurately capture the dynamics of job progress and to recognize the role played by supervisory personnel in the manufacturing environment. The problem also serves as a reminder that people are an important part of manufacturing systems.

Chapter 4 develops much of the mathematical framework that will be used to approach the posed problems. A class of stochastic processes called diffusion-threshold processes is introduced. The key features of these processes are a controlled diffusion process and a threshold boundary in the state space. When the diffusion first reaches the boundary, a significant event is considered to have occurred. This generally results in the re-initialization of the diffusion process and a new cycle of process evolution.

The goal in studying diffusion-threshold processes is to derive a probabilistic description of the threshold hitting time in terms of the process parameters and the control policy used. The control policy may include diffusion measurement feedback.

The treatment of general diffusion-threshold processes is difficult and not considered in this work. Attention is instead restricted to simpler classes of processes where solutions are obtainable. Included are the cases of constant infinitesimal coefficients; coefficients constant between diffusion measurement feedback; and generalizations to multidimensional diffusions with independent components.

The treatment of the piecewise constant coefficient case is considered in Appendix B. This development demonstrates the difficulty encountered when handling more general diffusion processes. The result for the piecewise constant coefficient case yields a partial solution to a Brownian motion curve crossing problem. This problem has appeared in the statistics literature.

Application of diffusion-threshold processes to several categories of problems is discussed. Furthermore, diffusion-threshold processes play a central role in all the example problems discussed. An important feature of diffusion-threshold processes is that they combine discrete event and continuous time phenomena into a single process. This allows a different viewpoint of discrete event systems to be formed. The importance of this is believed to extend far beyond the specific problems considered in this work.

Chapter 5 considers the drilling problem in detail. An important issue in the analysis of the drilling problem is the modeling of tool wear and tool life in a production environment. A brief history of this problem is presented. The model of tool wear and tool life must be compatible with the control theoretic approach in order to apply it to the drilling problem. A diffusion-threshold model is proposed which meets the requirements.

The introduction of the diffusion-threshold model allows the drilling problem to be placed in a stochastic control setting. Two classes of control policies are considered. First age replacement policies are examined. Age replacement policies assume fixed machining parameters and predetermined ages at which tools are replaced. These are the traditional policies assumed in machining economics problems. Procedures for the evaluation of two different performance measures under age replacement policies are derived.

Next, feedback policies are considered in which tool wear measurements are occasionally available. The analysis and evaluation of general feedback policies for the drilling problem is very difficult. This motivates the development of a class of cost functionals called one step. One step cost functionals assess the costs and profits associated with producing the next part. They also enjoy the distinct advantage of being computable.

The optimal policy for one step cost functionals is described. Implementation issues are discussed along with practical considerations that affect the control problem formulation. A limitation of the one step approach is that the long term performance of the manufacturing system cannot be easily evaluated in analytical terms.

Chapter 6 addresses the problem of performance evaluation for one step policies, and the comparison of age replacement and feedback based policies. A specific problem is formulated based on drilling data and realistic economic assumptions. Age replacement analysis is carried out, and the optimal age replacement policies under two different criteria are determined. Next, one step costs are evaluated and the optimal one step policies are determined.

Performance assessment is done by simulation. Two different production situations are simulated with varying part worth. Age replacement and feedback policies are implemented based on the optimization results. Variations of these policies are also simulated. Economic performance and secondary performance characteristics are compared.

Also given in this chapter are various algorithms used in the computation of the cost functionals and in the generation of random numbers for the simulations. The computational algorithms offer a significant speedup over numerical integration techniques.

Chapter 7 discusses the other example manufacturing problems introduced in Chapter 3. The first two problems rely on the results obtained for the drilling problem. The multi-tool machine problem is shown to be a generally straightforward extension of the drilling problem. Results are obtained for age replacement and one step policies.

The multi-machine problem is considerably more difficult. Attention is restricted to serial transfer lines with synchronous part transfer. A form of decentralized control based on local costs is proposed as one approach to the problem. Simulation is carried out for a two machine problem under decentralized age replacement and feedback policies. The performance of these policies is compared.

A heuristic dynamic programming type of approach to the supervisor's problem is considered. The result is the derivation of a quasi-variational inequality that the optimal policy must satisfy. This result is similar to results obtained by other researchers working on different problems. However, the supervisor's problem has unique features. This development suggests a different approach to the analysis of systems described by diffusion-threshold processes.

CHAPTER 2

A CONTROL THEORETIC APPROACH TO MANUFACTURING

2.1. Introduction

Manufacturing presents to the researcher and practitioner an enormous variety of problems, covering diverse fields of knowledge and a spectrum of complexity. The scope of manufacturing is so broad that the development of an all encompassing, unified approach to manufacturing problems is quite unlikely. Nonetheless, there are significant classes of manufacturing problems where sufficient commonality is present such that some general approaches might be developed. This work will only consider discrete part manufacturing. The word batch will also be used to describe this type of manufacturing, with the understanding that the size of the batch can be arbitrarily large.

Batch manufacturing is not a particularly restrictive classification; in general, the same breadth of variety and scope is still present. Consequently, batch manufacturing includes a diversity of products, technologies, techniques, machines, and skills. However, large segments of the batch manufacturing community are facing similar problems. Some of these problems will be considered here.

2.2. Problems in Manufacturing

Perhaps the most pervasive problem today in batch manufacturing is the need to increase productivity. Productivity can be considered one important measure of manufacturing performance. Unfortunately, there is no universal definition of this measure, but it will be understood to subsume the production of a product in an economic and timely way, while achieving some level of acceptable quality. A summary of the preceding statement reveals three fundamental problems in batch manufacturing: cost, availability, and quality. The relative contribution of each of these factors and the presence of other factors in the productivity measure is specific to the particular manufacturing environment.

Manufacturing problems can go far beyond operational issues. Such problems as product design and facility planning are clearly related to manufacturing. In order to try to further reduce the scope of manufacturing problems addressed here, only operational issues will be considered. Productivity will be considered a measure of operational performance.

There are other operational concerns that may be factors in productivity. These concerns include: utilization of resources consumed in producing a product; utilization of production equipment; rate of production; speed of response to changes in the product; speed of response to changes in demand; capability of producing a variety of products; choice of production process; amount of work in process; inventory size; and required labor to produce a product. The intent here is only to provide a sampling of the types of concerns faced in manufacturing.

An important motivation for research in manufacturing and in particular for this work is to address the problem of productivity improvement through improvement in one or more of the factors that may affect productivity. The question, then, is how to improve the productivity of a manufacturing system.

One of the major deterrents to successfully answering the above question, and thus another important manufacturing issue, is a general lack of a scientific knowledge base in manufacturing. The "science of manufacturing" has not yet evolved. The result of this lack of fundamental knowledge is the proliferation of ad hoc techniques for dealing with manufacturing problems. It comes as no surprise that much of manufacturing is essentially trial and error, drawing heavily on experience and experiment. Without analytic tools and techniques to offer guidance, solution synthesis becomes difficult and problem insight may require long observation, if it is ever attained.

The intention of this research is to find ways to improve the operational productivity of manufacturing systems through the use of analytic models and methods. The specific use of analytic methods will hopefully yield insight into the problems, result in solution synthesis techniques, and contribute to a scientific knowledge base for manufacturing.

2.3. Intelligent Manufacturing

Although manufacturing suffers from the previously indicated problems, it has by no means stagnated. Advances in supporting technologies have had a tremendous impact on all aspects of manufacturing. Of particular importance are the great advances made in computer and communication technologies. These technological achievements have helped to make the use of such devices as robots and CNC machines economically feasible in many manufacturing environments. Plant floor communication networks are becoming a reality. The computer related technologies are not alone in contributing to manufacturing. Major advances have also occurred in process technologies, material technologies, material handling, sensor technologies, and even in the basic structure of the manufacturing system: Flexible Machining and Flexible Assembly Systems (FMS/FAS).

The advances in the supporting technologies have allowed for new approaches to the productivity problem. That is, a re-thinking of how manufacturing is done is now both appropriate and necessary. Since most of manufacturing has evolved by trial and error, the present solutions are specific to the manufacturing environments of the past. Changes in the environment dictate investigation of alternate structures and methods for manufacturing. However, this does not imply that the existing manufacturing structures are wrong and must be replaced by new structures. As an example, the assumption that all new machining systems must be flexible is unfounded and limiting. Underlying this type of assumption is the belief that traditional manufacturing systems have fully evolved. This belief has no basis. Consideration should be given as to how advancing technologies can be used in all types of manufacturing. An additional benefit of this consideration is the discovery that much of existing manufacturing is not really well understood.

It should be clear that prior assumptions about the structure of advanced manufacturing systems will be avoided in this work. Instead, a more general and more powerful concept of manufacturing systems will be introduced. This concept will be called **intelligent manufacturing**. The concept of intelligent manufacturing is quite simple. Assume that there exists sufficient data collection, communication, processing, and storage facilities so that plant floor data can be used to aid in operational decision making. Furthermore, assume that the manipulation of the data can be done in a timely way. Given this data collection and processing structure, how can the performance of the manufacturing system be improved. Intelligent manufacturing assumes the advanced technological support without making assumptions about the particular manufacturing structure.

Given the availability of the requisite technology, the concept of intelligent manufacturing has a certain intuitive appeal. This appeal arises from the belief that an increase in the amount and quality of data available for decision making, and an increase in the speed at which it can be processed should result in better decisions, better performance, and greater productivity. This reasoning suffers from one flaw. The ability to efficiently handle data does not imply a knowledge of how to effectively use it. It is precisely a deficiency in the knowledge of how to use plant floor data effectively that has impeded greater advancement in intelligent manufacturing.

The central theme of this research is conveyed by the following question:

> *How can the feedback of information be used to improve the performance of the manufacturing system?*

The question presumes an intelligent manufacturing setting and the capability of influencing the system through on-line decision making. The methodology for approaching this question will be called control theoretic. Before explaining this methodology further, some motivation for the control theoretic approach will be given.

Fundamental to control theory is the concept of feedback. Feedback of information is used in the determination of what actions should be taken so as to influence the behavior of a system in a desirable way. The similarity between this point of view and the central theme motivates the control theoretic approach to manufacturing systems.

2.4. The Control Theoretic Approach

The control theoretic approach has both advantages and disadvantages for approaching manufacturing problems. These relative merits and demerits are reveled by describing what constitutes the control theoretic view.

As previously stated, fundamental to the control theoretic view is modification of system behavior. This modification can be accomplished by utilizing feedback of information regarding the condition of the system and exercising on-line strategies to influence the system behavior. The control theoretic view presumes the existence of means for exercising this influence. Clearly, the available means represent a limiting factor in influencing system behavior. Identification of the available means is an important modeling issue. The control theoretic view also recognizes some important issues regarding system performance modification. The first is that the effective use of information feedback is not trivial. The second issue is that information feedback can be detrimental as well as beneficial. Finally, the magnitude of performance improvement that can be achieved is limited.

The control theoretic view represents an analytic orientation. The advantage of this orientation is the potential for gaining insight into classes of mathematical systems that can be used to model many physical systems. From this insight, analysis and synthesis methods for constructing and determining the performance of controllers can be derived. However, the control theoretic view still recognizes that the performance of the physical system is what actually matters. The analytic orientation helps to reduce considerably the amount of trial and error required in order to attain the desired system performance.

Unfortunately, there are some difficulties that arise in attempting to apply the control theoretic approach to manufacturing. The most important of these is mathematical incompatibility. The most powerful control theoretic results assume mathematical structures (finite dimensional, linear differential/difference equations) that are not readily applicable to most manufacturing problems. This means that either we accept poor system models that are familiar and fit the classical framework, or search for new models that are more appropriate but for which new analytic methods must be developed. The latter approach will be emphasized in this work.

The problem of constructing compatible mathematical models for manufacturing systems is compounded by the relative lack of analytic models in manufacturing. Furthermore, those that do exist are generally not adequate for use with the control theoretic approach. Clearly, a major endeavor in using the control theoretic approach will be model construction.

One of the main difficulties in applying control theory to manufacturing problems is that from some viewpoints, manufacturing systems appear as collections of discrete events. The problem here is that a branch of control theory dealing with discrete event systems has not yet evolved (except for certain special cases). See [Ho1] for some recent approaches to the discrete event system problem. Also see [Br1] and [Cr1] for the special cases of controlled point processes and controlled queues. This fundamental inadequacy contributes greatly to the difficulty in applying control theory not only to manufacturing, but to many other classes of problems with similar attributes. One of the goals of this work is to present some ways of approaching certain types of systems that can be viewed as discrete event systems. Implicit in this effort is the belief that this has value beyond specific manufacturing problems.

Before attempting to construct suitable models of manufacturing systems, it is worthwhile reviewing some of the properties and characteristics of manufacturing. The following review is certainly not exhaustive, but it does present some features that are important. Some of these features are not often discussed in the literature.

The reference [Gel] also presents a control perspective on manufacturing. There are many similarities but also some important differences between the viewpoint presented here and the one presented in that reference. See also [Wh1].

2.5. Some Properties and Characteristics of Manufacturing Systems

Based on the previous discussion, it is apparent that modeling is a major issue if the control theoretic approach is to be used in manufacturing. The goals of this modeling effort are to characterize important features of manufacturing systems, identify the requirements of the control theoretic view, and to develop models that incorporate these features. In this section, the first two goals will be considered. The importance or relevancy of features will of course vary with the system being modeled. Nonetheless, the following provides a list of features shared by many manufacturing systems.

2.5.1. Complexity

Manufacturing systems, when viewed in whole or even in part, can be enormously complex. A manufacturing system can easily involve many facilities located internationally and employing (hundreds of) thousands of people. Furthermore, these facilities may in turn depend on thousands of other suppliers and their respective facilities. Decision making may involve time intervals from sub-second to decades.

The processes and technologies involved can be extremely diverse. Even when the view in restricted to a single facility, the complexity of the system can be overwhelming. Hundreds of machines and thousands of people can be producing thousands of products.

From a modeling standpoint, there are at least two approaches to the complexity problem. The first approach is to model the system using extensive aggregation and simplification. The second approach is to decompose the manufacturing system into simpler subsystems that are easier to model.

2.5.2. Decomposition

Fortunately, most manufacturing systems can be decomposed into simpler subsystems that have less complexity. In general, several levels of decomposition are possible. As an example, consider a manufacturing system which has a machining line as a subsystem. The machining line can be decomposed into sets of related processes or operations; these can be further decomposed into distinct operations or workstations; these individual operations can be decomposed into sequences of steps that are necessary to do the operation. This example is only representative; subsystems involving human workers can be similarly decomposed. This decomposition of manufacturing systems is based on levels of detail. It is not a unique approach. Another approach is decomposition based on time scales ([Ge1]). In this work, levels in the manufacturing system will refer to levels of detail and breadth of scope unless otherwise stated.

The impact of decomposition on modeling must be considered. Although decomposition makes the modeling problem more tractable, it can also lead to certain views of the system that are incomplete or misleading. That is, depending upon the level of decomposition chosen, there can be considerable aggregation, simplification, approximation, and omission of higher or lower level features. It is the author's opinion that this has resulted in the evolution of two disparate model classes for many types of manufacturing systems. More will be said about this later.

2.5.3. Uncertainty

Manufacturing and manufacturing systems cannot be described in a completely deterministic way. Elements of uncertainty enter into all aspects and at all levels of the system. Some examples of uncertainty will help demonstrate its pervasiveness. At the operation level uncertainty manifests itself in the form of machine reliability and production quality. At higher levels, uncertainty is present in the aggregated performance of a facility and in the quality and availability of raw materials. At even higher levels, uncertainty is present in future product demand, capital availability, and economic conditions.

Uncertainty is always present in the manufacturing system, but it is not necessarily independent of other factors. That is, uncertainty may affect the operation of a manufacturing system, but the operation decisions made and the actions taken can also affect the nature and magnitude of the uncertainty that arises. Consideration should be given to this potential coupling during the modeling effort.

A very important form of uncertainty in manufacturing is production quality. Each part may deviate somewhat from what is intended. These deviations may be inconsequentially small, or so large that the part is unacceptable (scrap or rework required). There are many sources of these deviations: uncertainty in the raw parts, the machines, damage in handling, etc. The production of unacceptable parts is a manufacturing reality. These issues should be considered during model construction.

2.5.4. Deterministic Influences

The presence of elements of uncertainty in the manufacturing system does not imply that the system is purely random. A manufacturing system is actually governed by an order or discipline imposed on it, and by natural physical principles. Deterministic influences are involved at all levels of the system. Examples of these influences at lower levels include selection of machining parameters, maintenance schedules for equipment, and the sequencing of processing steps. At even lower levels, the process evolution is presumed to obey physical principles, though they may not be well understood. At higher levels, influences include production rate decisions and activity scheduling. Long term planning decisions can influence the evolution of the system over long time horizons.

The inclusion of deterministic influences in manufacturing system models is particularly important when the control theoretic approach is used. This approach is naturally interested in identifying the means of influencing system behavior, and the effects of those influences.

2.5.5. Coupling

Decomposition of the manufacturing system for the purposes of modeling does not in general produce a set of independent subsystems. As stated earlier, decomposition is generally only a simplification of the view of the system. Coupling generally exists between the subsystems, regardless of how the decomposition is carried out. Of course, the magnitude of coupling is quite variable, and in many cases nothing is lost in ignoring it. Manufacturing requires cooperation among the components and subsystems, so many elements of the system are interdependent. As an example, when a particular operation is isolated for the purposes of modeling, it is important to remember that there are few stand alone operations in manufacturing and that the actual performance of the operation in a manufacturing context may be very different from its performance in isolation. The isolated characteristics may not be an adequate description of the operation's behavior. Generalizing, coupling may affect the performance of any subsystem.

2.5.6. Dynamical System

Manufacturing systems are dynamical systems, not static. Their temporal evolution depends upon physical laws, deterministic and uncertain influences, and interaction with other systems. This is a key feature that provides a logical link between manufacturing and control theory.

Whether or not the dynamics of the manufacturing system need to be modeled depends upon the level of the modeling viewpoint, the time scales of the features being modeled, and the time horizon for which the model is intended to hold. In some cases, static assumptions are reasonable approximations and are simpler to handle. However, static models fail to capture the true character of the system and thus are necessarily incomplete. Models that capture the dynamical character will in general be more complicated, though potentially more accurate. The dynamical character of the system to be modeled should be carefully considered so that the dynamics of the system with time scales relevant to the model intention are captured.

2.5.7. Control Theoretic Compatibility

Manufacturing models that are to be used in conjunction with the control theoretic approach require a degree of compatibility with the basic assumptions of the approach. Therefore, the model needs to include explicit control mechanisms that can be used to influence the system's evolution. Furthermore, in order to conform with the central theme of information feedback, those entities that can be measured or otherwise assessed also need to be identified.

The identification of the control mechanisms is not sufficient. Also required are direct and indirect consequences of the manipulation of these mechanisms. This includes deterministic influences as well as influences on the system uncertainty. The model should also specify limits or constraints imposed on the influences, and limitations on the available measurements. Measurement limitations include both quality and availability.

An example of a control input common to many manufacturing systems, but often overlooked in models, is the production rate or more generally the rate at which work is done. However, in order to use it effectively as a control input, the direct and indirect effects of production rate need to be understood and modeled. This particular control input will be used in this work.

Control inputs can also be of a discrete nature. Some examples of this type include part routing decisions, pass/fail inspection, and on/off actuator inputs.

Some control mechanisms are best modeled as discrete while others are best described as continuous. Since a typical manufacturing system has both types of control mechanisms the model may have to include both continuous and discrete valued control inputs.

2.5.8. Discrete Event vs. Continuous Process Viewpoint

A popular modeling view of manufacturing systems is to model only the occurrence of specific events that are of some significance. This is the discrete event representation of manufacturing systems. This modeling approach has the advantage of simplifying the representation of the manufacturing system and easing the burden of using the model in analysis, simulation, or design. [Ge1] has several examples of this viewpoint. This viewpoint has also influenced manufacturing simulation language design. See [Pe1] as an example.

The discrete event viewpoint has the drawback of neglecting the underlying dynamics (typically continuous processes) that give rise to the discrete events. Consequently, many discrete event models neglect entirely the additional structure present in the actual system. This neglect can result in missed potential for system control and information feedback that may actually be available.

Similarly, most models of continuous time processes fail to represent the occurrence of significant events (at discrete times) that impact system evolution. A possible explanation for this is that most instances of models of manufacturing systems where continuous processes are used are at the lowest levels (e.g., process dynamics) where the discrete event view is not considered important. Similarly, the discrete event view can be regarded as a way to ignore process details that higher level views are not concerned with.

An important modeling consideration, and one that will surface repeatedly in this work, is the possibility of hybrid models that capture important features of both continuous process and discrete event viewpoints. The advantage to be gained by this approach will include control compatibility while retaining the impact and importance of significant events.

2.6. A Brief Survey of Models of Manufacturing System

The intention of this section is to provide a brief survey of classes of models of manufacturing systems, and to compare and critique these models with respect to the features previously described as desirable. Refer to [Gel] for a more complete bibliography and also to [Su2].

Prior to surveying the models, some remarks are in order concerning the use of models of manufacturing systems. In particular, there can be a distinction between models that are used for analysis and those that can be used for synthesis. Analysis refers to analysis of system behavior. Synthesis refers to producing candidate system designs. Systems design includes both structural issues (e.g., number and type of machines) and operational issues. Models that can be used for synthesis, and which include the previously specified features, are desirable but scarce. Much more common are analysis models, with synthesis performed by trial and error based upon the analysis results.

The admittance by a manufacturing system of a decomposition based on levels of detail has been previously discussed. Detail based decomposition induces a hierarchy of views of the manufacturing system, ordered by the breadth of the view and the detail considered. The models discussed here can be considered as reflections of the view of the system considered important by the modeler. For descriptive convenience, a high-level view will correspond to breadth but lack of detail, and a low-level view will represent detail but lack of breadth. The models discussed here are restricted to system operational models.

The first class of models considered is the queueing models. These models rely on the network of queues formulation. In these models, the manufacturing system is composed of servers (work stations), queues (buffers), and routes (work paths through the system). These models have become popular for modeling flexible manufacturing systems. Queueing models seldom contain details of the processes involved, allowing for generality. As such, these models have a relatively high view among models of factory floor operations. See [Bu1], [Cr1], [Kl1], [So1], and [Wh1]. The reference [Cr1] is a bibliography of queueing research.

Analytic results for queueing models are only proven under strong assumptions about service times and input processes. Recently there has been some work suggesting that these models are robust to deviations from these assumptions, indicating a more general usefulness ([Su1]). Variations of queueing analysis allow for determination of certain performance characteristics (e.g., expected queue length) quite rapidly.

Several deficiencies of queueing models are evident. Probabilistic part routing and infinite queue capacities are often unrealistic assumptions. The probabilistic service time assumptions (exponential distributions) are also unrealistic for many systems. As an example, in machining the process times can be variable, but are certainly not exponentially distributed. Expanding on this criticism, queueing models are in some sense "too stochastic", and sacrifice lower level detail or knowledge that may be of use. When analytic tractability is sacrificed, representational accuracy may be improved, but the model is difficult to use for meaningful analysis of the system. A major drawback to the use of existing queueing models with the control theoretic approach is the lack of explicit control mechanisms, and identification of information that may be available for feedback. In fact, these models assume that the control policies are somehow "encoded" into the probabilistic descriptions and queue disciplines. This can be quite awkward to deal with. Queueing models do provide a dynamical representation, however.

The next class of models is the Markov chain or Markov process models. In both models the system is decomposed into machines and buffers. Each machine can assume one of two states: operational or under repair. The state of the system refers to the current inventory in the buffers and the state of each of the machines. In the discrete time version, production is usually viewed discretely, with parts produced at some multiple of the time quantum unit (usually one) ([Ge3]). In the continuous time model, production is usually modeled by a flow approximation ([Ge2], [Ki1]). The transition of a machine from operational to under repair and vice versa is assumed to be a Markov process, and thus is exponentially or geometrically distributed. These models can allow for machine blockage and starvation.

Closed form solutions are only available under very special circumstances. However, some approximation methods are known that can be more generally applied. These models do incorporate uncertainty through the failure/repair transitions. The binary state model is somewhat limiting, though, since it does not allow for machine deterioration due to aging. This comment applies to all models that assume exponential or geometric failure distributions. It is possible to incorporate production rate into some of these models. The failure mechanism is assumed to be independent of the production rate. These models do allow for some control theoretic approaches, since system state is assumed to be known to the controller ([Ge2], [Ki1]).

These models do incorporate many of the discussed features. Some assumptions are made that may not hold for many manufacturing systems, unfortunately. Some synthesis methods are available for some of these models. Typically, these are based on approximations.

Other classes of network models have been proposed as models of manufacturing systems. Most prevalent are the Petri net models, including the many variations and extensions of them ([Du1]). [Kam1] discusses various Petri net type models. These models offer a graphical representation of concurrent, asynchronous processes, which are properties of manufacturing systems. The variations that have been proposed include timing information, classes of activities, and stochastic phenomena. Some of the properties and results that are known about Petri nets can be applied to manufacturing systems. These include liveness, safeness and boundedness. Petri net models have also been used as the foundation for simulation models and for the design of control logic. The Petri net viewpoint is decidedly discrete event. Concepts of state and control pertain to the components of the Petri net. Lower level details are generally omitted.

The next category of models involves the specific modeling of individual processes. The focus here is at a lower level where process details become important but manufacturing aspects are often ignored. Detailed models of process behavior are constructed from physical principles or experimentation. Control mechanisms often enter naturally into these models. Generally these models are close to those types of models common in traditional control engineering: lumped parameter differential equations. References exist for many different manufacturing operations. Some specific references are [Da1], [Kan1], and [Ko1].

Drawbacks to these types of models include viewpoint (often ignoring manufacturing concerns) and neglect of uncertainty. This latter feature could be included but typically is not. The absence of uncertainty tends to result in the modeling of nominal behavior only, ignoring the abnormal behavior that occurs in the manufacturing environment. The significance of certain discrete events is usually ignored in these models.

Models that respect some process characteristics while including more manufacturing oriented concerns are also available. Examples of these models are found in machining economics problems with multi-stage machining systems. See [Er1], [Ha1], [Hi1], [Iw1], [Ke1], [Le1], [Le2], [Ph1], [Shk1], [Tak1], and [Tay1]. In these models, some low level details concerning tool wear, feedrate selection, and cutting constraints are used along with some high-level details such as profit and production rates. These models are used in conjunction with some optimization criteria to determine the (static) selection of certain parameters (e.g., feed speeds).

These models are typically deterministic and static. Without a dynamical system viewpoint, many of the control theoretic concepts are not readily applied. In some formulations of theses models, uncertainty is included via probabilistic descriptions of the process variables and constraints. This does not change the static character, however. The intermediate view taken by these models is notable, as it includes a mixture of levels of detail.

The next class of models is simulation models. Actually, simulation is best described as a technique for evaluation. However, since simulation does require that models for evaluation be constructed, it will be considered here. This is a varied class, and can take on essentially any viewpoint of the system desired. There is no explicit limit on the amount of detail that can be captured (although there are usually practical limits). Arbitrary degrees of complexity can be treated, but not without cost. As more detail is included, more effort is required in preparation, run time, and analysis. Also, more detail increases the chance of error. Simulation can include all of the desirable features mentioned, though manufacturing simulations usually don't. Simulation is the single most common method of analyzing manufacturing system performance (other than experimentation with the actual system). See [La1] as a general simulation reference and [Pe1] for a perspective on manufacturing simulation from a simulation language supplier.

The generality of simulation makes it difficult to critically evaluate without specifying a particular problem. Usually, simulation is only used for evaluation though some techniques have recently appeared that can aid in synthesis ([Ho1], [Ho2]). Simulation can represent a great deal of effort, and does not always result in useful insight into the behavior of the manufacturing system.

As a side comment, it is worth noting that there exists some dichotomy in the simulation community between discrete event and continuous model simulations. Although many simulation languages support both types of systems, it is perhaps not coincidental that the discrete event view prevails in manufacturing simulation ([Pe1]). Again, this can lead to some restriction in the modeling approaches considered.

With the exception of simulation (which will be excluded from the following discussion), none of the models discussed really captures all the desirable features and characteristics previously discussed. Clearly, each has some advantages and disadvantages. By extracting the appropriate features of each model type, it may be possible to develop models that satisfy the requirements. However, changes to each of these models can represent additional complexity not easily dealt with. Seemingly, modeling of manufacturing systems is still an open field.

2.7. Discrete Event Processes Arising From Underlying Processes

Discrete event processes play a prominent role in many manufacturing system models. This prominent role can have the unfortunate consequence of obscuring the nature of discrete event processes. It is typically the case that discrete events are the result of underlying processes that give rise to these events. As an example, consider the machining of a part. The process of metal removal, which is occurring continuously in time gives rise to the discrete event of part completion.

The discrete event viewpoint presents some compatibility problems with the control theoretic approach. First, the discrete event view can disguise the presence of underlying control mechanisms. Thus, available control inputs can be overlooked. Second, the potential utility of information feedback can be inhibited by the discrete event view, since the use of information concerning the underlying process is ignored.

The explicit introduction of underlying processes giving rise to the discrete event provides a convenient link between the discrete event view and the control theoretic approach. Presuming that the underlying process has control mechanisms and measurable components, the utility of the control theoretic approach can be explored. It should be expected, however, that the use of underlying processes to describe discrete events, particularly stochastic events, may not be simple. This will be considered in some detail in this work, with particular attention to some discrete event processes arising in manufacturing. However, the author believes there is considerable potential for the application of this approach in areas outside of manufacturing.

CHAPTER 3

DESCRIPTION OF SOME MANUFACTURING PROBLEMS

3.1. Introduction

In this chapter, several prototype manufacturing problems will be introduced and discussed. The problems are at once similar and dissimilar to problems typically discussed in manufacturing. These problems have been constructed with several intentions in mind. First, each problem embodies some important features or aspects of manufacturing. Second, the problems are not of the "black box" variety; rather, they represent realistic situations. This has been done in order to aid familiarization and problem visualization. That is, certain features are more clearly evident in a manufacturing setting that might otherwise be obscured in a purely mathematical setting. Third, the problems are set in an intelligent manufacturing context, with features that admit a control theoretic approach. These features are designed to be realistic or at least reasonable. However, the presence of these features results in a departure from typical related problem formulations. This will motivate a comparison between the traditional and the intelligent manufacturing problem formulations, and also help clarify the concept of the control theoretic approach. Fourth, each of the problems contains certain phenomena that can be modeled using a class of stochastic processes central to this work. This modeling relationship will be explored in some detail for one of the prototype problems.

The presentation of these prototype problems is not meant to imply that they embody all the important manufacturing problems, nor that they are the only problems to which the concepts developed in this work can be applied. However, each of the problems does have some importance in manufacturing. More importantly, each of the problems offers a viewpoint different than that which is usually presented. The author wishes to stress the importance of new viewpoints for manufacturing problems that more accurately convey the actual system or overcome limitations in past approaches. The primary objective in presenting these problems is to gain additional insight into some characteristics of manufacturing systems; insight that may ultimately yield productivity improvement.

Finally, each of the problems provides a vehicle for exploring the control theoretic approach to manufacturing. A framework will be developed, with particular emphasis on one of the problems, that demonstrates the application of the control theoretic approach. Again, the intent is to provide examples, but the framework is not believed to be limited to the problems considered here.

3.2. Problem 1: The Drilling Problem

The drilling problem represents a simple version of a problem found in metal-removal manufacturing operations, but with some new extensions. The heart of the problem is the question of how to operate a machining system in the best way subject to the limitations of uncertain physical phenomena. In this simplified version, only a single machine carrying out a single operation will be considered. It is important not to divorce the problem entirely from the surrounding manufacturing complex, however. That is, the problem formulation should represent at once aspects of manufacturing and aspects of drilling.

Assume there is a metal drilling manufacturing operation. The machine is a single tool drill, repetitively carrying out a single identical task on parts as they arrive. The task involves the drilling of a single hole in each part. An unlimited supply of parts for drilling is available. When a part is completed, another part is immediately available for drilling.

As holes are drilled, the tool wears and is susceptible to breakage. It is assumed that there is a limit to the wear beyond which the tool is unacceptable. A broken tool is considered unacceptable. Evolution of tool wear necessitates the occasional replacement of the tool. Replacement of the tool involves some costs for both time and material.

The spindle speed and the feed speed of the drill are variable. The machining rate, i.e. the rate of metal removal, is proportional to the feed speed. The time required to complete an operation is inversely proportional to the machining rate. Assuming that the tool does not break, the time to complete a part is inversely proportional to the feed speed. See Appendix A for a description of relevant drilling terminology.

The dynamics of tool wear are not completely understood. Experiments have been made from which empirical formulas for tool wear have been derived. These formulas are known to be only approximate, and considerable variation in tool life is evident. Furthermore, the material to be drilled has some uncertain characteristics that also affect tool wear. The empirical formulas for tool life suggest that under the assumption of a fixed feed (the ratio of feed speed to spindle speed), the tool wear rate is an increasing function of the cutting speed, and an increasing function of the feed speed.

The replacement of tools prior to breakage is feasible. If the tool breaks, it must be replaced before continuing. Furthermore, tool breakage may cause damage to the current part, possibly resulting in a scrap part or necessitating rework. Costs are therefore associated with tool breakage. Tool breakage is an unplanned event, so replacement of the tool due to breakage may take longer (and thus cost more) than planned replacement.

The drilling problem is to determine the best policies for operating the drill. The required decisions are when to replace the tool, and the selection of spindle and feed speeds used to drill the parts. The criteria for ranking policies is a function of the various economic considerations. Any of several different criteria may be reasonable.

The above description of the drilling problem is typical of problems of machining economics, and as such does not appreciably extend previous considerations of these problems. The on-line variability of the spindle and feed speeds is not usual, however. Similar problems can be found in many references, including [Er1], [Ha1], [Hi1], [Iw1], [Le1], [Le2], [Ph1], [Shk1], [Tak1], and [Tay1].

The problem will now be extended. Suppose that occasional measurements of the extent of tool wear are made. Continuous measurements cannot be made while the tool is engaged in the part however. Measurements may be made between parts, when the tool is disengaged. The method of wear measurement is not specified. This new information is to be incorporated into the decision policies.

The essential features of the drilling problem are:

(1) Mechanical Aspects: The drill is used in a fixed operation carried out repetitively on parts. A fixed volume of material is removed from each part. As metal is removed from the parts, the tool wears. The tool wear mechanism is only known empirically, and variation in tool life is evident. As the tool wears, it eventually breaks or becomes unacceptable and must be replaced. Failure of the drill necessitates replacement, but can also cause damage to the part being machined.

(2) Economic/Manufacturing Aspects: The production of parts results in profit, and the profit rate is related to the production rate. The profit rate is assumed to be an increasing function of the feed speed. Tool replacement is costly in terms of time and material. Tool breakage is in general more costly than simple replacement, due to possible part damage and the unplanned replacement that results.

(3) Control Aspects: The objective is to determine policies for tool replacement and for feed speed selection. Tool replacement and changes in feed speed may only occur between parts, unless there is a tool breakage.

(4) Information Feedback: Tool wear can be measured when the tool is disengaged from the part, but not during the machining.

The specific use of a drilling operation in this problem is only to facilitate the presentation of the concepts and is not meant to imply a restriction of the ideas to that type of machining operation.

There are several important features incorporated into the drilling problem that are common to many manufacturing systems. Since some treatments of manufacturing problems overlook these characteristics additional emphasis is warranted. First, the discrete nature of parts is represented. This property imposes certain restrictions on the way manufacturing system analysis can be carried out, and definitely imposes restrictions on the control policies and availability of information. Second, the rate of production is specifically cited as a decision control variable in the sense that the local or instantaneous rate of metal removal can be controlled. Local means in the absence of other effects (tool failure, tool replacement, etc.). Although in actual manufacturing processes the production rate will necessarily be constrained by other physical considerations, its availability as a control input should not be overlooked. Third, production rate selection impacts the system in other ways. In the drilling problem, the production rate affects the rate of resource consumption (tool wear) and affects the likelihood of system failure. Production rate can not be made arbitrarily large without adversely affecting other aspects of the system's performance. Fourth, an instance of failure can result in the production of scrap. That is, not every piece produced is acceptable, and when certain limits are exceeded the risk of producing unacceptable parts increases. In the drilling problem, this effect is captured in a particularly simple way. It should be noted that in actual systems this effect may be far more subtle.

3.3. Problem 2: Multi-tooled Machines

This problem is a simple but realistic extension of the drilling problem. It is not usually considered in the machining economics literature, although see [Shk1].

Suppose that the drill discussed in the drilling problem is actually multi-tooled. That is, suppose that several holes are drilled simultaneously. This type of machine is very common in high volume production systems. The tools are not independent and must be controlled as a group. Since all the tools are simultaneously engaged in the discrete nature of parts is represented. This property imposes certain restrictions on the way manufacturing system analysis can be carried out, and definitely imposes restrictions on the control policies and availability of information. Second, the rate of production is specifically cited as a decision control variable in the sense that the local or instantaneous rate of metal removal can be controlled. Local means in the absence of other effects (tool failure, tool replacement, etc.). Although in actual manufacturing processes the production rate will necessarily be constrained by other physical considerations, its availability as a control input should not be overlooked. Third, production rate selection impacts the system in other ways. In the drilling problem, the production rate affects the rate of resource consumption (tool wear) and affects the likelihood of system failure. Production rate can not be made arbitrarily large without adversely affecting other aspects of the system's performance. Fourth, an instance of failure can result in the production of scrap. That is, not every piece produced is acceptable, and when certain limits are exceeded the risk of producing unacceptable parts increases. In the drilling problem, this effect is captured in a particularly simple way. It should be noted that in actual systems this effect may be far more subtle.

3.3. Problem 2: Multi-tooled Machines

This problem is a simple but realistic extension of the drilling problem. It is not usually considered in the machining economics literature, although see [Shk1].

Suppose that the drill discussed in the drilling problem is actually multi-tooled. That is, suppose that several holes are drilled simultaneously. This type of machine is very common in high volume production systems. The tools are not independent and must be controlled as a group. Since all the tools are simultaneously engaged in the same part, the failure of any one tool can damage the part. The problem now is to determine the group machining parameters and tool replacement policies so that good performance is achieved. Note that group replacement is not in general required (though it may be specified for convenience reasons). Assume all other conditions are as given in the drilling problem. This includes the availability of occasional wear measurements.

3.4. Problem 3: Multi-machine Manufacturing Systems

Not many manufacturing systems consist of only one machine, as presented in the drilling problem. Typically, many machines are involved, with parts visiting several or all machines before their processing is completed. Although the machines can be multi-purpose, assume here that they are single purpose. Also assume that part routes are predetermined and fixed. This forces attention on operational issues other than scheduling and routing problems. In particular, the usual FMS problems are not being considered here. It is possible that buffers exist between some or all of the machinery, but it is not necessary.

The problem is to determine the appropriate machining parameters; e.g., spindle speeds and feed speeds, and thus the production rate for each of the machines, as well as to specify tool replacement policies in order to achieve good system performance. Assume that each machine is conceptually the same as that considered in the drilling problem: on-line control of the machining parameters is possible and occasional tool wear measurements for each machine may be available. The status of machines in the system and current inventory in the buffers may also be available.

Several important issues should be noted. For interconnected machines, tool failure, repair, and tool replacement decisions at one machine can affect other machines through blockage and starvation. Furthermore, the costs for repair and tool replacement may be dependent upon the system status, and the availability of required workers or other resources. The worth of a part depends in general on how close it is to completion.

The multi-machine system described here is representative of many manufacturing systems. However, it is possible that other types of machines are in the system as well, with different characteristics from those presented in the drilling problem. Consideration is restricted here to networks of machines that are conceptually similar to that described in the drilling problem.

An additional issue arises in the multi-machine problem that does not appear in the single machine problem. The structure of the controller also needs to be specified. Consideration should be given to both centralized and decentralized controllers. Each structure has certain advantages and disadvantages that need to be examined as part of the problem. As an example, a centralized controller may be difficult to implement for a large system because of the amount of data that has to be handled and the computational burden.

3.5. Problem 4: The Supervisor's Problem

A vital component in manufacturing is often omitted from problem formulations. That component is human workers. There are no manufacturing systems today that can operate for more than a few hours without human involvement. Many manufacturing systems are completely dependent upon humans. In order to bring recognition to the importance of human involvement, the next problem emphasizes human aspects of manufacturing and downplays the role of machines.

Consider a manufacturing facility (e.g., a job shop) that employs a group of workers and a group supervisor. A job is scheduled for the facility. This job will be viewed as consisting of a single task. This task may be arbitrarily complicated, however. A group of workers (one or more) is to be assigned to the task. A supervisor is responsible for overseeing the task and the workers. The task has associated with it various goals and economic considerations. Possible goals include minimizing costs/maximizing profits, meeting schedules (due dates), and minimizing time in shop. The task also has uncertainty associated with it. The sources of this uncertainty may include: how big the task is, how long the task will take, how fast the workers will work, mistakes in task specification, miscommunication between the workers and the supervisor and mistakes in carrying out the work.

The supervisor's role is to assess the work requirements for the task and assign workers and other resources to the task. The supervisor is allowed to make occasional inspections to assess the work progress on the task. These assessments may be imprecise. Furthermore, these inspections are costly or otherwise constrained and thus constant inspection is impossible. As examples, work progress may be stopped during an inspection, or inspections might only be possible once a day. Based on the assessments, the supervisor is allowed to make adjustments to the manpower and resources assigned to the task in order to influence the work rate. These adjustments have some costs or constraints associated with them.

The objective of the supervisor's problem is to develop policies the supervisor may use to determine when to inspect work progress and how to assign resources based upon work progress assessments in order to optimize some performance functional.

When the supervisor is not present, the above problem is related to typical stochastic scheduling and assignment problems. This class of problems has been considered by many researchers, and a variety of techniques are available for approaching the problems (e.g., dynamic programming, other mathematical programming, search techniques, heuristics, etc.). Solutions are not necessarily simple, however. The usual solution to these types of problems is to produce essentially static assignments and schedules. The problem is that this solution fails to capture the role of the supervisor. In particular, manufacturing environments are dynamic, and the role of the supervisor is to react to the dynamics in order to keep the facility operating in a reasonable way.

Fundamental to studying the supervisor's problem is the construction of a suitable model that incorporates the required features. Unfortunately, the traditional formulations of scheduling and assignment lack the mechanisms needed to capture the supervisor's role. A new model is needed.

The supervisor's problem is an attempt to capture features common to manufacturing problems that are not typically represented in traditional formulations of related problems. The supervisor's problem recognizes the dynamic nature of the manufacturing environment and thus extends the formulation beyond the usual static assumptions. The dynamics of the problem are not deterministic; there is uncertainty associated with the system. Most importantly, the problem recognizes the role of the supervisor in the manufacturing environment as a decision maker. Thus, the concepts of on-line control and information feedback are represented in the supervisor's problem.

3.6. Summary of the Problems

The prototype problems presented here are intended to convey aspects of manufacturing that are not typically considered in other problem formulations. As such, these problems depart from the traditional views of manufacturing found in the literature. The problems are also placed in an intelligent manufacturing setting, where the availability of information and the capability of influencing the system online is assumed.

The control theoretic approach has been previously proposed as a method for handling intelligent manufacturing problems. However, in order to utilize this approach, there is a need for new models of manufacturing systems that incorporate certain desired features as well as adhere to the control compatibility requirements. The next chapter is devoted to the study of a class of stochastic processes that will be used to model the problems proposed here. These stochastic processes are fundamental to this work. After introducing and describing these processes, the first prototype problem (the drilling problem) will be studied in some detail.

CHAPTER 4

MATHEMATICAL THEORY OF DIFFUSION-THRESHOLD PROCESSES AND THEIR APPLICATION

4.1. Introduction

In this chapter, a class of stochastic processes called diffusion-threshold processes will be introduced and discussed. The name diffusion-threshold is derived from the process construction. A diffusion process evolves in time until some boundary (called the threshold) is first achieved. The attainment of the threshold signifies an event, resulting in some discrete action being taken; e.g., re-initialization or suspension of the diffusion process. The time at which the threshold is first attained is a random variable called the threshold hitting time or threshold crossing time.

The discussion will begin by considering only real scalar diffusions. These diffusions can be controlled, however. The thresholds will be constant real values. The objective is to obtain probabilistic descriptions of the threshold crossing time based on the diffusion and threshold parameters. In particular, the probability distribution function is to be determined, and the moments are to be computed.

Solutions to this problem are difficult in the general case. This work will focus on the development of some special cases where analytical solutions can be obtained. The basic result assumes the case of constant infinitesimal coefficients. Some of the properties of this special case will be developed and discussed.

The result for constant coefficients is extended to the case of deterministic piecewise constant coefficients in Appendix B. The results for this case are significantly more complicated than the constant coefficient case. The piecewise constant coefficient case represents an approximation to the more general problem of deterministic time-varying coefficients. The determination of the distribution function for this more general problem is unsolved.

The incorporation of measurements of the diffusion process into the distribution function is next considered. These results show how process measurement feedback can be used to update the threshold crossing probability distribution function.

Multi-dimensional extensions of the diffusion-threshold process are next considered. Higher dimensional problems are generally difficult, but one special class of problems will be isolated and developed. Although the class appears to be restrictive, it will be shown later to have application.

A discussion of some reasons for using diffusion-threshold processes in modeling will follow. Emphasis is placed on the significance of discrete event generation by the diffusion-threshold process. Interpretations of the various parameters in the process will be given from a modeling viewpoint. Several categories of general application of diffusion-threshold models will help to clarify the potential usage of the process.

The chapter will conclude with alternative solution techniques and a discussion of generalizations of the process and its relation to impulsive control.

4.2. General Process Description

4.2.1. Preliminary Definitions

Let (Ω, F, P) be a probability space. Define $I\!R \triangleq (-\infty,\infty)$, $I\!R^{+} \triangleq [0,\infty)$, and $\overline{I\!R}^{+} \triangleq [0,\infty]$. A history on (Ω, F) is an increasing family of σ-algebras $\{F_t,\ t \geq 0\}$ such that for $0 \leq s < t$, $F_s \subset F_t \subset F$. Define $F_\infty \triangleq \bigvee_{t \geq 0} F_t$, the smallest σ-algebra containing the history. Note that $F_\infty \subset F$.

Let $\{W(\omega,t),\ \omega \in \Omega,\ t \in I\!R^{+}\}$ be a standard Brownian motion, (i.e. a Wiener Process), separable and measurable with respect to $F \otimes B^{+}$, where B^{+} is the Borel σ-algebra on $I\!R^{+}$, and $\otimes$ denotes the smallest product σ-algebra. A standard Brownian motion is a real valued scalar Gaussian process such that:

$$P[W(\omega,0) = 0] = 1 \tag{4.2.1.1a}$$

$$E[W(\omega,t)] = 0 \text{ for } t \in I\!R^{+} \tag{4.2.1.1b}$$

$$E[W(\omega,s)W(\omega,t)] = \min(s,t) \text{ for } s,t \in I\!R^{+} \tag{4.2.1.1c}$$

Recall that for a Gaussian process, the random variables $(W(\cdot,t_1), \ldots, W(\cdot,t_n))$ are jointly Gaussian for every finite collection $(t_1, \ldots, t_n) \in I\!R^{+}$.

Since $W(\cdot,\cdot)$ is assumed separable, every sample path is continuous, P-$a.s.$ Without loss of generality, assume $W(\omega,\cdot)$ to be continuous for all $\omega \in \Omega$. (Note that some authors specifically define Brownian motion to have this property.) The explicit dependence of the process on ω will be omitted when no confusion is created: $W(t)$ and $W_t \triangleq W(\omega,t)$.

An important property of Brownian motion is that it has independent increments ([Lo1], [Kar1], [Wo1]).

Let $\{F_t^W, t \in \mathbb{R}^+\}$ be the history generated by $\{W_t\}$. That is

$$F_t^W \triangleq \sigma\{W_s, 0 \leq s \leq t\} \tag{4.2.1.2}$$

the smallest σ-algebra generated by the set of random variables $\{W_s, 0 \leq s \leq t\}$.

Definition: A stopping time with respect to a history $\{F_t\}$ is an extended random variable $T : \Omega \rightarrow \overline{\mathbb{R}}^+$ such that $\{\omega : T(\omega) \leq t\} \in F_t$ for all $t \in \overline{\mathbb{R}}^+$. See [Br1], [Lo1], [Wo1].

A stopping time T induces a σ-algebra denoted F^T and defined as

$$F^T \triangleq \sigma\{S \in F : S \bigcap \{\omega : T(\omega) \leq t\} \in F_t, \text{ for all } t \in \overline{\mathbb{R}}^+\} \tag{4.2.1.3}$$

Remark: An important stopping time for Brownian motion is

$$T_{a,b} \triangleq \min\{t \geq 0 : W_t \notin (a, b)\} \text{ for } a < 0 < b \tag{4.2.1.4}$$

Definition: A strong Markov process is a process that exhibits the Markov property through stopping times. See [Kar2], [Lo1], [Wo1].

Remark: Brownian motion is a strong Markov process.

4.2.2. Diffusion Processes

Brownian motion can be thought of as the prototype for a class of stochastic processes called diffusion processes. Diffusion processes have continuous sample paths and the strong Markov property ([Kar2]).

Define a real-valued stochastic process $\{X_t, t \in \mathbb{R}^+\}$ by

$$X_t = X_0 + \int_0^t b(\tau, X_\tau) d\tau + \int_0^t \sigma(\tau, X_\tau) dW_\tau \tag{4.2.2.1}$$

The last integral is a stochastic integral and will in general be interpreted as an Ito integral, although other interpretations are also possible ([Fl1], [Wo1]). Then, under appropriate technical conditions on the coefficients b and σ, $\{X_t\}$ is called a diffusion. Growth and Lipschitz conditions (called the Ito conditions) on the coefficients are sufficient to assure existence and uniqueness. These are not necessary conditions, however ([Fl1], [Wo1]).

The functions $b(\cdot,\cdot)$ and $\sigma(\cdot,\cdot)$ will be called the infinitesimal coefficients of the diffusion, with b called the drift coefficient and σ called the diffusion coefficient.

The process X_t can be thought of as the solution to the stochastic differential equation

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t \tag{4.2.2.2}$$

although this is not precise except when defined in terms of the above integral equation. For more information on general diffusions, see [Fl1], [It1], [Kar2], [Wo1].

There are other ways of defining or characterizing diffusions. The above stochastic integral equation will be the operational definition in this work. Conceptually, though, a diffusion is a strong Markov process with continuous sample paths ([Kar2]).

The important relation between the diffusion and its infinitesimal coefficients is given by the following:

Theorem (4.2.2.1): Let $b(\cdot,\cdot)$ and $\sigma(\cdot,\cdot)$ be Borel measurable on $\mathbb{R}^+ \times \mathbb{R}$ Assume there exist positive σ_0 and K such that

$$|\ b(t, x)\ | \leq K\sqrt{1 + x^2} \tag{4.2.2.3a}$$

$$0 < \sigma_0 \leq \sigma(t, x) \leq K\sqrt{1 + x^2} \tag{4.2.2.3b}$$

$$|\ b(t, x) - b(t, y)\ | \leq K\ |\ x - y\ | \tag{4.2.2.3c}$$

$$|\ \sigma(t, x) - \sigma(t, y)\ | \leq K\ |\ x - y\ | \tag{4.2.2.3d}$$

(These are the Ito conditions.) Then the diffusion process X_t satisfies

$$\lim_{\eta \to 0} \frac{1}{\eta} E[X_{t+\eta} - X_t \mid X_t = x] = b(t,x) \tag{4.2.2.4a}$$

$$\lim_{\eta \to 0} \frac{1}{\eta} E[(X_{t+\eta} - X_t)^2 \mid X_t = x] = \sigma^2 (t,x) \tag{4.2.2.4b}$$

Proof: See [Wol].

Stopping times can be defined with respect to diffusions. A particular stopping time that will be used in this work is the following. Let $A > 0$ be a real value. Define

$$T_A : \Omega \to \overline{\mathbb{R}}^+$$

$$T_A = \inf\{t \geq 0 : X_t \geq A\} \tag{4.2.2.5}$$

Then T_A is a stopping time with respect to the history $\{F_t^X\}$. Note that T_A is well defined because of sample path continuity. Also

$$\{T_A = \infty\} = \{\omega : X_t < A \text{ for all } t \geq 0\} \tag{4.2.2.6}$$

which can be a non-null event. One property of Ito integrals with respect to stopping times that is used implicitly in this work is the following:

Theorem (4.2.2.2): Let X_t be a diffusion satisfying the Ito conditions. Let T be a stopping time with respect to $\{F_t^X\}$. Then the following holds:

$$\int_0^{(t \wedge T)} \sigma(\tau, X_\tau) dW_\tau = \int_0^t 1_{[\tau \leq T]} \sigma(\tau, X_\tau) dW_\tau \tag{4.2.2.7}$$

where $1_{[A]}$ is the indicator function for the event A.

Proof: See [Ell].

This result is quite intuitive, but not trivial.

4.2.3. Controlled Diffusions

The previous results can be generalized to controlled diffusions ([Fl1], [Kry1]). Define a real-valued stochastic process $\{X_t, t \in \mathbb{R}^+\}$ by

$$X_t = X_0 + \int_0^t b(\tau, X_\tau, u_\tau)d\tau + \int_0^t \sigma(\tau, X_\tau, u_\tau)dW_\tau \qquad (4.2.3.1)$$

where $u_t(\cdot)$ is a function which takes values in U, where U is a Borel measurable subset of $\mathbb{R}$ Assume that $u_t(\cdot)$ is measurable with respect to $\{F_t^X\}$. Then with appropriate conditions placed on the infinitesimal coefficients, the process $\{X_t\}$ is called a controlled diffusion. The diffusions discussed in this work will be controlled, but will have an especially simple structure.

4.2.4. Diffusion-Threshold Processes

By combining diffusion processes with stopping times based on threshold crossings, a diffusion-threshold process is constructed. Without loss of generality assume that $X_0 = 0$. Let $A > 0$ denote a positive threshold value, and T_A the stopping time generated by a controlled diffusion X_t:

$$T_A = \inf\{t \geq 0 : X_t \geq A\} \qquad (4.2.2.5)$$

At time T_A a significant event will be considered to have occurred. This will give rise to the following effects. If T_A occurs, the diffusion process will be re-initialized. That is, $X_{T_A+} = X_0$. The instance of the threshold crossing can be thought of as causing a jump in the state space (instantaneous return process [Kar2]). Alternatively, the threshold crossing can be thought of as causing a killing of the process, and the re-initialization as a new selection of $\omega \in \Omega$, with a time shift of T_A. In order for this to make sense, assume that $b(\cdot,u_t)$ and $\sigma(\cdot,u_t)$ are functions of $t - t_0$ only, where t_0 is the last re-initialization time. In order to simplify this, assume that a re-initialization results in a resetting of the diffusion clock to zero. Each process segment consisting of re-initialization to threshold crossing will be called a cycle. Thus, the diffusion-threshold process generates a sequence of cycles. In this work, the control will be restricted so that these cycles are probabilistically independent, though this may not be true in general. An example sample path for a diffusion-threshold process is shown in Fig. 4.2.4.1.

An additional discrete control action may also be allowed. It may be possible to force a re-initialization of the diffusion process at particular times prior to the diffusion achieving the threshold. This forced re-initialization will otherwise have the same effect as a re-initialization resulting from a threshold crossing.

The goal in studying diffusion-threshold processes is to obtain a probabilistic description of the threshold crossing time. That is, the goal is to determine the probability distribution function for the random variable T_A as well as compute the moments. Actually, the problem is somewhat more complicated because the diffusion-threshold process is controlled, and the relationship between the control and the threshold crossing time is to be determined. That is, it is actually a family of distributions parameterized by the control that is to be determined. Similarly, in

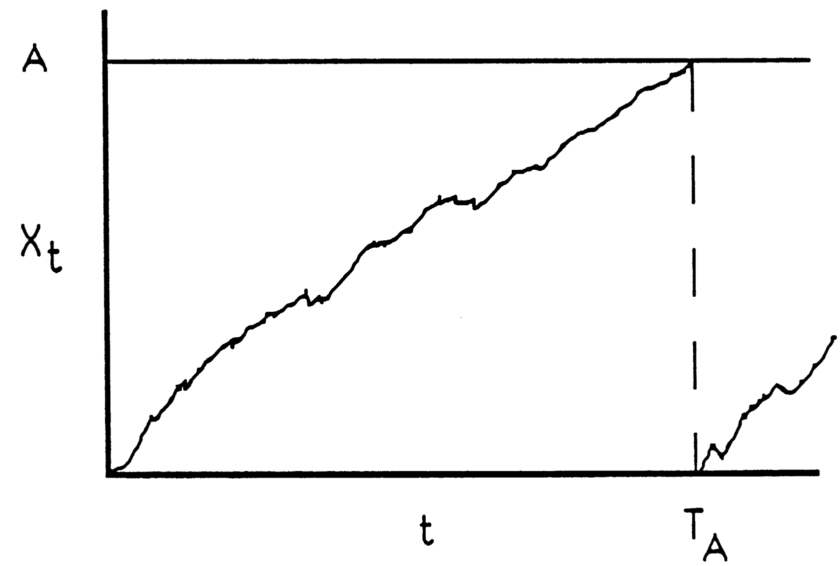

Fig. 4.2.4.1 Diffusion-Threshold Sample Path

order to accommodate measurement feedback a family of conditional distributions given the diffusion state is to be determined.

The general problem of determining the distribution function is hard ([Bl1], [Me1], [Krt1], [Si1]). It is known that for sufficiently smooth infinitesimal coefficients, the solution is associated with a partial differential equation called the backward equation. Closed form solutions to the backward equation are seldom available. In this work, the general problem will not be considered. Instead, the focus will be on certain classes of diffusions where the distribution function can be analytically obtained. The reference [Bl1] is a survey of various level (threshold) crossing problems.

4.3. Processes with Constant Coefficients

In this section, attention will be restricted to a particularly simple class of diffusions. This restriction will allow closed form computation of the probability density function and the moments of the threshold crossing time.

Consider those diffusions where b and σ are independent of X and t and thus depend only on the control u_t. Also assume that $u_t = u$ is constant. That is:

$$X_t = X_0 + \int_0^t b\,d\tau + \int_0^t \sigma\,dW_\tau \tag{4.3.1}$$

and

$$b = b(u),\ 0 \le b < \infty \tag{4.3.2}$$

$$\sigma = \sigma(u),\ 0 < \sigma < \infty \tag{4.3.3}$$

Also assume that $X_0 = 0 < A$. Under these assumptions, $\{X_t\}$ is a Brownian motion with drift. As before let T_A be the threshold crossing time. Note that in this case the stochastic integral can be equivalently evaluated as an Ito, Stratonovich, or

Riemann-Stieltjes integral, though never as a Lebesgue-Stieltjes integral ([El1], [Fl1], [Wo1]).

For diffusions in the above class, it is possible to compute the density function for the threshold crossing time. A preliminary result is needed first concerning the probability of a threshold crossing occurring. The following establishes that under appropriate conditions a threshold crossing will occur almost surely.

Lemma (4.3.1): Given $X_t = bt + \sigma W_t$, $b \geq 0$, $\sigma > 0$, $A > 0$ then

$$P[T_A < \infty] = 1 \tag{4.3.4}$$

Proof: It suffices to prove the result for $b = 0$, since

$$\{\omega : T_A(\omega) < \infty,\ b > 0\} \supset \{\omega : T_A(\omega) < \infty,\ b = 0\} \tag{4.3.5}$$

For $b = 0$,

$$\begin{aligned} \{\omega : T_A(\omega) < \infty\} &\supset \{\omega : \sup_{t \geq 0} W_t(\omega) > \frac{A}{\sigma}\} \\ &\supset \{\omega : \overline{\lim_{t \geq 0}}\, W_t(\omega) > \frac{A}{\sigma}\} \\ &\supset \{\omega : \overline{\lim_{t \geq 0}}\, W_t(\omega) = \infty\} \end{aligned} \tag{4.3.6}$$

But

$$P\{\omega : \overline{\lim_{t \geq 0}}\, W_t = \infty\} = 1 \tag{4.3.7}$$

is a standard result for Brownian motion. See [Lo1]. Therefore $P[T_A < \infty] = 1$.

This lemma asserts that under the assumed conditions $P[T_A = \infty] = 0$. Thus the distribution function for T_A is not defective on $\mathbb{R}^+$.

The main result can now be stated.

Theorem (4.3.1): Given $X_t = bt + \sigma W_t$, $b(u) \geq 0$, $\sigma(u) > 0$, $A > 0$ the probability density function for T_A, denoted $q_A(t \mid 0\, ;\, u)$ is given by:

$$q_A(t \mid 0\, ;\, u) = \begin{cases} \dfrac{A}{\sqrt{2\pi}\sigma t^{\frac{3}{2}}} \exp\left\{\dfrac{-(A - bt)^2}{2\sigma^2 t}\right\} & t > 0 \\ 0 & t \leq 0 \end{cases} \tag{4.3.8}$$

Remark: This is a previously known result ([Bl1]). However, a proof will be given here that the author has not previously seen in detail. This proof will allow the introduction of certain Laplace transforms that are useful for studying threshold crossing problems. See also [Me1], [Si1].

Proof: The process X_t is Gaussian with density function

$$f_{X_t}(x\, ;\, u) = \frac{1}{\sqrt{2\pi t}\,\sigma} \exp\left\{\frac{-(x - bt)^2}{2\sigma^2 t}\right\} \tag{4.3.9}$$

Choose $t > 0$, $x > A$, but otherwise arbitrary. By the strong Markov property

$$f_{X_t}(x\, ;\, u) = \int_0^t f_{X_t \mid X_\eta}(x \mid A\, ;\, u) q_A(\eta \mid 0\, ;\, u) d\eta \tag{4.3.10}$$

where

$$f_{X_t \mid X_\eta}(x \mid y\, ;\, u) = \frac{1}{\sqrt{2\pi(t-\eta)}\,\sigma} \exp\left\{\frac{-(x - y - b(t-\eta))^2}{2\sigma^2(t-\eta)}\right\} \tag{4.3.11}$$

is a conditional Gaussian density. The explanation of the integral equation (4.3.10) is the following ([Bl1]). By the strong Markov property, what happens before the threshold crossing is independent of what happens after. Further, the time of threshold crossing can be thought of as partitioning the set of sample paths. Thus the density function for the process must equal the integral of the conditional densities for processes with crossing times at η weighted by the density of processes with crossing times in $(\eta, \eta + d\eta)$. Note that t and η appear only as a difference in the conditional density function. Therefore the integral equation is a convolution and Laplace transform is a viable solution technique. Denote

$$\hat{q}_A(s \mid 0 ; u),\ \hat{f}(s ; x,u),\ \hat{f}(s ; x \mid A,u)$$

as the Laplace transforms (in t) of the density functions. Then

$$\hat{q}_A(s \mid 0 ; u) = \frac{\hat{f}(s ; x,u)}{\hat{f}(s ; x \mid A,u)} \tag{4.3.12}$$

Now

$$\hat{f}(s ; x,u) = L\{\frac{1}{\sqrt{2\pi t}\,\sigma} \exp\left\{\frac{-x^2}{2\sigma^2 t}\right\}\exp\left\{\frac{bx}{\sigma^2}\right\}\exp\left\{\frac{-b^2 t}{2\sigma^2}\right\}\}$$

$$= \frac{1}{\sqrt{2}\sigma} \exp\left\{\frac{bx}{\sigma^2}\right\} F(s - a) \tag{4.3.13}$$

where ([Ab1])

$$F(s) = L\{\frac{1}{\sqrt{\pi t}} \exp\left\{\frac{-x^2}{2\sigma^2 t}\right\}\}$$

$$= \frac{1}{\sqrt{s}} \exp\left\{\frac{-x}{\sigma}\sqrt{2s}\right\} \tag{4.3.14}$$

and

$$a = \frac{-b^2}{2\sigma^2} \tag{4.3.15}$$

Thus

$$\hat{f}(s ; x,u) = \exp\left\{\frac{bx}{\sigma^2}\right\} \frac{1}{\sqrt{2(s-a)}\,\sigma} \exp\left\{\frac{-x}{\sigma}\sqrt{2(s-a)}\right\} \tag{4.3.16}$$

Similarly,

$$\hat{f}(s ; x \mid A,u) = \exp\left\{\frac{b(x-A)}{\sigma^2}\right\} \frac{1}{\sqrt{2(s-a)}\,\sigma} \exp\left\{\frac{-(x-A)}{\sigma}\sqrt{2(s-a)}\right\} . \tag{4.3.17}$$

where a is given by (4.3.15). Therefore

$$\hat{q}_A(s \mid 0 ; u) = \exp\left\{\frac{bA}{\sigma^2}\right\}\exp\left\{\frac{-A}{\sigma}\sqrt{2(s-a)}\right\} \tag{4.3.18}$$

Now

$$L^{-1}\{\exp\left\{\frac{-A}{\sigma}\sqrt{2s}\right\}\} = \frac{A}{\sqrt{2\pi}\sigma t^{\frac{3}{2}}} \exp\left\{\frac{-A^2}{2\sigma^2 t}\right\} \tag{4.3.19}$$

from [Ab1]. So

$$q_A(t \mid 0 \,;\, u) = \frac{A}{\sqrt{2\pi}\sigma t^{\frac{3}{2}}} \exp\left\{\frac{-A^2}{2\sigma^2 t}\right\}\exp\left\{\frac{bA}{\sigma^2}\right\}\exp\left\{\frac{-b^2 t}{2\sigma^2}\right\}$$

$$= \begin{cases} \dfrac{A}{\sqrt{2\pi}\sigma t^{\frac{3}{2}}} \exp\left\{\dfrac{-(A - bt)^2}{2\sigma^2 t}\right\} & t>0 \\ 0 & t\leq 0 \end{cases} \tag{4.3.8}$$

Remark: The result holds for $b = 0$. It does not hold for $b<0$, since in this case $P[T_A = \infty] > 0$. That is, the distribution becomes defective.

Remark: From the proof of Theorem 4.3.1, note that

$$q_A(t \mid 0 \,;\, u) = \frac{A}{t} f_{X_t}(A \,;\, u) \tag{4.3.20}$$

This has been previously shown to be true for a class of processes by Borovkov. See [Bl1] for remarks and references.

Remark: From equation (4.3.8) it is easily seen that

$$\lim_{t \to 0^+} q_A(t \mid 0 \,;\, u) = 0 \,,\ b\geq 0,\ \sigma>0,\ A>0 \tag{4.3.21}$$

Theorem 4.3.1 generalizes to non-zero initial conditions, where $X_0 = x_0 \neq 0$.

Corollary (4.3.1): Given $X_t = x_0 + bt + \sigma W_t$, $b(u)\geq 0$, $\sigma(u)>0$, $A>x_0$ the probability density function for T_A, denoted $q_A(t \mid x_0 \,;\, u)$ is given by:

$$q_A(t \mid x_0 \,;\, u) = \begin{cases} \dfrac{A - x_0}{\sqrt{2\pi}\sigma t^{\frac{3}{2}}} \exp\left\{\dfrac{-(A - x_0 - bt)^2}{2\sigma^2 t}\right\} & t>0 \\ 0 & t\leq 0 \end{cases} \tag{4.3.22}$$

Proof: Simple extension of the proof used in Theorem 4.3.1.

Remark: The corollary follows, since the infinitesimal coefficients don't depend on the state. Therefore the state space is homogeneous to the diffusion. The same argument applies to the time variable as well. Note that

$$q_A(t \mid x_0 \,;\, u) = q_\alpha(t \mid 0 \,;\, u) \tag{4.3.23}$$

where $\alpha = A - x_0$.

From the above, the distribution function for the threshold crossing time given u is immediate. Define the distribution function for the random variable T_A by:

$$Q_A(t \mid 0 \,;\, u) \triangleq P\{\, T_A \leq t \mid X_0 = 0 \,;\, u\}$$

$$= \int_0^t \frac{A}{\sqrt{2\pi}\sigma\tau^{\frac{3}{2}}} \exp\left\{\frac{-(A - b\tau)^2}{2\sigma^2\tau}\right\} d\tau \tag{4.3.24}$$

The moments of T_A can be evaluated from the Laplace transform of the density for $b > 0$. However, $b = 0$ is a special case.

Theorem (4.3.2): For $b = 0$, $E[T_A] = \infty$.

Proof: (See also [Ro1])

$$E[T_A] = \int_0^\infty \eta q_A(\eta \mid 0 \,;\, u) d\eta = \int_0^\infty \frac{A}{\sqrt{2\pi}\sigma\eta^{\frac{1}{2}}} \exp\left\{\frac{-A^2}{2\sigma^2\eta}\right\} d\eta \tag{4.3.25}$$

Substituting

$$y = \frac{A}{\sigma\sqrt{\eta}} \qquad dy = \frac{-A}{2}\eta^{-\frac{3}{2}} d\eta \tag{4.3.26}$$

into (4.3.25) gives

$$E[T_A] = \int_0^\infty \frac{2A^2}{\sqrt{2\pi}\sigma^3 y^2} \exp\left\{\frac{-y^2}{2}\right\} dy$$

$$= \frac{2A^2}{\sqrt{2\pi\sigma^3}} \int_0^\infty \frac{1}{y^2} \exp\left\{\frac{-y^2}{2}\right\} dy$$

$$\geq \frac{2A^2}{\sqrt{2\pi\sigma^3}} \exp\left\{\frac{-1}{2}\right\} \int_0^1 \frac{1}{y^2} dy = \infty \qquad (4.3.27)$$

Although the threshold crossing will occur with certainty, it may take an arbitrarily long time to happen with significant probability. This is analogous to a null-recurrent Markov chain.

For $b > 0$, the moments of T_A are finite and can be computed from the Laplace transform:

$$E[T_A^n] = (-1)^n \frac{d^n}{ds^n} \hat{q}_A(s \mid 0 ; u) \Big|_{s=0}, \; n = 1,2,... \qquad (4.3.28)$$

The central moments follow easily. The first four central moments are:

$$E[\, T_A \,] = \frac{A}{b} \qquad (4.3.29a)$$

$$\mathrm{var}[T_A] = \frac{A\sigma^2}{b^3} \qquad (4.3.29b)$$

$$\mu_3 = \frac{3A\sigma^4}{b^5} \qquad (4.3.29c)$$

$$\mu_4 = \frac{15A\sigma^6}{b^7} + \frac{3A^2\sigma^4}{b^6} \qquad (4.3.29d)$$

where μ_3 and μ_4 are the third and fourth central moments.

Remark: The mean of T_A is simply the distance divided by the rate.

The density function q_A is known in the statistics literature as an inverse Gaussian density ([Ch1], [Shr1], [Tw1]). However, a different parameterization is usually used. Define the parameters

$$\mu(u) = \frac{A}{b} \qquad (4.3.30)$$

$$\lambda(u) = \frac{A^2}{\sigma^2} \qquad (4.3.31)$$

Substitution into the formula for the density q_A gives:

$$q_A(t\,|\,0\,;\,u) = \sqrt{\frac{\lambda}{2\pi t^3}}\exp\left\{\frac{-\lambda(t-\mu)^2}{2\mu^2 t}\right\} \tag{4.3.32}$$

Other parameterizations are also possible.

In this work, the original parameterization will be used in order to retain the explicit dependence on the diffusion coefficients and the threshold value.

Remark: The inverse Gaussian density is not a version of any of the commonly known density functions.

Properties of the inverse Gaussian distribution and density can be found in [Ch1], [Shr1], [Tw1]. However, in each of these works, the statistical viewpoint prevails and the origin of the distribution in terms of threshold crossings in not exploited.

In the following graph (Fig. 4.3.1) a comparison is made of the inverse Gaussian density to some other common density functions.

One particularly remarkable result is the relationship between the inverse Gaussian distribution and the normal distribution. Since this result is not well known but is used in this work, it is presented here parameterized to conform with this work.

Theorem (4.3.3): Let $\Phi(x) \triangleq P[X \le x]$ when $X \sim N[0,1]$. Then

$$Q_A(t\,|\,0\,;\,u) = \Phi\left[\frac{A}{\sigma}\sqrt{\frac{1}{t}}\left(\frac{tb}{A}-1\right)\right] + \exp\left\{\frac{2Ab}{\sigma^2}\right\}\Phi\left[\frac{-A}{\sigma}\sqrt{\frac{1}{t}}\left(\frac{tb}{A}+1\right)\right] \tag{4.3.33}$$

Proof: See [Ch1].

The utility of this equivalence is that it allows the use of many known results for the normal distribution to be used in the analysis of the inverse Gaussian distri-

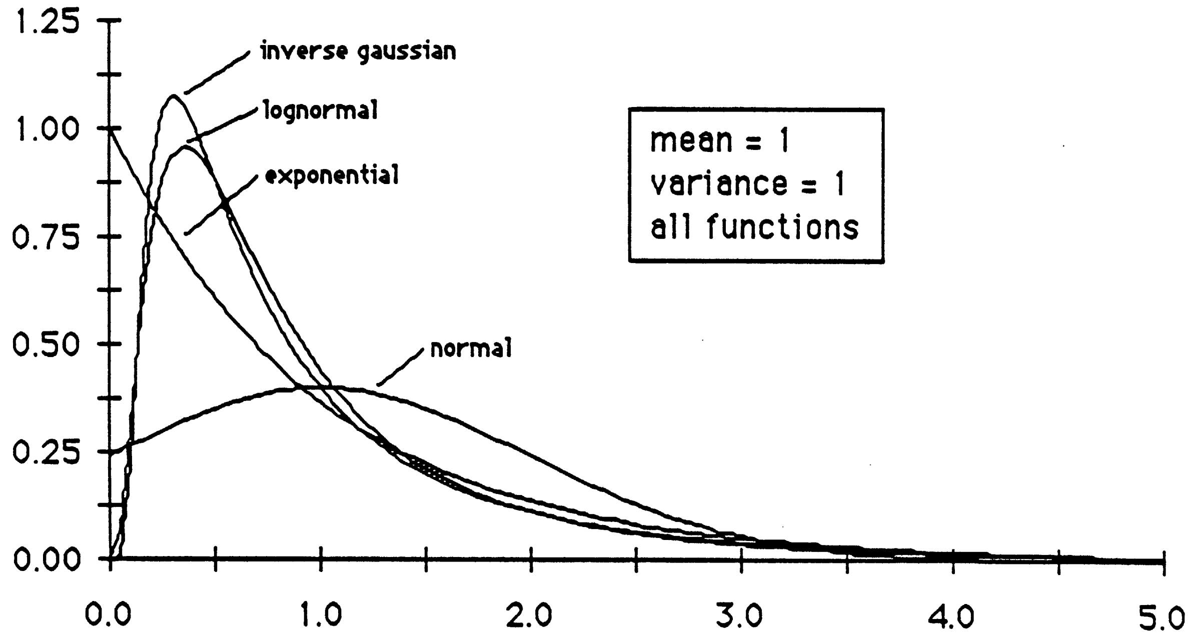

Fig. 4.3.1 Comparison of Density Functions

bution. In particular, bounds, limiting properties, and computational procedures for the normal distribution can be utilized.

It is possible to compute the threshold crossing probability for the case where the coefficients are piecewise constant. Since this is not used elsewhere in the thesis, the development is presented in Appendix B. However, the following theorem will be used.

Theorem (4.3.4): Given $X_t = bt + \sigma W_t$, $b(u) \geq 0$, $\sigma(u) > 0$. Define

$$M_t \triangleq \sup_{s \in [0,t]} X_s \tag{4.3.34}$$

Then the joint density of (X_t, M_t) is given by:

$$g_{X_t, M_t}(\eta, \nu ; t, u) =$$

$$\begin{cases} \frac{2(2\nu-\eta)}{\sqrt{2\pi\sigma^3} t^{\frac{3}{2}}} \exp\left\{\frac{-(2\nu-\eta-bt)^2}{2\sigma^2 t}\right\} \exp\left\{\frac{-2(\nu-\eta)b}{\sigma^2}\right\} & \text{for } \nu \geq \eta,\ \nu \geq 0,\ t > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.3.35}$$

Proof: See Appendix B.

4.4. Processes with Measurement Feedback

In the previous section, the availability of information feedback was not considered. Assume now that occasional measurements of the diffusion process are available. These measurements are to be used to update the threshold crossing probability. As will be seen, the threshold crossing distribution Q_A gives the updated probability as well.

Theorem (4.4.1): Given

$$X_t = \int_0^t b(u(\tau))d\tau + \int_0^t \sigma(u(\tau))dW_t \tag{4.4.1}$$

such that for $t' \geq 0$

$$u(\tau) \quad \text{arbitrary, } \tau < t' \tag{4.4.2}$$

$$u(\tau) = u',\ t' \leq \tau \tag{4.4.3}$$

$$b(u') = b' \geq 0 \tag{4.4.4}$$

$$\sigma(u') = \sigma' > 0 \tag{4.4.5}$$

and

$$X_{t'} = x' < A > 0 \tag{4.4.6}$$

Then

$$P[t' < T_A \leq t \mid X_{t'} = x', T_A > t'] = Q_A(t-t' \mid x' \ ; u') \tag{4.4.7}$$

Proof: From the Markov property

$$
\begin{aligned}
P[t' < T_A \le t \mid X_{t'} = x',\, T_A > t'] &= P[t' < T_A \le t \mid X_{t'} = x'] \\
&= P[T_A \le t - t' \mid X_0 = x' \; ; u'] \\
&= Q_A(t - t' \mid x' \; ; u') \qquad (4.4.8)
\end{aligned}
$$

This theorem generalizes to the more important case of feedback control where the control $u(t)$ is F_t^X measurable. This class of control laws preserves the Markov property so the result still applies. In particular, suppose that u_t is a Borel measurable function of the current state. That is

$$u(\tau) = u(X_\tau),\ 0 \le \tau < t' \qquad (4.4.9)$$

$$u(\tau) = u(X_{t'}),\ \text{constant for } t' \le \tau \qquad (4.4.10)$$

Then

$$P[t' < T_A \le t \mid X_{t'} = x',\, T_A > t' \; ; u] = Q_A(t - t' \mid x' \; ; u') \qquad (4.4.11)$$

That is, measurement of the process decouples the past and future, making future behavior of the process independent of past control laws. Of course, the future control must be constant in order to apply the results previously developed (or piecewise constant to use the results of Appendix B). In this work, the future control will generally be constant up to the time of the next measurement. This can be thought of as a type of sample-hold control law.

Remark: The result also holds when t' is a stopping time.

4.5. Multi-dimensional Processes

The treatment of multi-dimensional diffusion-threshold processes is generally even more difficult than the scalar case. One class of processes for which the distribution function can be obtained will be treated here. Although the class might seem restrictive, it does in fact have application.

The general theory of multi-dimensional diffusions will not be discussed here. It suffices to say that most of the scalar concepts generalize to higher dimensions. The class of processes considered here is sufficiently simple that the general theory is not required.

Consider the class of n-dimensional processes defined as follows. Let:

$$\underline{W}(t) \triangleq (W_1(t), \ldots, W_n(t))^T \tag{4.5.1}$$

be an n-dimensional standard Brownian motion with independent components where superscript T denotes transpose;

$$\underline{b}(\underline{u}(t)) \triangleq (b_1(u_1(t)), \ldots, b_n(u_n(t)))^T,\ b_i(u_i) \geq 0,\ i=1, \ldots, n \tag{4.5.2}$$

be an n-dimensional drift vector;

$$\underline{\sigma}(\underline{u}(t)) \triangleq \mathrm{diag}(\sigma_1(u_1(t)), \ldots, \sigma_n(u_n(t))),\ \sigma_i(u_i) > 0,\ i=1, \ldots, n \tag{4.5.3}$$

be an $n \times n$ diagonal diffusion matrix;

$$\underline{u} \triangleq (u_1, \ldots, u_n)^T \tag{4.5.4}$$

be the control vector. Define the n-dimensional diffusion process

$$\underline{X}(t) = \int_0^t \underline{b}(\underline{u}(\tau))d\tau + \int_0^t \underline{\sigma}(\underline{u}(\tau))d\underline{W}(\tau) \tag{4.5.5}$$

Next define a threshold in $\mathbb{R}^n$ as the union of hyperplanes

$$\underline{A} \triangleq \bigcup_{i=1}^{n} \{\underline{x} \in \mathbb{R}^n : x_i = A_i\},\ A_i > 0,\ i=1, \ldots, n \tag{4.5.6}$$

Let $T_{\underline{A}}$ be the time when the process $\underline{X}_t$ first hits the threshold $\underline{A}$.

$$T_{\underline{A}} = \inf\{t \geq 0 : \underline{X}_t \in \underline{A}\} \tag{4.5.7}$$

For fixed $\underline{u}$ the distribution function for the threshold crossing time $T_{\underline{A}}$ can be computed.

Theorem (4.5.1): Given $\underline{X}(t)$ and $\underline{A}$ as above, and fixed control $\underline{u}$, then:

$$P[T_{\underline{A}} < \infty] = 1 \tag{4.5.8}$$

and the threshold crossing probability distribution function is given by

$$Q_{\underline{A}}(t \mid 0 \,;\, \underline{u}) = 1 - \prod_{i=1}^{n}[1 - Q_{A_i}(t \mid 0 \,;\, u_i)\,] \tag{4.5.9}$$

Proof: Note that under the assumptions the process $\underline{X}(t)$ has independent components. The first statement follows easily from Lemma 4.3.1 since $T_{\underline{A}} \leq T_{A_i}$ where T_{A_i} is the threshold crossing time for component i and threshold A_i. To show the second part observe that

$$T_{\underline{A}} = \min_i(T_{A_i}) \tag{4.5.10}$$

Thus

$$\begin{aligned} P[T_{\underline{A}} > t] &= P[\min_i(T_{A_i}) > t] \\ &= P[(T_{A_i}) > t,\, i=1, \ldots, n] \\ &= \prod_{i=1}^{n} P[(T_{A_i}) > t] \\ &= \prod_{i=1}^{n}[1 - Q_{A_i}(t \mid 0 \,;\, u_i)] \end{aligned} \tag{4.5.11}$$

by independence. Therefore

$$Q_{\underline{A}}(t \mid 0 \,;\, \underline{u}) = 1 - \prod_{i=1}^{n}[1 - Q_{A_i}(t \mid 0 \,;\, u_i)\,] \tag{4.5.9}$$

Remark: This result also generalizes to allow the incorporation of measurement feedback, provided that all measurements are available at the same time. If this is not the case, the computation becomes more complicated.

4.6. Application of Diffusion-Threshold Processes

In this section several modeling applications of diffusion-threshold processes will be discussed. As will be seen, diffusion-threshold processes can be used to model a variety of phenomena. The applications will not be specific problems, but general categories of problems. Some of the motivation for these applications comes from manufacturing related problems, but certainly the applications are not restricted to manufacturing. More specific applications will be described in following chapters. An important aspect of each application is that the use of the diffusion-threshold model allows the problem to be viewed in a broader context than is typically considered. The ideas of control and feedback enter the problem in a natural way, allowing the problems to be placed in a control theoretic framework.

One of the most important characteristics of diffusion-threshold processes is the generation of discrete events by hitting the threshold. Assuming a threshold hit is followed by a process re-initialization, a sequence of discrete events is generated. If the cycles are independent, then the durations of the cycles are also independent random variables.

The diffusion-threshold process can be thought of as a discrete event generator. What makes it more powerful than simply assuming a sequence of random variables is the presence of the underlying diffusion process. This underlying diffusion is a continuous process whose state represents the "progress" made toward the threshold. Furthermore, the presence of this underlying process allows control and measurement feedback to naturally enter the discrete event generation through the diffusion.

Consider the following interpretation. Let the threshold value represent some critical value, such as the amount of work to do a job, or the extent of deterioration prior to failure for some device. The diffusion state then represents the current amount of work done, or the current extent of deterioration. The drift coefficient takes on the interpretation of mean work rate or mean rate of deterioration. The diffusion coefficient represents the magnitude of the uncertainty in the rate of state evolution. Conceptually, the diffusion-threshold model overlays a deterministic model and a noisy model to describe the system evolution.

When underlying processes are used to describe discrete events, the whole viewpoint towards control of discrete event systems can change. In particular, the essentially binary viewpoint traditionally taken (the event has or hasn't occurred) is replaced by a viewpoint that recognizes that the system evolution is in fact continuous. Feedback of progress measurement now makes sense. Furthermore, the memoryless feature commonly assumed is replaced by a system that definitely possesses memory. It is asserted that many physical systems have memory. The following application examples and discussions will hopefully clarify these points.

4.6.1. Failure and Reliability Model

One branch of reliability theory is concerned with the following types of problems: How long will a component last until it fails? What is the lifetime of a component? What are appropriate models for the lifetime and time to failure of a component? Each of the above questions can be generalized by substituting "system" for "component".

The approach usually taken is to consider such questions in a stochastic process setting. In this approach one describes the lifetime and time to failure of components by random variables. Reliability of components (and systems) becomes in part a problem of assignment of probability distributions and in part a problem of analyzing various properties related to these distributions.

The problem of assignment is a modeling issue. Sometimes a probability distribution can be selected on the basis of underlying physical properties, but often it will be selected on an empirical basis. In this case a distribution is assigned that fits the observed data. This assignment problem is at least bilevel. First a distribution family is chosen, then the parameters of the distribution function are chosen (estimated).

Reliability theory often places a special emphasis on the following problem. Given that a component has survived (gone without failure) for some period of time, what is the probability distribution of its remaining life, or time to failure? The conditional probability density that describes this conditioned behavior is called the hazard or failure rate function, and plays a prevalent role in the reliability literature. Formally, the failure rate function is the conditional density of the failure time, given that the failure time is greater than some value. There exists much work relating the failure rate function and the physical characteristics of the components being modeled. Also related to the characteristics of the failure rate function are the concepts of aging and memory. See [Mu1].

There are many devices, components, and systems for which the above formulation and description of reliability is a reasonable one. The claim made here is that there are other systems for which this formulation is neither adequate, nor entirely satisfying. Some of the deficiencies and limitations of this usual formulation of failure models will be examined.

First of all, an empirical formulation is never entirely satisfactory. One would prefer a description that incorporates at least some features derived from or motivated by a physical basis. Sometimes this is not possible due to a lack of understanding and then an empirical description has to suffice. This does not mean that an empirical formulation is not useful.

The second issue is concerned with the use of a single distribution to describe failure. Some devices are adequately described by one mode of use or operation. This means that the device is utilized in an essentially uniform way, and thus can be described by a single failure distribution. Some devices and systems do not readily admit such a description. For these devices a range of modes of use is possible. Furthermore, the mode of use influences the rate of deterioration (wear) of the device, and thus affects the failure rate of the device. When a single distribution is used to describe the failure probability of such a device, the description is an aggregate one. A more specific characterization of device failure would be a parameterized family of distributions, with the parameter a function of the mode of usage. This would be particularly useful if the mode of usage can be measured and/or controlled.

The third issue is the conditioning used in the failure rate approach. The information structure assumed by the usual formulation is that inspection of the device reveals only that the device has or has not failed. For many devices, it is reasonable to assume that this is indeed the only information readily available. Measurement of deterioration may be physically impossible, impractical, or perhaps the measurement process is destructive. Some devices, though, can be inspected, and the extent of deterioration and wear can be determined or at least approximated. For these devices, it would be desirable to use this additional information in determining the probability of failure in the future.

These various ideas can be unified into a control theoretic view of reliability that extends the usual formulation of reliability models. Suppose that the mode of use of a device is in fact a function of some control input. This usage will be interpreted as a rate of wear. This same control input also affects the performance of the system in some way. Examples might be speed of operation or time in use of a device or system. Assume that this rate of wear and performance are conflicting: higher or better performance necessitates a faster rate of wear. This rate of wear affects the probability of failure of the device or system. How can a good compromise between these two conflicts be determined? Suppose that occasional measurements of the extent of the wear of the device or system are available. This information (feedback) is to be used in the determination of a control value that achieves an appropriate (according to some criteria) balance between performance, rate of wear, and probability of failure, given some set of wear measurements. This is a controlled reliability system.

4.6.2. Diffusion-Threshold Model of Controlled Reliability System

The diffusion-threshold process can be used as a model for the controlled reliability system. Let X_t represent the wear, or deterioration of the system at time t. Let the threshold level A correspond to the wear level at which failure or unacceptable performance is considered to have occurred. The drift coefficient corresponds to a mean local rate of wear, and as such will be assumed to be strictly positive. Assume that the drift coefficient is a function of some control variable u. The diffusion coefficient will also be assumed to be strictly positive and a function of the control. Reference [Krt1] also considers the reliability problem in a similar way.

When the coefficients do not depend on the process X_t, the deterministic component of the wear at time t is given by:

$$\int_0^t b(u_\tau)d\tau \tag{4.6.2.1}$$

and the stochastic component of the wear at time t is given by:

$$X_0 + \int_0^t \sigma(u_\tau)dW_\tau \tag{4.6.2.2}$$

A simplifying assumption would be $X_0 = 0$ P-$a.s.$ This is equivalent to saying that new components have the same wear, assigned the value zero.

A special case of interest results when b and σ depend only on u and u is a constant function. This generates a family of distributions for the random variables $\{T_A^*\}$. The control u corresponds to the mode of usage and is assumed to remain constant over the lifetime of the component. In this case the parameterized family of conditional first crossing densities is given by Corollary 4.3.1. Similarly, a parameterized family of failure rate functions can also be calculated:

$$r(t \mid x_0\,;\,u) = \frac{q_A(t \mid x_0\,;\,u)}{1 - Q_A(t \mid x_0\,;\,u)} = \frac{q_A(t \mid x_0\,;\,u)}{\int_t^\infty q_A(\eta \mid x_0\,;\,u)d\eta} \tag{4.6.2.3}$$

The use of the inverse Gaussian density as a reliability model has been previously proposed. See [Ch1], and [Shr1]. However, these references do not consider the underlying diffusion-threshold model, and instead focus on the use of the inverse Gaussian distribution as an empirical model. The concepts of control and feedback are not considered. Under the assumption of constant distribution parameters, estimators for the parameters are developed. See also [Tw1] for estimator properties.

In [Ch1], the failure rate function associated with the inverse Gaussian distribution is examined. Important properties of this failure rate function are nonmonotonicity and nonzero asymptotic tail. Comparison is made to the lognormal distribution in particular. The authors in [Ch1] conclude that the inverse Gaussian distribution is a reasonable reliability model for many systems, including those with early failure where an increasing failure rate is not appropriate.

In [Shr1], the inverse Gaussian distribution is considered for a tool life model. The authors show that this distribution gives a slightly better fit than the lognormal distribution for a specific tool life data set.

In none of these references is the choice of the inverse Gaussian distribution strongly motivated by its relation to the diffusion-threshold model. The authors of [Ch1] do mention this connection, however. In this work, the suitability of the inverse Gaussian or other threshold crossing distribution is regarded as arising from the underlying diffusion-threshold processes which generates them, and the natural connection between wear and these processes.

4.6.3. Server Models

Queueing theory is concerned with the modeling of the interaction of servers and customers, and in particular the waiting of customers when services must be shared. In modeling this type of problem, queueing theory typically uses a probabilistic description of customer arrivals and service times. Queueing theory, formulated in this way, represents an important modeling technique for many types of systems: communication networks, computer systems, customer service facilities. See [Br1] and [Kl1] for detailed discussions of queueing theory, and [Cr1] for a controlled queueing bibliography.

Sometimes queueing theory is extended into an area where it is not evident that the above formulation represents an appropriate system description. One of the deficiencies that arises is the service time description. The purely probabilistic description is devoid of underlying structure. That is, what are the dynamics of the server that give rise to the service time distribution? If the probabilistic structure of the service time can be generated based on the dynamics of the server, it would allow the introduction of additional information, parameters, and controls regarding the server into the problem that are not usually considered. The use of a service time distribution to describe a server system results in the following viewpoint: either the service is or it is not complete. No significance is attached to a service that is partially completed. Knowledge of partially completed work does not normally enter the information structure of the problem. Therefore the conditional distributions of the service time given the amount of partially completed work are usually not considered.

In many actual service situations, the rate at which work is done by the server is influenced by many factors. These factors may include effects associated with the queueing system such as type of service required, number of customers waiting for service, and the performance of other servers in the system. Some influences may be thought of as control inputs: speed control of a machine or influencing an employee's work rate through various motivational schemes. For many systems, the rate at which work is done is a natural input for influencing the system. This rate may have a deterministic or stochastic interpretation. The deterministic interpretation would be a true work rate; the stochastic interpretation would be a mean work rate. How can this be incorporated into a probabilistic description of the service time? One approach would be to parameterize the service time distribution, and let the parameter be some (possibly random) function. The problem with this approach is determining a reasonable distribution and a method of parameterizing it which preserves the intuitive sense of work rate.

In some systems, it is possible to measure or approximate the amount of work required for a particular service in advance. When such knowledge is available, it would be desirable to include it in the service time description. Furthermore, it may be possible to determine the amount of work remaining on a particular service, while the service is in progress. This should also be incorporated into the model.

For some systems, it would be appropriate to extend the usual formulation of queueing problems so as to create a truly controlled queueing model. This model would include descriptions of the dynamics of the underlying service mechanisms. From these descriptions, the probabilistic structure for the service times would be determined. These models would also admit the concepts of partial work completion and the control of the rate of work.

4.6.4. Diffusion-Threshold Models of Servers

The diffusion-threshold process is now proposed as one model of a service mechanism. The intention is to model the underlying dynamics of a server by a diffusion with threshold and show that this formulation allows for the inclusion of work progress feedback and rate of work as an input. Let the diffusion process X_t correspond to the work completed on the current service at time t. Let the threshold value $A > 0$ correspond to the amount of work required to complete the service. Reaching A corresponds to completing the current service. At the completion of a service the process either resets or sojourns at A for an arbitrary time (until the next customer arrives).

The coefficient b takes on the physical interpretation of being the (local) mean work rate. This is presumed to be a function of some control variable u. The coefficient σ represents the magnitude of the uncertainty in the work rate, and the magnitude of uncontrolled influences on the server.

Suppose u is taken to be constant for the duration of each cycle of the diffusion-threshold process. This means that the rate of work for each service can be varied, but not in the middle of a service. Thus the service times are associated with the family of random variables $\{T_A^u\}$. The distribution of T_A is now parameterized by u and can be conditioned on the state of the current task $X_t = x < A$. This new model now admits the concept of a controlled queue as described before. The control input is a (mean, local) rate of work and information feedback may include the state of the current service which is the amount of work completed on the current service. Presumably, a controller could use this information and capability in order to achieve some level of performance, subject to some constraints. [Cr1] includes a list of references where service control is considered. However, none of the listed works presents a viewpoint similar to the one given here.

The service time distribution generated by the diffusion-threshold model can be compared to other distributions commonly used to describe service times, in particular exponentially distributed service times. For the diffusion-threshold model the probability of two events occurring "very close" (relative to the mean) in time is essentially zero. This is not so with the exponential distribution. For many physical systems, it is not likely that a service can take place arbitrarily fast. For these systems, then, the diffusion threshold model is a potentially more realistic model. Also, the diffusion-threshold model is not memoryless, and thus better describes service mechanisms with nearly deterministic service times. Many operations in manufacturing are of this type, particularly repetitive operations found in production and assembly systems.

4.6.5. System Failure Models

This application represents an extension to the application discussed in Section 4.6.1.

Consider the problem of trying to model the failure of systems, where attention is no longer restricted to individual components making up the system. In the spirit of Section 4.6.1, suppose also that the system admits a range of modes of usage or operation. Each mode of system usage causes each of the components to be subjected to some individual mode of usage. As before, the component's failure rate depends upon the mode of usage. Is it possible to formulate a failure model of the entire system that captures these features? Assume failure of the system to be indicative of the failure of any component which causes the system performance to degrade to an unacceptable level. Redundant systems or systems with intermediate degradation are not considered here. Two modeling approaches using diffusion-threshold processes are proposed.

The first approach is to use a scalar diffusion-threshold model as a reduced order, aggregated failure model. The representation is essentially the same as before, although the physical significance of some parts of the model may now be obscured. Since a scalar random variable is being used to describe the aggregated wear of the system, there may no longer be a physically meaningful relationship between the diffusion state and the system. Of course, the threshold value suffers in a similar way: the assignment of the maximum aggregated wear is no longer a simple problem.

There is a second approach that preserves some of the physical significance of the models in Section 4.6.2. Suppose the scalar diffusion assumption is removed by allowing the diffusion to be a vector process in $\mathbb{R}^n$ and the threshold to be a boundary in $\mathbb{R}^n$. Assume that each component of the diffusion represents an actual system component. In this case, the boundary might be a union of hyperplanes, as described by equation (4.5.6). Failure is indicated by the diffusion achieving any part of the threshold boundary. Since redundant components are not considered here, only those components whose failure will result in a system failure have an associated diffusion component and threshold hyperplane. Each component of the system has an associated component in the vector drift coefficient. The diffusion coefficient is now a matrix expressing the interaction of all the components of the system. Wear measurements consist of a vector of wear measurements, one for each component of the system. For the particularly simple case of full decoupling the probability distribution is given by Theorem 4.5.1. This corresponds to independent component wear in the system, but equal capability for causing system failure.

4.6.6. Project Models

This category can be thought of as an extension to the service models discussed in Section 4.6.3.

Consider a project or job requiring the cooperation of workers, and perhaps the coordination of equipment, resources, deliveries, etc. The completion of the project requires the completion of a certain amount of work. How can the progress of the work, the rate at which work is done, and the effect of uncontrolled or unanticipated influences on the project be modeled? A deterministic model is not generally correct; this is substantiated by practical experience and observation of the progress of several large projects, although some projects are less likely to have large variations in their schedules than others. What is required is the development of a model for the progress of projects that includes stochastic effects and the impact of work rate influence, and from which the probabilistic descriptions for the completion time for the project conditioned on the current state of the project and the amount of resources committed can be computed.

A diffusion-threshold process is again proposed as a model. Let the diffusion state X_t, represent the work completed at time t, and the threshold A represent the amount of work necessary to complete the project. As an aggregated variable, the scalar random variable X_t may not have a simple physical interpretation. Let the control represent the amount of resources committed to the project, effectively controlling the work rate on the project. Thus the drift coefficient would normally be an increasing function of the control. The control may affect the diffusion coefficient as well, reflecting the coordination problem associated with large task forces. Projects with large suspected variance in their schedules can be accommodated by suitable selection of the diffusion coefficient. Projects with large diffusion coefficients correspond to risky projects.

An interesting situation in this model is that $X_t < 0$, or locally decreasing X_t has a physical interpretation. Observe that errors or mistakes in the project cause a possible increase in the amount of work required to complete the project. This is no doubt familiar to the reader.

4.7. Relation to Other Work and Generalizations

The diffusion-threshold processes discussed in this chapter admit several extensions, some of which have been considered by other authors. Furthermore, there are related problems which have also been previously proposed by other authors. This section will try to briefly summarize some of this existing work.

The use of Brownian motion and diffusions to model various physical phenomena is not new. In fact, Brownian motion was first an observed physical process and then later a mathematical process. The deep mathematical properties of these processes offer several advantages to the modeler looking for ways to capture certain types of uncertain, erratic, noisy, or chaotic behavior. Consequently, the list of applications that have been considered is far too long to attempt to list here.

The concept of controlled diffusion processes is more recent. The references [Kry1] and [Fl1] present many of the basic ideas of controlled diffusions. However, both of these references, and in fact most of the existing literature on controlled diffusions assume continuous perfect measurements of the process and continuously variable control. These assumptions may be overly restrictive for some applications. Unfortunately, departure from these assumptions can cause many complications. In this work, these assumptions will be relaxed. See also [Be1].

Most references on Brownian motion and diffusions include discussions on stopping times, particularly exit times. The threshold crossing times considered in this work are exit times. [Kry1], [Fl1] and [Be1] discuss the relation between stopping times and controlled diffusions.

Another class of problems that are related to this work are inspection problems (called surveillance problems by Savage). See [An1], [An2], [Sa1]. In these problems, the control is usually of the discrete type: decide when to observe the process; decide when to re-initialize the process. The stochastic process being observed is often Brownian motion (though not always) but is otherwise usually not influenced. These problems are just examples of the very large class of stopping time problems.

In most of the literature, including the references cited here ([Kry1], [Fl1], [Be1]) the controlled diffusion is placed in a stochastic optimal control setting. The solution techniques generally rely on solving non-linear partial differential equations that arise from dynamic programming methods. By taking this approach, more general diffusions than those considered in this work can be handled. However, solutions are extremely difficult to obtain. Part of the motivation for the work here is to explore reasonable alternatives to solving partial differential equations.

There is an alternate method to solving PDE's, however, that still allows a fairly general class of diffusions. The method uses boundary layer methods to approximate the probabilistic descriptions of the threshold crossing times. See [Mat1]. The requirements for this approximation technique are that the diffusion coefficient be small enough. For many applications, this is in fact the case. This technique represents a good alternative to solving PDE's, and yet accommodates more general diffusion processes.

An important extension to the diffusion-threshold process is the inclusion of arbitrary jumps in the state. The diffusion-threshold process allows for a re-initialization action. This can be generalized to arbitrary (and random) jumps under the category of Impulsive Control [Be2]. These authors consider many examples of processes where both coefficient influence and impulsive control actions are permitted. Some of the applications considered in their work are related to applications discussed here. However, these authors prefer to place the problems in a quasi-variational setting. This necessitates the solution of a nonlinear partial differential inequality, which is very difficult. Furthermore, these authors generally assume continuous exact measurements and continuously variable controls. Nonetheless, [Be2] has many common features with the work presented here. The quasi-variational approach will be used for one of the problems considered in this thesis.

CHAPTER 5

THE DRILLING PROBLEM

5.1. Introduction

In this chapter, one manufacturing problem will be considered in some detail using the control theoretic approach. This problem is called the drilling problem, and has already been introduced in Chapter 3. Some simplifying assumptions will be made here in order to facilitate the presentation of the essential concepts. Crucial to the development of the drilling problem is the use of diffusion-threshold processes to model tool wear. The use of this modeling method allows the control theoretic approach to be taken.

The drilling problem is properly a machining economics problem. However, the features of the problem and the modeling and control methodologies used in this work are considerably different than the usual machining economics approach. The control theoretic approach imparts a distinctively different viewpoint.

Actually, two different approaches will be considered in this chapter. The first approach will use the diffusion-threshold model of tool wear, but will otherwise assume a traditional control policy and information structure. The second approach will explore the potential of additional control capability and the significance of tool wear measurement feedback. An example problem comparing both approaches will be discussed in the next chapter.

The drilling problem is intended as a prototype manufacturing problem. The specific use of a drilling operation in this problem is only to facilitate the presentation of the concepts and is not meant to imply a restriction of the ideas to that type of machining operation. The features of the problem have been designed to reflect important characteristics of manufacturing systems. In an actual system, some of these features may differ, typically resulting in a more complex problem. Further, there are important manufacturing issues that are not represented in the drilling problem. Nonetheless, the problem is an attempt to convey how the features of manufacturing problems affect the kinds of problem approaches that can be taken, and how new approaches can arise from the intelligent manufacturing setting.

5.2. Problem Description

A description of the drilling problem has been previously presented in Chapter 3. Only a brief review of the problem, along with more detailed discussions concerning specific features and assumptions will be given here.

A drill is used in a fixed operation carried out repetitively on parts. A fixed volume of material is removed from each part. As metal is removed from the parts, the tool wears. The tool wear mechanism is only known empirically, and variation in tool life is evident. As the tool wears, it eventually breaks or becomes unacceptable and must be replaced. Failure of the drill necessitates replacement, but can also cause damage to the part being machined.

The production of parts results in profit and thus the profit earned is related to the production rate. However, there are other factors affecting the profit, such as tooling costs and scrapped parts. Tool replacement is costly in terms of time and material. Tool breakage is in general more costly than simple replacement, due to possible part damage and the unplanned replacement that results.

Tool wear measurements may be occasionally available. If they are available, these measurements are to be used in the determination of operational policies. If tool wear measurements are available, they are only available when the tool is disengaged from the part. Tool wear measurements are not available during the drilling operation.

Certain features of the drilling problem are common to many manufacturing problems. These features will be briefly reviewed here. The discrete nature of parts is represented and imposes restrictions on the problem. The rate of production is specifically cited as a decision control variable. Tool failure affects the quality and quantity of the production output. Production rate and tool failure rate are not independent phenomena.

The objective in the drilling problem is to determine policies for tool replacement and for feed speed selection to obtain a suitable tradeoff between production rate and tool failure rate.

Appendix A has a description of the drilling terminology used in this work. Drilling is a machining process with two parameters: feed speed and spindle speed. For this work, the drilling problem will be reduced to a problem with one independent parameter. This parameter will be the feed speed. The spindle speed will be assumed to vary in proportion to the feed speed in order to maintain a constant ratio; i.e., the feed is assumed constant. This assumption simplifies the analysis, but does not represent a limitation of the model or the approaches used.

The term tool failure will mean breakage or wear past an unacceptable limit. Breakage and in particular early breakage of the tool is arguably a discontinuous process, and thus should be modeled by processes that can accommodate discontinuities. In this work, all tool failures will be aggregated with the understanding that early breakage (due to defective tools) is a relatively rare event. This is true of most production situations.

Feed speed control, tool replacement decisions, and tool wear measurements will be constrained depending upon the control approach being taken. For the age replacement policies, the feed speed will be fixed and tool replacement will be based upon the number of parts drilled. For feedback policies with variable feed speed, tool wear measurements will be available prior to starting a part. At that time, the tool may be replaced, and/or the feed speed may be changed. The feed speed will remain constant throughout the operation on a part.

In all cases, an instance of tool failure will cause the current part to be scrapped. Part repair will not be considered in this work. Furthermore, tool failure will be considered to be immediately detectable. This is not a hard assumption, but it simplifies the analysis. A tool replacement must occur after a tool failure before continuing.

There are many possible performance measures that can be used for the drilling problem. In fact, even if a single performance measurement is specified, there are usually secondary considerations that are important but not represented in the primary measure.

In this work, the primary performance measures will be economic. In particular, cost per time and cost per part will be used to evaluate control strategies. However, there are many secondary considerations that may be significant in a manufacturing setting. These include: amount of scrap produced, frequency of tool failures, frequency of tool changes, etc. The importance of these secondary considerations is due to the burden they may place on the manufacturing system through resource requirements, additional congestion, or disruption. As an example, frequent tool changes may necessitate greater tool handling capability and greater tool setup capability. This is over and above the pure cost of the tool and tool change. Although the economic issues are important, they are not the entire picture. Unfortunately, some of the existing literature in machining economics ignores this point.

5.3. Survey of Related Work

Aspects of the drilling problem have been considered by many authors. The work of Taylor [Tay1] is generally considered the first extensive treatment of tool wear and machining economics. One result of his work is the Taylor tool life formula relating tool life to cutting speed.

Much of the work after Taylor presumed that the tool life is deterministically related to the cutting speed and other machining parameters, using variations of the Taylor formula. Under the assumptions of a deterministic relation, simple calculus can be used to arrive at optimal machining parameters for any of several criteria. See also [Dr1] and [Hi1] for examples. Many refinements and extensions are possible, including the consideration of multiple machines, constraints on machining parameters due to finish and power requirements, and tool geometry. Usually these treatments assume that the tool is changed at the end of its life, as given by the tool life formula, and that tool breakage and scrap parts do not occur.

More recent work in the area has recognized the stochastic aspects of tool life and the impact that tool life uncertainty has on machining operation productivity. In these works, tool life is viewed as a random variable whose distribution is parameterized by cutting speed and other factors. Some researchers have assumed particular distributions, while a few have only specified moments of the tool life random variable. See [De1], [Er1], [Ke1], and [Shk1].

The machining economics problem, whatever the assumption about tool life, usually considers as performance measures production rate, cost, or time to produce a part. Various types of constraints have been proposed for inclusion in the problem. Several things are not usually considered, however. The concept of information feedback and on-line control of machine operation is almost never considered. As a result, most machining optimization problems have assumed constant machining parameters. On-line variability of machining parameters based on tool wear measurements and other information is not usually considered. Policies for tool change are invariably defined in terms of information such as the number of parts machined, or change upon failure. Costs due to scrap and damage, or the status of other machines in the manufacturing system are not considered in tool replacement policies. Most importantly, tool wear information is not incorporated into the policies.

5.4. The Tool Wear Subproblem

The problem of tool wear is central to the drilling problem. Without an appropriate model that captures the essential features of tool wear, the drilling problem cannot be readily approached in a control theoretic way. Consequently, tool wear model development is very important. Some background material is presented in this section that summarizes existing approaches to the tool wear subproblem.

The tool wear problem has been the subject of much study, and an extensive body of literature reflects this. There are at least two views of the problem that can be taken. The first is to explain on a physical basis the mechanisms of tool wear. The attempts in this area often lead to very complicated formulations for tool wear as a function of several variables. See [Da1], [Kan1] and [Ko1] as examples. Unfortunately, these models are not complete, describing at best the normal evolution of some aspects of tool wear under restricted conditions. Since the manufacturing environment is uncertain, and the information requirements of these models can be considerable, their current utility in the machining economics problem is debatable.

The second view is that usually taken in machining economics. Precise description of tool wear phenomena is forsaken for empirical formulas that are easier to use, and that capture the essential character of tool life as a function of machining conditions. The empirical formulation of tool wear can be considered either deterministically or stochastically. Deterministic formulations are prevalent in the earlier literature, but more recently the stochastic nature of tool life has been recognized. The stochastic formulations usually assume that the deterministic formulas properly represent mean values of tool life. References [De1], [Hi2], [Ke1], and [Wa1] give various viewpoints of the tool wear problem. Also, the references on machining economics usually give some tool wear model.

The stochastic formulation brings with it another level of complexity. If tool life is a random variable, its distribution must be specified. Actually, a family of distributions parameterized by machining conditions must be specified. Many distributions have been proposed for tool life including exponential, normal, and lognormal. In general these distributions are arrived at from purely empirical considerations, and not from any physical basis. This is not an entirely satisfying situation. It should be mentioned that some authors have avoided this problem by parameterizing only the moments of the tool life (usually the mean and variance) and not specifying a distribution.

It is asserted that a new model of tool wear is required in order to accommodate the intelligent manufacturing approach. This new model must incorporate certain features. The stochastic behavior of tool life must be recognized. The influence of machining conditions must enter into the model in a clear way. The distributions should arise from physical considerations of the process. The results of the model should agree with observed behavior in a statistical sense. This leads to the development of a diffusion-threshold model for tool wear.

5.5. Diffusion-Threshold Model of Tool Wear

5.5.1. Model Formulation

The diffusion-threshold process discussed in Chapter 4 is now proposed as a model for tool wear in the drilling problem. The following analogy is made. Let the diffusion process $\{X_t\}$ represent an aggregated wear variable for the tool. It is understood that all of the categories of tool wear are somehow represented by a single variable, and X_t therefore represents the magnitude of wear at time t. The feed speed will be regarded as the control input u_t, with the drift coefficient $b(\cdot)$ being a positive increasing function of the feed speed. The drift denotes a local mean rate of wear. The diffusion coefficient is the square root of the local rate of change of the variance. Let the maximum allowable tool wear correspond to the threshold A. When the wear reaches the threshold, the tool is considered unacceptable and must be replaced by a new tool. The new tool is assumed to have zero wear. This is the re-initialization action. Likewise, the decision to replace the tool results in a re-initialization prior to reaching the threshold. The decision times will take place between parts; i.e., just prior to the start of a new part. At these decision times, denoted $\{t_k\}$, information concerning drill wear may become available. In this work, assume that exact measurements of the drill wear, X_{t_k}, become available if tool wear feedback is allowed.

5.5.2. Rapproachment with Taylor Tool Life Formula

For the drilling problem and other tool wear problems, a comparison can be made between the diffusion-threshold model and other more usual models. A comparison to the (simple) Taylor tool life formula [Dr1] will be considered. The Taylor formula, in use for many years, is a strictly empirical formula for calculating tool life. As is shown in Appendix A, for fixed feed the Taylor formula may be written as

$$uT^n = C_1 \tag{5.5.2.1}$$

where:

$u \triangleq$ the feed speed

$T \triangleq$ life of tool

$n \triangleq$ an empirical constant dependent upon the part material, tool type, and machine type

$C_1 \triangleq$ an empirical constant dependent upon the part material, tool type, and machine type

Assume that replacement is indicated by having reached a certain level of wear that is fixed for a given tool type. Let this level of wear be denoted by W and let the wear of the tool at time t be denoted $w(t)$. Also, define the instantaneous feed speed by $u(t)$. If the tool life given by the Taylor equation is interpreted as a mean value, a diffusion threshold model with the same expectation can be constructed. Let the drift coefficient be given by

$$b(u) = \frac{W}{C_1^{\frac{1}{n}}} u^{\frac{1}{n}} \tag{5.5.2.2}$$

so that

$$E[w(t)] = \int_0^t \frac{W}{C_1^{\frac{1}{n}}} u(\tau)^{\frac{1}{n}} d\tau \tag{5.5.2.3}$$

When the feed speed is fixed at a constant value u, the expected value for the the tool life can be computed as

$$E[T] = \frac{W}{b(u)} = \left(\frac{C_1}{u}\right)^{\frac{1}{n}} \tag{5.5.2.4}$$

in agreement with the Taylor tool life formula. No restrictions are placed on the diffusion coefficient σ by the Taylor formula.

In this work, σ will be assumed constant. Note however that this does not imply that the variance of the tool life is independent of the feed speed.

5.5.3. Explanation, Justification and Limitations of the Model

Some explanation of the diffusion-threshold model of tool wear is required in order to justify its use. Further, the restrictions of the model need to be presented so that limitations on its use are understood.

The diffusion-threshold model of tool wear is motivated by several issues. First, the difficulty associated with trying to model tool wear deterministically using physical laws. This difficulty arises from the nature of the tool wear process: it is complex and chaotic, and it takes place in an uncertain environment with unmeasurable conditions. These statements are particularly oriented toward the production environment. Second, there are observed differences in tool wear evolution under similar operating conditions. Third, the system has a hybrid characterization; that is, there is tool wear and there is tool failure.

There are several goals that the modeling effort should try to achieve, if possible. First, the model should capture a suitable analytic description of system evolution. This requires presenting an appropriate view of the phenomena (macroscopic vs. microscopic) and that the level of detail incorporated into the model be compatible with the intended use of the model. Second, the model should use an established mathematical base, and the model should be in some way tractable to be of value. Third, the model should have a certain intuitive appeal, although this is hard to quantify. Fourth, the model must be appropriate for use with the control theoretic viewpoint.

The important characteristics of the diffusion-threshold model are as follows. Brownian motion represents the prototype chaotic system or system driven by noise. Without deeper understanding of the problem, Brownian motion is often a reasonable first approximation for systems that exhibit chaotic or noisy behavior. Diffusions are then the natural generalizations of Brownian motion. Properties of diffusions that facilitate mathematical analysis include the strong Markov property and continuous sample paths. Furthermore, the diffusion-threshold model has a certain intuitive appeal: a system evolves in time until some limit is reached. Finally, there is some agreement between experimentally observed tool wear processes and the diffusion-threshold model. This is evident in the similarity between the inverse Gaussian density and densities that have been proposed for tool life without any physical basis;

e.g., lognormal.

The diffusion-threshold model also invites extension to more general diffusions with more complex coefficients. This does introduce computational difficulties, however. The model is thus extensible as additional properties of tool wear are discovered or proposed. This extensibility is very important.

The model is not without limitations, however. Brownian motion (the mathematical process) is an abstraction and not a true characterization of the phenomena at the tool/part interface. Brownian motion is too erratic (since sample paths have unbounded variation) to truly represent a physical system. The Markov property inherent in diffusions is probably not valid for tool wear. As an example, material properties can depend on past histories and not just on current conditions. Unfortunately, deviation from the Markov assumption is hard to handle mathematically. Negative excursions of the diffusion are possible, but this has no physical interpretation in terms of tool wear. However, the probability of sustained negative values will in general be very small. This is similar to assuming Gaussian distributions for random variables that in fact are constrained in magnitude. The lack of known properties for the diffusion coefficient is disturbing, but not really a model limitation. It is more a limitation of available data. Finally, it is not clear whether or not the model captures certain types of catastrophic failure correctly. Since the diffusion process is continuous, it cannot literally represent jump phenomena. However, it can be argued that small jumps are still accommodated in a probabilistic way. Whether or not large jumps are accommodated probabilistically depends upon the diffusion construction (the coefficient selection). Generally, the model will not be assumed to represent early tool failure due to defective tools or parts. It is possible to extend the model to allow for random reinitializations. This extension may capture premature catastrophic phenomena.

5.6. Age Replacement Policies for the Drilling Problem

In this section, traditional control policies for machining economics problems will be considered. Traditional policies assume a fixed feed speed and replacement of the tool after either a tool failure or a predetermined number of parts has been processed. In other words, replacement takes place after a fixed machining time; hence the name age replacement. The diffusion-threshold model of tool wear will be used in this section. Furthermore, the features of the drilling problem will remain as described. This results in a problem formulation and solution method that is different than typical machining economics problems. Therefore, even though traditional policies are studied in this section, new results are obtained.

The section will begin with a discussion of age replacement policies for diffusion-threshold processes, and a discussion on renewal theory. The application of these ideas to the drilling problem specifically will follow.

5.6.1. Diffusion-Threshold Processes and Age Replacement Policies

Under age replacement policies, the evolution of the diffusion-threshold process is influenced by selection of the variable u, and by allowing the process to be reinitialized after a certain age has been reached. In this section, all reinitializations will be to zero. A control policy π consists of a pair (u,t_r) where u determines the infinitesimal coefficients in the diffusion and t_r is the replacement age. If the current cycle reaches age t_r without a threshold crossing, the cycle is terminated. If the threshold is reached prior to t_r then that event terminates the cycle. This class of policies will be denoted age replacement, but the availability of the selection of u

generalizes the usual use of this term.

In the analysis of the drilling problem under age replacement policies, it will become convenient to alter slightly the policy definition to an equivalent description. The reasons for doing this will become evident.

The probabilistic structure of a cycle under an age replacement policy can be developed. Assume an age replacement policy $\pi = (u, t_r)$. Let T be the random duration of one cycle under this policy. Then

$$T = \min(T_A, t_r) \tag{5.6.1.1}$$

where T_A is the threshold crossing time and t_r is the replacement age. The distribution function F_T^π is then given by:

$$F_T^\pi(t) = \begin{cases} Q_A(t \mid 0 \,;\, u) & \text{if } t < t_r \\ 1 & \text{if } t \geq t_r \end{cases} \tag{5.6.1.2}$$

where

$$Q_A(t \mid 0 \,;\, u) = \int_0^t \frac{A}{\sqrt{2\pi}\sigma\eta^{\frac{3}{2}}} \exp\left\{\frac{-(A - b\eta)^2}{2\sigma^2\eta}\right\} d\eta \tag{5.6.1.3}$$

is just the threshold crossing probability distribution. Note in particular that

$$P^\pi[\, T = t_r \,] = 1 - Q_A(t_r \mid 0 \,;\, u) > 0 \tag{5.6.1.4}$$

so that $T = t_r$ with positive probability.

The expected value of T can be computed from the above:

$$\mu_T \stackrel{\Delta}{=} E^\pi[\, T \,] = \int_0^\infty t dF_T^\pi(t) = \int_0^{t_r} t dQ_A(t \mid 0 \,;\, u) + t_r(1 - Q_A(t_r \mid 0 \,;\, u))$$

$$= \int_0^{t_r} \frac{A}{\sqrt{2\pi}\sigma\eta^{\frac{1}{2}}} \exp\left\{\frac{-(A - b\eta)^2}{2\sigma^2\eta}\right\} d\eta + t_r \cdot \int_{t_r}^\infty \frac{A}{\sqrt{2\pi}\sigma\eta^{\frac{3}{2}}} \exp\left\{\frac{-(A - b\eta)^2}{2\sigma^2\eta}\right\} d\eta \tag{5.6.1.5}$$

5.6.2. Renewal Processes in the Drilling Problem

The drilling problem, along with many other manufacturing processes, can be described in terms of a cyclic sequence of events. The nature and number of events as well as their temporal relationship in any given cycle may be random, however. Furthermore, more than one cycle of events may be present. In the drilling problem, two obvious event cycles are the part processing cycle and the tool cycle. Sometimes it is possible to identify a sequence of cycles that may be adequately modeled as probabilistically independent and with identically distributed duration. For the drilling problem under age replacement policies, tool cycles can be so modeled. When such a sequence is available, and is such that it can be used to describe the long term performance of the system, then a renewal theory approach is suggested. Placing the problem in a renewal setting has the great advantage of allowing the use of several powerful theoretical results.

5.6.3. Application of Renewal Theory

In this section several key renewal theory results will be applied to the drilling problem. For a more complete treatment of renewal theory see [Fel], [Karl], [Rol].

Let $\{T_i\}$ be the set of tool cycle durations. That is, the random variable T_i is the duration of the i^{th} tool cycle. Assume that these random variables are independent and identically distributed with common probability distribution function $F_T^x(\cdot)$. Let

$$E[\ T_i\ ] = \mu_T < \infty \tag{5.6.3.1}$$

Define the set of random variables $\{S_n\}$ by

$$S_0 = 0 \tag{5.6.3.2}$$

$$S_n = \sum_{i=1}^{n} T_i \tag{5.6.3.3}$$

The random variable S_n then corresponds to the cumulative time of the first n tool cycles, or the time at which the n^{th} tool cycle completes. The first cycle is understood to begin at $t = 0$.

Define the counting process $\{N_t : t \geq 0\}$ by

$$N_t = \max\{n \geq 0 : S_n \leq t\} \tag{5.6.3.4}$$

The random variable N_t is the number of complete tool cycles that have occurred up to time t. Note that the cycle on-going at time t is not counted unless t is its completion time. The process $\{N_t\}$ is a renewal counting process.

Define the function $m : \mathbb{R}^+ \to \mathbb{R}^+$ by

$$m(t) = E[\, N_t \,] \tag{5.6.3.5}$$

This function is the renewal function. The value of the renewal function at time t is the expected number of completed tool cycles that have occurred by time t.

The advantage to be gained by placing the drilling problem in a renewal setting is that long term performance can be assessed using only statistics for a single tool cycle. Since the computation of single cycle statistics is generally easier (though not necessarily easy) than the computation of long term statistics, the attraction of the renewal setting is evident.

The number of tool cycles per unit time satisfies the relation

$$\lim_{t \to \infty} \frac{N_t}{t} = \frac{1}{\mu_T} \text{ with probability 1} \tag{5.6.3.6}$$

by the strong law of large numbers. The elementary renewal theorem [Ro1] states that the expected number of tool cycles per unit time is similarly described:

$$\lim_{t \to \infty} \frac{m(t)}{t} = \frac{1}{\mu_T} \tag{5.6.3.7}$$

Assume now that for each tool cycle, some cost is incurred. Assume that the cost incurred in one tool cycle is independent of all other cycles and their associated costs. A dependence between the duration of a tool cycle and the cost for that tool cycle is permitted. Under these circumstances, the expected long term cost per time can be computed on the basis of single tool cycle statistics.

Let $\{c_j\}$ be the set of tool cycle costs, with c_j the cost for the j^{th} tool cycle. Assume that these random variables are independent and identically distributed with

$$E[\ c_j\ ] = \mu_c < \infty \tag{5.6.3.8}$$

Define the cumulative cost function

$$C_t = \sum_{j=1}^{N_t} c_j \tag{5.6.3.9}$$

The random variable C_t is the cost incurred through the first N_t tool cycles. Then the long term cost per unit time is given by the following [Ro1]:

$$\lim_{t \to \infty} \frac{C_t}{t} = \frac{\mu_c}{\mu_T} \text{ with probability 1} \tag{5.6.3.10}$$

and

$$\lim_{t \to \infty} \frac{E[\ C_t\ ]}{t} = \frac{\mu_c}{\mu_T} \tag{5.6.3.11}$$

That is, the expected long term average cost per time converges to a value that may be computed by considering only statistics for a single tool cycle.

Remark: The above result holds even when the cost is not lump sum. That is, even if costs are distributed throughout the cycle, the results remain valid ([Kar1], [Ro1]).

An extension is required in order to allow for a control structure in the problem.

5.6.4. Controlled Renewal Cost Processes

The renewal theoretic view starts with the collection $\{T_i\}$ of random times. It is reasonable to inquire into the nature of the physical processes that give rise to these random variables. If these physical processes can be influenced by external inputs, then a means exists for controlling probabilistically these random variables. This gives rise to the concept of a controlled renewal process.

Assume that there are inputs available that affect the probability distribution of the random variables $\{T_i\}$, and the associated costs $\{c_i\}$. Then as long as the control policy preserves the renewal structure, the renewal results can be used to evaluate the performance of the policy. Of course, if the policy does not preserve the renewal structure, other evaluative means must be employed.

It is sufficient to identify the controls available and their effect on the indicated random variables. This is sufficient, but not entirely satisfying because it lacks a sense of physical cause. Instead, consider a different approach. Suppose that an underlying process is identified that gives rise to the renewal cycles. That is, this underlying process is a model of the dynamics of the system with an explicit representation of how external inputs affect its evolution. If the underlying process is such that the probabilistic structure of the renewal can be computed, then the nature and effect of the control can be inferred instead of being hypothesized. This motivates the use of diffusion-threshold models to describe the underlying dynamics of renewal processes.

5.6.5. Analysis of the Drilling Problem under Age-Replacement Policies

Assume a tool cycle begins with a new tool, and ends after the tool is replaced. During the tool cycle the following events can occur: parts are produced causing tool wear; the tool reaches the replacement age or else it fails; tool failure produces scrap; finally, a new tool is installed. Assume that the tools are independent, and tool wear is described by a diffusion-threshold process. Under age replacement policies, the tool cycles generate a renewal process.

In order to preserve the interpretation of the variables under control, a slight alteration in the definition of a policy is required. The continuous variable under control is the feed speed u. The drift and diffusion coefficient are then functions of u. Using the Taylor tool life approximation, an appropriate form for the drift is $b = \beta u^m$. The diffusion coefficient σ will be assumed constant. In order to conform with the drilling problem formulation, age will be expressed in terms of number of parts completed, denoted by n_r. This also conforms with typical specifications in an actual manufacturing environment. Therefore, a policy will be defined as $\pi = (u, n_r)$ where $u \in (0,\infty)$ and $n_r \in \{1,2,...,\infty\}$. An implicit assumption is that $n_r \geq 1$; i.e., that at least one part can be completed by one tool on the average. The special case $n_r = \infty$ will denote a failure replacement policy. Under a failure replacement policy, the tool is replaced only when it fails. The time of replacement is now a function of both u and n_r. Let the processing time for one part be given by $t_f = \dfrac{V}{u}$ where V is the depth of the hole to be drilled. Then $t_r = n_r t_f = n_r \dfrac{V}{u}$.

Associated with each tool cycle is a cost. This cost consists of any resources used (including time) minus any profit earned for completed parts. Define the following parameters:

$G \triangleq$ profit earned per completed part

$R \triangleq$ cost of a tool replacement, including tool and time costs

$B \triangleq$ cost of a tool breakage, not including replacement cost

$K \triangleq$ overhead cost rate for machine in operation

$t_R \triangleq$ replacement time

$t_B \triangleq$ repair time

Overhead costs incurred during tool replacement and repair are assumed to be included in R and B respectively. Note that a total cost of B+R is associated with a threshold crossing (breakage or unacceptable wear). In the following formulation it is assumed that *all* profit for the current part is lost when a breakage occurs. This conforms to the assumption that failed tools produce scrap.

The replacement time t_R will be a part of every cycle. The repair time t_B will be a part of those cycles with tool failure. It is convenient to introduce the random variable $\widetilde{P}$ as the number of parts produced in one tool cycle (not counting scrap parts). Let $\widetilde{T}$ denote the random duration of one tool cycle, and let $T = \min(T_A, n_r \frac{V}{u})$. The random variable $\widetilde{T}$ is the duration of the total tool cycle, including repair and replacement time. The random variable T is the tool life. As stated earlier, the long term expected cost per unit time can be approximated by the expected cost per expected time for one tool cycle. The expected long term cost per time is given by:

$$J_t^{\pi} \triangleq \frac{E^{\pi}[\text{ cost/cycle }]}{E^{\pi}[\text{ duration of cycle }]} = \frac{E^{\pi}[\, R + KT + B1_{[T=T_A]} - G\widetilde{P}]}{E^{\pi}[\, \widetilde{T}\,]} \qquad (5.6.5.1)$$

Also of interest will be the expected long term cost per part:

$$J_p^\tau \triangleq \frac{E^\tau[\text{ cost/cycle }]}{E^\tau[\text{ part/cycle }]} = \frac{E^\tau[\, R + KT + B 1_{[T=T_A]} - G\tilde{P}]}{E^\tau[\, \tilde{P}\,]} \tag{5.6.5.2}$$

Strictly speaking, there is an ambiguity that arises when $T_A = n_r \dfrac{V}{u}$; since this occurs with probability 0, the assumption that the part is scrapped presents no difficulties. Also note that overhead costs for repair and replacement could have been explicitly included in the cost function.

In order to compute the performance statistics of interest, the following expectations must be determined:

$$E^\tau[\, T\,],\ E^\tau[\, \tilde{T}\,],\ E^\tau[\, \tilde{P}\,]$$

Note that

$$E^\tau[\, \tilde{T}\,] = E^\tau[\, T\,] + E^\tau[\, t_B 1_{[T=T_A]}\,] + E^\tau[\, t_R\,]$$

$$= E^\tau[\, T\,] + t_B\, Q_A(n_r \frac{V}{u} \mid 0\, ;\, u) + t_R \tag{5.6.5.3}$$

and $E^\tau[\, T\,]$ has been previously computed:

$$E^\tau[\, T\,] = \int_0^{t_r} \frac{A}{\sqrt{2\pi}\sigma\eta^{\frac{1}{2}}} \exp\left\{\frac{-(A - \beta u^m \eta)^2}{2\sigma^2\eta}\right\} d\eta$$

$$+ t_r \cdot \int_{t_r}^{\infty} \frac{A}{\sqrt{2\pi}\sigma\eta^{\frac{3}{2}}} \exp\left\{\frac{-(A - \beta u^m \eta)^2}{2\sigma^2\eta}\right\} d\eta \tag{5.6.5.4}$$

The computation of $E^\tau[\, \tilde{P}\,]$ is somewhat more complex. Observe that:

$$\tilde{P} = \sum_{j=0}^{n_r-1} j \cdot 1_{[T\in(jt_f,(j+1)t_f]]} + n_r 1_{[n_r t_f < T_A]} \tag{5.6.5.5}$$

Therefore,

$$E^\tau[\, \tilde{P}\,] = \sum_{j=0}^{n_r-1} j \cdot [\, Q_A(\,(j+1)t_f \mid 0\, ;\, u) - Q_A(jt_f \mid 0\, ;\, u)\,] + n_r \cdot [\, 1 - Q_A(n_r t_f \mid 0\, ;\, u)\,]$$

$$= n_r[\, 1 - Q_A(n_r t_f \mid 0\, ;\, u)\,] + [n_r - 1][\, Q_A(n_r t_f \mid 0\, ;\, u) - Q_A((n_r-1)t_f \mid 0\, ;\, u)\,] + \cdots$$

$$\cdots + [\, Q_A(2t_f \mid 0\, ;\, u) - Q_A(t_f \mid 0\, ;\, u)\,]$$

$$= n_r - \sum_{j=0}^{n_r-1} Q_A(jt_f \mid 0 ; u) \tag{5.6.5.6}$$

Using these expressions, the performance statistics J_t^π and J_p^π can be computed. Substituting into (5.6.5.1) and (5.6.5.2) yields

$$J_t^\pi = \frac{R + KE^\pi[\ T\ ] + BQ_A(t_r \mid 0 ; u) - GE^\pi[\widetilde{P}]}{E^\pi[\ T\ ] + t_B Q_A(t_r \mid 0 ; u) + t_R} \tag{5.6.5.7}$$

Also of interest will be the expected long term cost per part:

$$J_p^\pi = \frac{R + KE^\pi[\ T\ ] + BQ_A(t_r \mid 0 ; u)}{E^\pi[\ \widetilde{P}\ ]} - G \tag{5.6.5.8}$$

with $E^\pi[\ T\ ]$ given by (5.6.5.4) and $E^\pi[\ \widetilde{P}\ ]$ given by (5.6.5.6). In general, numerical integration can be used to compute these functions for a given policy π. In Chapter 6, algorithms will be given that allow the evaluations to be carried out without numerical integration. This facilitates rapid evaluation of a policy's performance.

5.6.6. Optimal Age Replacement Policies

Since the economic cost functionals for the drilling problem under age replacement policies can be evaluated, it is possible to consider optimal age replacement policies with respect to the cost functionals. An optimal policy π^* consists of a pair (u^*, n_r^*) such that $J^{\pi^*} \leq J^\pi$ for all admissible policies π.

Since the variable n_r is an integer, the optimization procedure is a mixed integer programming problem and is not straightforward. There are several possible approaches. The first approach is to use one of the available optimization algorithms for handling mixed programs. Since the system is of low order, this is not unreasonable. A second approach is to assume that n_r is a positive real number and simply roundoff the result. Although no statement of the reasonability of this scheme is presented here, some numerical examples suggest that it is not a bad approach. A third approach, and one that is used in a later chapter, is an explicit calculation of the cost functionals for fixed age replacement to determine optimal costs for that age, followed by a search among the ages. This can be done graphically and has proven to be quite effective and very fast. This does rely on fast algorithms for evaluating the various functions. These will be discussed in a later chapter.

The optimal policy is very simple to describe. All drilling takes place at the optimal feed speed u^*. A counter maintains the number of parts completed by the current tool. The tool is replaced when the tool fails or when the counter reaches n_r^*, whichever occurs first. The replacement of the tool resets the counter. The simplicity of the policy allows a variety of possible implementations.

5.7. Feedback Policies for the Drilling Problem

In this section, the drilling problem will be examined in an intelligent manufacturing setting. In this setting, information and control assumptions will differ from the usual machining economics problem formulations. This will allow the exploration of a new class of decision policies. In the next chapter, a comparison between the two approaches (in Sections 5.6 and 5.7) will be made for an actual problem. Since tool wear measurements will be assumed available for the policies considered here, they will be called feedback policies. However, variability of the feed speed is also assumed.

5.7.1. An Optimal Control Problem: General Formulation

Consider the drilling problem previously described. Assume that pieces are being machined, and that the time to complete each piece depends on the selected feed speed. Furthermore, assume that measurements of tool wear can be made at the completion of each piece. At each measurement time, a decision can be made to adjust the feed speed, and/or change the drill. Each of these decisions has some cost associated with it. In addition, there is a penalty cost associated with breaking the drill. The optimization problem is to make these decisions in such a way as to minimize some cost functional.

The formulation of cost functionals for feedback policies presents some difficulties. The cost functional used must capture the important manufacturing concerns and allow for the additional information and control capability indicated. Ideally, cost functionals similar to those introduced for age replacement policies would be used. These cost functionals capture the long term performance of the system. In order to use the renewal theory setting, it is necessary to restrict attention to feedback policies which preserve the renewal structure. This is not a serious restriction. A far more serious problem is the computational difficulties encountered when feedback policies are considered. The probability functions involved are extremely difficult to compute.

An alternate approach might be to formulate a dynamic programming problem over one tool cycle (or an even shorter horizon). This again presents major complications because it involves programming over stopping times with incomplete information. Furthermore, the probability functions are again very difficult to compute.

The difficulties associated with these approaches motivates the consideration of alternate cost functionals. An alternate formulation should yield good performance, but be computationally tractable. In particular, on-line computability would be desirable. This is because the actual manufacturing system and its parameters are stochastically varying in time. Long term optimality would be willingly sacrificed. These considerations have resulted in the formulation of a class of cost functionals different than those introduced previously in Section 5.6.

The class of cost functionals considered in this analysis are called one step. One step in this context means one part ahead. One step costs incorporate the profits earned and the costs incurred in producing the next part. The formulation of one step costs must be carefully considered. In particular they must be formulated so that behavior extremely detrimental to successive parts is not encouraged.

For this problem, define a control vector $\underline{u} \triangleq (u,v)$ where $u \in (0,\infty)$ is the feed speed and $v \in \{0,1\}$ is the replacement decision, where 0 corresponds to not replacing the tool and 1 corresponds to replacing the tool.

Define the following economic parameters:

R $\triangleq$ cost of a tool replacement, including tool and time costs

B $\triangleq$ cost of a tool breakage, not including replacement cost

K $\triangleq$ overhead cost rate for machine in operation

and the functions:

g(u) $\triangleq$ profit rate for selected feed speed u

$T_f(u) \triangleq$ processing time for a part corresponding to a constant feed speed u

and the random variables:

$M_{[t_0,t_1]} \triangleq \sup_{t_0 \le t \le t_1} \{X_t\}$ the maximum excursion of the process over $[t_0,t_1]$

$h(X_{t_k+T_f}, M_{[t_k,t_k+T_f]}, x_k) \triangleq$ tool utilization cost for processing a part

Overhead costs incurred during tool replacement are assumed to be included in R. Note that a total cost of B+R is associated with a threshold crossing (breakage or unacceptable wear). A candidate function for h will be presented later. In the following formulation it is assumed that all profit for the current part is lost when a breakage occurs.

A cost functional can now be formulated:

$$\text{Minimize } J(\underline{u},x_k) = (1-v)\cdot J_1(u,x_k) + v\cdot J_2(u,0) \tag{5.7.1.1}$$

where

$$J_1(u,x_k) = E[\,-g(u)\cdot T_f\cdot 1_{[t_k+T_f<T_A]} + (B+R)\cdot 1_{[T_A\le t_k+T_f]}$$
$$+\, h(X_{t_k+T_f}, M_{[t_k,t_k+T_f]}, x_k) + K\cdot(T_f \wedge (T_A-t_k)) \mid X(t_k)=x_k\,] \tag{5.7.1.2a}$$

is the expected one step cost if the tool is not replaced, and

$$J_2(u,0) = E[\,-g(u)\cdot T_f\cdot 1_{[T_f<T_A]} + R + (B+R)\cdot 1_{[T_A\le T_f]}$$
$$+\, h(X_{T_f}, M_{[0,T_f]}, 0) + K\cdot(T_f \wedge T_A) \mid X(0)=0\,] \tag{5.7.1.2b}$$

is the expected one step cost if the tool is replaced. Note that $(T_1 \wedge T_2) \triangleq \min(T_1,T_2)$ and that x_k is the measured wear. The explicit use of 0 in J_2 is only a reminder of its conditioning on zero wear.

Before examining a special case of this problem, there remains the question of the reasonability of this cost functional. In response to the earlier comment, this cost functional contains terms which discourage initial detrimental behavior (high feed speeds). In particular, high feed speeds may be discouraged through the tool utilization cost term, and by the increased probability of a threshold crossing. A threshold crossing causes a complete loss of profit for the current part, and incurs the additional costs of replacement due to breakage (R+B).

5.7.2. Restricted Formulation

A restricted formulation of interest can be obtained under the following assumptions. Without loss of generality, assume that $t_k=0$. Let the threshold value A = W, the maximum tool wear. Let $b(u) = \beta u^m$, where $\beta>0$ and $m>1$ are constants. This form of the drift function is motivated by the Taylor tool life formula with $m = \frac{1}{n}$ and $\beta = \frac{W}{C_1^m}$. Assume that σ is constant. Let $g(u) = g \cdot u$, where $g>0$ is a constant, and $T_f(u)=\frac{V}{u}$ where $V>0$ is the depth of the hole to be drilled in each part. In this case, $g(u)\cdot T_f(u) = gV = G$, the profit earned per completed part. Let

$$h = D\cdot(X_{T_f}-x_0)\cdot 1_{[M_{T_f}<A]} + D\cdot(A-x_0)\cdot 1_{[M_{T_f}\geq A]} \tag{5.7.2.1}$$

That is, the tool utilization is proportional to the amount of tool consumed, with a maximum tool utilization given by $A-x_0$, the remaining tool utility. The constant D is the tool utilization cost rate. Further note the equivalence of the random variables:

$$1_{[M_{T_f}<A]} = 1_{[T_f<T_A]} \tag{5.7.2.2}$$

Summarizing, the new economic parameters are:

G $\triangleq$ profit earned for a completed part

D $\triangleq$ tool utilization cost rate

The restricted formulation becomes

$$\text{Minimize } J(\underline{u},x_0) = (1-v)\cdot J_1(u,x_0) + v\cdot J_2(u,0) \tag{5.7.2.3}$$

where

$$J_1(u,x_0) = E[\,-G\cdot 1_{[T_f<T_A]} + (B + R)\cdot 1_{[T_A\leq T_f]} + D\cdot(X_{T_f}-x_0)\cdot 1_{[T_f<T_A]}$$

$$+ D\cdot(A-x_0)\cdot 1_{[T_A \le T_f]} + K\cdot(T_f \wedge T_A) \mid X(0)=x_0] \qquad \textbf{(5.7.2.4a)}$$

and

$$J_2(u,0) = E[-G\cdot 1_{[T_f < T_A]} + R + (B + R)\cdot 1_{[T_A \le T_f]} + D\cdot X_{T_f}\cdot 1_{[T_f < T_A]}$$

$$+ D\cdot A\cdot 1_{[T_A \le T_f]} + K\cdot(T_f \wedge T_A) \mid X(0)=0] \qquad (5.7.2.4b)$$

and

$$T_f(u) = \frac{V}{u} \qquad (5.7.2.5)$$

The term $(T_f \wedge T_A)$ can be alternatively expressed as

$$(T_f \wedge T_A) = T_f\cdot 1_{[T_f < T_A]} + T_A\cdot 1_{[T_A \le T_f]} \qquad (5.7.2.6)$$

As a consequence of this restricted formulation, the following functions become relevant.

$$s_A(u \mid x_0) = E[1_{[T_A \le \frac{V}{u}]} \mid x_0] = P[T_A \le \frac{V}{u} \mid x_0] \qquad (5.7.2.7)$$

$$r_A(u \mid x_0) = E[(X_{\frac{V}{u}} - x_0)\cdot 1_{[\frac{V}{u} < T_A]} \mid x_0] \qquad (5.7.2.8)$$

$$z_A(u \mid x_0) = E[T_A\cdot 1_{[T_A \le \frac{V}{u}]} \mid x_0] \qquad (5.7.2.9)$$

The function $s_A(u \mid x)$ is the conditional probability of hitting the threshold before part completion. The function $r_A(u \mid x)$ is the conditional expected change in wear for the tool for those sample paths that do not have a threshold crossing prior to finishing the part, weighted by the probability of a threshold crossing not occurring. The function $z_A(u \mid x)$ is the expected time of the threshold crossing conditioned on sample paths that do have a threshold crossing prior to part completion, weighted by the probability of a threshold crossing occurring.

The functions s_A, r_A, and z_A can each be expressed in integral forms using the results of Chapter 4. The functions s_A and r_A follow easily from the threshold crossing density:

$$s_A(u \mid x) = Q_A(\frac{V}{u} \mid x \; ; \; u) = \int_0^{\frac{V}{u}} q_A(\eta \mid x \; ; \; u) d\eta \tag{5.7.2.10}$$

$$z_A(u \mid x) = \int_0^{\frac{V}{u}} \eta q_A(\eta \mid x \; ; \; u) d\eta \tag{5.7.2.11}$$

where

$$q_A(\eta \mid x \; ; \; u) = \frac{\alpha}{\sqrt{2\pi}\sigma\eta^{\frac{3}{2}}} \exp\left\{\frac{-(\alpha - \beta u^m \eta)^2}{2\sigma^2\eta}\right\} = q_\alpha(\eta \mid 0 \; ; \; u) \tag{5.7.2.12}$$

and $\alpha = A - x$, $x < A$.

The function r_A is somewhat more complicated to compute. Let $t_f = \dfrac{V}{u}$.

Without loss of generality assume that $x = 0$. Then

$$\begin{aligned} r_A(u \mid 0) &= E[\, X_{\frac{V}{u}} \cdot 1_{[\frac{V}{u} < T_A]} \mid X_0 = 0 \,] \\ &= \int_0^{\infty} \int_{-\infty}^{\infty} \eta 1_{[\nu < A]} g_{X_{t_f}, M_{t_f}}(\eta, \nu) d\eta d\nu \\ &= \int_0^{A} \int_{-\infty}^{\nu} \eta g_{X_{t_f}, M_{t_f}}(\eta, \nu) d\eta d\nu \end{aligned} \tag{5.7.2.13}$$

where g is the joint density function from Theorem 4.3.4. Equation (5.7.2.13) can be further simplified by splitting the integral into two parts

$$\int_{-\infty}^{0} \eta \int_0^{A} g_{X_{t_f}, M_{t_f}}(\eta, \nu) d\nu d\eta + \int_0^{A} \eta \int_{\eta}^{A} g_{X_{t_f}, M_{t_f}}(\eta, \nu) d\nu d\eta \tag{5.7.2.14}$$

and evaluating. The result is given by

$$r_A(u \mid 0) = \int_{-\infty}^{A} \eta p(A, u, \eta) d\eta \tag{5.7.2.15}$$

where

$$p(\alpha, u, \eta) =$$

$$\frac{\sqrt{u}}{\sqrt{2\pi V}\,\sigma} \exp\left\{\frac{-u(\eta - \beta V u^{m-1})^2}{2\sigma^2 V}\right\} \left[1 - \exp\left\{\frac{-2u\,\alpha(\alpha - \eta)}{\sigma^2 V}\right\}\right] \tag{5.7.2.16}$$

When $A > x \neq 0$, the result is given by

$$r_A(u \mid x) = \int_{-\infty}^{A-x} \eta p(A-x,u,\eta)d\eta \tag{5.7.2.17}$$

Note that the function p is a Gaussian density times a weighting function. This is more easily seen by substituting $t_f = \frac{V}{u}$ into (5.7.2.16).

The functions J_1 and J_2 can be expressed in terms of these functions as follows.

$$J_1(u,x_0) = (B + R + D\cdot(A-x_0))\cdot s_A(u \mid x_0)$$

$$+ (K\cdot\frac{V}{u} - G)\cdot(1-s_A(u \mid x_0)) + D\cdot r_A(u \mid x_0) + K\cdot z_A(u \mid x_0) \tag{5.7.2.18a}$$

$$J_2(u,0) = (B + R + D\cdot A)\cdot s_A(u \mid 0) + R$$

$$+ (K\cdot\frac{V}{u} - G)\cdot(1-s_A(u \mid 0)) + D\cdot r_A(u \mid 0) + K\cdot z_A(u \mid 0) \tag{5.7.2.18b}$$

Clearly, the existence of minima is tied to the properties of the functions s_A, r_A, z_A and to the choice of parameters in the cost functional. The following gives a sufficient condition for the existence of a minimum.

Theorem (5.7.2.1): Assume $m>1$ and that B, R, D, K, V, G are all strictly positive. Suppose there exists u' such that $J_1(u',x_0) < 0$ or $J_2(u',0) < 0$. Then $J(\underline{u}, x_0)$ has a minimum on $(0,\infty)\times\{0,1\}$.

Proof: The proof of this theorem uses the following properties of the cost functionals:

$$\lim_{u\to\infty} J_1(u,x_0) = B + R + D(A - x_0) > 0 \tag{5.7.2.19a}$$

$$\lim_{u\to\infty} J_2(u,0) = B + 2R + DA > 0 \tag{5.7.2.19b}$$

$$\lim_{u\to 0^+} J_1(u,x_0) = \infty \tag{5.7.2.19c}$$

$$\lim_{u\to 0^+} J_2(u,0) = \infty \tag{5.7.2.19d}$$

The proof of these properties can be found in Appendix C. Note also that the functionals J_1 and J_2 are continuous in u for $u \in (0,\infty)$.

Suppose there exists u' such that $J_1(u',x_0) < 0$. By the above properties, there exists an interval $U \triangleq [u_1,u_2]$ containing u' such that for $u \notin U$, $J_1(u,x_0) > 0$. Now J_1 has a minimum on U (compact set) and

$$\inf_{u\in U} J_1(u,x_0) < 0 < \inf_{u\notin U} J_1(u,x_0) \tag{5.7.2.20}$$

Either J_1 is the active functional in J in which case the minimum is established or else there exists u'' such that

$$J_2(u'',0) < \inf_u J_1(u,x_0) < 0 \tag{5.7.2.21}$$

For this case, apply the first argument to establish a minimum for J_2 and thus for J. The same proof applies when there exists u' such that $J_2(u',0) < 0$.

Remark: The theorem says that if the operation can be performed profitably for a given wear value a minimum exists.

Corollary (5.7.2.1): Assume that the feed speed is additionally constrained such that $u \in (0,u_m]$. Then $J(\underline{u},x_0)$ has a minimum on $(0,u_m]\times\{0,1\}$ for $0\leq x_0<A$.

Proof: Using the limiting properties discussed in the previous theorem, there exists u_1 such that for $0<u<u_1$, $v \in \{0,1\}$, $U \triangleq [u_1,u_m]$

$$J((u,v),x_0) > \min(\ J_1(u_m,x_0),\ J_2(u_m,0)\ ) \geq \inf_{u\in U,\ v\in\{0,1\}} J(\underline{u},x_0) \tag{5.7.2.22}$$

and J has a minimum on $U\times\{0,1\}$.

Remark: The corollary says that under practical constraints on the feed speed a minimum exists for all wear values in $[0,A)$.

Other more complicated sufficient conditions are also available. In general, if the piece profit is sufficiently high or the overhead rate is sufficiently low, a minimum exists.

The assumption $m > 1$ is used in Appendix C in the proofs of properties (5.7.2.19a) and (5.7.2.19b). This assumption is satisfied according to Taylor tool life data for metal machining with all commonly used tools. Reference [Mac1] gives typical values for n in the Taylor tool life formula (5.5.2.1) in the range 0.1 to 0.4, resulting in 2.5 to 10.0 as a range of values for m.

5.7.3. Optimal Feedback Policies

The simple way in which the control variable v enters the problem allows the optimal policy to be given in terms of the functions J_1 and J_2. The optimal one step policy is of the following form (assuming existence of all minima). Let x be the measured tool wear. Denote

$$J_1^*(x) \triangleq \min_u J_1(u,x) \qquad u_1^*(x) \triangleq \arg\min_u J_1(u,x)$$

$$J_2^* \triangleq \min_u J_2(u,0) \qquad u_2^* \triangleq \arg\min_u J_2(u,0)$$

i.e., the minimizing values of $J_1(\cdot,x)$ and $J_2(\cdot,0)$ are $u_1^*(x)$ and u_2^* respectively. Let $v^*(x)$ be the optimal replacement decision. The optimal one step policy is given by:

(i) if $J_1^*(x) < J_2^*$, do not replace the tool ($v^*(x)=0$) and continue at feed speed $u_1^*(x)$.

(ii) if $J_1^*(x) > J_2^*$, replace the tool ($v^*(x)=1$) and continue at feed speed u_2^*.

(iii) if the tool breaks, replace the tool and continue at feed speed u_2^*.

The values $u_1^*(x)$ and u_2^* must be computed numerically even in the restricted formulation. There is no known closed form solution. Depending upon the parameters, $J_1(\cdot,x)$ and $J_2(\cdot,0)$ may exhibit large flat valleys, and so gradient methods may not always be appropriate. As will be seen in an example, reasonable suboptimal policies may exist for an interval of feed speeds. Therefore, it may be possible to maintain constant feed speed for a range of wear measurement values with negligible degradation in performance. This will become evident later.

Implementation of the policy may be carried out in the following way. J_2^* and u_2^* are constants and may be computed in advance. Initially, assume that a new tool is in place. The feed speed for the first part is u_2^*. For each part thereafter, a wear measurement x is made, and $J_1^*(x)$ and $u_1^*(x)$ must be computed numerically based on the wear measurement. A table lookup would be a reasonable alternative. Recall that under the assumption of fixed feed, each selection of the feed speed necessitates a corresponding selection of the spindle speed as well. The optimal policy given previously is now used to determine the proper action for the next part.

The above implementation can be extended to the realistic situation of (stochastically) time-varying costs and profit rates. In this case, $J_1^*(x)$, J_2^*, $u_1^*(x)$, and u_2^* must all be computed for each part on the basis of the current cost functional parameters and wear measurement. It is assumed that the cost functional parameters are supplied by some higher level control and/or authority. In this situation, the higher level authority could influence tool replacement and higher feed speeds through manipulation of the costs and profit rate. In particular, tool replacement could be vetoed, in which case the next part would be machined at feed speed $u_1^*(x)$. Also, tool replacement could be strongly encouraged by a sufficient reduction in the tool replacement cost. Rationale for manipulating tool replacement policies and feed speed selection could include tool availability, service availability, the status of other machines and components in the manufacturing system, part and supplies inventory levels and part demand. This manipulation represents an indirect feedback of other information in the manufacturing system to the drill controller.

In actual practice, the selected feed speed will necessarily be constrained by other considerations: surface finish, spindle and drive power, and machine, tool, and part characteristics.

In summary, the optimal feedback policy can be implemented in a manufacturing setting, provided sufficient computational capability is present, and tool wear measurement feedback is available. By allowing time variable cost and profit parameters, indirect feedback of other information about the manufacturing system can be introduced into the policy. The controller might be implemented in a local processor at the machine site.

CHAPTER 6

POLICY COMPARISON FOR THE DRILLING PROBLEM: NUMERICAL EXAMPLES AND SIMULATIONS

6.1. Introduction

In this chapter the methods developed in Chapter 5 for the drilling problem are applied to an example problem. The example is based on an actual drilling problem found in [Mac1]. Analyses of age replacement policies and feedback policies for this problem are carried out. The problem parameters are held fixed except for the part profit value. This parameter is allowed to vary in order to explore the ramifications of relatively more and less expensive parts. Simulations of the drilling problem under different operational policies are conducted in order to compare the performance of different policies. Both primary and secondary performance measures are considered in the comparison. Two different production scenarios are considered: a fixed batch size and a fixed time production run.

Algorithms to support the analysis and the simulations are also discussed in this chapter. Included are computational methods for calculating the various functions in the cost functionals and random number generators for use in the simulations. In particular, a random number generator for inverse Gaussian random variables is presented.

6.2. Problem Description

The drilling operation parameters for this problem have been taken from [Mac1]. They are as follows:

Part Material:	4340 Alloy Steel 341 BHN
Drill:	M2 HSS Twist Standard Point 0.25 in x 4.00 in 29° Helix 118° Point Angle 7° Lip Relief
Operation:	0.5 in through hole heavy oil lubrication
Feed:	0.002 in/rev
Tool life:	0.015 in end point wear

Tool life data is also given in [Mac1] for different cutting speeds for the feed indicated above. In terms of feed speed, the data was in the range $u \in [2.1, 3.0]$ in/min. It is assumed here that the tool life data represents mean values. However, the reference does not comment on this.

The tool life is assumed to agree with the Taylor tool life formula in the mean over the data range. That is, the mean tool life and feed speed are assume to be related by (5.5.2.1):

$$uT^n = C_1 \tag{5.5.2.1}$$

A least square fit of the data resulted in coefficient values of

$$C_1 = 3.966 \tag{6.2.1}$$

$$n = 0.194 \tag{6.2.2}$$

Recall that these coefficients are derived under the assumption of constant feed. The spindle speed is presumed to vary in proportion to u so as to maintain constant feed.

The tool life is modeled by a diffusion-threshold process with drift coefficient $b(u) = \beta u^m$ and diffusion coefficient $\sigma(u)$. The drift coefficient parameter are easily computed from the Taylor tool life coefficients and are given by:

$$m = \frac{1}{n} \approx 5.2 \tag{6.2.3}$$

$$\beta = \frac{W}{C_1^m} \approx 1.2 \text{ x } 10^{-6} \tag{6.2.4}$$

These values were used in all the analyses and simulations.

No information concerning the variance of the tool life is given in [Mac1]. This necessitated estimation of the diffusion coefficient σ based upon assumptions. First, assume that the diffusion coefficient is constant and independent of the control u. This assumption is motivated by the lack of information regarding alternative functional forms. Now assume a nominal feed speed of 2.6 in/min. This is approximately mid-range for the available tool life data. At this feed speed, the mean tool life is about 8.7 minutes. Suppose that the standard deviation for the tool life at this nominal speed is about 20% of the mean. This corresponds to a standard deviation of about 1.7 minutes. From (4.3.30b), the diffusion coefficient can be computed:

$$\sigma = \left[\frac{b^3}{A} \operatorname{var}(T_A)\right]^{\frac{1}{2}} \approx 0.001 \tag{6.2.5}$$

It would be very desirable to have information concerning tool life variance. This information could be used to determine (empirical) functional forms for the diffusion coefficient. Note that the diffusion-threshold model for tool wear does allow an arbitrary dependence of variance on the feed speed by appropriate choice of the diffusion coefficient.

Summarizing, the diffusion threshold parameters used for this problem are given by:

$$A = 0.015$$
$$\beta = 1.2 \text{ x } 10^{-5}$$
$$m = 5.2$$
$$\sigma = 0.001$$

The economic parameters used in the drilling problem are based upon the following assumptions. Let the overhead cost rate be 30.00 dollars/hour which equals 0.50 dollars/min. Let the cost of a drill be 1.00 dollar. Assume that the time to change a drill under normal conditions is 0.5 minutes and that an additional 0.5 minutes is required if a tool failure occurs. In order to determine the effect of part worth, let the piece profit take on the values 1.00, 10.00, and 100.00 dollars. Summarizing, the parameters are given by:

K (overhead rate) = 0.50 dollars/min
R (replacement cost) = 1.00 + 0.5 x 0.5 = 1.25 dollars
B (repair cost) = 0.5 x 0.5 = 0.25 dollars
G (piece profit) $\in$ {1, 10, 100} dollars

Additional operation parameters are given by:

V (depth of hole) = 0.5 in
t_R (tool replacement time) = 0.5 min
t_B (additional repair time) = 0.5 min

6.3. Age Replacement Policy Analysis

6.3.1. Evaluation of Age Replacement Policies

The results of Section 5.6 can be applied to this problem in order to evaluate the performance of age replacement policies. From Section 5.6, two performance measures can be computed. The expected long term cost per time is given by

$$J_t^\tau = \frac{E^\tau[\, R + KT + B\,1_{[T=T_A]} - G\widetilde{P}]}{E^\tau[\, \widetilde{T}\,]} \tag{5.6.5.1}$$

and the expected long term cost per part is given by

$$J_p^\pi = \frac{E^\pi[\, R + KT + B\,1_{[T=T_A]} - G\widetilde{P}\,]}{E^\pi[\, \widetilde{P}\,]} \tag{5.6.5.2}$$

where a policy $\pi \triangleq (u, n_r)$ and

$T \triangleq$ tool life

$\widetilde{T} \triangleq$ duration of one tool cycle

$\widetilde{P} \triangleq$ number of parts produced in one tool cycle

The expected values of these random variables were previously shown to be given by (5.6.5.3) - (5.6.5.6):

$$E^\pi[\, T\,] = \int_0^{n_r t_f} \eta q_A(\eta \mid 0\,;\, u)d\eta + n_r t_f \cdot (1 - Q_A(n_r t_f \mid 0\,;\, u) \tag{5.6.5.4}$$

$$E^\pi[\, \widetilde{T}\,] = E^\pi[\, T\,] + t_B\, Q_A(n_r t_f \mid 0\,;\, u) + t_R \tag{5.6.5.3}$$

$$E^\pi[\, \widetilde{P}\,] = n_r - \sum_{j=0}^{n_r - 1} Q_A(jt_f \mid 0\,;\, u) \tag{5.6.5.6}$$

with $t_f = \dfrac{V}{u}$.

Evaluation of the performance of age replacement policies requires the evaluation of these expected values. In general, numerical integration techniques can be used. However, expeditious evaluation motivates the design of alternate algorithms for computing these functions. These will be discussed in Section 6.6.

The expected long term cost per time is plotted in Figs. 6.3.1.1 - 6.3.1.3 against feed speed for several age replacement values, given in number of parts. Fig. 6.3.1.1 is for a piece profit of 1.00 dollars, Fig. 6.3.1.2 is for a piece profit of 10.00 dollars, and Fig. 6.3.1.3 is for a piece profit of 100.00 dollars.

The expected long term cost per part is plotted in Fig. 6.3.1.4 for a piece profit of 1.00 dollar. Since the piece profit only enters this cost functional in an additive way, the plots for different piece profit values are not shown.

6.3.2. Discussion of Age Replacement Policies

Examination of the graphs (Figs. 6.3.1.1 - 6.3.1.4) reveals several interesting features of this drilling problem under age replacement policies.

The first observation is that the two performance criteria, expected cost per time and expected cost per part, do not lead to the same optimal policies. In fact, the optimal policies are distinctly different. This result is somewhat surprising, even though there is no reason to expect different cost functionals to yield similar optimal policies. The surprise is perhaps due to the intuitive idea that cost per part and cost per time should be closely related measures. For this problem, it is seen that minimizing expected cost per time leads to higher feed speeds and earlier tool replacement ages than those obtained by minimizing expected cost per part. Closer examination of Fig. 6.3.1.4 reveals that minimizing expected cost per part is achieved by letting the age of replacement $n_r \rightarrow \infty$. This corresponds to a failure replacement policy. However, this conclusion may be problem specific and should not be generalized.

As indicated earlier, the optimal policy for expected cost per part does not depend upon the piece profit. Note, though, that the class of policies considered to give good performance (relative to the optimum) will depend upon the piece profit.

The optimal policy for expected cost per time does have some dependence on the piece profit. As the profit increases, the trend is towards more conservative tool replacement ages, but slightly faster feed speeds. Generally, there exists a range of policies that will give near optimal performance.

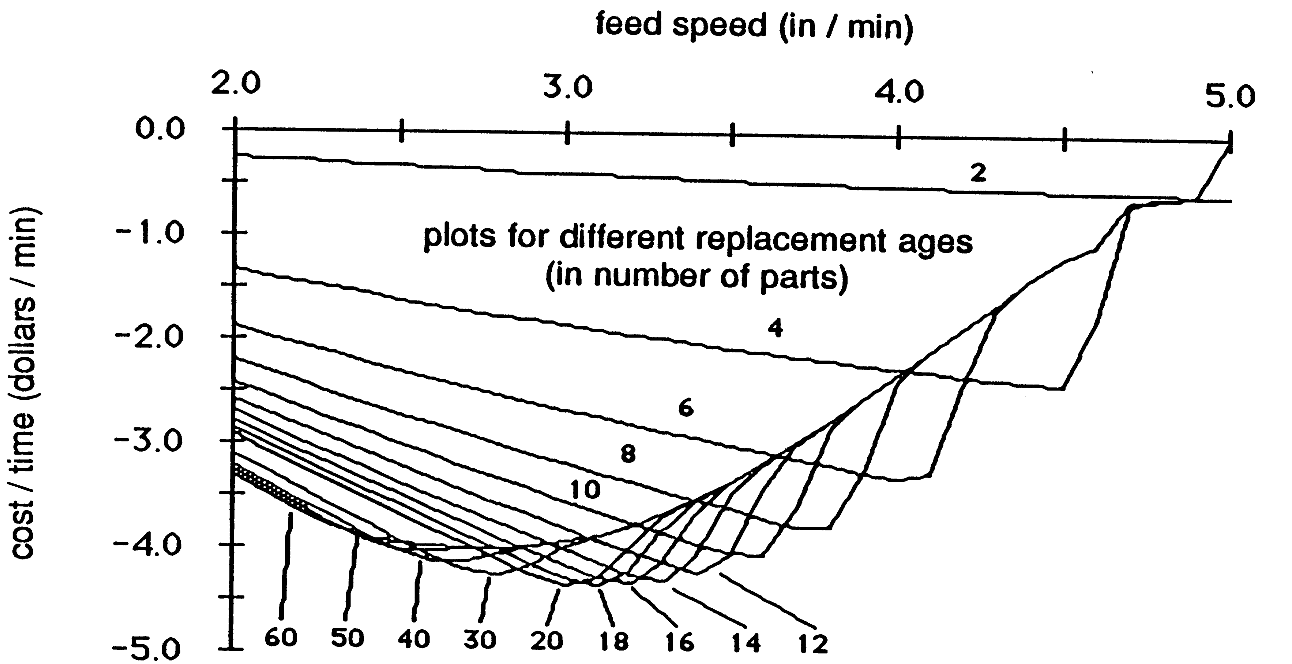

Fig. 6.3.1.1 Expected Cost per Time under Age Replacement for G = 1

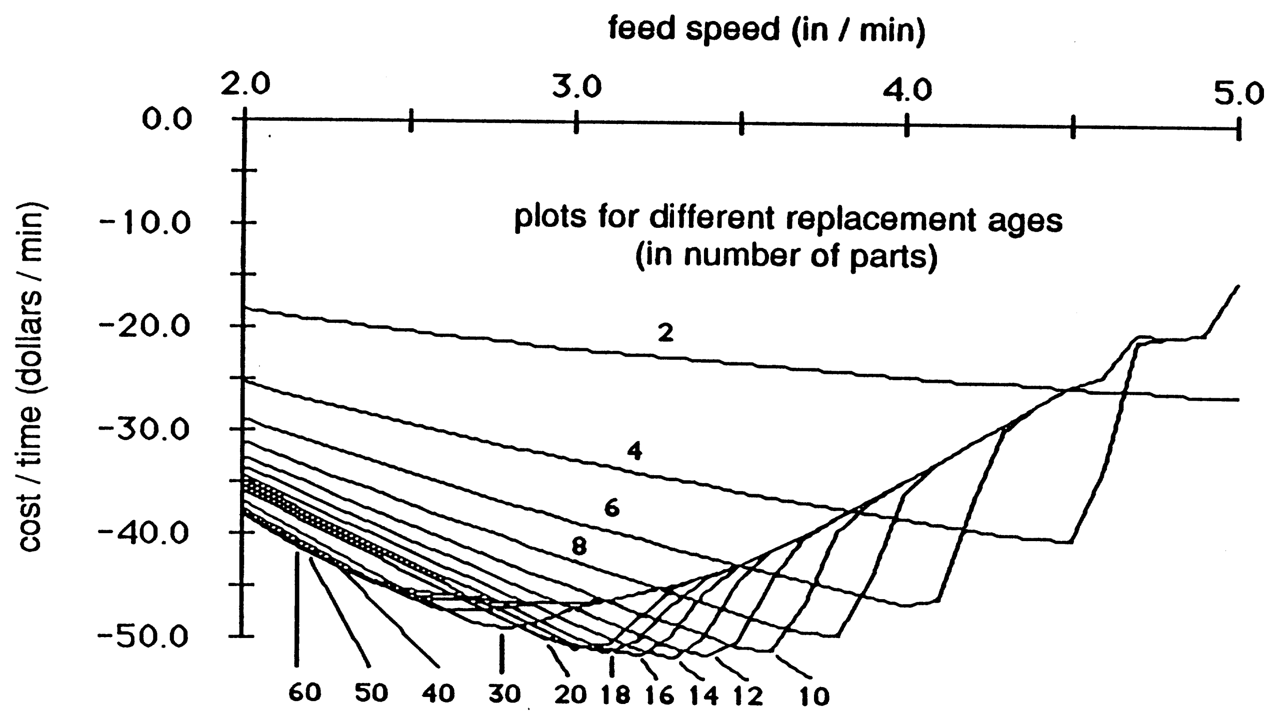

Fig. 6.3.1.2 Expected Cost per Time under Age Replacement for G = 10

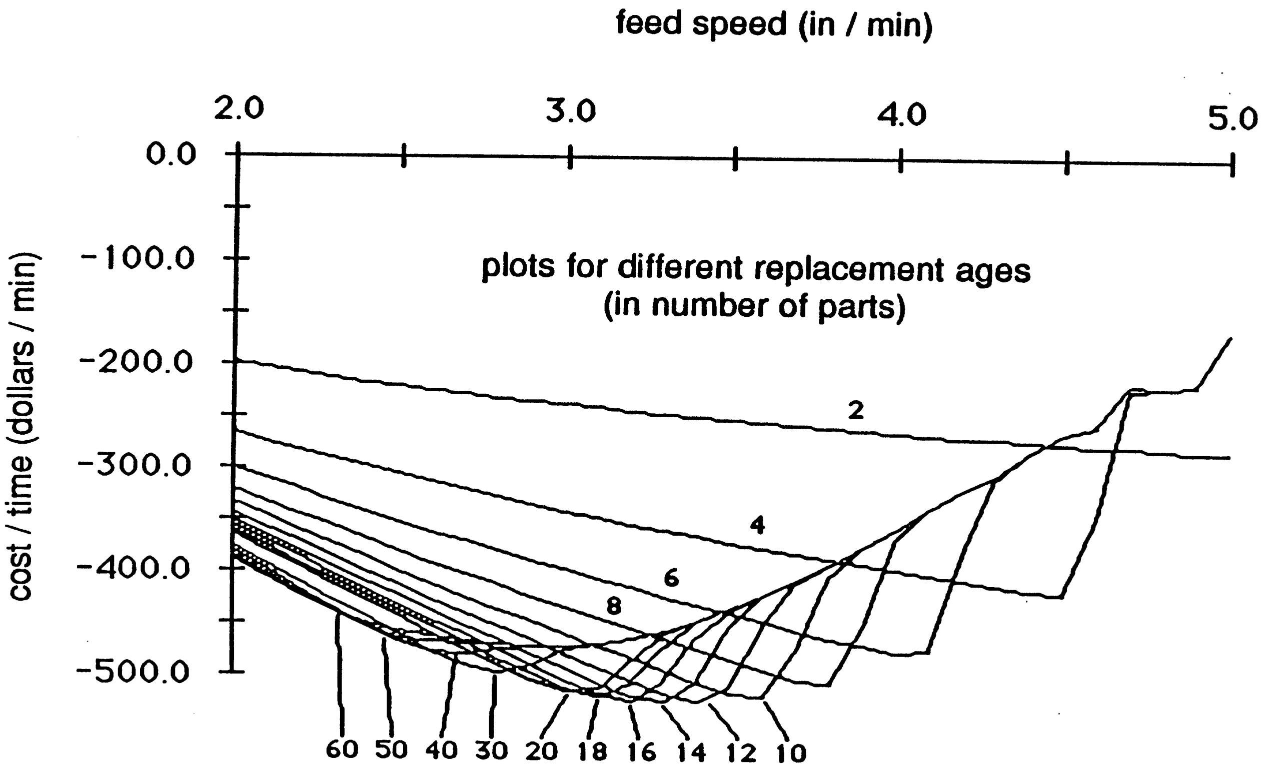

Fig. 6.3.1.3 Expected Cost per Time under Age Replacement for G = 100

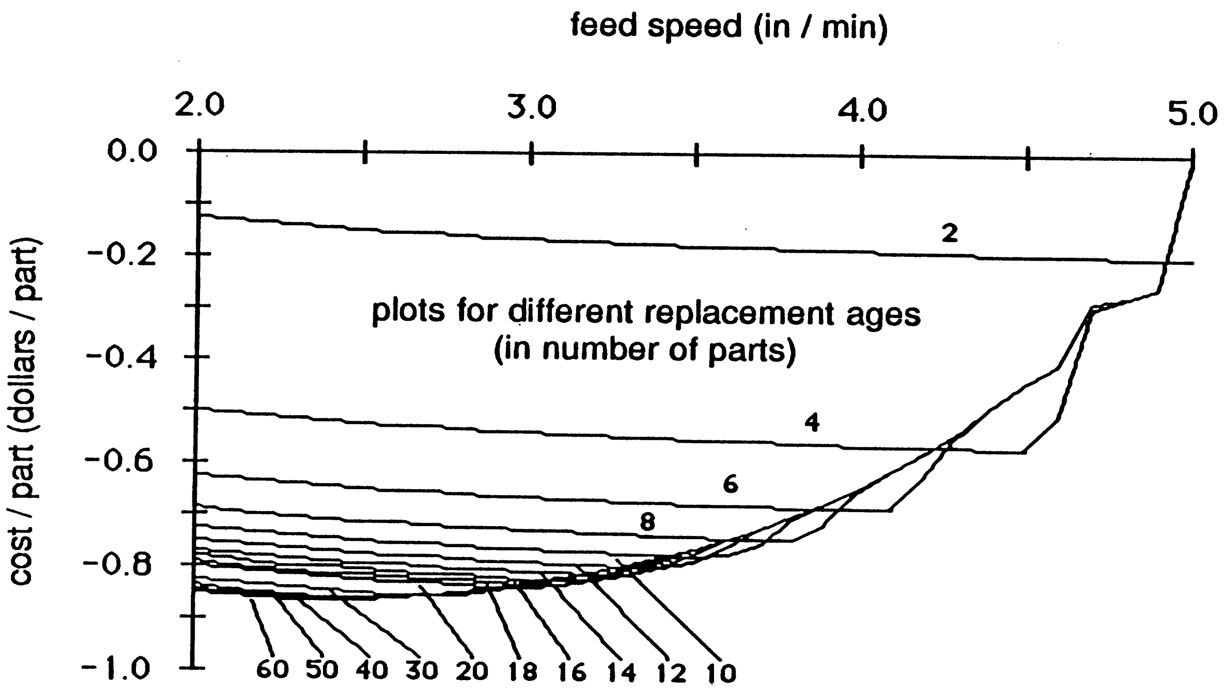

Fig. 6.3.1.4 Expected Cost per Part under Age Replacement for G = 1

6.4. Feedback Policy Analysis

6.4.1. Evaluation of Feedback Policies

The direct calculation of expected cost per part and expected cost per time for feedback policies is very difficult. Consequently, the one step cost was proposed in Section 5.7 as an alternative performance measure for feedback policies. In this section, good one step policies will be determined. Performance of these policies will be considered in the simulation section.

The results of Section 5.7 can be applied to this problem in order to compute one step costs as a function of tool wear, feed speed, and replacement decision. From Section 5.7.2, the equations

$$J_1(u,x_0) = (B + R + D\cdot(A-x_0))\cdot s_A(u \mid x_0) +$$

$$(K\cdot\frac{V}{u} - G)\cdot(1-s_A(u \mid x_0)) + D\cdot r_A(u \mid x_0) + K\cdot z_A(u \mid x_0) \qquad (5.7.2.18a)$$

$$J_2(u,0) = (B + R + D\cdot A)\cdot s_A(u \mid 0) + R +$$

$$(K\cdot\frac{V}{u} - G)\cdot(1-s_A(u \mid 0)) + D\cdot r_A(u \mid 0) + K\cdot z_A(u \mid 0) \qquad (5.7.2.18b)$$

need to be evaluated in order to determine the one step cost. First, though, there is an additional parameter in the cost functional that must be determined. This parameter is the tool utilization cost rate (D). For this problem, this parameter was determined on the basis of a normal tool replacement cost:

D (tool utilization cost rate) = 1.25/0.015 = 83.33 dollar/in

Calculation of the one step cost was done by discretizing both the feed speed and tool wear. The wear value was taken from the set

$\{$ 0.0, 0.003, 0.006, 0.009, 0.012, 0.014, 0.0145, 0.0146, 0.0147 $\}$ inches

The feed speed was discretized at intervals of 0.2 in/min.

Examination of the pair of equations (5.7.2.18a) and (5.7.2.18b) show that the functions s_A, r_A, and z_A need to be computed in order to evaluate the cost functionals J_1 and J_2. For each fixed piece profit (G), the functions s_A, r_A, and z_A were computed for each feed speed, for each of the tool wear values. In general, numerical integration can be used to compute these values. However, this is slow computationally, and so algorithms were designed to speed up the procedure. These algorithms will be discussed in Section 6.6.

Using the computed values for the functions s_A, r_A, and z_A, the function $J_1(\cdot,\cdot)$ can be calculated. For this problem, the family of functions $J_1(\cdot,x_0)$ was determined by linearly interpolating between discrete feed speed values. The wear value x_0 was allowed to take on values from the discrete set previously given. The function $J_2(\cdot,0)$ is easily determined from the zero wear $J_1(\cdot,0)$ function by adding the replacement cost.

Using graphs of the family of functions $J_1(\cdot,x_0)$ and $J_2(\cdot,0)$, the optimal feed speed selection and tool replacement decision can be approximated for each tool wear value. In order to better understand the one step policy, a somewhat different approach to policy determination was taken.

For each wear value, regions of feed speeds were determined, rather than just an optimal value. The regions were classified as near optimal (within 1% of optimal performance) and suboptimal (within 10% of optimal performance). These regions were plotted versus tool wear, with linear interpolation between wear points. The wear value threshold for tool replacement was also determined. At this wear value, a tool replacement is called for. Finally, critical boundaries were determined where possible. A critical boundary indicates the economic break even point where a crossover from profitable to costly operation takes place. These results are presented in Figs. 6.4.1.1 - 6.4.1.3 for the different piece profit values.

6.4.2. Discussion of Feedback Policies

Examination of the graphs reveals several interesting properties. When a tool has little or no wear, the near optimal operating regions are relatively wide, indicating a relatively large interval from which feed speed selection can be made with near optimal performance. As the tool wear increases to the high wear region, however, the near optimal region narrows appreciably. In general, it would seem that the choice of feed speed is more critical towards the end of the tool's life than it is for new and little worn tools. What is perhaps most surprising is the observation that in some cases constant feed speed can be used throughout the tool's life while maintaining near optimal performance.

Two effects of piece profit (G) on the policy are readily apparent. First, as G increases the width of the near optimal region increases for most wear values. Furthermore, as G increases tool replacement becomes more conservative; i.e., replacement of the tool occurs at lower wear levels. In particular, for sufficiently profitable parts and cheap tools, one could argue for tool replacement after very little wear (i.e., every part). Conversely, for cheap tools and cheap parts, replacement is close to the wear threshold. This is reminiscent of a "run until it breaks" policy. However, the region of operation for small G is generally narrower, so even a run until breakage replacement policy must be coupled with carefully chosen feed speeds. This seems to agree with the observation that manufacturing products with low profit margins requires more careful control of the operation in order to maintain (reasonable) profitability.

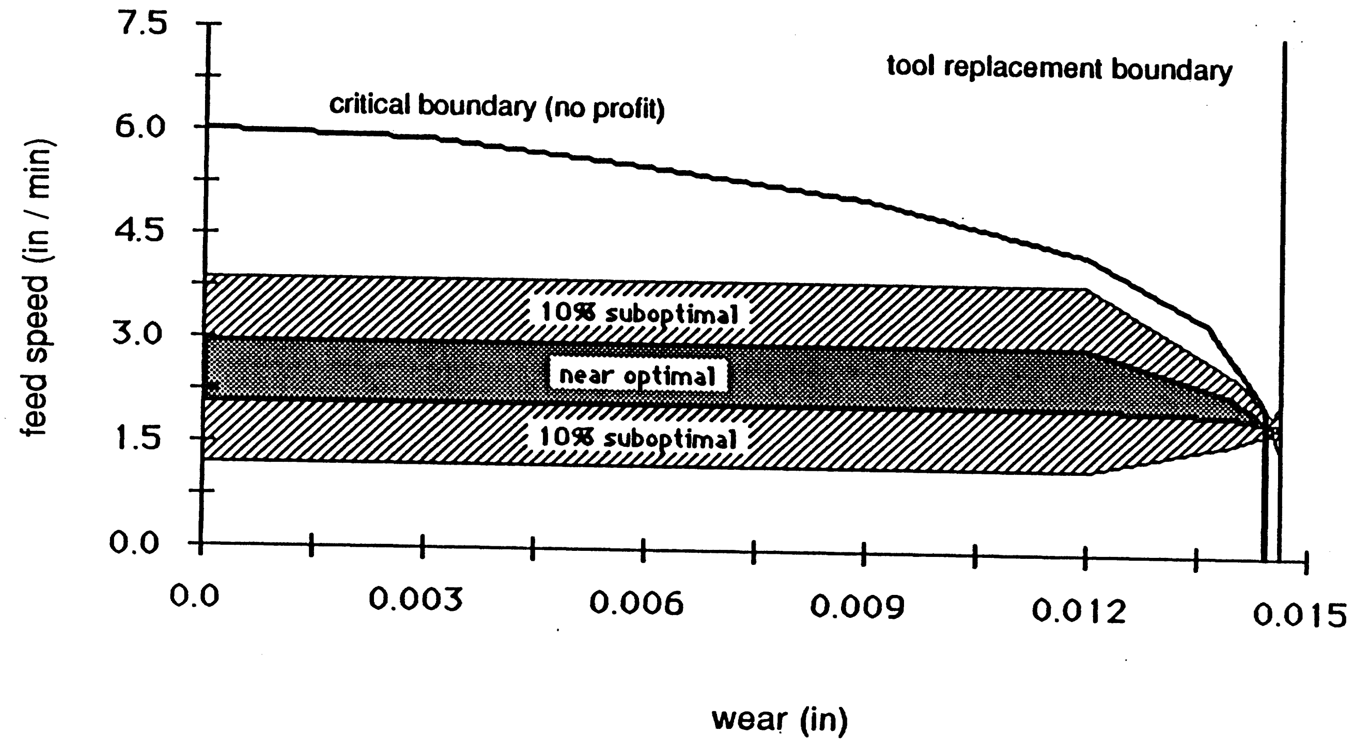

Fig. 6.4.1.1 One Step Results for G = 1

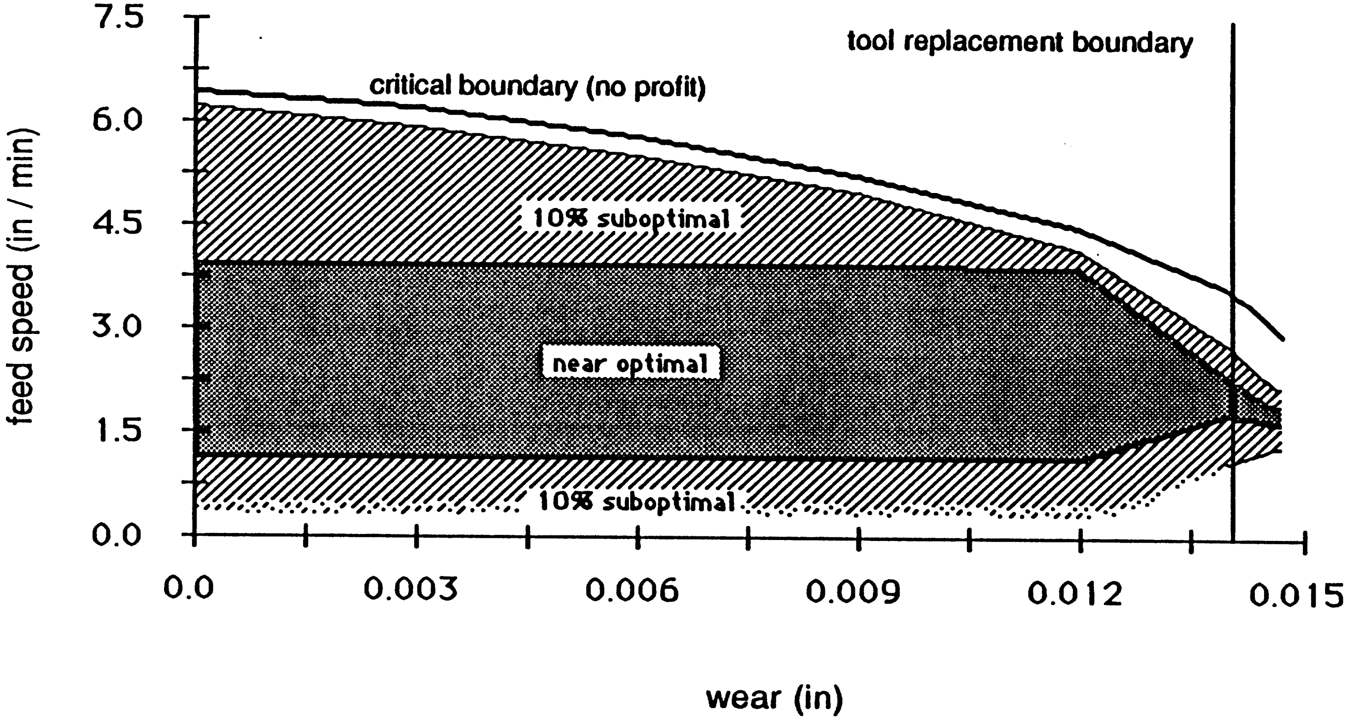

Fig. 6.4.1.2 One Step Results for G = 10

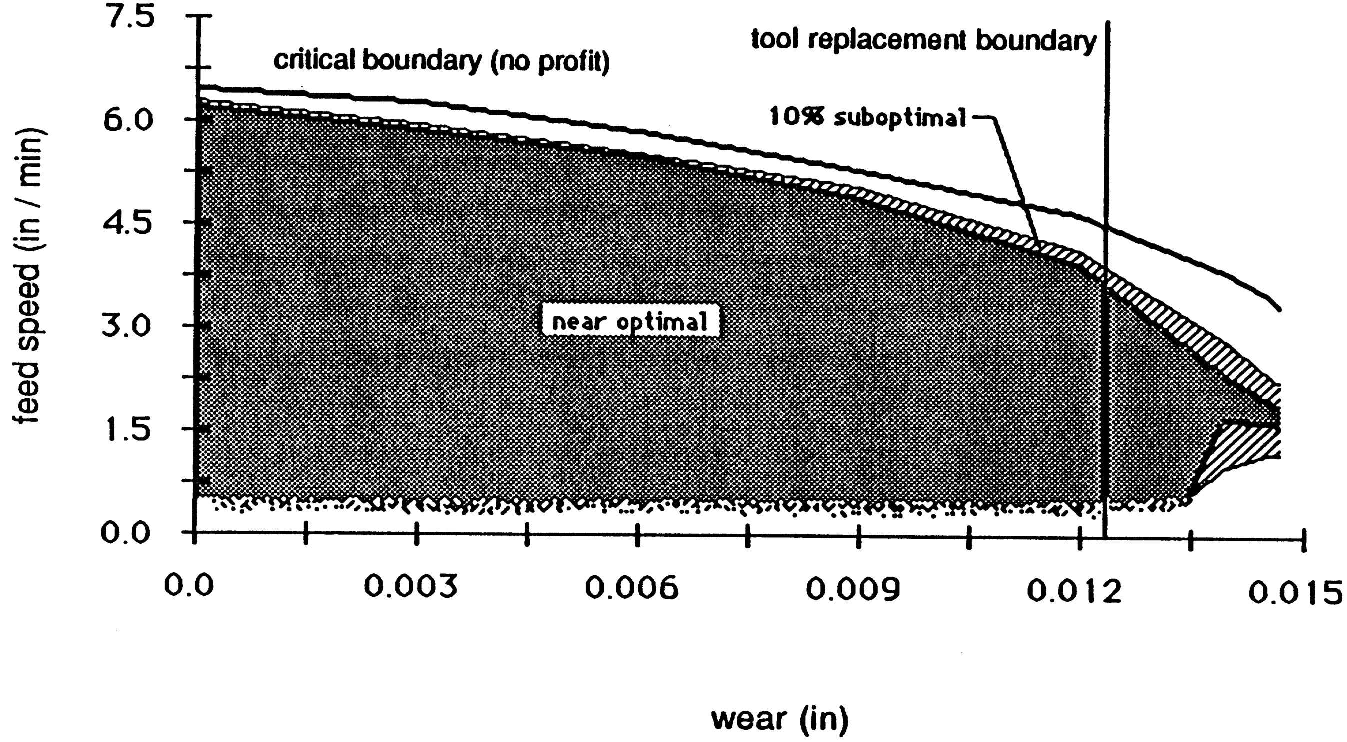

Fig. 6.4.1.3 One Step Results for G = 100

6.5. Simulation and Policy Comparison

Simulations of the drilling problem were conducted in order to evaluate the performance of different policies in different production situations. Because long term performance of the one step policies cannot easily be determined analytically, simulation represents the most direct way of evaluating and comparing policies. Simulation allows secondary performance characteristics to be compared as well. Simulation also provides an environment that allows policy experimentation. Thus, policy tuning can be carried out in an effort to find policies that perform well and are easily implemented.

6.5.1. Simulation Description

Simulation of the drilling problem is straightforward, with only the random number generation presenting technical difficulties.

The simulation written uses a tool cycle as an outer event loop. Recall that a tool cycle begins with a new tool and concludes when the next tool is installed. Tool failure necessitates tool replacement, but a tool can be replaced voluntarily prior to failure.

The simulation uses a part cycle as an inner event loop. When a part arrives, the subroutine responsible for implementing the control policy is given a set of information regarding the current situation. This set includes: length of time the current tool has been in use; number of parts drilled by the current tool; wear measurement of the current tool. The control subroutine responds with a tool replacement decision and a feed speed. If the tool is to be replaced, the current tool cycle statistics are updated, and a new tool cycle begins. The part cycle proceeds with the updated tool wear measurement and the selected feed speed. First, it is determined whether or not a tool failure occurs in this part cycle. This is done by generating an inverse Gaussian random number and comparing it to the completion time for the part. If a tool failure occurs, the time of occurrence is noted, the tool cycle statistics are updated, and a new tool cycle begins. If a failure does not occur, the new tool wear measurement is determined. The current tool cycle statistics are updated and the next part cycle begins.

Different production situations dictate different stopping criteria. In this work, two production situations are investigated. The first is a fixed time production run of 4800 minutes. This represents two 8 hour shifts running for 5 days. The second production situation calls for the production of 1000 good parts. This typifies a small or medium batch run. As will be seen, operational policies do not necessarily give comparable performance in both of these production situations.

Statistics collected for a simulation run include: time duration of production run, number of good parts produced, number of scrapped parts (number of tool failures), and number of tool changes. From these statistics, various measures of policy performance are computed and recorded.

Algorithms for the generation of the required random variables will be discussed in a Section 6.6.

6.5.2. Simulation Policy Descriptions and Results

The policies used in the simulations are summarized in Tables 6.5.2.1 - 6.5.2.3. The key for policy type is

AR	age replacement
FR	wear feedback replacement
FR/FS	wear feedback replacement and feed speed selection

The results from the simulations are summarized in Tables 6.5.2.4 - 6.5.2.9.

Policy Descriptions: G = 1			
Policy	Type	Replacement	Feed Speed
a	AR	20 parts	3.0
b	AR	40 parts	2.6
c	AR	failure	2.5
d	FR/FS	0.0147	$l_1(x)$
e	FR/FS	0.0145	$l_1(x)$
f	FR/FS	0.0140	$l_1(x)$
g	FR/FS	0.0135	$l_1(x)$
h	FR/FS	0.0147	$l_2(x)$
i	FR/FS	0.0145	$l_2(x)$
j	FR/FS	0.01425	$l_2(x)$
k	FR/FS	0.0140	$l_2(x)$
l	FR	0.0147	2.25
m	FR	0.0145	2.25
n	FR	0.0147	2.6
o	FR	0.0145	2.6
p	FR	0.0147	3.0
q	FR	0.0145	3.0
r	FR	0.0140	3.0
s	FR	0.0135	3.0

Table 6.5.2.1 Policies for G = 1

where the policies l_1 and l_2 are given in terms of the wear measurement x by:

$$l_1(x) = \begin{cases} 3.0 & x \leq 0.012 \\ 3.0 - 375(x - 0.012) & x > 0.012 \end{cases} \tag{6.5.2.1}$$

$$l_2(x) = \begin{cases} 3.0 & x \leq 0.012 \\ 3.0 - 400(x - 0.012) & 0.012 < x \leq 0.0135 \\ 2.4 - 533(x - 0.0135) & x > 0.0135 \end{cases} \tag{6.5.2.2}$$

Policy Descriptions: G = 10			
Policy	Type	Replacement	Feed Speed
a	AR	14 parts	3.25
b	AR	16 parts	2.6
c	AR	failure	2.5
d	FR/FS	0.0140	$l_3(x)$
e	FR/FS	0.0135	$l_3(x)$
f	FR	0.0140	3.0
g	FR	0.0135	3.0
h	FR	0.0140	3.25
i	FR	0.0135	3.25

Table 6.5.2.2 Policies for G = 10

where the policy l_3 is given in terms of the wear measurement x by:

$$l_3(x) = \begin{cases} 3.25 & x \leq 0.012 \\ 3.25 - 583.3(x - 0.012) & x > 0.012 \end{cases} \tag{6.5.2.3}$$

Policy Descriptions: G = 100			
Policy	Type	Replacement	Feed Speed
a	AR	14 parts	3.3
b	AR	16 parts	3.2
c	AR	failure	2.5
d	FR/FS	0.0125	$l_4(x)$
e	FR	0.0125	2.5
f	FR	0.0125	3.0
g	FR	0.0125	3.25
h	FR	0.0125	3.5

Table 6.5.2.3 Policies for G = 100

where the policy l_4 is given in terms of the wear measurement x by:

$$l_4(x) = \begin{cases} 4.5 & x \leq 0.009 \\ 4.5 - 342.9(x - 0.009) & x > 0.009 \end{cases} \tag{6.5.2.4}$$

	AR			FR/FS				FR/FS		
	a	b	c	d	e	f	g	h	i	j
Parts Produced	24754	22853	21726	23098	23875	24888	25253	22889	23594	24411
Profit	21096	19867	18908	19877	20602	21528	21797	19696	20355	21112
Scrap Parts	91	181	414	494	279	52	0	478	292	113
Tools Used	1246	593	414	810	848	954	1045	784	829	889
Parts/Tool	19.87	38.49	52.45	28.52	28.12	26.11	24.17	29.20	28.48	27.47
Production Rate	5.157	4.761	4.526	4.812	4.971	5.185	5.261	4.769	4.915	5.086

		FR		FR		FR				
	k	l	m	n	o	p	q	r	s	
Parts Produced	24803	20587	20760	22598	23011	23549	24182	25258	25478	
Profit	21448	17952	18125	19694	20098	20165	20774	21787	21960	
Scrap Parts	51	151	83	323	181	650	443	98	7	
Tools Used	945	252	256	506	517	970	999	1061	1104	
Parts/Tool	26.26	81.72	81.34	44.63	44.49	24.27	24.20	23.82	23.08	
Production Rate	5.167	4.289	4.325	4.708	4.794	4.906	5.038	5.262	5.308	

Key: AR - age replacement
FR - wear feedback replacement
FR/FS - wear feedback replacement and feed speed selection

Table 6.5.2.4 Production for 1 Week (4800 min) for G = 1

	AR			FR/FS				FR/FS		
	a	b	c	d	e	f	g	h	i	j
Production Time	194	210	221	208	201	193	190	210	203	197
Profit	852	869	870	861	863	865	863	861	863	865
Scrap Parts	4	8	19	21	12	2	0	21	12	5
Tools Used	51	26	19	35	36	39	42	35	36	37
Parts/Tool	19.87	38.49	52.45	28.52	28.12	26.11	24.17	29.20	28.48	27.47
Production Rate	5.157	4.761	4.526	4.812	4.971	5.185	5.261	4.769	4.915	5.086

		FR		FR		FR				
	k	l	m	n	o	p	q	r	s	
Production Time	194	233	231	212	209	204	198	190	188	
Profit	865	872	873	872	873	856	859	863	862	
Scrap Parts	2	7	4	14	8	28	18	4	(0.3)	
Tools Used	38	13	13	23	23	42	42	42	44	
Parts/Tool	26.26	81.72	81.34	44.63	44.49	24.27	24.20	23.82	23.08	
Production Rate	5.167	4.289	4.325	4.708	4.794	4.906	5.038	5.262	5.308	

Key: AR - age replacement
FR - wear feedback replacement
FR/FS - wear feedback replacement and feed speed selection

Table 6.5.2.5 Production of 1000 Parts for G = 1

	AR			FR/FS		FR		FR		
	a	b	c	d	e	f	g	h	i	
Parts Produced	25180	24942	21726	25104	25689	25260	25478	25506	26013	
Profit	247570	245410	214550	247360	253020	248990	251130	251140	256130	
Scrap Parts	45	292	414	50	0	98	7	202	22	
Tools Used	1802	1587	414	1262	1444	1060	1104	1499	1570	
Parts/Tool	13.97	15.72	52.45	19.89	17.78	23.82	23.08	17.01	16.57	
Production Rate	5.246	5.196	4.526	5.230	5.352	5.262	5.308	5.314	5.419	

Key: AR - age replacement
FR - wear feedback replacement
FR/FS - wear feedback replacement and feed speed selection

Table 6.5.2.6 Production for 1 Week (4800 min) for G = 10

	AR			FR/FS		FR		FR		
	a	b	c	d	e	f	g	h	i	
Production Time	191	192	221	191	187	190	188	188	185	
Profit	9832	9839	9875	9853	9849	9857	9857	9846	9846	
Scrap Parts	2	12	19	2	0	4	(0.3)	8	1	
Tools Used	72	64	19	51	57	42	44	59	61	
Parts/Tool	13.97	15.72	52.45	19.89	17.78	23.82	23.08	17.01	16.57	
Production Rate	5.246	5.196	4.526	5.230	5.352	5.262	5.308	5.314	5.419	

Key: AR - age replacement
FR - wear feedback replacement
FR/FS - wear feedback replacement and feed speed selection

Table 6.5.2.7 Production of 1000 Parts for G = 10

	AR			FR/FS	FR	FR	FR	FR		
	a	b	c	d	e	f	g	h		
Parts Produced	25301	25189	21726	21233	22745	25274	25765	25703		
Profit (1000's)	2525.8	2514.7	2171	2116.2	2272.6	2523	2572.3	2565.6		
Scrap Parts	102	122	414	0	0	0	0	0		
Tools Used	1817	1586	414	4600	498	1180	1674	2258		
Parts/Tool	13.92	15.88	52.45	4.616	45.72	21.41	15.39	11.38		
Production Rate	5.271	5.248	4.526	4.424	4.739	5.265	5.368	5.355		

Key: AR - age replacement
FR - wear feedback replacement
FR/FS - wear feedback replacement and feed speed selection

Table 6.5.2.8 Production for 1 Week (4800 min) for G = 100

	AR			FR/FS	FR	FR	FR	FR		
	a	b	c	d	e	f	g	h		
Production Time	190	191	221	226	211	190	186	187		
Profit	99828	99834	99926	99667	99917	99827	99834	99820		
Scrap Parts	4	5	19	0	0	0	0	0		
Tools Used	72	63	19	217	22	47	65	88		
Parts/Tool	13.92	15.88	52.45	4.616	45.72	21.41	15.39	11.38		
Production Rate	5.271	5.248	4.526	4.424	4.739	5.265	5.368	5.355		

Key: AR - age replacement
FR - wear feedback replacement
FR/FS - wear feedback replacement and feed speed selection

Table 6.5.2.9 Production of 1000 Parts for G = 100

6.5.3. Discussion of the Simulation Results

For each production situation, several different policies were tested. Included in the testing were policies based on age replacement, approximations to the feedback policies based on one step costs, and variations of these feedback policies. In all cases, the simulation results for age replacement policies were in good agreement with the previously computed analytic results. This provided some assurance that things were working correctly.

The results are presented in tabular form in Tables 6.5.2.4 - 6.5.2.9. The different columns correspond to different policies used. The policies used are described separately. Time units are in minutes and monetary units are in dollars.

The discussion here will focus mostly on the case where the piece profit (G) is 1.00 dollar (Tables 6.5.2.4, 6.5.2.5). This case is in some ways the most interesting. This is because the results from the feedback analysis indicate a fairly narrow region of near optimal performance. Furthermore, this case represents best the challenge of producing inexpensive parts profitably. However, many of the remarks also apply to the other cases.

The age replacement policies are *a*, *b*, and *c*. The remaining policies use tool wear feedback, though some have fixed feed speed. Policy *a* is based on the expected cost per time results, and policies *b* and *c* are based on the expected cost per part analysis. Note that policy *c* is failure replacement. As might be expected, each does well in a particular production situation. Minimizing cost per part results in good batch performance, whereas minimizing cost per time yields good long time performance. A general comment on age replacement policies is that if they are properly chosen for the production situation they do very well from an economic standpoint.

However, there are secondary performance measures that must be compared as well. Also, although the age replacement policies shown here perform well in particular production situations, it is the author's experience that in most manufacturing systems the same policies are used regardless of the production situation and that these policies are often very conservative.

The feedback policies show considerable variation in performance. The policies evaluated were constructed by approximating the true one step optimal policy by piecewise linear functions. That is, the feed speed was chosen to be a piecewise linear function of tool wear. Examination of the one step graphs (Figs. 6.4.1.1 - 6.4.1.3) shows that this should be a reasonable approximation. The policy closest to the true one step policy for the case G = 1 is policy *h*. As can be seen, it does not give a distinguished performance. In particular, the scrap count is too large. This observation gives some insight into the one step cost functional, and also into how some good policy variations can be constructed.

The problem with the one step cost functional is that the cost of tool replacement is too high. This is especially true when the piece profit is relatively low as is the case when G = 1. By forcing one part to absorb all the replacement cost, replacement is avoided until it is typically too late. Note that for the case under consideration it is cheaper to scrap the part (1.00 dollar) than to replace the tool (1.25 dollars). Thus, replacement is generally avoided. It would be more desirable to pro rate the tool replacement cost so that the burden of replacement is more evenly distributed among the parts that benefit from replacement. The easiest way to effect this cost disburdenment is to lower the tool wear threshold for replacement. This is precisely what is done in the feedback policy variations. The policies *d*, *e*, *f*, *g* and *h*, *i*, *j*, *k* represent policies with decreasing tool wear replacement thresholds.

There are other possible policy variations as well. An important practical question is what is the contribution of variable feed speed? That is, does tool wear feedback used only for replacement coupled with constant feed speed give good performance? In an effort to address this issue, some of the feedback policies use constant feed speed. Note that the one step graphs (Figs. 6.4.1.1 - 6.4.1.3) suggest that this may in fact be reasonable. Polices *l* through *s* are of this type.

Consider first the timed run production situation (1 week of production). For this case, the best age replacement policy is *a* (minimize cost per time). The feedback policies that will be examined closely are *g* (variable speed) and *s* (constant speed). Policy *s* gives the best economic performance, with an improvement of 4.1% over age replacement. However, policy *g* is probably the best choice. Although this policy only gives a 3.3% improvement over age replacement, it has much better characteristics. In particular, the drop in scrap production is *very* significant. Further note that the number of parts per tool increases and the total number of tool changes decreases. The effect of the variable feed speed appears to be to trade off production rate and pure economic gain for less scrap production and fewer tool changes. Policy *g* represents a good example of the potential performance of a modified one step policy. However, if variable feed speed is not practical, than policy *s* still gives good performance. Finally, note that the other age replacement policies give mediocre performance.

Now consider the batch production situation (production of 1000 parts). In this situation, the best age replacement policies are *b* and *c*. These policies minimize cost per part. Although policy *c* (failure replacement) gives the best economic performance, policy *b* is probably the one of choice since it has better secondary characteristics. Also, some managers and engineers may find failure replacement distasteful. Also note that age replacement policy *a* now gives mediocre performance. As can be seen, the candidate feedback policy *g* is economically worse representing a 0.7% decrease in profit. However, this policy still exhibits superior characteristics in terms of scrap production and production time. This would generally justify the slight economic loss. The reason for the slightly poorer performance is that policy *g* uses too many tools for this performance measure. None of the variable feed speed policies considered betters the age replacement policy economically. This motivates exploration of other feedback policies. Policies *n* and *o* were constructed by taking the feed speed used by the age replacement policy combined with tool wear feedback replacement. Policies *l* and *m* are similar with a slower feed speed, and policies *p*, *q*, *r*, and *s* use a faster feed speed. As can be seen, policies *l*, *m*, *n*, and *o* are economically superior, although by less than 0.5%. Policy *o* is the one of choice, since it has equal scrap production, is slightly faster, and uses fewer tools than the age replacement policy. An important point is that the age replacement policies can do *very* well, but must be selected for the batch production environment. This suggests that a properly chosen age replacement policy can be a good choice for some production situations. It also suggests that other cost functionals should be considered for determining feedback policies in some batch production situations. Nonetheless, the merits of feedback policy *g* include good performance in both production situations.

All of the preceding discussion is problem specific. Nonetheless, some important issues have been raised. First, age replacement policies can give good performance if properly chosen for the production situation. Second, the one step approach to feedback policies can yield improved performance. However, modification of the tool replacement threshold may be required in order to reduce the burden of replacement. Third, the one step approach does not always yield improved performance over optimal age replacement. Fourth, policies with constant feed speed and tool wear based replacement can give good performance as well. Fifth, it is important to look at characteristics other than just economic performance when evaluating a policy. Sixth, performance can vary with the production situation. When the production situation is uncertain, a feedback policy may give good all around performance, but an age replacement policy may not.

There is no reason to expect dramatic improvement by using feedback policies in the drilling problem. What has been seen is that modest improvements can be realized. This does not imply that feedback policies are not useful. Since a typical manufacturing system has many machines, small improvements in the operation of each machine may be collectively significant.

6.6. Algorithms

Several algorithms have been used to carry out the analysis in this chapter. These algorithms will be summarized in this section.

6.6.1. Algorithms for Age Replacement Analysis

The performance statistics for the drilling problem under age replacement policies can be expressed in terms of integrals involving the inverse Gaussian density as a kernel. In general, numerical integration could be used to evaluate each of the integrals. However, for problems with large age replacement values (n_r), this requires evaluation of many integrals ($n_r + 2$ integrals). This motivates the investigation of alternative methods for evaluating the integrals. As will be shown, it is possible to transform the integrals and use approximations to avoid all numerical integration.

The distribution function for the threshold crossing time $Q_A(t \mid 0\,;\,u) = P[T_A \leq t]$ has been previously given as

$$Q_A(t \mid 0\,;\,u) = \int_0^t q_A(\eta \mid 0\,;\,u)\,d\eta = \int_0^t \frac{A}{\sqrt{2\pi}\sigma\eta^{\frac{3}{2}}} \exp\left\{\frac{-(A-b\eta)^2}{2\sigma^2\eta}\right\} d\eta \qquad (4.3.24)$$

It can be shown that $Q_A(t \mid 0\,;\,u)$ can be expressed in terms of the normal distribution [Ch1]. Let $\Phi(x) = P[X \leq x]$ when $X \sim N(0,1)$. That is

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{\eta^2}{2}\right\} d\eta \qquad (6.6.1.1)$$

Then

$$Q_A(t \mid 0\,;\,u) = \Phi\left[\frac{A}{\sigma}\sqrt{\frac{1}{t}}\left(\frac{tb}{A}-1\right)\right]$$

$$+ \exp\left\{\frac{2Ab}{\sigma^2}\right\}\Phi\left[-\frac{A}{\sigma}\sqrt{\frac{1}{t}}\left(\frac{tb}{A}+1\right)\right] \qquad (4.3.33)$$

Using a procedure similar to that given in [Ch1], it is possible to show that the truncated mean can also be expressed in terms of the normal distribution.

Corollary (6.6.1.1):

$$\int_0^t \eta q_A(\eta \mid 0\,;\,u)\,d\eta =$$

$$\frac{A}{b}\left\{\Phi\left[\frac{A}{\sigma}\sqrt{\frac{1}{t}}\left(\frac{tb}{A}-1\right)\right] - \exp\left\{\frac{2Ab}{\sigma^2}\right\}\Phi\left[-\frac{A}{\sigma}\sqrt{\frac{1}{t}}\left(\frac{tb}{A}+1\right)\right]\right\} \qquad (6.6.1.2)$$

The above transformations show that all integrals required for evaluating the performance statistics can be expressed in terms of the normal distribution function. Although there is no known non-integral closed form expression for the normal distribution, there are several known approximations that are relatively easy to compute.

For this paper, the following approximation is used [Ab1]:

Approximation (6.6.1.1):

$$\Phi(x) = 1 - f(x)[a_1 t + a_2 t^2 + a_3 t^3] + \epsilon(x), \quad \text{for } 0 \leq x < \infty \tag{6.6.1.3}$$

where

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} \tag{6.6.1.4}$$

is the normal density function and

$$t = \frac{1}{1 + a_0 x} \tag{6.6.1.5}$$

$$a_0 = 0.3326700 \tag{6.6.1.6a}$$

$$a_1 = 0.4361836 \tag{6.6.1.6b}$$

$$a_2 = -0.1201676 \tag{6.6.1.6c}$$

$$a_3 = 0.9372980 \tag{6.6.1.6d}$$

$$|\epsilon(x)| < 1 \times 10^{-5} \tag{6.6.1.7}$$

Using this approximation, the performance statistics can be computed without numerical integration.

6.6.2. Algorithms for Feedback Analysis

The evaluation of the one step cost functional requires the evaluation of the three functions s_A, r_A, z_A. Note that

$$s_A(u \mid x) = Q_A(t_f \mid x \; ; u) \tag{6.6.2.1}$$

and that

$$z_A(u \mid x) = \int_0^{t_f} \eta q_A(\eta \mid x \; ; u) d\eta \tag{6.6.2.2}$$

Thus, the results from the previous section can be applied to evaluate these two functions without integration.

Numerical integration was required to evaluate the function r_A. However, approximations are necessary in order to reduce the integration interval to one with finite endpoints. Smaller integration intervals also reduce the computational burden. This reduction was done in the following way. From (5.7.2.17)

$$r_A(u \mid x) = \int_{-\infty}^{\alpha} \eta p(\alpha,u,\eta) d\eta \tag{6.6.2.3}$$

where $\alpha = A - x$ and

$$p(\alpha,u,\eta) = f(\eta\,;\,\alpha,u) \cdot w(\eta\,;\,\alpha,u) \tag{6.6.2.4}$$

with f a Gaussian density and w a weighting function. Now

$$\lim_{\eta \to -\infty} w(\eta) = 1 \tag{6.6.2.5}$$

for all admissible α and u. So, for $0 < \epsilon$ there exists γ such that

$$1 - \epsilon \leq w(\eta) < 1\,, \quad \text{for } \eta \leq \gamma \tag{6.6.2.6}$$

Assume that $\gamma \leq \alpha$. Then

$$r_A(u \mid x) = \int_{-\infty}^{\gamma} \eta p(\alpha,u,\eta) d\eta + \int_{\gamma}^{\alpha} \eta p(\alpha,u,\eta) d\eta \tag{6.6.2.7}$$

The second term must be computed numerically. The first term can be approximated, however.

$$\int_{-\infty}^{\gamma} \eta p(\alpha,u,\eta) d\eta \approx (1 - \epsilon) \int_{-\infty}^{\gamma} \eta f(\eta;\alpha,u) d\eta \tag{6.6.2.8}$$

where f is a Gaussian density. The integral in (6.6.2.8) can be evaluated as follows. Substituting

$$\xi = \frac{\eta - bt_f}{\sigma\sqrt{t_f}} \tag{6.6.2.9}$$

into (6.6.2.8) yields

$$\int_{-\infty}^{\gamma} \eta f(\eta;\alpha,u)d\eta = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\gamma - bt_f}{\sigma\sqrt{t_f}}} (\sigma\sqrt{t_f}\,\xi + bt_f) \exp\left\{\frac{-\xi^2}{2}\right\} d\xi$$

$$= \frac{bt_f}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\gamma - bt_f}{\sigma\sqrt{t_f}}} \exp\left\{\frac{-\xi^2}{2}\right\} d\xi - \frac{\sigma\sqrt{t_f}}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\gamma - bt_f}{\sigma\sqrt{t_f}}} -\xi \exp\left\{\frac{-\xi^2}{2}\right\} d\xi$$

$$= bt_f\,\Phi\left[\frac{\gamma - bt_f}{\sigma\sqrt{t_f}}\right] - \sigma\sqrt{t_f}\, f\left(\frac{\gamma - bt_f}{\sigma\sqrt{t_f}}\right) \tag{6.6.2.10}$$

Φ can be computed using Approximation 6.6.1.1. In the case where $\gamma > \alpha$, no numerical integration is required.

6.6.3. Random Number Generation

The simulations used to evaluate policy performance require the generation of random numbers. Two distribution families are involved. First, an inverse Gaussian random number (with variable parameters) is required. Second, a conditioned Gaussian random number is needed. In order to generate random numbers from these distribution families, two algorithms are required.

6.6.3a. Generation of Inverse Gaussian Random Numbers

The basic idea of this algorithm is to use the acceptance-rejection method [La1] on a transformed inverse Gaussian random variable.

Let (A, b, σ) be the parameters of the inverse Gaussian distribution from which random numbers are to be generated. Denote $\mu \triangleq \frac{A}{b}$, $\lambda \triangleq \frac{A^2}{\sigma^2}$. Let Y be a random variable with density

$$g_Y(y) = \left(1 - \frac{y}{\sqrt{\frac{4\lambda}{\mu} + y^2}}\right)\left(\frac{1}{\sqrt{2\pi}} \exp\left\{\frac{-y^2}{2}\right\}\right), \quad y \in \mathbb{R} \tag{6.6.3.1}$$

Then

$$T = \frac{\mu}{2\lambda}\left[(2\lambda + \mu Y^2) + Y\sqrt{4\mu\lambda + \mu^2 Y^2}\right] = h(Y) \tag{6.6.3.2}$$

is inverse Gaussian with parameters (A, b, σ) ([Chl]). Now acceptance-rejection can easily be used to generate Y because of its relationship to a Gaussian density:

$$g_Y(y) < \frac{2}{\sqrt{2\pi}} \exp\left\{\frac{-y^2}{2}\right\}, \quad y \in \mathbb{R} \tag{6.6.3.3}$$

Algorithm (6.6.3.1): Let

$$Z \sim N(0,1) \qquad U \sim U(0,1) \text{ be independent}$$

$$\text{if } U \leq \frac{1}{2}\left(1 - \frac{Z}{\sqrt{\frac{4\lambda}{\mu} + Z^2}}\right) \tag{6.6.3.4}$$

then $Y = Z$, else generate new (Z,U) and repeat. Then $T = h(Y)$ is inverse Gaussian.

In practice, acceptance-rejection methods can perform poorly since a random number of (Z,U) pairs must be generated for each random number. However, this algorithm performed well in the simulations.

6.6.3b. Generation of Conditional Gaussian Random Numbers

Assuming that a tool failure does not occur, it is necessary to generate the new tool wear value. This is done by generating the conditional change in tool wear:

$$Y \triangleq (X_{t_f} - x_0) \text{ given } T_A > t_f \tag{6.6.3.5}$$

The random variable Y has a density function f_Y given by

$$f_Y(\eta) = \begin{cases} \dfrac{p(\alpha,u,\eta)}{1 - Q_A(t_f \mid x_0 \,;\, u)} & \eta < \alpha \\ 0 & \eta \geq \alpha \end{cases} \tag{6.6.3.6}$$

where the function p is given by (5.7.2.16). Now p is majorized everywhere by a Gaussian density with mean bt_f and variance $\sigma^2 t_f$. Thus acceptance-rejection with

Gaussian random numbers is possible.

Algorithm (6.6.3.2): Let

$$Z \sim N(bt_f\,,\,\sigma^2 t_f\,) \qquad U \sim U(0,1) \text{ be independent}$$

$$\text{if } U \leq \frac{p(\alpha,u,Z)}{f(Z\,;\,\alpha,u)} = w(\eta\,;\,\alpha,u) \tag{6.6.3.7}$$

then $Y = Z$, else generate new (Z,U) and repeat.

CHAPTER 7

APPLICATION OF DIFFUSION-THRESHOLD MODELS TO OTHER MANUFACTURING PROBLEMS

7.1. Introduction

This chapter will consider the application of diffusion-threshold models and the control theoretic approach to other manufacturing problems. The problems that are considered here were introduced in Chapter 3. Each of these problems conveys a viewpoint different than that which is typically taken in manufacturing problems. The emphasis here will be on how these different viewpoints can be accommodated using the concepts developed earlier in this work. The problems considered in this chapter will not be discussed in as much detail as the drilling problem. The intention is to develop a framework for approaching certain types of manufacturing problems using diffusion-threshold models and control theoretic techniques. In so doing, it is hoped that the potential for new approaches to manufacturing problems is exemplified.

The problems in this chapter do not have simple solutions. They are not simple problems. In the first problem, a solution is presented under restricted assumptions. In the second problem, a simulation based approach is described. In the third problem, a dynamic programming type of condition is derived. More than anything, these problems reveal the difficulty of manufacturing problems and the need for new analytic methods for approaching manufacturing problems.

7.2. Multi-tooled Machines

The multi-tooled machine version of the drilling problem has been previously described in Section 3.3. The reference [Shk1] also considers multi-tool problems. In this problem the machine is allowed to have multiple tools that simultaneously machine the part. As was previously stated, this is a common design for high volume special machines. The tools are generally not independently controllable. For this problem, some simplifying assumptions will be made. The independent control variable will again be the feed speed. The spindle speed of each tool will be assumed to vary so as to maintain constant feed. This is a more restrictive assumption for the multi-tooled machine than it was for the single tooled machine. There are practical examples that adhere to these assumptions, however.

A multi-dimensional diffusion-threshold model of multiple tool wear and failure is proposed. Assume that each tool can be described by a diffusion-threshold model, and that the wear processes for each tool are independent. The wear processes of all tools can be thought of as an n dimensional diffusion, where n is the number of tools. The drift vector $\underline{b}$ and the diffusion matrix $\underline{\sigma}$ are functions of the feed speed u, which is a scalar.

The failure of each tool corresponds to a threshold crossing of the diffusion process component associated with that tool. Failure of the operation is considered to have occurred if there is a failure in any tool. Thus, a crossing of any component threshold corresponds to an operation failure.

The parameters in each diffusion-threshold component may be different. This would be the case if the tools are different, as an example. However, the component drift and diffusion coefficients are assumed to be strictly positive for all values of the control variable u. The component threshold values are assumed to be strictly positive as well. In this work, it will be assumed that all tools begin and end their operations at the same time. This is to simplify the calculations.

The results of Section 4.5 can now be applied in order to determine the distribution function for an operation failure. Let

$$\underline{X}(t) \triangleq (X_1(t), \ldots, X_n(t))^T \tag{7.2.1}$$

be an n-dimensional diffusion process given by

$$\underline{X}(t) = \int_0^t \underline{b}(u(\tau))d\tau + \int_0^t \underline{\sigma}(u(\tau))d\underline{W}(\tau) \tag{4.5.5}$$

where

$$\underline{W}(t) \triangleq (W_1(t), \ldots, W_n(t))^T \tag{4.5.1}$$

$$\underline{b}(u(t)) \triangleq (b_1(u(t)), \ldots, b_n(u(t)))^T,\ b_i(u) \geq 0,\ i = 1, \ldots, n \tag{4.5.2}$$

$$\underline{\sigma}(u(t)) \triangleq \mathrm{diag}(\sigma_1(u(t)), \ldots, \sigma_n(u(t))),\ \sigma_i(u) > 0,\ i = 1, \ldots, n \tag{4.5.3}$$

and u the feed speed is a scalar. Let the threshold be given by a union of hyperplanes

$$\underline{A} = \bigcup_{i=1}^{n} \{\underline{x} \in \mathbb{R}^n : x_i = A_i\},\ A_i > 0,\ i = 1, \ldots, n \tag{4.5.6}$$

Then the threshold crossing probability distribution function is given by

$$Q_{\underline{A}}(t \mid \underline{x}\,;\, u) = 1 - \prod_{i=1}^{n}[1 - Q_{A_i}(t \mid x_i\,;\, u)] \tag{4.5.9}$$

for $x_i < A_i$, $i = 1, \ldots, n$. The density function $q_{\underline{A}}(t \mid \underline{x}\,;\, u)$ is given by

$$q_{\underline{A}}(t \mid \underline{x}\,;\, u) = \frac{d}{dt} Q_{\underline{A}}(t \mid \underline{x}\,;\, u)$$

$$= \sum_{j=1}^{n} \left(q_{A_j}(t \mid x_j\,;\, u) \prod_{i=1,\, i \neq j}^{n} [1 - Q_{A_i}(t \mid x_i\,;\, u)] \right) \tag{7.2.2}$$

The control analysis methods used in the drilling problem can now be extended to the multi-tool case. Age replacement and one step analyses will be separately considered.

7.2.1. Age Replacement Analysis

Age replacement analysis for the multi-tooled drilling problem can be carried out under the assumption of block tool replacement. This means that all tools are replaced at the same time, whether the replacement is voluntary or necessitated by failure. The analysis is considerably more complicated when individual tool replacement is allowed because the renewal framework is then lost. This case will not be discussed here. Under the assumption of block tool replacement the renewal framework is preserved; thus the results of Section 5.6 can be used.

Assume that a policy π consists of the pair (u, n_r), $u \in \mathbb{R}^+$, $n_r \in \{1,2,...\}$ where u is the fixed feed speed and n_r is the age in number of parts at which the block tool replacement is made. Proceeding as in Section 5.6, the distribution function for the tool life T under policy $\pi = (u, n_r)$ is given by:

$$F_T^{\pi}(t) = \begin{cases} Q_{\Delta}(t \mid 0\,;\,u) & \text{if } t < t_r \\ 1 & \text{if } t \geq t_r \end{cases} \tag{7.2.1.1}$$

where $t_r = n_r \dfrac{V}{u}$. Here V is the depth to be drilled by the tools. The expected value of the tool life T is given by:

$$\mu_T \triangleq E^{\pi}[\,T\,] = \int_0^{\infty} t dF_T^{\pi}(t) = \int_0^{t_r} t dQ_{\Delta}(t \mid 0\,;\,u) + t_r(1 - Q_{\Delta}(t_r \mid 0\,;\,u)) \tag{7.2.1.2}$$

Similarly, the expected duration of a tool cycle $E[\widetilde{T}]$ and the expected number of completed parts in a tool cycle $E[\widetilde{P}]$ are given by:

$$E^{\pi}[\,\widetilde{T}\,] = E^{\pi}[\,T\,] + t_B\, Q_{\Delta}(n_r t_f \mid 0\,;\,u) + t_R \tag{7.2.1.3}$$

$$E^{\pi}[\,\tilde{P}\,] = n_r - \sum_{j=0}^{n_r-1} Q_{A}(jt_f \mid 0\,;\,u) \tag{7.2.1.4}$$

Therefore, from (5.6.5.1) and (5.6.5.2), both the long term expected cost per time, and the long term expected cost per part can be computed under policy π.

$$J_t^{\pi} = \frac{E^{\pi}[\,R + KT + B\,1_{[T=T_A]} - G\tilde{P}]}{E^{\pi}[\,\tilde{T}\,]} \tag{7.2.1.5}$$

$$J_p^{\pi} = \frac{E^{\pi}[\,R + KT + B\,1_{[T=T_A]} - G\tilde{P}]}{E^{\pi}[\,\tilde{P}\,]} \tag{7.2.1.6}$$

Although the assumption of block tool replacement is restrictive, it is a common policy in practice. Some exceptions may occur in the case of early tool failure, however. Analytically, the assumption greatly simplifies the problem.

7.2.2. Feedback Analysis

Feedback analysis of the multi-tooled drilling problem using a one step cost functional is also possible. As in the age replacement case, block tool replacement will be assumed. Some modification to the one step cost functional will be required in order to accommodate n dimensional diffusions. Assume that wear measurements of all tools are available.

Define a control $\underline{u} = (u, v)$, $u \in \mathbb{R}^+$, $v \in \{0,1\}$ where u is the feed speed and v is the block tool replacement decision. Let $\underline{x} = (x_1, \ldots, x_n)$ be the initial wear measurement vector, with $x_i < A_i$, $i = 1, \ldots, n$.

Let the one step cost functional be defined as in Section 5.7 with the following changes. The replacement cost is now the cost of replacing all tools, and the repair cost is the cost of repair due to any failure. As can be seen by examining (5.7.2.4a) and (5.7.2.4b), the tool utilization cost term needs to be redefined. That is, equivalents for $D(A - x_0)$ and $D(X_{t_f} - x_0)$ are required.

Consider a normalized tool consumption for each tool:

$$Y_i \triangleq \frac{X_i(t_f) - x_i}{A_i} \tag{7.2.2.1}$$

with

$$0 < Y_i \leq \frac{A_i - x_i}{A_i}, \quad i = 1, \ldots, n \tag{7.2.2.2}$$

Y_i is the percent of tool i consumed during the operation. A suitable measure of tool utilization is the maximum percent consumption among all the tools. This is a measure of the worst case tool utilization, which is in fact what is wanted. Define

$$Y \triangleq \max_{i=1,\ldots,n} \{Y_i\} \tag{7.2.2.3}$$

Then the tool utilization cost becomes $R \cdot Y$ under this criteria. Note that this is based on a normal replacement cost, the same as in the single tool case.

Similarly, the worst case tool consumption in the event of a tool failure is given by $R \cdot y_{\max}$, where

$$y_{\max} \triangleq \max_{i=1,\ldots,n} \left\{ \frac{A_i - x_i}{A_i} \right\} \tag{7.2.2.4}$$

The variable $y_{\max}$ corresponds to the worst case percent tool consumption resulting from a tool failure. Recall that a tool failure results in replacement of all tools under the block replacement assumption.

The one step cost functional for the multi-tooled problem becomes:

$$\text{Minimize } J(\underline{u}, \underline{x}) = (1-v) \cdot J_1(u, \underline{x}) + v \cdot J_2(u, 0) \tag{7.2.2.5}$$

where

$$\begin{aligned} J_1(u, \underline{x}) = E[& -G \cdot 1_{[T_f < T_\Delta]} + (B + R) \cdot 1_{[T_\Delta \leq T_f]} + R \cdot Y \cdot 1_{[T_f < T_\Delta]} \\ & + R \cdot y_{\max} \cdot 1_{[T_\Delta \leq T_f]} + K \cdot (T_f \wedge T_\Delta) \mid \underline{X}(0) = \underline{x}] \end{aligned} \tag{7.2.2.6a}$$

and

$$J_2(u,0) = E[\,-G\cdot 1_{[T_f < T_{\underline{A}}]} + R + (B+R)\cdot 1_{[T_{\underline{A}} \le T_f]} + R\cdot Y\cdot 1_{[T_f < T_{\underline{A}}]}$$

$$+ R\cdot y_{\max}\cdot 1_{[T_{\underline{A}} \le T_f]} + K\cdot (T_f \wedge T_{\underline{A}}) \mid \underline{X}(0) = 0\,] \qquad (7.2.2.6b)$$

Similar to the scalar case, evaluation of the one step cost involves the computation of three functions:

$$s_{\underline{A}}(u \mid \underline{x}) = E[\,1_{[T_{\underline{A}} \le \frac{V}{u}]} \mid \underline{x}\,] = P[\,T_{\underline{A}} \le \frac{V}{u} \mid \underline{x}\,] \qquad (7.2.2.7)$$

$$r_{\underline{A}}(u \mid \underline{x}) = E[\,Y\cdot 1_{[\frac{V}{u} < T_{\underline{A}}]} \mid \underline{x}\,] \qquad (7.2.2.8)$$

$$z_{\underline{A}}(u \mid \underline{x}) = E[\,T_{\underline{A}}\cdot 1_{[T_{\underline{A}} \le \frac{V}{u}]} \mid \underline{x}\,] \qquad (7.2.2.9)$$

The functions $s_{\underline{A}}$ and $z_{\underline{A}}$ can be computed since $Q_{\underline{A}}$ and $q_{\underline{A}}$ are known. Thus, only the function $r_{\underline{A}}$ needs to considered here.

From the definition,

$$r_{\underline{A}}(u \mid \underline{x}) = \int_{-\infty}^{y_{\max}} \eta\, dP(Y{=}\eta,\ t_f < T_{\underline{A}} \mid \underline{x}\,;\, u) \qquad (7.2.2.10)$$

where

$$dP(Y = y,\ t_f < T_{\underline{A}} \mid \underline{x}\,;\, u) = \frac{\partial}{\partial y} P(Y \le y,\ t_f < T_{\underline{A}} \mid \underline{x}\,;\, u)\,dy \qquad (7.2.2.11)$$

Now

$$P(Y \le y,\ t_f < T_{\underline{A}} \mid \underline{x}\,;\, u) = P(Y_i \le y,\ t_f < T_{A_i},\ i = 1, \ldots, n \mid \underline{x}\,;\, u)$$

$$= P(Y_i \le y,\ M_i(t_f) < A_i - x_i,\ i = 1, \ldots, n \mid 0\,;\, u) \qquad (7.2.2.12)$$

where M_i is the component maximum random variable. Since the components of the diffusion are independent, (7.2.2.12) can be expressed as

$$\prod_{i=1}^{n} P(Y_i \le y,\ M_i(t_f) < A_i - x_i \mid 0\,;\, u) \qquad (7.2.2.13)$$

Each term of this product

$$P(Y_i \le y,\ M_i(t_f) < A_i - x_i \mid 0\,;\, u)\,,\quad i = 1, \ldots, n \qquad (7.2.2.14)$$

can be computed from the component joint density function, as given in Section 4.3.

For each i,

$$P(Y_i \le y, M_i(t_f) < A_i - x_i \mid 0 ; u) = \int_{-\infty}^{yA_i} p(\alpha_i, u, \eta) d\eta \tag{7.2.2.15}$$

where $p(\alpha_i, u, \eta)$ is given by (5.7.2.16) and $\alpha_i = A_i - x_i$. Thus

$$\frac{\partial}{\partial y} \int_{-\infty}^{yA_i} p(\alpha_i, u, \eta) d\eta =$$

$$\begin{cases} \dfrac{A_i}{\sqrt{2\pi t_f}\,\sigma_i} \exp\left\{\dfrac{-(yA_i - b_i t_f)^2}{2\sigma_i^2 t_f}\right\} \left[1 - \exp\left\{\dfrac{-2\alpha_i(\alpha_i - yA_i)}{\sigma_i^2 t_f}\right\}\right] & \text{for } y \le \dfrac{\alpha_i}{A_i} \\ 0 & \text{for } y > \dfrac{\alpha_i}{A_i} \end{cases} \tag{7.2.2.16}$$

So

$$r_{\underline{A}}(u \mid \underline{x}) = \sum_{j=1}^{n} \int_{-\infty}^{\frac{\alpha_j}{A_j}} \frac{yA_j}{\sqrt{2\pi t_f}\,\sigma_j} \exp\left\{\frac{-(yA_j - b_j t_f)^2}{2\sigma_j^2 t_f}\right\} \left[1 - \exp\left\{\frac{-2\alpha_j(\alpha_j - yA_j)}{\sigma_j^2 t_f}\right\}\right]$$

$$\times \prod_{i=1,\, i \ne j}^{n} \left[\int_{-\infty}^{yA_i \wedge \alpha_i} \frac{1}{\sqrt{2\pi t_f}\,\sigma_i} \exp\left\{\frac{-(\xi - b_i t_f)^2}{2\sigma_i^2 t_f}\right\} \left[1 - \exp\left\{\frac{-2\alpha_i(\alpha_i - \xi)}{\sigma_i^2 t_f}\right\}\right] d\xi\right] dy \tag{7.2.2.17}$$

This can be simplified somewhat by substituting $\theta = yA_j$ in the first term. Then the expression becomes:

$$r_{\underline{A}}(u \mid \underline{x}) = \sum_{j=1}^{n} \int_{-\infty}^{\alpha_j} \frac{\theta}{\sqrt{2\pi t_f}\,\sigma_j} \exp\left\{\frac{-(\theta - b_j t_f)^2}{2\sigma_j^2 t_f}\right\} \left[1 - \exp\left\{\frac{-2\alpha_j(\alpha_j - \theta)}{\sigma_j^2 t_f}\right\}\right]$$

$$\times \prod_{i=1,\, i \ne j}^{n} \left[\int_{-\infty}^{\theta\frac{A_i}{A_j} \wedge \alpha_i} \frac{1}{\sqrt{2\pi t_f}\,\sigma_i} \exp\left\{\frac{-(\xi - b_i t_f)^2}{2\sigma_i^2 t_f}\right\} \left[1 - \exp\left\{\frac{-2\alpha_i(\alpha_i - \xi)}{\sigma_i^2 t_f}\right\}\right] d\xi\right] d\theta \tag{7.2.2.18}$$

This gives the expected maximum percent tool consumption. The expression is very complicated, but in principle computable. Thus the one step cost functional under block tool replacement can be computed.

Suppose now that individual tool replacements are allowed. For the case of feedback policies, there are 2^n possible replacement decisions. For each replacement decision, there may be a different replacement cost. It is also possible that there are different repair costs for different tool failures. Consequently, the problem becomes enormously complex as n gets large. This will not be further discussed in this work.

The multi-tooled drilling problem can be analyzed under both age replacement and feedback policies in a manner similar to the one tool case. In general, the computations become much more complicated, even under a block tool replacement assumption.

7.2.3. Numerical Example

The ideas of the multi-tooled problem are demonstrated here in a two tool age replacement analysis. Assume that each tool is a drill identical to the one considered in Chapter 6. Thus, each of the two component diffusions has the same parameter values. In this case, the threshold crossing distribution and density functions are given in terms of the scalar functions by:

$$Q_{\underline{A}}(t \mid 0 \,;\, u) = Q_A(t \mid 0 \,;\, u)[2 - Q_A(t \mid 0 \,;\, u)] \tag{7.2.3.1}$$

$$q_{\underline{A}}(t \mid 0 \,;\, u) = 2q_A(t \mid 0 \,;\, u)[1 - Q_A(t \mid 0 \,;\, u)] \tag{7.2.3.2}$$

The long term expected cost per part and long term expected cost per time were computed using the same economic parameter values used in Chapter 6. The piece profit (G) was chosen to be 1.00 dollar. The results are shown in Figs. 7.2.3.1 and 7.2.3.2 respectively.

Comparison of these figures to those for a single tool (Figs. 6.3.1.1 and 6.3.1.4) with the same economic conditions reveals only small differences. The expected cost per part graphs (Figs. 6.3.1.4 and 7.2.3.1) indicate that the optimal policy for two tools uses a slightly slower feed speed than for the one tool case. The expected cost per time graphs (Figs. 6.3.1.1 and 7.2.3.2) indicate an earlier replacement age for the two tool case: 16 - 18 parts vs. 18 - 20 parts.

Generally, the optimal policies for the two tool case are somewhat more conservative than the comparable policies for the one tool case.

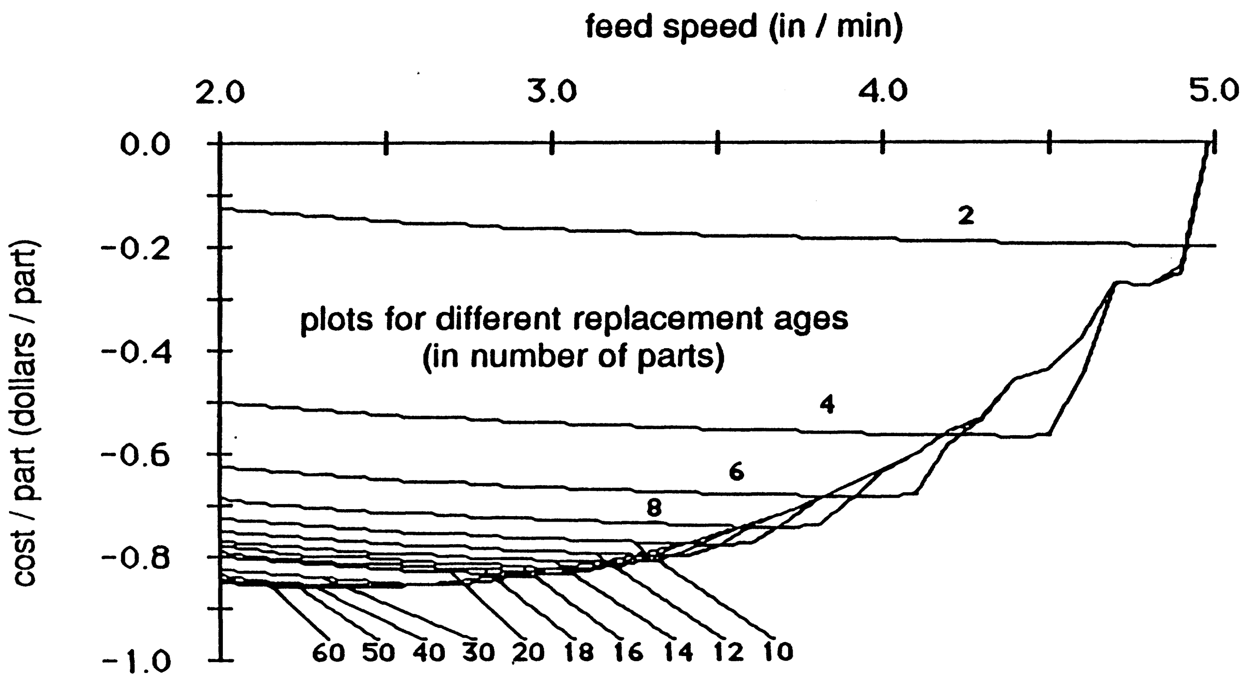

Fig. 7.2.3.1 Expected Cost per Part for 2 Tools under Age Replacement for G = 1

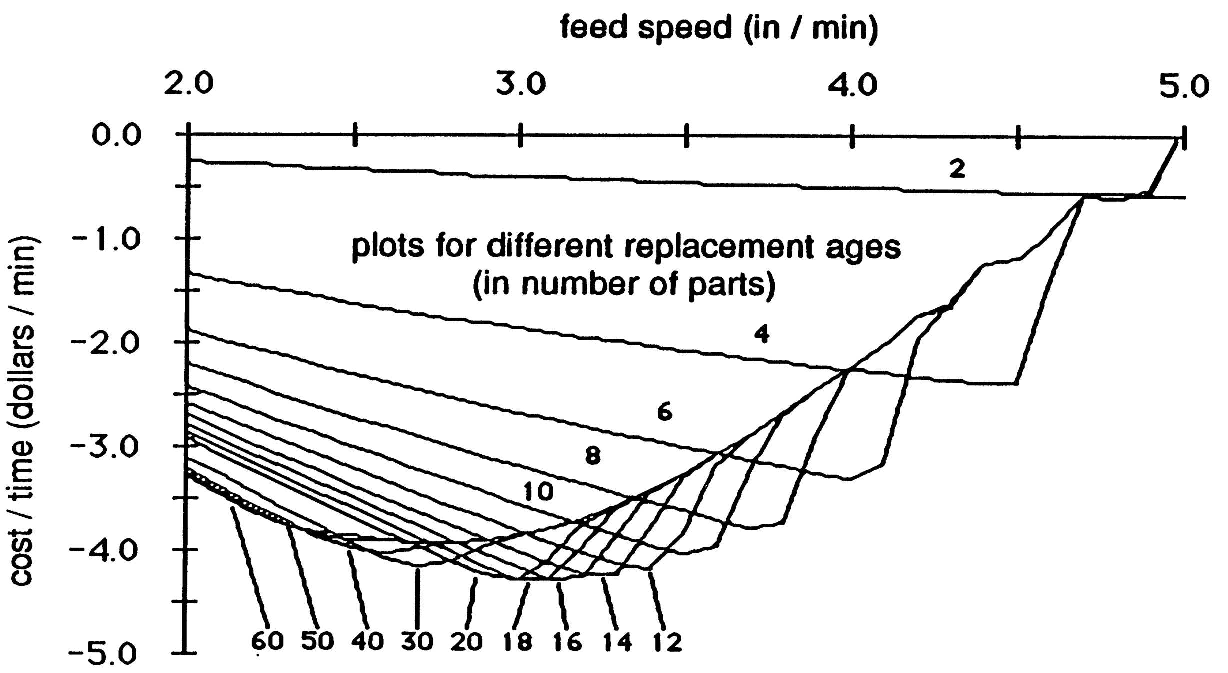

Fig. 7.2.3.2 Expected Cost per Time for 2 Tools under Age Replacement for G = 1

7.3. The Multi-Machine Problem: A Serial Transfer Line Example

Manufacturing systems are rarely as simple as the system considered in the drilling problem. Typical manufacturing systems involve many operations on many machines, with a variety of support equipment such as material handling and buffer storage. All of these factors make the analysis of realistic manufacturing systems much more complex than the analysis of a single machine.

The intent here is to provide one framework for approaching more general manufacturing systems, using the concepts developed for the single machine case. The framework will be demonstrated through an example problem. The approach taken is strongly motivated by practical concerns for dealing with real manufacturing problems. That is, the concern is to develop techniques for controlling real systems in such a way as to improve the performance of the system. Optimality is not an issue here. The direction taken is very much an extension of the concepts laid down in the development of the drilling problem. However, the presentation is only one example of what the author believes is the true potential of the control theoretic approach to manufacturing.

Consider a multi-machine manufacturing system, where each machine in the system exhibits properties similar to those present in the drilling problem. That is, each machine has tooling that wears with use and must be occasionally replaced. Tool failure can result in scrapped parts. Also, the operational speed of each machine is a controllable variable. The main difference between this problem and the single machine problem is that the machines are now interconnected. Thus, the performance of any one machine may affect the performance of the other machines and the performance of the overall system. The problem is to devise ways of operating the system so as to obtain good overall performance.

It is intended that the multi-machine problem be viewed in an intelligent manufacturing setting. That is, assume that information regarding the system (tool wear, buffer levels, etc.) is available. This information is to be incorporated into the operational decision process. There is another issue of great significance, however, that arises in this problem. The issue is the distribution of control and information in the manufacturing system. This issue is important both analytically and practically.

The complexity and size of manufacturing systems makes the design and implementation of systems with centralized control and information difficult. Furthermore, the volume of information that can arise can easily overwhelm the capabilities of a communication system, especially if the information is required frequently for control purposes. Also, centralized control can impact system reliability and expandability. Finally, the design and implementation of such centralized systems is very complex and error prone.

A more desirable situation might be attained with distributed control and information. This has many advantages in terms of design, implementation, expandability, and reliability. Unfortunately, distributed systems lead to decentralized control, and this can greatly complicate the problem analytically. Also, it is not immediately clear how the control and information should be distributed and organized. This issue distinguishes the single machine and multi-machine problems.

The author believes that the practical advantages of distributed control and information far outweigh the analytical disadvantages. However, this position can only be supported if there are ways of eliciting good performance out of a distributed system. The problem that will be considered here will have that as its goal.

Returning now to the multi-machine problem, a diffusion-threshold model of each machine is proposed. This effectively gives rise to a network of diffusion-threshold processes, and possibly buffers and material handling equipment. This is still far too general a problem to deal with, so only a very special class of systems will be considered. This is the class of serial transfer lines.

The serial transfer lines considered here consist of a sequence of machines that perform operations on the parts. It is assumed that there is no intermediate buffer storage between machines. The material handling system (the transfer system) is synchronous; all parts are transferred at the same time to the next machine in sequence. The machines themselves are independent, however. It should be noted that this is a common configuration for high volume production systems.

The implications of restricting attention to this class of manufacturing systems are as follows. First, the production rate of the system can be no faster than the production rate of the slowest machine in the system. Second, a failure in any one machine results in a system stoppage until that machine is repaired. Consequently, each machine in the transfer line can strongly influence the total system.

Analytic treatment of transfer lines under some restricted control policies (i.e., traditional) has been considered by other authors. That will not be the goal here. Rather, an approach that builds upon the results of the single machine problem while maintaining a distributed structure will be proposed. Since analytic treatment is not immediately available, simulation will be used to test the approach for one simple case.

Machining economics problems for serial transfer lines have been considered in the literature. Typically, these works have dealt with deterministic formulations of the problem. In these formulations, tool life usually obeys a Taylor type formula. The phenomena of tool failure is not generally represented. The policies proposed in these deterministic analyses are of the age replacement type. One method that has been suggested [Hil] is the following. The optimization procedure is algorithmic. Optimize each machine individually with respect to the overall system criteria. From this, determine the worst case (controlling) machine or machines. Call these the critical machines, since they determine the production rate. Now relax the operation speeds of the non-critical machines in order to improve their economic performance. Whether or not this is the best way to select machining parameters for serial transfer lines will not be discussed here. Rather, the intent here will be to implement policies in a related manner for purposes of comparison. That is, suppose that an essentially local optimization is carried out for each machine in the transfer line, and the optimal local policies are then implemented at each machine in a decentralized way. The local policy may be age replacement, or feedback, or something else. This concept gives rise to the following question. How do decentralized age replacement and decentralized feedback policies compare?

Simulations have been carried out for an especially simple transfer line operating under decentralized age replacement and decentralized feedback policies. This will be described shortly. First, a few remarks are required.

It might be expected that most serial transfer lines will have one or more machines that dominate the system performance. Similarly, there may be one or more machines that have a relatively weak influence on the overall line performance.

As an example, a 10 machine line with 4 critical failure prone machines and 6 unimportant reliable machines may behave in a manner similar to a line made up of just the 4 critical machines. This is not proven here, but seems intuitively plausible. In order to avoid this phenomenon, the simulated transfer line is composed of only 2 identical machines. This also keeps the simulation simple.

7.3.1. Simulation Description

A 2 machine serial transfer line with identical machines is considered. Part transfer is synchronous. Each machine is similar to the one described in the drilling problem. That is, feed speed is considered a variable that affects the rate of operation and the rate of tool wear. Each tool is susceptible to failure which necessitates repair and causes the current part to be scrapped. There are a few differences from the single machine case, however. Repair and maintenance work takes place serially on a first request first served basis. This corresponds to a single worker/operator. If a tool failure occurs on a machine, the other machine may continue until part completion. At that time, a repair on the failed machine may commence. A repair may not begin while the other machine is still running. If both machines fail, repair begins after the second failure. If a failure occurs in the first machine, the part is scrapped and no part will be transferred to the second machine. That is, a hole appears in the part stream. If a failure occurs at the second machine, the part is simply scrapped. The first machine has an unlimited supply of parts, and so holes never appear there. A machine does not cycle if no part is present and hence no tool wear occurs. A part is completed and credited toward production only if it makes it through both machines successfully. Without loss of generality, the part transfer is assumed to take no time.

The control policies for the two machines are completely decentralized and share no information. Although the second machine knows when a failure occurs in the first machine because of the hole, this information is not used in decision making.

The tool wear/tool life models for each machine are identical to the one considered in Chapter 6. That is, each tool is modeled by a diffusion-threshold process with coefficients as determined in Section 6.2. The economic parameters are identical to those in Chapter 6 with a piece profit (G) value of 1.00 dollar. No partial credit for a part is allowed.

7.3.2. Simulation Results

Simulations were run for two production situations under three control policies. In all cases, both machines used identical policies. The three policies were taken from the results in Chapter 6 for the case of a piece profit of 1.00 dollar. It was found there that age replacement policy *a* worked best for the long term production situation and age replacement policy *b* was recommended for the batch production situation. Feedback policy *g* was recommended for overall performance. The results of the simulations of a two machine serial transfer line are shown in Tables 7.3.2.1 and 7.3.2.2.

For the batch production situation (Table 7.3.2.1), policy *b* continues to give the best economic performance. As before, this performance is gained at the expense of scrap production and production time. Feedback policy *g* results in less profit, but has much better scrap characteristics and is 17 minutes faster. Policy *a* gives the worst economic performance.

The long term (4800 minutes) production situation (Table 7.3.2.2) shows policy *g* to be clearly better in all characteristics except number of tools used. Note that policy b now gives better economic performance than policy a. This is an unexpected result. In the single machine case policy g resulted in a 3.3% improvement over policy a. Here, though, a 6.8% improvement is realized. Policy g also gives a 4.5% improvement over policy b as well as producing over 1500 more parts.

Based on the above results, it appears that the decentralized approach to implementing feedback policies has some merit. In particular, for longer term production runs, the results here indicate that this approach yields appreciable improvement over the recommended age replacement policy. Furthermore, the improvement is relatively greater than that seen in the single machine case. This is encouraging. Of course, neither the age replacement nor the feedback policies considered here are optimal, but they do represent a reasonable engineering compromise between complexity, good performance, and practical concerns. Clearly, this approach warrants further investigation in the future.

	AR		FR/FS
	a	b	g
Production Time	222	230	213
Profit	788	834	810
Scrap Parts	7	16	0
Tools Used	101	52	83
Parts/Tool	9.917	19.17	11.99
Production Rate	4.509	4.356	4.685

Table 7.3.2.1 Production in 2 Machine Serial Transfer Line of 1000 Parts for G = 1

	AR		FR/FS
	a	b	g
Parts Produced	21641	20911	22488
Profit	17049	17428	18205
Scrap Parts	158	337	0
Tools Used	2182	1091	1876
Parts/Tool	9.917	19.17	11.99
Production Rate	4.509	4.356	4.685

Table 7.3.2.2 Production in 2 Machine Serial Transfer Line for 1 Week (4800 min) for G = 1

7.4. The Supervisor's Problem

The supervisor's problem was introduced in Chapter 3. The gist of the problem is the determination of when to inspect and how to assign resources to a project. The problem formulation allows for the possibility of occasional progress assessment (i.e., feedback) and re-assignment of resources (control).

The supervisor's problem is at once dissimilar and similar to the other manufacturing problems considered. The dissimilarity is perhaps most apparent. People have replaced machines; reliability and failure are no longer central to the problem; the time to complete a task is now stochastic. However, the problems do have some similarity. All represent problem situations that arise in manufacturing. All of the problems capture viewpoints different than those present in usual problem formulations. Finally, all of the problems admit a diffusion-threshold model allowing each problem to be placed in a stochastic control theoretic setting.

The supervisor's problem attempts to capture the true role of the supervisor in scheduling problems. That is, most scheduling problems fail to include the underlying dynamics of the system being scheduled. The supervisor's problem extracts these hidden dynamics and brings them to the front.

7.4.1. Diffusion-Threshold Model of the Supervisor's Problem

A model of use in the supervisor's problem using diffusion-threshold processes was introduced in Section 4.6.6. The idea is to model the progress on a project as a diffusion process, with the threshold representing the required amount of work for project completion. The infinitesimal coefficients are functions of the amount of resources committed to the project.

Project assessment can now be represented as occasional measurement of the diffusion process. The supervisor has the role of a controller. Based upon measurement feedback of project progress, resources are re-assigned to the project. However, the supervisor has the additional duty of scheduling the next assessment. Thus, the interval between measurements is in general random.

Some simplifying assumptions will be made. Assume that the amount of resources committed to the project is constant between assessments. Thus, the coefficients of the diffusion are piecewise constant. Assume that project assessment can be done perfectly. Assume also that project assessment and resource assignment can be done in negligible time.

7.4.2. Optimal Control Formulation

The supervisor's problem can be placed in a stochastic optimal control formulation. The goal will be to complete the project at minimal cost.

Define the following given constants

$t_d \triangleq$ project due date

$h \triangleq$ cost rate for overdue project

$g \triangleq$ cost of an inspection

the given function

$f(\cdot) \triangleq$ cost rate function for resources

the random stopping time

$T_A \triangleq$ project completion time and the control variable

$u \triangleq$ resources assigned to project

Let U be the set of available resources. Assume here that the set U is finite and that h, g, and f are positive. Let a policy $\pi \triangleq (\tau, u)$ where $\tau = (\tau_0, \tau_1, \ldots, \tau_N)$ is an inspection time sequence, $u = (u_0, u_1, \ldots, u_N)$ is a control sequence and N is a random integer. The stopping time τ_i is $F^{X_{\tau_{i-1}}}$ measurable and u_i is $F^{X_{\tau_i}}$ measurable, with $\tau_0 = 0$ and $X_0 = 0$. Let Π be the set of admissible policies. As before, the diffusion process is given by:

$$X_t = \int_0^t b(u(\eta))d\eta + \int_0^t \sigma(u(\eta))dW_\eta \tag{7.4.2.1}$$

where

$$u(\eta) = u_k, \quad \tau_k \leq \eta < \tau_{k+1} \tag{7.4.2.2}$$

The project completion time T_A is a threshold crossing time defined as before.

$$T_A = \inf\{t \geq 0 : X_t \geq A\} \tag{4.2.2.5}$$

The threshold level A corresponds to the amount of work required to complete the task.

Define a cost functional

$$J^\pi = E[\int_0^{T_A} f(u(\eta))d\eta + \sum_{n=0}^{\infty} g \cdot 1_{[\tau_n < T_A]} + h \cdot (T_A - t_d)^+] \tag{7.4.2.3}$$

Then the objective is to minimize J^π for $\pi \in \Pi$. Assume here that there exists $\pi \in \Pi$ such that $J^\pi < \infty$.

The cost functional J^π captures the cost of resource usage, the cost of inspection, and the cost of lateness associated with a policy π. Of course, there are other candidate cost functionals that could be used as well.

The objective here will be to develop a heuristic dynamic programming approach to the supervisor's problem. The result is quite similar to results obtained in [An1], [An2], [Be1], [Be2]. However, there are some differences that make the supervisor's problem somewhat unique. Rigorous justification in a more technical setting of the dynamic programming approach can be found in the indicated references. Some solution techniques can also be found there. The intent here is only to show how the supervisor's problem can be placed in a mathematical setting.

Define $J^{\pi}(t, s, x, u)$ as the cost to go under policy π at current time t if the last inspection was at time s and resulted in a progress measurement of x and a control u. This is just the standard embedding technique used in dynamic programming. Now define an optimal cost to go as

$$V(t, s, x, u) \triangleq \inf_{\pi \in \Pi} J^{\pi}(t, s, x, u) \qquad (7.4.2.4)$$

This function represents the best that can be done from now on given the current state.

Given the current state (t, s, x, u), there are only two available choices: continue as is, or inspect and possibly change the control u. Consider each choice in turn.

Suppose that continuation is chosen for some additional time Δt. Assume that $t < T_A$. Then

$$V(t, s, x, u) \leq E^{(s,x,u)}[\int_{t}^{t+\Delta t} f(u) 1_{[\eta \leq T_A]} d\eta + V(t+\Delta t, s, x, u)] \qquad (7.4.2.5)$$

if $t + \Delta t < t_d$. If $t \geq t_d$ then

$$V(t, s, x, u) \leq$$

$$E^{(s,x,u)}[\int_t^{t+\Delta t} f(u)1_{[\eta \le T_A]}d\eta + h\cdot(\Delta t \wedge (T_A - t)) + V(t+\Delta t, s, x, u)] \quad \textbf{(7.4.2.6)}$$

Consider first the case $t+\Delta t < t_d$. Note that if $t < t_d$ such a Δt can always be found.

$$V(t, s, x, u) \le f(u) \int_t^{t+\Delta t} E^{(s,x,u)}[1_{[\eta \le T_A]}]d\eta + E^{(s,x,u)}[V(t+\Delta t, s, x, u)]$$

$$= f(u) \int_t^{t+\Delta t} P[\eta \le T_A \mid T_A > t, X_s = x, u]d\eta$$

$$+ V(t+\Delta t, s, x, u)\cdot P[T_A \ge t+\Delta t \mid T_A > t, X_s = x, u] \quad (7.4.2.7)$$

Define the function

$$\lambda(r \mid x ; u) \triangleq \frac{q_A(r \mid x ; u)}{1 - Q_A(r \mid x ; u)} \quad (7.4.2.8)$$

Then

$$P[t < T_A \le t+dt \mid T_A > t, X_s = x, u] \approx \lambda(t - s \mid x ; u)dt \quad (7.4.2.9)$$

and

$$P[T_A \ge t+dt \mid T_A > t, X_s = x, u] \approx 1 - \lambda(t - s \mid x ; u)dt \quad (7.4.2.10)$$

Note that $\lambda(\cdot \mid x ; u)$ is just the failure rate function associated with the random variable T_A given the initial condition x and control u.

Now for small Δt,

$$P[T_A \ge t+\Delta t \mid T_A > t, X_s = x, u] \approx 1 - \lambda(t - s \mid x ; u)\Delta t \quad (7.4.2.11)$$

and

$$\int_t^{t+\Delta t} P[\eta \le T_A \mid T_A > t, X_s = x, u]d\eta \approx \Delta t \quad (7.4.2.12)$$

So for small Δt, the inequality is approximately

$$V(t, s, x, u) \le$$

$$f(u)\Delta t + V(t+\Delta t, s, x, u) - V(t+\Delta t, s, x, u)\cdot\lambda(t - s \mid x ; u)\Delta t \quad (7.4.2.13)$$

Dividing by Δt and taking the limit as $\Delta t \rightarrow 0$ results in

$$\frac{-\partial}{\partial t} V(t, s, x, u) + V(t, s, x, u) \cdot \lambda(t - s \mid x ; u) - f(u) \leq 0 \qquad (7.4.2.14)$$

In the case where $t \geq t_d$ a similar inequality can be obtained:

$$\frac{-\partial}{\partial t} V(t, s, x, u) + V(t, s, x, u) \cdot \lambda(t - s \mid x ; u)$$

$$- f(u) - h(1 - \lambda(t - s \mid x ; u)) \leq 0 \qquad (7.4.2.15)$$

A combined statement of (7.4.2.14) and (7.4.2.15) is

$$\frac{-\partial}{\partial t} V(t, s, x, u) + V(t, s, x, u) \cdot \lambda(t - s \mid x ; u)$$

$$- f(u) - h 1_{[t \geq t_d]}(1 - \lambda(t - s \mid x ; u)) \leq 0 \qquad (7.4.2.16)$$

Suppose now that an inspection is made. Then

$$V(t, s, x, u) \leq \min_{u' \in U} E^{(s,x,u)}[\, g + V(t, t, X_t, u')\,] \qquad (7.4.2.17)$$

Since one of these two choices must be made, the function V must satisfy

$$[\, \frac{-\partial}{\partial t} V(t, s, x, u) + V(t, s, x, u) \cdot \lambda(t - s \mid x ; u) - f(u) - h 1_{[t \geq t_d]}(1 - \lambda(t - s \mid x ; u))\,] \cdot$$

$$[\, V(t, s, x, u) - \min_{u' \in U} E^{(s,x,u)}[\, g + V(t, t, X_t, u')\,]\,] = 0 \qquad (7.4.2.18)$$

and

$$V(t, s, x, u) = 0 \quad \text{for } x \geq A \qquad (7.4.2.19)$$

This extremely complicated set of equations ((7.4.2.16), (7.4.2.17), (7.4.2.18), (7.4.2.19)) is a form of a so called quasi-variational inequality ([An1], [An2], [Be1], [Be2]) and has been studied by these and other authors. Closed form solution is generally impossible, but numerical solution techniques have been discussed. Characterization of the solutions for some types of quasi-variational problems has also been discussed. Needless to say, much work is still needed in the area.

As stated earlier, the goal was not to solve the problem explicitly, so no further attempt to resolve the supervisor's problem will be given here. The example shows how a problem arising in manufacturing might be approached by diffusion-threshold modeling and control theoretic techniques.

CHAPTER 8

SUMMARY AND CONCLUSIONS

The pressing need for improvement that confronts manufacturing today has inspired the composition of the central theme of this work: How can the feedback of information be effectively used to improve the manufacturing system?

A control theoretic approach has been proposed as a methodology for addressing this problem. Unfortunately, the application of control theory concepts to manufacturing problems is not immediate. Several obstacles must be overcome in order to effect the application. Of particular concern in this work has been the question of modeling manufacturing problems in such a way so as to capture the important features while maintaining a compatibility with the control theory framework.

The control theoretic approach has been conveyed through four example manufacturing problems. These problems have been formulated to capture important manufacturing features and to show viewpoints different from existing manufacturing problems.

The problem of modeling has led to the development of a class of stochastic processes called diffusion-threshold processes. The objective in the study of diffusion-threshold processes is to obtain a probabilistic description of the stopping time associated with a threshold crossing by a diffusion. Although generally difficult, several cases of interest have been developed where a distribution function can be computed.

An important feature of diffusion-threshold processes is that they combine continuous and discrete phenomena into a single process. This allows diffusion-threshold processes to be used in the modeling of discrete events generated by underlying continuous processes. The implication of viewing discrete event systems this way is that information feedback and control can now be incorporated into the system naturally through the continuous process.

Several examples of the application of diffusion-threshold processes to general classes of problems have been given. Diffusion-threshold processes are central to the study of the example manufacturing problems as well.

The most developed example problem is the drilling problem. The problem is to determine feed speeds and make tool replacement decisions for a drilling operation. A stochastic model for tool wear, based on a diffusion-threshold process, has been introduced. This model allows the drilling problem to be formulated as a stochastic optimal control problem.

The general control problem is difficult, so attention has been focused on two types of policies: age replacement and one step. Age replacement policies are open loop whereas one step policies use tool wear feedback.

A comparison of age replacement and one step policies has been carried out through simulation. Two production situations have been considered: a batch run of 1000 parts and a longer term run of two shifts a day for a week. Various age replacement and feedback policies have been compared in terms of economic performance and secondary characteristics.

Two important conclusions can be drawn from this simulation study. The first is that age replacement policies can perform very well, provided that they are chosen (tuned) for the particular production situation. The second conclusion is that feedback policies based on the one step cost functional can give good performance in different production situations. However, some tuning of the policy may be required. For the problems simulated in this work, the importance of feedback in making tool replacement decisions seems to outweigh the importance of feedback in feed speed selection. In many cases, constant feed speed performs very well.

Policies based on one step cost functionals performed better in the longer term production situation than in the batch situation. For one case considered, economic improvement in the longer term production situation over optimal age replacement policies was in the 3% - 4% range. These results are problem specific, but dramatic improvement cannot really be expected. Modest economic improvement and improvement in secondary characteristics is more likely.

The concepts developed in the drilling problem have been extended to the case where multiple tools are used. The computations become more complicated, however. Results have been obtained under a block tool replacement assumption. An example age replacement problem with two tools has been worked out and the results show a somewhat more conservative policy as compared to the one tool problem.

The multiple machine problem remains open. For the case of serial transfer lines, some simulation evidence suggests that decentralized policies based on local one step cost functionals perform better than local age replacement policies. The latter have been proposed by previous authors for transfer lines. The most encouraging observation is that for the two machine case, the percent improvement of one step policies over age replacement policies more than doubled compared to the single machine case. This is clearly an area that deserves further future investigation.

Finally, a different application of diffusion-threshold processes has been considered in the supervisor's problem. The problem attempts to capture the role of a supervisor in overseeing jobs. A heuristic dynamic programming argument leads to a quasi-variational inequality that the optimal cost function must satisfy. This development demonstrates both an alternative interpretation of diffusion-threshold models and a different analytic approach.

This work is only an introduction to the control theoretic approach to manufacturing. Clearly there is a need for much more research. It is hoped that this work contributes to the understanding of how control theoretic ideas can be applied to manufacturing problems, and motivates others to pursue problems in this area.

APPENDICES

APPENDIX A

DRILLING TERMINOLOGY

This appendix summarizes relevant terms used in drilling and in this work. These terms are typical of what is found in machining data handbooks. U.S. customary units and typical metric units are indicated.

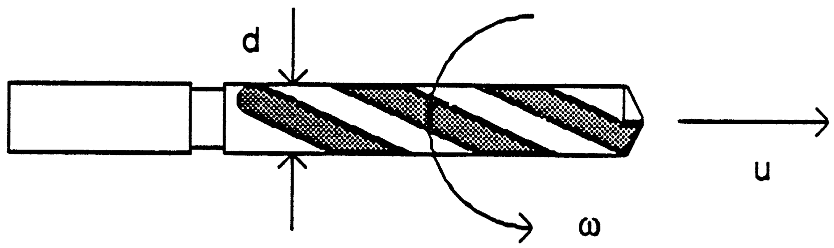

Fig. A.1 Drill

$d \triangleq$ drill diameter (in or mm)

$\omega \triangleq$ spindle speed (rpm)

$u \triangleq$ feed speed (or feedrate) (in/min or mm/min)

$V \triangleq$ cutting speed - the tool/material relative velocity at the outer tool edge (surface feet per minute (sfm) or m/min)

$f \triangleq$ feed - the amount of linear travel of the tool per revolution (in/rev or mm/rev)

In U.S. customary units the cutting speed is given by

$$V = \frac{\pi d \omega}{12} \quad \text{(sfm)} \tag{A.1}$$

The feed is given by

$$f = \frac{u}{\omega} \tag{A.2}$$

Note that

$$V = \frac{\pi d u}{12 f} \tag{A.3}$$

so that only two of the variables u, f, and V are independent. Machining data guides (e.g., [Mac1]) often give recommended values for f and V.

The simple Taylor tool life formula for fixed feed is given by [Dr1]:

$$VT^n = C \tag{A.4}$$

which can be expressed as (5.5.2.1):

$$uT^n = \frac{12 f C}{\pi d} = C_1 \tag{A.5}$$

APPENDIX B

DIFFUSION-THRESHOLD PROCESSES WITH PIECEWISE CONSTANT COEFFICIENTS

In this appendix, the class of diffusion processes is extended to allow for piecewise constant coefficients. This class of processes has not been previously discussed in the literature (to the author's knowledge).

The diffusion-threshold problem is related to a problem in statistics concerned with Brownian motion crossing a curve. Based on the results for piecewise constant coefficients, a partial solution to the Brownian motion curve crossing problem is obtained.

B.1 Calculation of Distribution Function

Consider those diffusions where b and σ depend only on u and where u and thus b and σ are piecewise constant, i.e. there exists a set of times $0=t_0<t_1< \cdots <t_n$ such that

$$u_\tau = u_k,\ t_k \leq \tau < t_{k+1} \text{ for } k = 0,1,\ldots,n-1 \tag{B.1.1}$$

$$u_\tau = u_n,\ t_n \leq \tau \tag{B.1.2}$$

$$b(u_k) = b_k \geq 0 \tag{B.1.3}$$

$$\sigma(u_k) = \sigma_k > 0 \tag{B.1.4}$$

for $k = 0,1,\ldots,n$. Under these assumptions, $\{X_t\}$ is a Brownian motion with piecewise constant drift and piecewise linear variance.

In order to develop the threshold crossing distribution for this process, a result is needed that pertains to the constant coefficient case of Section 4.3.

Theorem (4.3.4): Given $X_t = bt + \sigma W_t$, $b(u) \geq 0$, $\sigma(u) > 0$. Define

$$M_t \triangleq \sup_{s \in [0,t]} X_s \tag{B.1.5}$$

and for $0 \leq t_0 < t_1$

$$M_{[t_0, t_1]} \triangleq \sup_{s \in [t_0, t_1]} X_s \tag{B.1.6}$$

Then the joint density of (X_t, M_t) is given by:

$$g_{X_t, M_t}(\eta, \nu \,;\, t, u) =$$

$$\begin{cases} \dfrac{2(2\nu-\eta)}{\sqrt{2\pi\sigma^3} t^{\frac{3}{2}}} \exp\left\{\dfrac{-(2\nu-\eta-bt)^2}{2\sigma^2 t}\right\} \exp\left\{\dfrac{-2(\nu-\eta)b}{\sigma^2}\right\} & \text{for } \nu \geq \eta,\ \nu \geq 0,\ t > 0 \\ 0 & \text{otherwise} \end{cases} \tag{B.1.7}$$

Proof: The proof of this result uses a technique due to Lévy in his study of driftless Brownian motion, as reported in [It1].

Define the random variable

$$\chi_t(\alpha, \beta) = 1_{[X_t \leq \alpha,\ M_t \geq \beta]} \quad \text{for } \alpha \leq \beta,\ \beta \geq 0,\ t > 0 \tag{B.1.8}$$

Then

$$E[\chi_t(\alpha, \beta)] = P[X_t \leq \alpha\ ,\ M_t \geq \beta\ ] \tag{B.1.9}$$

Now

$$\begin{aligned} \chi_t(\alpha, \beta) &= 1_{[X_t \leq \alpha]} \cdot 1_{[M_t \geq \beta]} \\ &= \phi_t(\alpha) \cdot \psi_t(\beta) \end{aligned} \tag{B.1.10}$$

so that

$$\begin{aligned} \int_0^\infty \exp\{-st\} E[\chi_t(\alpha, \beta)]\, dt &= E\left[\int_0^\infty \exp\{-st\} \phi_t(\alpha) \cdot \psi_t(\beta)\, dt\right] \\ &= E\left[\int_{T_\beta}^\infty \exp\{-st\} \phi_t(\alpha)\, dt\right] \end{aligned}$$

$$= E\left[\exp\{-sT_\beta\}\int_0^\infty \exp\{-st\}\phi_{t+T_\beta}(\alpha)dt\right] \qquad \textbf{(B.1.11)}$$

where T_β is the β-threshold crossing time. By using the smoothing property, (B.1.11) can be written as

$$E\left[\exp\{-sT_\beta\}\; E\left[\int_0^\infty \exp\{-st\}\phi_{t+T_\beta}(\alpha)dt \mid F^{T_\beta}\right]\right] \qquad \text{(B.1.12)}$$

where F^{T_β} is the σ-algebra induced by T_β. Using the strong Markov property and time shift (B.1.12) becomes

$$E\left[\exp\{-sT_\beta\}E\left[\int_0^\infty \exp\{-st\}\phi_t(\alpha)dt \mid X_0{=}\beta\right]\right]$$

$$= E\left[\exp\{-sT_\beta\}\right]\; E\left[\int_0^\infty \exp\{-st\}\phi_t(\alpha)dt \mid X_0{=}\beta\right] \qquad \text{(B.1.13)}$$

again using the strong Markov property to decouple past and future of X_t by $X_{T_\beta} = \beta$. But (B.1.13) is just

$$\hat{q}_\beta(s \mid 0\,;\, u)\cdot E\left[\int_0^\infty \exp\{-st\}\phi_t(\alpha)dt \mid X_0{=}\beta\right] \qquad \text{(B.1.14)}$$

where the first term is the Laplace transform of the threshold crossing density. The second term can be evaluated as follows.

$$E\left[\int_0^\infty \exp\{-st\}\phi_t(\alpha)dt \mid X_0{=}\beta\right] = \int_0^\infty \exp\{-st\}E\left[\phi_t(\alpha) \mid X_0{=}\beta\right]dt$$

$$= \int_0^\infty \exp\{-st\}\int_{-\infty}^{\alpha} f_{X_t \mid X_0}(\eta \mid \beta)d\eta\; dt$$

$$= \int_{-\infty}^{\alpha}\int_0^\infty \exp\{-st\}f_{X_t \mid X_0}(\eta \mid \beta)dt\; d\eta \qquad \text{(B.1.15)}$$

where f is a conditional Gaussian density. But

$$\int_0^\infty \exp\{-st\}f_{X_t \mid X_0}(\eta \mid \beta)dt \qquad \text{(B.1.16)}$$

is just the Laplace transform of a conditional Gaussian density and is given by

$$\frac{\exp\left\{\frac{(\eta-\beta)b}{\sigma^2}\right\}\exp\left\{\frac{-\mid\eta-\beta\mid}{\sigma}\sqrt{2(s-a)}\right\}}{\sqrt{2(s-a)}\sigma} \tag{B.1.17}$$

where

$$a = \frac{-b^2}{2\sigma^2} \tag{B.1.18}$$

Now substitute (B.1.17) into (B.1.14) and differentiate with respect to α and β and negate to obtain the Laplace transform of the joint density

$$\begin{cases} \frac{2}{\sigma^2}\exp\left\{\frac{\alpha\beta}{\sigma^2}\right\}\exp\left\{\frac{-(2\beta-\alpha)}{\sigma}\sqrt{2(s-a)}\right\} & \text{for } \beta \geq \alpha \\ 0 & \text{for } \beta < \alpha \end{cases} \tag{B.1.19}$$

Taking the inverse transform and simplifying gives the desired result:

$$\begin{cases} \frac{2(2\beta-\alpha)}{\sqrt{2\pi\sigma^3}t^{\frac{3}{2}}}\exp\left\{\frac{-(2\beta-\alpha-bt)^2}{2\sigma^2 t}\right\}\exp\left\{\frac{-2(\beta-\alpha)b}{\sigma^2}\right\} & \text{for } \beta\geq\alpha,\ \beta\geq 0,\ t>0 \\ 0 & \text{otherwise} \end{cases} \tag{B.1.20}$$

Remark: As before, the result holds for $b = 0$ (studied by Lévy) but does not hold for $b < 0$.

The result easily generalizes to the case of nonzero initial condition $X_0 = x_0$, with $\nu \geq x_0$ as given in the next corollary

Corollary (B.1.1): Given $X_t = x_0 + bt + \sigma W_t$, $b(u) \geq 0$, $\sigma(u) > 0$. Then the joint density of (X_t, M_t) is given by:

$$g_{X_t,M_t}(\eta,\nu \mid x_0\,;\,t,u) = \begin{cases} g_{X_t,M_t}(\eta-x_0,\nu-x_0\,;\,t,u) & \text{for } \nu\geq\eta,\ \nu\geq x_0,\ t>0 \\ 0 & \text{otherwise} \end{cases} \tag{B.1.21}$$

Remark: The result extends to non-zero initial time as well.

It is possible to verify that the marginal densities from the above joint density agree with the known results. Verifying the density for X_t is straightforward, but verifying the density for M_t is more difficult.

Using the joint density for the maximum and final value of the process, the piecewise constant coefficient threshold crossing distribution can be developed. The term multi-step will be used synonymously with piecewise constant.

Theorem (B.1.2): (Multi-step Threshold Crossing Distribution) Let

$$X_t = \int_0^t b(u(\tau))d\tau + \int_0^t \sigma(u(\tau))dW_\tau \tag{B.1.22}$$

where

$$u(\tau) = u_k,\ t_k \leq \tau < t_{k+1} \text{ for } k = 0,1,\ldots,n-1 \tag{B.1.23}$$

$$u(\tau) = u_n,\ t_n \leq \tau \tag{B.1.24}$$

$$u^n = (u_0, \ldots, u_n) \tag{B.1.25}$$

$$b(u_k) = b_k \geq 0 \tag{B.1.26}$$

$$\sigma(u_k) = \sigma_k > 0 \tag{B.1.27}$$

$$0 = t_0 < t_1 < \cdots < t_n \tag{B.1.28}$$

and let $A > 0$ be a threshold value. Then the threshold crossing probability distribution is given by:

$$Q_A^n(t \mid 0 ; u^n) = P\{ T_A \leq t \mid X_0 = 0 ; u^n\} =$$

$$\begin{cases} Q_A^{n-1}(t \mid 0 ; u^{n-1}) & \text{for } 0 \leq t < t_n \\ Q_A^{n-1}(t_n \mid 0 ; u^{n-1}) + \int_{-\infty}^{A} Q_A(t-t_n \mid x ; u_n) G_A^{n-1}(x, t_n ; u^{n-1}) dx & \text{for } t_n \leq t \end{cases} \tag{B.1.29}$$

where

$$Q_A^0(t \mid 0 \, ; \, u) = Q_A(t \mid 0 \, ; \, u) \tag{B.1.30}$$

and

$$G_A^n(x, t_{n+1} \, ; \, u^n) = \int_{-\infty}^{A} G_{A-y}^0(x-y, t_{n+1}-t_n \, ; \, u_n) G_A^{n-1}(y, t_n \, ; \, u^{n-1}) dy \tag{B.1.31}$$

$$G_A^0(x, t_1 \, ; \, u) = \int_{(0 \vee x)}^{A} g(x, \nu \, ; \, t_1, u) d\nu \tag{B.1.32}$$

Remark: The distribution for n steps (changes in u) can be calculated recursively from the distribution for $n-1$ steps, assuming the same control sequence $(u_0 , \ldots , u_{n-1})$ and time of changes $(t_0 , \ldots , t_{n-1})$.

Remark:

$$G_A^n(x, t_{n+1} \, ; \, u^n) = \frac{\partial}{\partial x} P[X_{t_{n+1}} \leq x \, , \, M_{t_{n+1}} < A] \tag{B.1.33}$$

Proof: Let

$$u^{n-1} = (u_0 , \ldots , u_{n-1}) \tag{B.1.34}$$

$$u^n = (u^{n-1}, u_n) \tag{B.1.35}$$

$$0 = t_0 < t_1 < \cdots < t_{n-1} < t_n \tag{B.1.36}$$

Then for $t \leq t_n$, the processes under u^{n-1} and u^n are identical. Therefore

$$Q_A^n(t \mid 0 \, ; \, u^n) = Q_A^{n-1}(t \mid 0 \, ; \, u^{n-1}) \text{ for } 0 \leq t \leq t_n \tag{B.1.37}$$

and it suffices to prove the update part of the formula. Let $t_n \leq t$. Then

$$\begin{aligned} P[T_A \leq t] &= P[T_A \in [0, t_n]] + P[T_A \in (t_n, t]] \\ &= Q_A^{n-1}(t_n \mid 0 \, ; \, u^{n-1}) + P[T_A \in (t_n, t]] \end{aligned} \tag{B.1.38}$$

Now

$$P[T_A \in (t_n, t]] = \int_{-\infty}^{A} P[T_A \in (t_n, t] \mid X_{t_n} = x, \, M_{t_n} < A] dP[X_{t_n} = x, \, M_{t_n} < A]$$

$$= \int_{-\infty}^{A} P[T_A \in (t_n, t] \mid X_{t_n} = x] dP[X_{t_n} = x,\ M_{t_n} < A]$$

$$= \int_{-\infty}^{A} Q_A(t_n - t \mid x\ ;\ u_n) dP[X_{t_n} = x,\ M_{t_n} < A] \tag{B.1.39}$$

Denote

$$\frac{\partial}{\partial x} P[X_{t_n} \leq x,\ M_{t_n} < A] = G_A^{n-1}(x, t_n\ ;\ u^{n-1}) \tag{B.1.40}$$

and this is the first part of the result. Next, the update formula for G_A^n must be determined. Let $n=1$. Then

$$G_A^0(x, t_1\ ;\ u_0) = \frac{\partial}{\partial x} P[X_{t_1} \leq x,\ M_{t_1} < A]$$

$$= \frac{\partial}{\partial x} \int_{(0 \vee x)}^{A} \int_{-\infty}^{x} g(\eta, \nu\ ;\ t_1, u_0) d\eta\, d\nu$$

$$= \int_{(0 \vee x)}^{A} g(x, \nu\ ;\ t_1, u_0) d\nu \tag{B.1.41}$$

where $g(\cdot, \cdot\ ;\ t_1, u_0)$ is the joint density function from Theorem 4.3.4. For $n > 1$

$$G_A^n(x, t_{n+1}\ ;\ u^n) = \frac{\partial}{\partial x} P[X_{t_{n+1}} \leq x,\ M_{t_{n+1}} < A]$$

$$= \frac{\partial}{\partial x} \int_{-\infty}^{A} P[X_{t_{n+1}} \leq x,\ X_{t_n} = y,\ M_{t_n} < A,\ M_{(t_n, t_{n+1}]} < A] dy$$

$$= \frac{\partial}{\partial x} \int_{-\infty}^{A} P[X_{t_{n+1}} \leq x,\ M_{(t_n, t_{n+1}]} < A \mid X_{t_n} = y,\ M_{t_n} < A] P[X_{t_n} = y,\ M_{t_n} < A] dy$$

$$= \frac{\partial}{\partial x} \int_{-\infty}^{A} P[X_{t_{n+1}} \leq x,\ M_{(t_n, t_{n+1}]} < A \mid X_{t_n} = y] G_A^{n-1}(y, t_n\ ;\ u^{n-1}) dy$$

$$= \int_{-\infty}^{A} G_{A-y}^0(x - y, t_{n+1} - t_n\ ;\ u_n) G_A^{n-1}(y, t_n\ ;\ u^{n-1}) dy \tag{B.1.42}$$

which is the required update formula.

Remark: The formula assumes a deterministic sequence

$$\{u_0, u_1, \ldots, u_n\}$$

for the control. That is, this is the threshold crossing distribution for open loop finite control sequences. A similar formula for closed loop Markov control sequences is considerably more complicated.

B.2 Curve Crossing Problems for Brownian motion

A problem of some interest in the statistics literature is the following. Given a Brownian motion and an arbitrary curve $\gamma : \mathbb{R}^+ \to \mathbb{R}$ as a function of time determine the distribution and moments for the time of first crossing of the curve by a Brownian motion. See [Brel], [Led1], [Rob1].

As will be shown, the curve crossing problem for Brownian motion is related to the multi-step problem previously considered. Based on the results for the multi-step case, a partial solution to the curve crossing problem is given. In the general formulation of the curve crossing problem, discontinuous curves are allowed provided that the discontinuities are points of increase. Only continuous curves will be considered here.

Let

$$\gamma : \mathbb{R}^+ \to \mathbb{R} \tag{B.2.1}$$

be a continuous function. Let W_t be a standard Brownian motion. Define

$$T_\gamma \triangleq \inf\{t \geq 0 : W_t \geq \gamma(t)\} \tag{B.2.2}$$

T_γ is the time at which the Brownian motion first hits the curve γ, and is a stopping time. The problem is to find the distribution function for T_γ. This is an essentially unsolved problem, except for the case where γ is a decreasing straight line with positive initial value.

Definition: Let C^n be the class of decreasing concave piecewise linear curves such that for $\gamma \in C^n$

$$\gamma : \mathbb{R}^+ \rightarrow \mathbb{R} \tag{B.2.3}$$

$$\gamma(t) = A - \int_0^t b(\tau) d\tau \tag{B.2.4}$$

where

$$A > 0 \tag{B.2.5}$$

$$b(\tau) = b_k \geq 0, \; t_k \leq \tau < t_{k+1} \text{ for } k = 0,1,\ldots,n-1 \tag{B.2.6}$$

$$b(\tau) = b_n \geq 0, \; t_n \leq \tau \tag{B.2.7}$$

$$0 = t_0 < \cdots < t_n \tag{B.2.8}$$

Theorem (B.2.1): Let $\gamma \in C^n$. Then

$$P[T_\gamma < \infty] = 1 \tag{B.2.9}$$

and

$$P[T_\gamma \leq t] = Q_A^n(t \mid 0 \, ; u^n) \tag{B.2.10}$$

where

$$b(u_k) = b_k \tag{B.2.11}$$

$$\sigma(u_k) = 1 \tag{B.2.12}$$

Proof: Let

$$X_t = \int_0^t b(\tau) d\tau + W_t \tag{B.2.13}$$

Then

$$X_t \geq A \iff W_t \geq A - \int_0^t b(\tau) d\tau \tag{B.2.14}$$

so the results of the multi-step threshold crossing problem apply immediately. Thus the first statement follows from Lemma 4.3.1. For the second statement, observe that

$$\{\omega : T_\gamma \le t\} = \{\omega : W_t \ge A - \int_0^t b(\tau)d\tau\}$$

$$= \{\omega : X_t \ge A\} = \{\omega : T_A \le t\} \tag{B.2.15}$$

where T_A is the threshold crossing time for the process X_t. Therefore

$$P[T_\gamma \le t] = P[T_A \le t] = Q_A^n(t \mid 0 ; u^n) \tag{B.2.16}$$

Remark: Clearly there exist curves not in C^n such that $P[T_\gamma = \infty] > 0$.

Remark: The curves considered here ($\in C^n$) are reasonable approximations to the class of decreasing concave asymptotic curves. The following example (Fig. B.2.1) illustrates this:

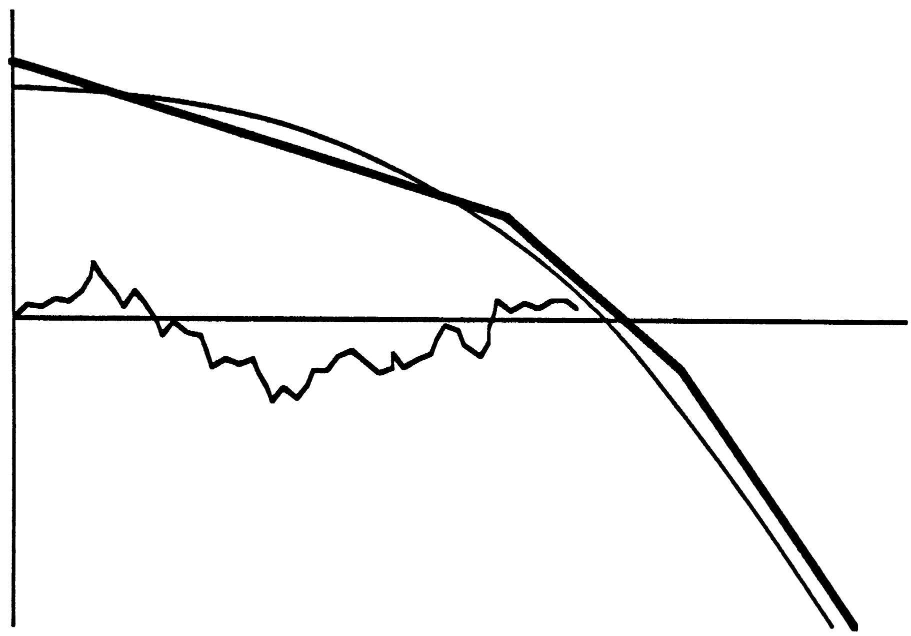

Fig. B.2.1 Example Curve Crossing Problem

APPENDIX C

PROPERTIES OF THE ONE STEP COST FUNCTIONAL

In this appendix, properties of the one step cost functional used in Theorem 5.7.2.1 will be proven. It is sufficient to consider the limiting properties of the functions s_A, r_A, and z_A.

Assume $b(u) = \beta u^m$, $\sigma > 0$ is constant, $\alpha = A - x_0 > 0$, and $t_f = \dfrac{V}{u}$. Without loss of generality, it is sufficient to consider the functions $s_\alpha(\cdot\,|\,0)$, $r_\alpha(\cdot\,|\,0)$, and $z_\alpha(\cdot\,|\,0)$.

Lemma (C.1):

$$\lim_{u \to 0^+} s_\alpha(u\,|\,0) = 1 \tag{C.1}$$

Proof: Note that $s_\alpha(u\,|\,0)$ is bounded

$$0 \leq s_\alpha(u\,|\,0) \leq 1 \tag{C.2}$$

because

$$s_\alpha(u\,|\,0) = P[T_\alpha \leq t_f \;;\; u] \tag{C.3}$$

Now

$$s_\alpha(u\,|\,0) = \int_0^{t_f} q_\alpha(\eta\,|\,0\;;\;u)\,d\eta = \int_0^{\infty} 1_{[\eta \leq t_f]} q_\alpha(\eta\,|\,0\;;\;u)\,d\eta \tag{C.4}$$

By Fatou's lemma

$$\begin{aligned}\varliminf_{u \to 0^+} s_\alpha(u\,|\,0) &= \varliminf_{u \to 0^+} \int_0^{\infty} 1_{[\eta \leq t_f]} q_\alpha(\eta\,|\,0\;;\;u)\,d\eta \\ &\geq \int_0^{\infty} \varliminf_{u \to 0^+} 1_{[\eta \leq t_f]} q_\alpha(\eta\,|\,0\;;\;u)\,d\eta\end{aligned}$$

$$= \int_0^\infty q_\alpha(\eta \mid 0 \,;\, 0) d\eta = 1 \qquad \textbf{(C.5)}$$

Since s_α is bounded above by 1, the result follows.

Remark: This result is true for $m > 0$.

Lemma (C.2): For $m > 1$

$$\lim_{u \to \infty} s_\alpha(u \mid 0) = 1 \qquad (C.6)$$

Proof: This is most easily seen by using the normal form transformation for the inverse Gaussian distribution given in Section 4.3.

$$s_\alpha(u \mid 0) =$$

$$\Phi\left[\frac{\alpha}{\sigma}\sqrt{\frac{u}{V}}\left(\frac{V\beta u^{m-1}}{\alpha} - 1\right)\right] + \exp\left\{\frac{2\alpha\beta u^m}{\sigma^2}\right\}\Phi\left[\frac{-\alpha}{\sigma}\sqrt{\frac{u}{V}}\left(\frac{V\beta u^{m-1}}{\alpha} + 1\right)\right] \qquad (C.7)$$

Then

$$\lim_{u \to \infty} s_\alpha(u \mid 0) =$$

$$\lim_{u \to \infty}\left(\Phi\left[\frac{\alpha}{\sigma}\sqrt{\frac{u}{V}}\left(\frac{V\beta u^{m-1}}{\alpha} - 1\right)\right] + \exp\left\{\frac{2\alpha\beta u^m}{\sigma^2}\right\}\Phi\left[\frac{-\alpha}{\sigma}\sqrt{\frac{u}{V}}\left(\frac{V\beta u^{m-1}}{\alpha} + 1\right)\right]\right)$$

$$\geq 1 + \lim_{u \to \infty}\left(\exp\left\{\frac{2\alpha\beta u^m}{\sigma^2}\right\}\Phi\left[\frac{-\alpha}{\sigma}\sqrt{\frac{u}{V}}\left(\frac{V\beta u^{m-1}}{\alpha} + 1\right)\right]\right) \qquad (C.8)$$

The result follows since s_α is bounded.

Remark: It is relatively easy to show that

$$\lim_{u \to \infty}\left(\exp\left\{\frac{2\alpha\beta u^m}{\sigma^2}\right\}\Phi\left[\frac{-\alpha}{\sigma}\sqrt{\frac{u}{V}}\left(\frac{V\beta u^{m-1}}{\alpha} + 1\right)\right]\right) = 0 \qquad (C.9)$$

Remark: The result in not true for $m \leq 1$. It can be shown that for $m = 1$

$$\lim_{u \to \infty} s_\alpha(u \mid 0) = \begin{cases} 1 & \dfrac{\alpha}{V\beta} < 1 \\ \dfrac{1}{2} & \dfrac{\alpha}{V\beta} = 1 \\ 0 & \dfrac{\alpha}{V\beta} > 1 \end{cases} \tag{C.10}$$

and for $0 < m < 1$

$$\lim_{u \to \infty} s_\alpha(u \mid 0) = 0 \tag{C.11}$$

Lemma (C.3):

$$\lim_{u \to 0^+} z_\alpha(u \mid 0) = \infty \tag{C.12}$$

Proof: By Fatou's lemma

$$\begin{aligned} \varliminf_{u \to 0^+} z_\alpha(u \mid 0) &= \varliminf_{u \to 0^+} \int_0^\infty 1_{[\eta \leq t_f]} \eta q_\alpha(\eta \mid 0 \,;\, u) d\eta \\ &\geq \int_0^\infty \varliminf_{u \to 0^+} 1_{[\eta \leq t_f]} \eta q_\alpha(\eta \mid 0 \,;\, u) d\eta \\ &= \int_0^\infty \eta q_\alpha(\eta \mid 0 \,;\, u) d\eta = \infty \end{aligned} \tag{C.13}$$

where the last result is from Theorem 4.3.2.

Lemma (C.4):

$$\lim_{u \to \infty} z_\alpha(u \mid 0) = 0 \tag{C.14}$$

Proof:

$$\begin{aligned} \lim_{u \to \infty} z_\alpha(u \mid 0) &= \lim_{u \to \infty} \int_0^\infty 1_{[\eta \leq t_f]} \eta q_\alpha(\eta \mid 0 \,;\, u) d\eta \\ &\leq \lim_{u \to \infty} \int_0^\infty \eta q_\alpha(\eta \mid 0 \,;\, u) d\eta \end{aligned}$$

$$= \lim_{u \to \infty} \frac{\alpha}{\beta u^m} = 0 \tag{C.15}$$

But $z_\alpha(u \mid 0) \geq 0$ for all $u \in \mathbb{R}^+$. Therefore the result follows.

Lemma (C.5):

$$\lim_{u \to 0^+} r_\alpha(u \mid 0) = -\alpha \tag{C.16}$$

Proof: Recall that

$$r_\alpha(u \mid 0) = E[X_{t_f} 1_{[t_f < T_\alpha]}] \tag{C.17}$$

Now

$$E[X_{t_f}] = E[X_{t_f} 1_{[t_f < T_\alpha]}] + E[X_{t_f} 1_{[t_f \geq T_\alpha]}] \tag{C.18}$$

So

$$\lim_{u \to 0^+} E[X_{t_f}] = \lim_{u \to 0^+} E[X_{t_f} 1_{[t_f < T_\alpha]}] + \lim_{u \to 0^+} E[X_{t_f} 1_{[t_f \geq T_\alpha]}] \tag{C.19}$$

assuming this is well defined. Since $X_{t_f} \sim N(\beta V u^{m-1}, \sigma^2 t_f)$

$$\lim_{u \to 0^+} E[X_{t_f}] = \lim_{u \to 0^+} \beta V u^{m-1} = 0 \tag{C.20}$$

for $m > 1$. Therefore

$$\lim_{u \to 0^+} r_\alpha(u \mid 0) = - \lim_{u \to 0^+} E[X_{t_f} 1_{[t_f \geq T_\alpha]}] \tag{C.21}$$

By the smoothing property for strong Markov processes

$$E[X_{t_f} 1_{[t_f \geq T_\alpha]}] = E[\, E[X_{t_f} 1_{[t_f \geq T_\alpha} \mid T_\alpha = t] \,]$$

$$= \int_0^{t_f} E[X_{t_f} 1_{[t_f \geq T_\alpha} \mid T_\alpha = t] q_\alpha(t \mid 0 \,;\, u) dt$$

$$= \alpha \int_0^{t_f} q_\alpha(t \mid 0 \,;\, u) dt + \beta u^m \int_0^{t_f} (t_f - t) q_\alpha(t \mid 0 \,;\, u) dt \tag{C.22}$$

using the fact that

$$E[X_{t_f} 1_{[t_f \geq T_\alpha} \mid T_\alpha = t] = \alpha + \beta u^m (t_f - t) \tag{C.23}$$

Taking the limit of each term in (C.22) gives

$$\lim_{u \to 0^+} \alpha \int_0^{t_f} q_\alpha(t \mid 0 \,;\, u)dt = \alpha \lim_{u \to 0^+} s_\alpha(u \mid 0) = \alpha \tag{C.24}$$

$$\lim_{u \to 0^+} \beta u^m \int_0^{t_f} (t_f - t) q_\alpha(t \mid 0 \,;\, u)dt \le \lim_{u \to 0^+} \beta u^m t_f \int_0^{t_f} q_\alpha(t \mid 0 \,;\, u)dt$$

$$= \lim_{u \to 0^+} \beta V u^{m-1} = 0 \tag{C.25}$$

But for $u \in \mathbb{R}^+$

$$\beta u^m \int_0^{t_f} (t_f - t) q_\alpha(t \mid 0 \,;\, u)dt \ge 0 \tag{C.26}$$

Thus the limit of this term is 0. The result now follows.

Remark: Note that

$$\lim_{u \to 0^+} (\, r_\alpha(u \mid 0) + \alpha s_\alpha(u \mid 0) \,) = 0 \tag{C.27}$$

Lemma (C.6):

$$\lim_{u \to \infty} r_\alpha(u \mid 0) = 0 \tag{C.28}$$

Proof:

$$\lim_{u \to \infty} \mid r_\alpha(u \mid 0) \mid^2 = \lim_{u \to \infty} \mid E[X_{t_f} 1_{[t_f < T_\alpha]}] \mid^2$$

$$\le \lim_{u \to \infty} E[X_{t_f}^2] \cdot E[1_{[t_f < T_\alpha]}^2] = \lim_{u \to \infty} E[X_{t_f}^2] \cdot P[t_f < T_\alpha]$$

$$= \lim_{u \to \infty} (\beta^2 V^2 u^{2(m-1)} + \frac{\sigma^2 V}{u}) \cdot (1 - Q_\alpha(t_f \mid 0 \,;\, u))$$

$$= \lim_{u \to \infty} \beta^2 V^2 u^{2(m-1)} \cdot (1 - Q_\alpha(t_f \mid 0 \,;\, u))$$

$$= \lim_{u \to \infty} \beta^2 V^2 u^{2(m-1)}$$

$$\times \left(\Phi\left[\frac{-\alpha}{\sigma} \sqrt{\frac{u}{V}} \left(\frac{V \beta u^{m-1}}{\alpha} - 1\right)\right] - \exp\left\{\frac{2\alpha\beta u^m}{\sigma^2}\right\} \Phi\left[\frac{-\alpha}{\sigma} \sqrt{\frac{u}{V}} \left(\frac{V \beta u^{m-1}}{\alpha} + 1\right)\right] \right)$$

$$\le \lim_{u \to \infty} \beta^2 V^2 u^{2(m-1)} \cdot \Phi\left[\frac{-\alpha}{\sigma} \sqrt{\frac{u}{V}} \left(\frac{V \beta u^{m-1}}{\alpha} - 1\right)\right] \tag{C.29}$$

Now for $x > 0$ ([Ab1]):

$$\Phi[-x] \leq \frac{1}{\sqrt{2\pi}} \frac{\exp\left\{\frac{-x^2}{2}\right\}}{x} \tag{C.30}$$

Using (C.30) in (C.29) it can be shown that

$$\lim_{u \to \infty} \beta^2 V^2 u^{2(m-1)} \cdot \Phi\left[\frac{-\alpha}{\sigma}\sqrt{\frac{u}{V}}\left(\frac{V\beta u^{m-1}}{\alpha} - 1\right)\right] = 0 \tag{C.31}$$

The result then follows.

Using these properties of the functions s_α, r_α, and z_α, the limits of the one step cost functional can be computed. The limits are stated in Theorem 5.7.2.1.

$$\lim_{u \to \infty} J_1(u,0) = \lim_{u \to \infty}\Big[(B + R)\cdot s_\alpha(u \mid 0) + D(\alpha s_\alpha(u \mid 0) + r_\alpha(u \mid 0))$$

$$+ (K \cdot \frac{V}{u} - G)\cdot(1-s_\alpha(u \mid 0)) + K \cdot z_\alpha(u \mid 0)\Big]$$

$$= B + R + D\alpha \tag{C.32}$$

$$\lim_{u \to 0^+} J_1(u,0) = \lim_{u \to 0^+}\Big[(B + R)\cdot s_\alpha(u \mid 0) + D(\alpha s_\alpha(u \mid 0) + r_\alpha(u \mid 0))$$

$$+ (K \cdot \frac{V}{u} - G)\cdot(1-s_\alpha(u \mid 0)) + K \cdot z_\alpha(u \mid 0)\Big]$$

$$\geq \lim_{u \to 0^+} K \cdot z_\alpha(u \mid 0) = \infty \tag{C.33}$$

The results for J_2 follow similarly giving equations (5.7.2.19a) - (5.7.2.19d).

REFERENCES

REFERENCES

[Ab1] Abramowitz, M. and Stegun, I.A., *Handbook of Mathematical Functions*, Dover, N.Y., 1965, 9th printing

[An1] Anderson, R.F. and Friedman, A., "Optimal Inspections in a Stochastic Control Problem with Costly Observations", Math. of Op. Res., Vol. 2, No. 2, May 1977, pp. 155-190

[An2] Anderson, R.F. and Friedman, A., "Optimal Inspections in a Stochastic Control Problem with Costly Observations, II", Math. of Op. Res., Vol. 3, No. 1, Feb. 1978, pp. 67-81

[Be1] Bensoussan, A. and Lions, J.-L., *Applications of Variational Inequalities in Stochastic Control*, North-Holland, 1982 (trans. from 1978 French ed.)

[Be2] Bensoussan, A. and Lions, J.-L., *Impulse Control and Quasi-Variational Inequalities*, Bordas, Paris, 1984 (trans.)

[Bl1] Blake, I.E. and Lindsey, W.C., "Level-Crossing Problems for Random Processes", IEEE Trans. Info. Theory, Vol. 19, May 1973, pp. 295-315

[Bre1] Breiman, L., "Exit Times from a Square Root Boundary", in 5th Berkeley Symp. on Mathematical Statistics and Probability, Vol. 2, Pt. 2, 1966, pp. 9-16

[Br1] Bremaud, P., *Point Processes and Queues*, Springer-Verlag, New York, 1981

[Bu1] Buzacott, J.A. and Shanthikumar, J.G., "Models for Understanding Flexible Manufacturing Systems", Trans. AIIE, Vol. 12, No. 4, Dec. 1980, pp. 339-350

[Ch1] Chhikara, R.S. and Folks, J.L., "The Inverse Gaussian Distribution as a Lifetime Model", Technometrics, Vol. 19, No. 4, Nov. 1977, pp. 461-468

[Cr1] Crabill, T.B., Gross, D., and Magazine, M.J., "A Classified Bibliography of Research on Optimal Design and Control of Queues", Operations Research, Vol. 25, No. 2, March-April 1977, pp. 219-232

[Da1] Danai, K. and Ulsoy, A.G., "A Dynamic State Model for Tool Wear Estimation in Turning", Sensors and Controls for Manufacturing, ASME Winter Conf., Nov. 1985, pp. 137-148

[De1] DeVor, R.E., Anderson, D.L., and Zdeblick, W.J., "Tool Life Variation and Its Influence on the Development of Tool Life Models", ASME Jour. Eng. Ind., Vol. 99, Aug. 1977, pp. 578-584

[Dr1] Drozda, T.J. and Wick, C. (eds.), *Tool and Manufacturing Engineers Handbook, Vol. I*, Society of Manufacturing Engineers, Dearborn, Mi., 1983

[Du1] Dubois, D. and Stecke, K., "Using Petri Nets to Represent Production Processes", Proc. 22nd CDC, Dec. 1983

[El1] Elliot, R.J., *Stochastic Calculus and Applications*, Springer-Verlag, N.Y., 1982

[Er1] Ermer, D.S., "A Bayesian Model of Machining Economics for Optimization by Adaptive Control", ASME Jour. Eng. Ind. Vol. 92, Aug. 1970, pp. 628-632

[Fe1] Feller, W., *An Introduction to Probability Theory and Its Applications, Vol. II*, Wiley, New York, 1971

[Fl1] Fleming, W.H. and Rishel, R.W., *Deterministic and Stochastic Optimal Control*, Springer-Verlag, New York, 1975

[Ge1] Gershwin, S., Hildebrant, R., Suri, R., and Mitter, S., "A Control Theorist's Perspective on Recent Trends in Manufacturing Systems", Proc. 23rd CDC, Dec. 1984

[Ge2] Gershwin, S., Akella, R., and Choong, Y., "Short Term Production Scheduling of an Automated Manufacturing Facility", Proc. 23rd CDC, Dec. 1984

[Ge3] Gershwin, S., from Tutorial Notes: "Manufacturing Systems Modeling and Control", IEEE Conf. on Robotics and Automation, April, 1986

[Ha1] Hati, S.K. and Rao, S.S., "Determination of Optimum Machining Conditions - Deterministic and Probabilistic Approaches", ASME Jour. Eng. Ind. Vol. 98, Feb. 1976, pp. 354-359

[Hi1] Hitomi, K., "Optimization of Multistage Machining System: Analysis of Optimal Machining Conditions for the Flow-Type Machining System", ASME Jour. Eng. Ind. Vol. 93, May 1971, pp. 498-506

[Hi2] Hitomi, K., Nakamura, N., and Inoue, S., "Reliability Analysis of Cutting Tools", ASME Jour. Eng. Ind. Vol. 101, May 1979, pp. 185-190

[Ho1] Ho, Y.C. and Cassandras, C., "A New Approach to the Analysis of Discrete Event Dynamic Systems", Automatica, Vol. 19, No. 2, 1983

[Ho2] Ho, Y.C., Suri, R., Cao, X.R., Diehl, G.W., Dille, J.W., and Zazanis, M., "Optimization of Large Multi-Class (Non-Product Form) Queueing Networks Using Perturbation Analysis", Large Scale Systems Journal, Vol. 7, No. 2/3, Dec. 1984, pp. 165-180

[It1] Ito, K. and McKean, H.P., *Diffusion Processes and Their Sample Paths*, Springer-Verlag, N.Y., 1965

[Iw1] Iwata, K., Murotsu, Y., Iwatsubo, T., Fujii, S., "A Probabilistic Approach to the Determination of the Optimum Cutting Conditions", ASME Jour. Eng. Ind. Vol. 94, Nov. 1972, pp. 1099-1107

[Kam1]

Kamath, M. and Viswanadham, N., "Applications of Petri Net Models in the Modelling and Analysis of Flexible Manufacturing Systems", Proc. IEEE Conf. on Robotics and Automation, April, 1986

[Kan1]

Kannatey-Asibu, Jr., E., "The Metal Cutting Optimal Control Problem - A State Space Formulation", *Control of Manufacturing Processes and Robotic Systems*, Hardt, D.E. and Book, W.J. (eds.), ASME, New York, 1983

[Kar1] Karlin, S. and Taylor, H., *A First Course in Stochastic Processes*, 2nd ed., Academic Press, New York, 1975

[Kar2] Karlin, S. and Taylor, H., *A Second Course in Stochastic Processes*, Academic Press, New York, 1981

[Ke1] Kendall, L.A. and Sheikh, A.K., "Coefficient of Variation: An Index of Tool Reliability and Its Significance in Automated Production", SME Proc. NAMRC, 1979, pp. 362-369

[Ki1] Kimemia, J. and Gershwin, S.B., "An Algorithm for the Computer Control of a Flexible Manufacturing System", IIE Trans., Vol. 15, No. 4, Dec. 1983, pp. 353-362

[Kl1] Kleinrock, L., *Queueing Systems Vol. I*, Wiley, New York, 1975

[Ko1] Koren, Y., "Flank Wear Model of Cutting Tools Using Control Theory", ASME Jour. Eng. Ind. Vol. 100, No. 1, Feb. 1978, pp. 103-109

[Krt1] Krtolica, R., "A Singular Perturbation Model of Reliability in Systems Control", Automatica, Vol. 20, No. 1, p. 51, 1984

[Kry1] Krylov, N.V., *Controlled Diffusion Processes*, Springer-Verlag, New York 1980

[La1] Law, A.M. and Kelton, W.D., *Simulation Modeling and Analysis*, McGraw-Hill, N.Y., 1982

[Led1] Leadbetter, M.R., "On Crossings of Levels and Curves by a Wide Class of Stochastic Processes", Ann. Math. Stat., Vol. 37, 1966, pp. 260-267

[Le1] Levi, R. and Rossetto, S., "Machining Economics and Tool Life Variation, Part 1", ASME Jour. Eng. Ind. Vol. 100, Nov. 1978, pp. 393-396

[Le2] Levi, R. and Rossetto, S., "Machining Economics and Tool Life Variation, Part 2", ASME Jour. Eng. Ind. Vol. 100, Nov. 1978, pp. 397-402

[Lo1] Loève, M., *Probability Theory I, II*, 4th ed., Springer-Verlag, N.Y., 1978

[Mac1]

Machining Data Handbook, Second Ed., compiled by Machinability Data Center, Metcut Research Assoc., Inc., Cincinnati, Ohio, 1972

[Mat1]

Matkowsky, B.J. and Schuss, Z., "The Exit Time Problem for Randomly Perturbed Dynamical Systems", SIAM J. of Appl. Math., Vol. 33, 1977, pp. 365-382

[Me1] Mehr, C.B. and McFadden, J.A., "Certain Properties of Gaussian Processes and their First-Passage Times", J. Royal Stat. Soc. (B), No. 3, 1965, pp. 505-522

[Mu1] Muth, E.J., "Memory as a Property of Probability Distributions", IEEE Trans. Rel., Vol. 29, No. 2, June 1980, pp. 160-164

[Pe1] Pegden, C.D., *Introduction to SIMAN*, Systems Modeling Corp., State College, PA, 1982

[Ph1] Philipson, R.H. and Ravindran, A., "Application of Goal Programming to Machinability Data Optimization", ASME Jour. Eng. Ind. Vol. 100, April 1978, pp. 286-291

[Rob1]

Robbins, H. and Siegmund, D., "Boundary Crossing Probabilities for the Wiener Process and Sample Sums", Ann. Math. Stat., Vol. 41, 1970, pp. 1410-1429

[Ro1] Ross, S., *Stochastic Processes*, Wiley, New York, 1983

[Sa1] Savage, R.I., "Surveillance Problems", Naval Research Logistics Quarterly, Vol. 9, 1962, pp. 187-209

[Shk1] Sheikh, A.K., Kendall, L.A., and Pandit, S.M., "Probabilistic Optimization of Multitool Machining Operations", ASME Jour. Eng. Ind. Vol. 102, Aug. 1980, pp. 239-246

[Shr1] Sherif, Y.S., and Smith, M.L., "First Passage Time Distribution of Brownian Motion as a Reliability Model" IEEE Trans. Rel., Vol. 29, No. 5, Dec. 1980, pp. 425-426

[Si1] Siegert, A., "On the First Passage Time Probability Problem", Phys. Review, Vol. 81, Feb. 1951, pp. 617-623

[So1] Solberg, J., "Analytical Performance Evaluation for the Design of Flexible Manufacturing Systems", Proc. CDC, 1979

[Su1] Suri, R., "Robustness of Queueing Network Formulae", J. ACM, Vol. 30, No. 3, July, 1983, pp. 564-594

[Su2] Suri, R., "An Overview of Evaluative Models for Flexible Manufacturing Systems", Proc. First ORSA/TIMS Conf. on FMS, Aug., 1984

[Tak1]

Takahashi, N. and Kahng, C.H., "Stochastic Optimization of Production Speed", SME Proc. NAMRC, 1983, pp. 461-465

[Tay1]

Taylor, F.W., "On the Art of Cutting Metals", Trans. ASME, Vol. 28, 1906, pp. 31-350

[Tw1] Tweedie, M.C.K., "Statistical Properties of Inverse Gaussian Distributions I", Ann. Math. Stat., Vol. 28, 1957, pp. 362-377

[Wa1] Wager, J.G. and Barash, M.M., "Study of the Distribution of the Life of HSS Tools", ASME Jour. Eng. Ind. Vol. 93, Nov. 1971, pp. 1044-1050

[Wh1] Whitney, C., "Control Problems in Flexible Manufacturing", Proc. 1984 ACC, San Diego, 1984

[Wo1] Wong, E. and Hajek, B., *Stochastic Processes in Engineering Systems*, Springer-Verlag, New York, 1985