AN ADAPTIVE CONTROL TECHNIQUE FOR SYSTEMS WITH LIGHTLY DAMPED RESONANCES

Thesis by

Elliot Neil Pinson

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1961

ACKNOWLEDGMENT

I wish to express my gratitude to Professor F. J. Mullin for his encouragement and suggestions during the preparation of this thesis. I am also grateful to Professor C. H. Wilts for his careful reading of several chapters of the manuscript and his helpful criticism thereof.

I am indebted to Space Technology Laboratories for the generous financial support received under the terms of an S. T. L. Fellowship. In addition, personal contact and discussions with several members of the Technical Staff at S. T. L. were greatly appreciated. In this respect I should like in particular to express my thanks to Dr. R. K. Whitford, Mr. N. W. Trembath and Mr. E. P. Blackburn.

Thanks are also extended to the art department of the Aerospace Corporation where the figures were prepared; to Mrs. Joe Mullins, who typed the final text; and to my wife, Stefi, who was encouraging and helpful throughout.

ABSTRACT

An Adaptive Controller capable of stabilizing dynamic systems containing multiple lightly damped resonances is synthesized. The controller acts to stabilize the dynamic system by introducing cascade compensation which has zeros of transmission very close to the critical resonant frequencies. Very little a priori knowledge is needed about the frequencies at which the resonances occur because the Adaptive Controller itself measures these frequencies while the system is operating. It then adjusts its internal parameters on the basis of these measurements to insure that the overall system performance is satisfactory. Since the measurement process can be performed continually, this adaptive control technique is applicable to systems whose resonant frequencies change slowly with time.

Both the measurement and compensation functions are performed by a digital computer. The resonant frequencies are measured by cross-correlating a signal generated by the dynamic system with a set of periodic signals whose frequencies span the frequency intervals in which the resonances are known to occur. The necessary compensation is instrumented in a set of difference equations stored in the digital computer. Certain coefficients which appear in these difference equations are adjusted according to logic programmed into the computer.

Necessary and sufficient conditions are derived to describe the conditions under which the proposed system can be successful. The fact that the system can perform successfully is demonstrated by a detailed digital simulation of an adaptive autopilot for a highly flexible ballistic missile.

TABLE OF CONTENTS

AN ADAPTIVE CONTROL TECHNIQUE FOR SYSTEMS
WITH LIGHTLY DAMPED RESONANCES

CHAPTER	SECTION	TITLE	PAGE
I		ADAPTIVE CONTROL SYSTEMS	1
	1.1	Introduction	1
	1.2	Previous Adaptive Systems	8
	1.3	Scope of the Present Investigation	18
II		THE FLEXIBLE MISSILE CONTROL PROBLEM	23
	2.1	Introduction	23
	2.2	Essential Characteristics	24
	2.3	The Present Design Approach	45
III		LIGHTLY DAMPED RESONANCES AND ADAPTIVE CONTROL	50
	3.1	Introduction	50
	3.2	Detection and Frequency Measurement	57
	3.3	Digital Compensation	99
	3.4	Summary	119
IV		SIMULATION OF THE SYSTEM	122
	4.1	Introduction	122
	4.2	Results of the Simulation	124
	4.3	Description of the Simulation	138
V		FINIS	
	5.1	Summary and Conclusions	147
	5.2	Suggestions for Further Study	149
Appendix		Approximate Equations of Motion for a Flexible Ballistic Missile	151

CHAPTER	SECTION	TITLE	PAGE
Partial List of Symbols			156
References			163

CHAPTER I

ADAPTIVE CONTROL SYSTEMS

1.1 Introduction

In the past generation the field of Feedback Control System Engineering has developed to the point where, for large classes of systems, it is much more a science than an art. A decade ago the principle tools used in system synthesis were a large scale analogue computer and extensive simulation studies. Today the important techniques available for linear system synthesis associated with such names as Nyquist, Bode, Nichols, Evans and Wiener are familiar to most graduating engineers who have taken a course in servomechanism analysis (1)(2). The field of linear sampled-data systems has been thoroughly developed in works of Linvill, Ragazzini, Jury and others (3)(4). Significant strides have been taken, too, toward the development of techniques for handling simple non-linear systems via describing function and/or phase-plane techniques. Very recently the rediscovered works of the Russian mathematician Liaponov have caused excitement for, in the eyes of some, it appears that through Liaponov methods of stability analysis it may be possible to develop a general approach to the problem of non-linear system design (5)(6).

Despite the happy state of affairs which a reading of the preceding paragraph seems to imply, there is much room for improving and adding to the contents of the control designer's little bag of analytical techniques. The additions would be most welcome in regard to the treatment of non-linear systems, where what is available lacks the unity and generality evident in the treatment of constant-coefficient

linear systems. A good deal of effort is being directed toward the solution or amelioration of precisely this problem.

However, another area of investigation has captured the imaginations of many engineers interested in the field of automatic controls. The class of systems included in this area has been given the rather glamorous title of "Adaptive Control Systems." Similar systems have also been referred to as "self-adapting," "self-optimalizing," "self-adjusting," and in a few other ways as well. The glamour in the word "adaptive" lies in part in the fact that adaptability is an attribute which is traditionally associated with living organisms. It would be exciting indeed to develop a mechanism which exhibited an appreciable capacity for organized learning and the ability to alter its characteristics or function depending upon sensory information received from its environment. The ability of a system to evaluate its own performance and to take action to improve this performance is a concept which is close to the foundation of adaptive control principles.

A good deal of time and effort and a great number of words have been consumed in an effort to arrive at a satisfactory definition of the term "adaptive control system." Three difficulties seem to block attempts at reaching this goal.

First, words such as "adaptive" and "adaptation" have been part of the English vocabulary for years, long before they were introduced into the technical vocabulary of control engineering. Thus a bias exists regarding what connotation the word should have when applied to a particular mechanism. This problem is compounded because, as is true with many familiar words, although most of us use the word more or

less correctly, we do not know its precise definition and thus the same word means slightly different things to different people.

Second, the word "adaptation" already possesses two relatively distinct implications depending upon the field in which it is being used. The following definition from Webster's Unabridged New International Dictionary (Second Edition) demonstrates this point:

> Adaptation: Adjustment to environmental conditions, specifically:
>
> a) Physiology and Psychology, adjustment of a sense organ, as the eye or receptors in the skin, to the intensity or quality of stimulation, as of light, temperature or pressure prevailing at the moment, by changes in sensitivity.
>
> b) Biology, modification of an animal or plant (or of its parts or organs) fitting it more perfectly for existence under the conditions of its environment; applied especially to a process of evolutionary change in structure and function, in organisms of a group or race.

Evidently, what is required of the control engineer's lexicographer is the addition of a subheading "c" to the above definition.

Third, and perhaps most important, has been a desire to form a definition which will exclude from the select ranks of "adaptive control systems" all control systems which were designed before this new word was added to the technical vocabulary. If the "adaptive" concept is something new and exciting it must surely follow, the argument goes, that "conventional" control systems do not exhibit the "adaptive" feature. Unfortunately, this "exclusion principle" is not easy to implement in a definition, nor is it necessarily a desirable objective. This writer is not going to add to the growing list of definitions of adaptive systems. The significant point is that the success of any

particular control system, "adaptive" or "conventional," depends on whether or not it performs the function for which it was designed. Whether or not it conforms to the requirements of some arbitrary definition of a particular class of systems is irrelevant.

A survey of the adaptive control field up to the beginning of 1959 is available in the literature (7). A broad classification of adaptive systems is stated in this article. The classes selected are somewhat arbitrary and there is considerable overlapping between classes. Essentially all feedback control or regulatory systems fall in at least one category. The article serves as an excellent introduction to the field of adaptive controls and it is worthwhile to repeat here some of its major points. The five classes of adaptive systems suggested are:

1. Passive Adaptation
2. Input Signal Adaptation
3. Extremum Adaptation
4. System-variable Adaptation
5. System-characteristic Adaptation

Systems which display passive adaptation (Class 1) do not change their internal parameters in response to variations in environment or performance. Rather, they have been designed to perform satisfactorily over a wide range of environmental variations. Feedback itself illustrates the principle. The reduction of effects of element changes through the use of feedback is well known.

The next order of complexity is displayed by systems which are input-signal adaptive (Class 2). In this form of system, the "measurement function" makes its first appearance. Measurements are made of the environment in which the system is operating (i.e., the

inputs to the system form the environment) and internal system parameters are altered to improve the performance of the over-all system based on the results of these measurements. This measurement function is frequently performed by a portion of the control system which is distinct from the main loop.

In extremum-adaptive systems (Class 3), adjustments are made to minimize or maximize some system variable. Actually these systems form a sub-class of classification 4 (system variable adaptation) but the authors of the referenced article felt this type of system to be important enough to warrant a class of its own.

In system-variable adaptive systems (Class 4), internal parameters may be changed on the basis of measurements made of variables occurring within the system. The logical relationship between the measured quantities and the selected parameter variations is frequently, but not necessarily, incorporated in a separate portion of the control loop which enforces the adaptive action.

In system-characteristic adaptive systems (Class 5), a measure of the actual system performance is obtained based on a comparison made between the actual and the desired system response. Thus such indirectly available characteristics as damping ratio or peak overshoot may be used as performance criteria. A good deal of overlap occurs between Classes 4 and 5. The logical computing function for systems of Class 5 is ordinarily somewhat more complex than that employed in Class 4 systems.

For a bibliography of adaptive systems which have been classified according to the groupings discussed above the reader is referred once again to Reference (7).

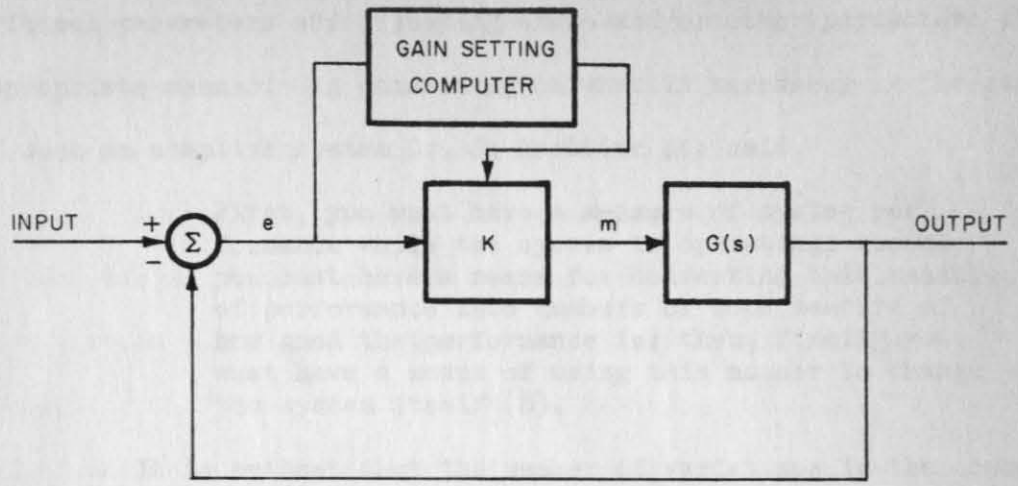
One of the difficulties apparent in this and any attempt to classify adaptive systems which is not mathematically precise is the fact that a particular system, when viewed one way, will fall under one heading while the same system, perhaps with its block diagram drawn slightly differently, will appear to fall in a different category. For example, the reader may look at Figure 1.1a. Depending upon the criterion used by the gain setting computer this system appears to fall in either Class 4 or Class 5. A closer look at what the gain setting computer does might reveal the true situation to be what is shown in Figure 1.1b, which is nothing more than a saturation non-linearity. This would be considered to fall into Class 1 by someone following the suggestions above. Many people, however, would like to withhold the magic term "adaptive" and not use it at all in connection with a simple non-linearity of this type. It is easy to see that much time can be spent trying to resolve such semantic problems, but it is hardly time well spent.

A remark made by Dr. J. G. Truxal seems appropriate.

> An adaptive feedback system [is] ... one which is designed with an adaptive viewpoint. This sounds superficial when you first hear it but there is really considerable merit because nobody has any idea how to design a system with an intentional non-linearity introduced into the system to obtain desirable results. By this adaptive viewpoint one obtains a logical, simple, and straightforward technique towards the inclusion of a non-linear element within the system to obtain some reasonable performance specifications or meet some reasonable optimization criteria (8).

The adaptive viewpoint seems to be particularly useful when applied to systems whose characteristics and parameters change slowly

(a)



(b)

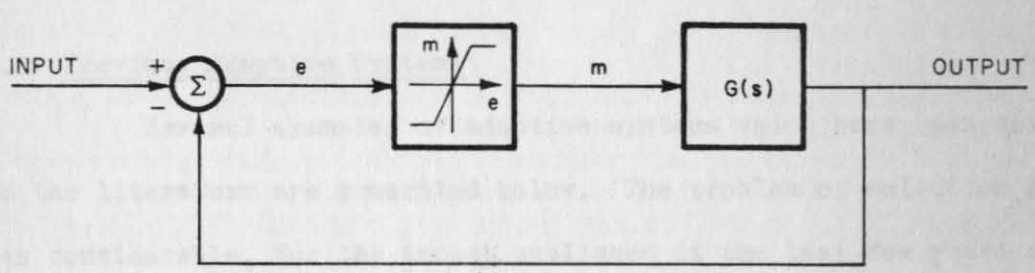


Figure 1.1: Is this system adaptive?

compared with the response time of the system. Under these conditions it may be possible to improve system performance by monitoring certain critical parameters and adjusting these and/or other parameters in an appropriate manner. In commenting on what is necessary in the design of such an adaptive system Dr. J. Aseltine has said

> First, you must have a measure of system performance while the system is operating; second, you must have a means for converting this measure of performance into numbers or some measure of how good the performance is; then, finally, you must have a means of using this number to change the system itself (8).

It is evident that the number of variations in the concept of adaptive control systems is approximately equal to the number of people who have thought about the problem. In accordance with the belief that a single example is worth a thousand words, the next section is devoted to a discussion of several adaptive systems which have been discussed in the literature. Following that, this chapter concludes with an outline of the problem and a description of those adaptive techniques which are the principle subjects of this paper.

1.2 Previous Adaptive Systems

Several examples of adaptive systems which have been described in the literature are presented below. The problem of selection here was considerable, for the amount published in the last few years on this subject would, if collected, fill many large volumes. The examples selected illustrate the feature the writer feels to be most important in an adaptive system. Specifically, some degree of uncertainty must exist regarding the structure of the system being controlled or regarding the values of various parameters describing the system and its

environment. Measurements are therefore made while the system is operating to reduce this uncertainty to the point where it is possible to design a control system which will perform in a satisfactory manner. Furthermore, this design is accomplished automatically during system operation by the adaptive control system itself.

Interest in adaptive systems antedates the current activity in the area of control systems by many years. It is not too surprising that workers in the life sciences (neurology, physiology, biology, etc.) recognized the importance of adaptive behavior long ago. After all, one of the fundamental characteristics of living organisms, at least those which are likely to survive, is precisely their adaptability. Ashby has devoted a book (9) to the problem of "the origin of the nervous system's unique ability to produce adaptive behavior." A form of behavior is adaptive, he asserts, "if it maintains the essential variables within physiological limits." Furthermore, the essential characteristics of adaptive behavior in a living organism are achieved by a trial and error process. "The basic rule for adaptation by trial and error is: If the trial is unsuccessful, change the way of behaving; when and only when it is successful, retain the way of behaving." A system which behaves in this way Ashby has termed an "ultra-stable" system, i.e., it does not have to be designed to be stable but will automatically seek a stable state.

As an example of a mechanism which is adaptive and ultra-stable Ashby suggested and built a relatively simple system which he called the "homeostat." The device had four principle variables, $X_1 \ldots X_4$, which were the angular deflections of four heavily damped

magnet coils. Each coil was driven by a current equal to a linear combination of the deflections of each of the four coils. Because of the heavy damping, the equations of motion of the system are obtained by equating the turning torques (proportional to the coil currents) on each coil to the damping torque (proportional to the angular rate of the coil) on each coil. The resulting system of equations is

$$\frac{dX_i}{dt} = \sum_{j=1}^{4} a_{ij}X_j \qquad (i = 1,2,3,4) \qquad (1.1)$$

It is evident that the stability of the over-all system depends upon the values of the sixteen coefficients a_{ij}. In the homeostat, twelve of these coefficients were fixed while four coefficients (one for each magnet) were determined by the settings of four twenty-five position stepping-switches. The coefficients corresponding to each setting of a stepping-switch were preselected from a table of random numbers. The i^{th} stepping-switch will increase its position by one unit every T seconds (where T can be chosen anywhere between one and ten) if $X_i > 45^{\circ}$. Thus, if the system is unstable it changes its parameters in a random fashion until it finds a setting which keeps all the variables within the limits prescribed. The number of possible different systems, then, is $25^4 = 390625$.

The homeostat exhibits several undesirable properties which make it a rather inefficient adaptive system. First, because it has no memory it cannot learn from experience. Second, in cases where only a small fraction of the totality of possible states is stable it takes the homeostat excessively long to reach a stable equilibrium condition. It may be argued that with no a priori knowledge of how the

parameters which are being changed affect the system's performance, adaptation in a random fashion is as good as any other method. However, there are few, if any, engineering applications in which a priori information is not available about the dynamics of the system under consideration and about how certain control parameters affect the system. Indeed, a mathematical model of the system is usually specified, although the precise values of certain critical parameters may not be known. Since random adaptation does not seem particularly useful it follows that changes in control parameters must be made on the basis of measurements made of the state of the system.

The problem of measurement is very critical. What are the measurements necessary to obtain information in the most convenient way, and what quantities should be determined? Much thought has been given to the problem of determining process dynamic characteristics when very little is known a priori of what the system is (10)(11). Kalman has suggested a self-optimalizing control system which automatically measures the pulse transfer function of the dynamic process being controlled and, on the basis of these measurements, automatically selects the coefficients of a digital controller which result in optimum over-all system performance. Optimum in this case was defined to mean a controller which forced the error resulting from a step input to become zero in minimum time and remain zero for all time thereafter (dead beat control)(12).

Kalman's system configuration is shown in Figure 1.2. If the dynamic process is linear and time invariant it is well known that the sampled outputs $C^*(kT)$ can be related to the sampled inputs $M^*(kT)$

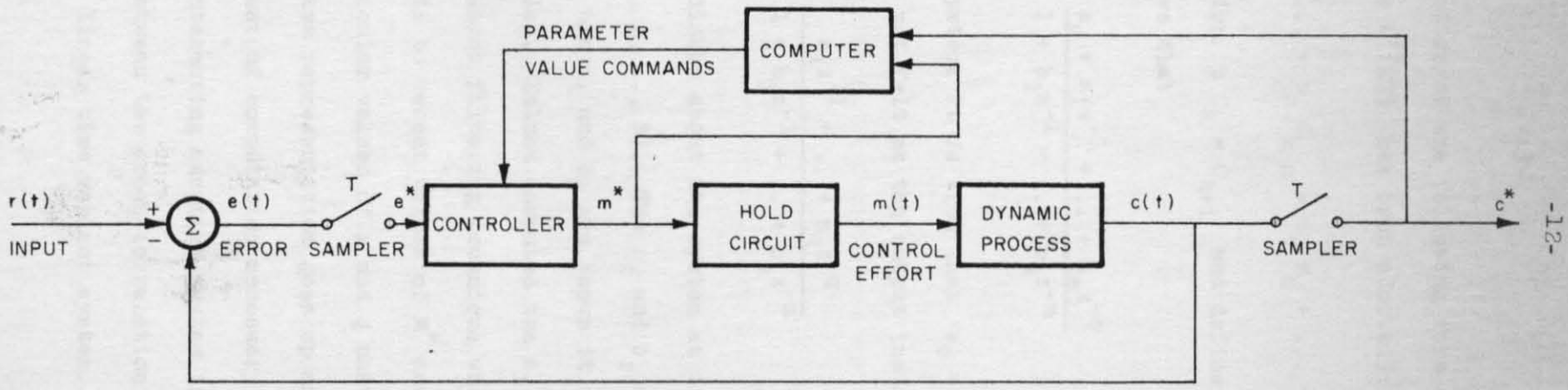


Figure 1.2: Kalman's self-optimalizing system configuration.

by a difference equation of the following form (note that the notation $M_k = M^*(kT)$ and $C_k = C^*(kT)$ has been adopted):

$$C_k + b_1 C_{k-1} + \ldots + b_n C_{k-n} = a_o M_k + \ldots + a_q M_{k-q} \qquad (1.2)$$

If we use the notation $z^i C_k = C_{k+i}$ and define $C(z) = C_k$ and $M(z) = M_k$ it follows that

$$\frac{C(z)}{M(z)} = \frac{a_o + a_1 z^{-1} + .. + a_q z^{-q}}{1 + b_1 z^{-1} + .. + b_n z^{-n}} \qquad (1.3)$$

For most physical systems it is true that $a_o = 0$ since the effect of an input is usually not felt at the output instantaneously. Therefore,

$$\frac{C(z)}{M(z)} = \frac{a_1 z^{-1} + .. + a_q z^{-q}}{1 + b_1 z^{-1} + .. + b_n z^{-n}} \qquad (1.4)$$

The information available about the system at time NT are the values of C_k and M_k ($k = 0, 1, .., N$). The a_i and b_i must be computed from these values. Once the a_i and b_i are known it is possible to design the optimum controller. Kalman computed the a_i and b_i on the basis of a weighted least squares filtering technique where the weighting assigned higher weights to recent values of M^* and C^* than to older measurements. Particular values of n and q must be selected in advance. The accuracy of system representation goes up as n and q are increased, but so does the amount of computation necessary.

Another interesting adaptive system is based on the relationship which exists between the cross-correlation function between the input and output of a linear time variant system. If a physical system

having an impulse response g(t) is excited by a stationary noise signal having an auto-correlation function $\phi_{ii}(\tau)$, then the cross-correlation, $\phi_{io}(\tau)$, between input and output is

$$\phi_{io}(\tau) = \int_{-\infty}^{\infty} g(x)\, \phi_{ii}(\tau - x)\, dx \qquad (1.5)$$

If the excitation noise has a bandwidth considerably larger than that of the system being tested, $\phi_{ii}(\tau)$ is effectively an impulse and, from Equation 1.5

$$\phi_{io}(\tau) = g(\tau) \qquad (1.6)$$

i.e., the cross-correlation between the system's input and output is identically equal (within the validity of the approximation) to the system's impulse response. Anderson, Aseltine, Mancini and Sarture have described a self-adjusting system based on this principle (13). Their system configuration is shown in Figure 1.3. White noise is introduced into the system along with the signal which is to be followed. The system output is cross-correlated with the white noise input to obtain the system's impulse response. The impulse response is converted to a figure of merit (a number) which is used to adjust system parameters. The technique was applied successfully to a second order system. The system automatically adjusted a single parameter in order to maintain a constant closed loop damping ratio.

The final example treated in this section is a system suggested by Staffin (14)(15). The method is applicable to systems whose open loop transfer function has a lightly damped dominant pole pair whose natural frequency is not known precisely. The configuration is

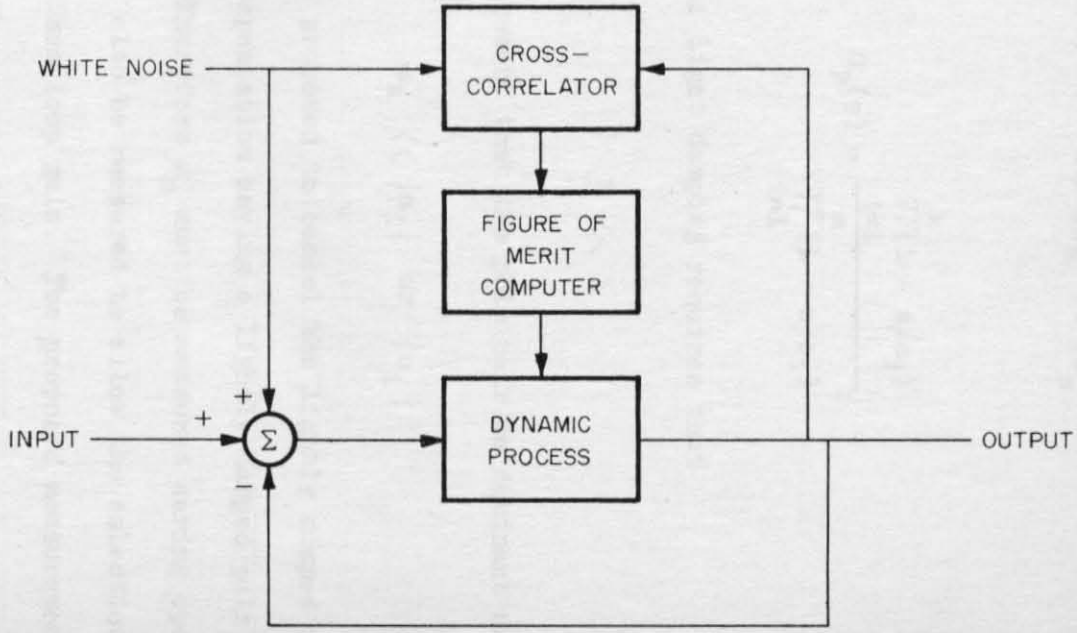


Figure 1.3: Self-adjusting system due to Anderson et al.

shown in Figure 1.4. The dynamic process is assumed to have a transfer function of the form

$$\frac{C(s)}{M(s)} = \frac{K\, G_p(s)}{1 + 2\zeta s/\omega_n + s^2/\omega_n^2} \tag{1.7}$$

where

$$G_p(s) = \frac{\prod_{i=1}^{k}(1 - s/a_i)}{\prod_{i=1}^{m}(1 - s/b_i)} \tag{1.8}$$

The condition of light damping requires that

$$\zeta << 1 \tag{1.9}$$

while the requirement that the pole pair be dominant means that

$$\omega_n << |a_i| \quad \text{or} \quad |b_i| \tag{1.10}$$

It is proposed to cancel the lightly damped pole pair by using tandem compensation having a lightly damped pair of zeros at frequency ω_n. Therefore ω_n must be measured during operation of the system. K will also be measured to allow the selection of a gain K_c to yield an optimum loop gain. The proposed measurement technique is quite simple. Choose two frequencies ω_1 and ω_2 such that

$$3\omega_1 < \omega_n < \omega_2 << |a_i| \text{ or } |b_i| \tag{1.11}$$

Then it follows that

$$|K| \cong \frac{|C(\omega_1)|}{|M(\omega_1)|} \tag{1.12}$$

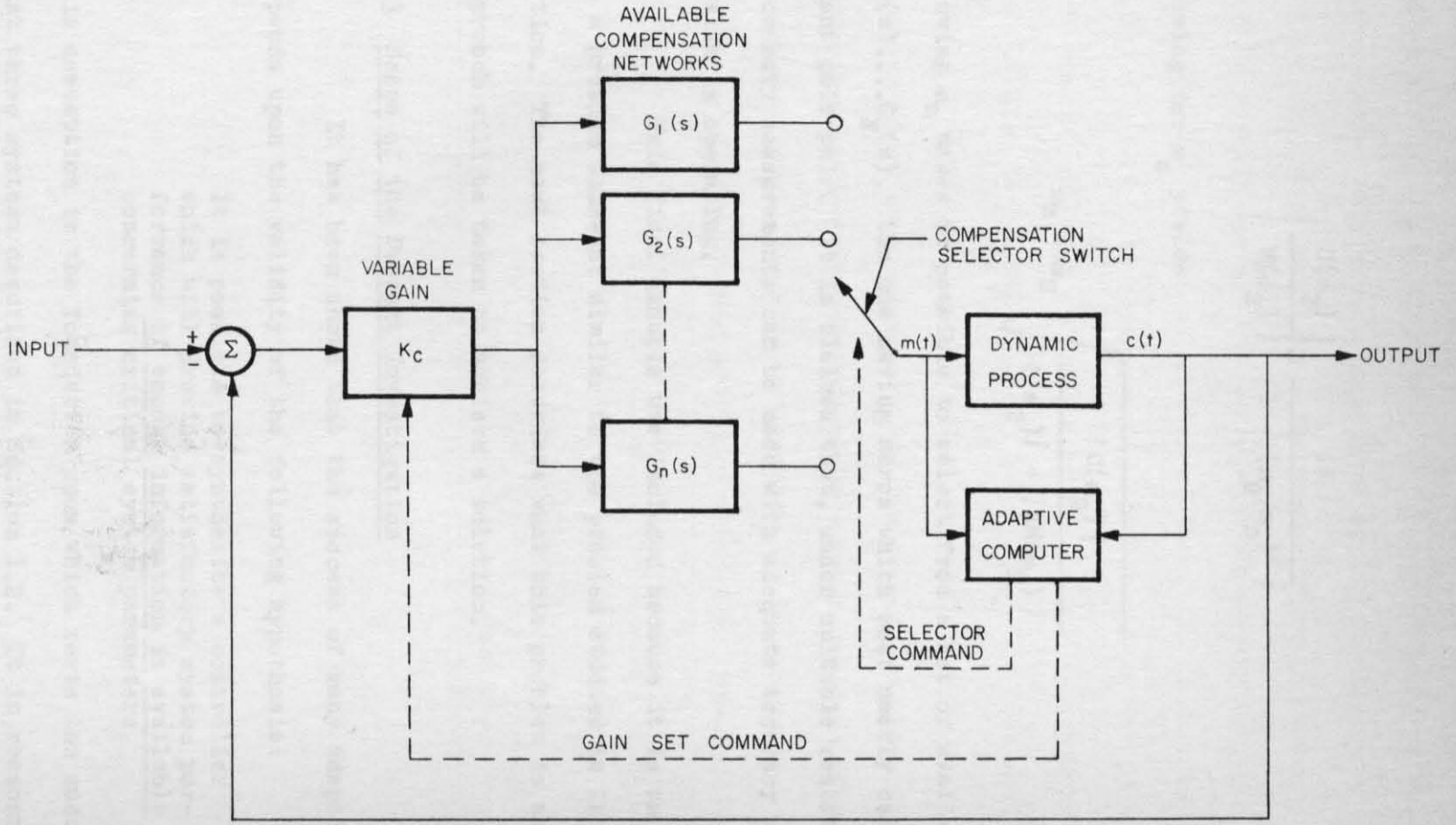


Figure 1.4: Staffin's adaptive system.

and

$$\left| \frac{C(\omega_2)}{M(\omega_2)} \right| \cong \frac{|K|}{|1 - (\omega_2/\omega_n)^2|} \tag{1.13}$$

Solving for ω_n yields

$$\omega_n \cong \omega_2 \sqrt{\frac{|C(\omega_2)|}{|C(\omega_2)| + |KM(\omega_2)|}} \tag{1.14}$$

Knowing ω_n makes it possible to select from a set of available filters, $G_1(s)....G_N(s)$, the one having zeros which most nearly cancel the dominant pole pair. It is claimed that, under suitable restrictions, the necessary measurements can be made with adequate accuracy while the system is operating.

This final example was included because it is an approach to a problem somewhat similar to the problem studied in this investigation. The next section describes what this problem is and what approach will be taken to achieve a solution.

1.3 Scope of the Present Investigation

It has been shown that the success of many adaptive techniques depends upon the validity of the following hypothesis:

> It is possible to synthesize a controller which will provide satisfactory system performance if enough information is available concerning critical system parameters.

This assumption is the foundation upon which rests the success of the last three systems described in Section 1.2. It is reasonable to say that if all critical system parameters are known and do not vary, adaptability in the sense used here (i.e., the ability to change

parameters if system performance is unsatisfactory) is an unnecessary and expensive attribute. Nothing is gained by incorporating a sophisticated measuring computer in the system since it can only tell us what is already known. If some system parameters vary but are well specified as functions of time or as functions of directly measurable quantities (e.g., dynamic pressure of Mach number for an aircraft control system) an adaptive system may or may not be preferable depending upon whether or not its over-all reliability and probability of successful operation exceed that of a non-adaptive system. In the case where critical parameters vary and furthermore are not predictable with adequate accuracy between members of an ensemble of systems in which they occur, the ability to adapt may be imperative.

It is a dynamic system of the latter type with which this investigation is principally concerned. The general adaptive control problem may be specified analytically as follows:

Given: A dynamic system specified by the following system of equations:

$$\frac{dX_i}{dt} = F_i(X_1 \ldots X_n, \alpha_1 \ldots \alpha_m, \beta_1 \ldots \beta_\ell, t) \qquad (1.15)$$

where

$X_i(t)$ = i^{th} state variable ($i = 1, \ldots, n$)

$\alpha_i(t)$ = i^{th} control parameter ($i = 1, \ldots, m$)

$\beta_i(t)$ = i^{th} critical system parameter ($i = 1, \ldots, \ell$)

t = time

The system is completely determined by this set of equations for, if all the α_i and β_i are known functions of time and initial conditions

are specified on the X_i, then the X_i are determined for all future times.

Required: To select the proper values for $\alpha_1 \ldots \alpha_m$ at all times to insure that the over-all system response satisfies some specified performance criterion.

It will be assumed that the adaptive system considered is a member of the class of systems that satisfy the hypothesis stated at the beginning of this section, i.e., if all the $\beta_1 \ldots \beta_\ell$ are known at a particular time it is possible to assign values to the $\alpha_1 \ldots \alpha_m$ which insure satisfactory system performance.

In the following chapters a special case of this type is treated in detail. The author's interest in this problem arose during the course of his association with Space Technology Laboratories where he expended considerable effort on the problem of designing autopilots for highly flexible missile systems. The characteristic feature of the flexible missile design problem is the presence of many extremely lightly damped resonances due to the structural flexibility of the vehicle. Actually, of course, a complete representation would require an infinite number of such modes. Fortunately, the frequencies associated with the higher modes can be attenuated by proper design and only a finite number of modes need be considered. This, then, is the origin of the system with multiple lightly damped resonances. A block diagram of the system described is shown in Figure 1.5. Although the problem arose in connection with ballistic missile autopilot design, it is evident that the configuration shown could arise in a number of other ways.

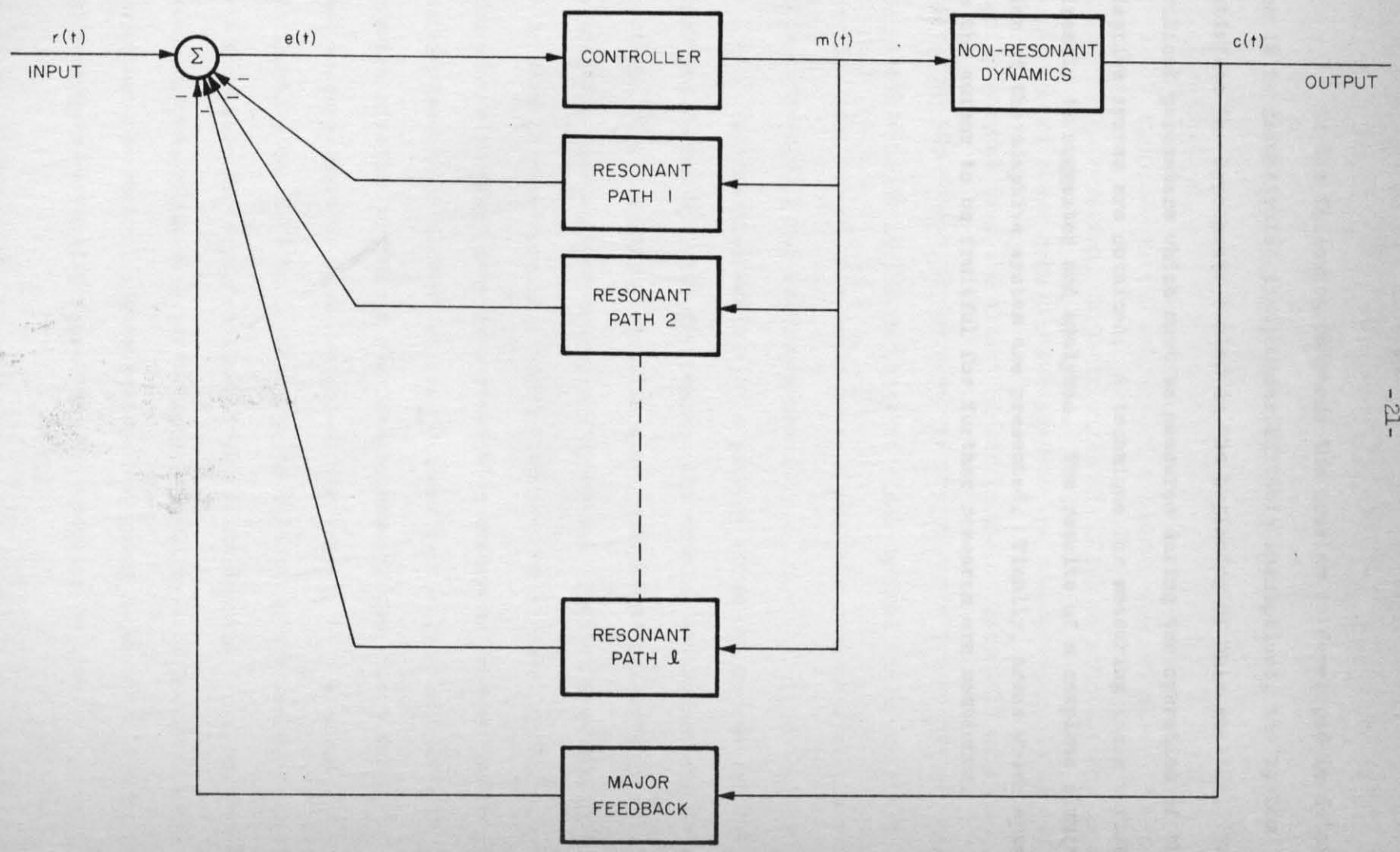


Figure 1.5: Multiply resonant feedback structure.

In the following chapters the problem is developed in detail and it is demonstrated that, under suitable assumptions, the system satisfies the hypothesis stated in the beginning of this section. The critical parameters which must be measured during the operation of the adaptive system are obtained. A technique for measuring these variables directly is suggested and analyzed. The results of a complete simulation of the adaptive system are presented. Finally, areas which appear to the author to be fruitful for further research are mentioned.

CHAPTER II

THE FLEXIBLE MISSILE CONTROL PROBLEM

2.1 Introduction

Textbooks on the subject of automatic control theory traditionally have stressed, in their examples, first or second order systems. Even in cases where higher order systems are touched upon, it is frequently assumed that a dominant closed loop pole pair will exist which determines the essential character of the system's transient response and stability. It is assumed that if close control can be maintained over the location of this dominant pole pair the remainder of the poles of the system will not be troublesome.

This chapter deals with a problem which has become extremely important during the last few years. The problem is that of designing an autopilot for a highly flexible space vehicle or ballistic missile. In addition to the great practical importance of this space age problem, it is also of considerable academic interest as an excellent illustration of a high order complex system which cannot be reduced to the convenient example of a first or second order servomechanism. High performance missiles are quite flexible because of the effort which is made to keep the structural weight of the vehicle to a minimum. Every additional pound used to stiffen the vehicle structure can mean a pound less available for a payload or several miles decrease in range for a fixed payload. Thus with performance capability at a premium it appears that the control system designer is going to have to live with low structural natural frequencies for some time to come.

The remainder of this chapter contains a detailed discussion of the dynamics pertinent to the design of a control system for a highly flexible missile. Design criteria are specified and the conventional approach to the problem is discussed. The limitations of this approach are pointed out. It is intended that this chapter will be the foundation on which an understanding of the adaptive approach described in later chapters can be based.

2.2 Essential Characteristics

Approximate dynamic equations describing the perturbed motion of a ballistic missile about its nominal trajectory are presented in the Appendix. Although these equations are formidable as they stand, they do not do justice to the complexity of the problem and neglect several effects of considerable importance. The model used in the derivation does not, for example, represent the effects of propellent sloshing in liquid fueled vehicles. Nor does it represent the sizable inertial reaction forces which act on the body of the vehicle due to the swivelling of the massive rocket engine which generates the vehicle's thrust. These effects were neglected in order to enable a detailed simulation of the system to be made on an Electrodata-220 digital computer. Had a much more detailed model been used, the solution time would have been prohibitive. A complete list of the approximations and assumptions implicit in the form of the equations is given in the Appendix. In spite of these approximations, however, the system does exhibit the feature of principle interest in this investigation, i.e., the presence of multiple lightly damped resonances.

The lightly damped resonances arise directly from the structural characteristics of the vehicle which, in flight, looks like a beam with a transverse distributed forcing function acting on it. The transverse forces are due to aerodynamic effects, the component of the rocket engine's thrust which acts normal to the vehicle, and inertial reaction forces acting at the point where the engine is attached to the main airframe. The boundary conditions at the ends of the beam are the usual "free-free" end conditions, i.e., there can be no internal shear and no internal moment at either end of the beam.

Figure 2.1 shows the geometry of the situation being described. Below is a list defining the symbols appearing in the figure.

$\vec{e}_X$, $\vec{e}_Y$ unit vectors defining an orthogonal set of inertial reference axes. $\vec{e}_Y$ is oriented in the direction of the local vertical. $\vec{e}_X$ is directed "downrange." $\vec{e}_X$ and $\vec{e}_Y$ define the pitch plane of the missile's trajectory.

$\vec{e}_x$, $\vec{e}_y$ unit vectors defining a set of orthogonal coordinate axes whose origin is fixed at the center of mass of the missile. $\vec{e}_x$ lies along the undeformed elastic axis and $\vec{e}_y$ is normal to this direction.

c.m. the location of the center of mass

c.p. the location of the center of pressure

ξ the distance from the c.m. of the vehicle to any point on the rigid body axis. ξ is positive for points forward of the c.m. and negative for points aft of the c.m.

M the total mass of the vehicle

I the total inertia of the vehicle about an axis passing through the c.m. normal to the pitch plane

$\vec{V}$ the instantaneous velocity of the c.m.

$\vec{g}$ gravity vector

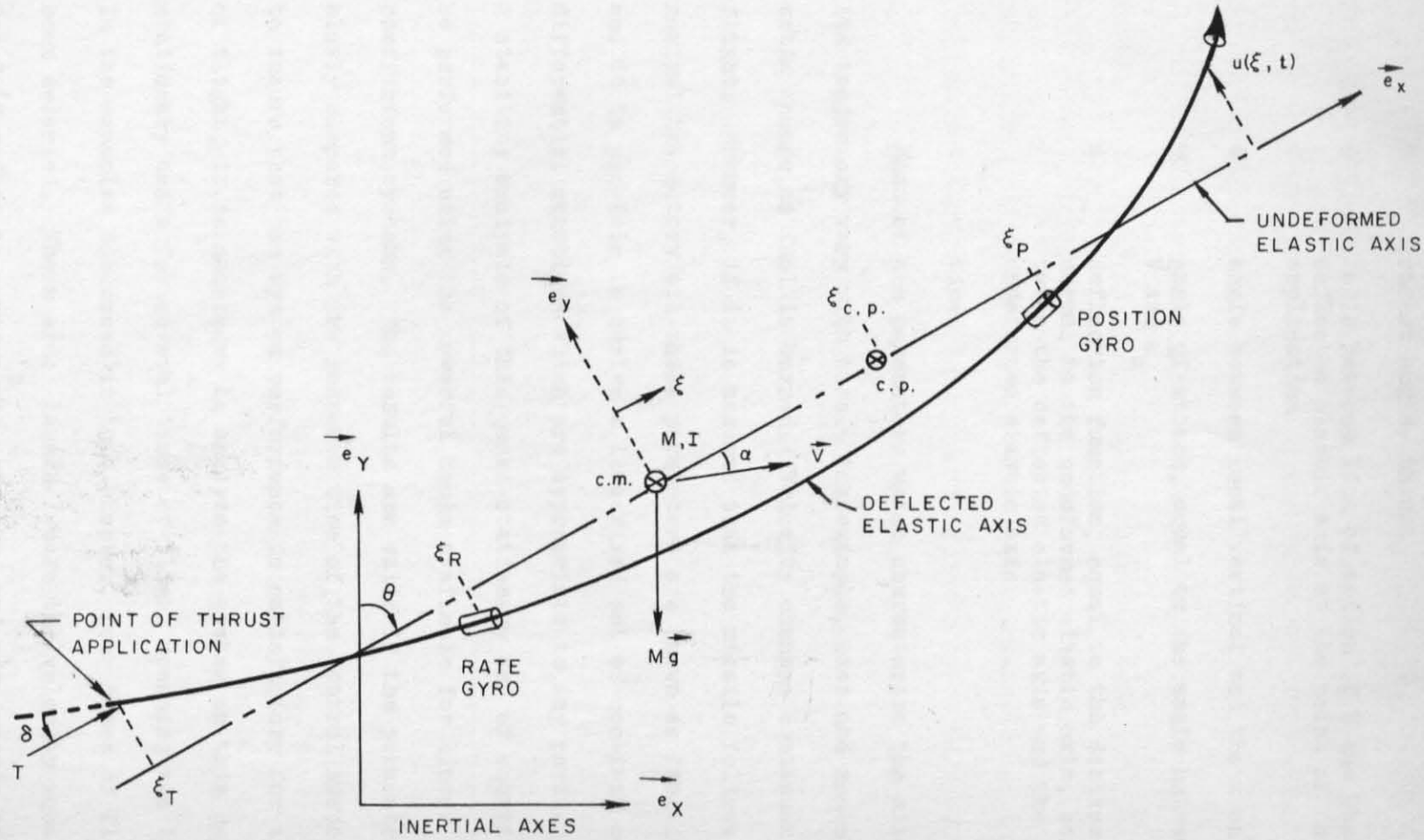


Figure 2.1: Geometry of the missile system.

T	rocket engine thrust
δ	angle between line of action of T and the deflected elastic axis at the point of thrust application
θ	angle between local vertical and the x axis
α	angle of attack, equal to the angle between $\vec{V}$ and e_x
u	deflection function, equal to the distance, normal to the undeformed elastic axis, between the deflected elastic axis and the undeformed elastic axis
t	time

Most of the parameters which characterize the missile and its trajectory vary with time. For example, mass and moments of inertia change as fuel is burned. Velocity changes continuously during flight. However, if it is assumed that the missile follows a known nominal trajectory all these parameters are known as functions of time and it is possible to derive a linearized set of constant coefficient differential equations which are appropriate to any particular time. A stability analysis of this quasi-stationary set of equations can be performed using the powerful tools available for linear constant coefficient systems. The results are valid if the parameters change slowly compared with the response time of the control system. In order to insure that the system performance is satisfactory for all times of flight, it is necessary to analyze the system on this quasi-stationary basis for several times of flight spanning the trajectory. In the examples discussed in this chapter, four times of flight have been selected. These are: launch (where the velocity equals zero); max Q (Q refers to aerodynamic pressure, and aerodynamic effects are

of great importance when Q is at its maximum value); midburn (velocity is high and the aerodynamic effect is still of some importance); and burnout (assumed to occur out of the atmosphere and the aerodynamic effect is negligible).

Using the model of the dynamic system developed in the Appendix, it is possible to derive transfer functions between the various significant variables. A normal mode representation is used for the deflection function $u(\xi,t)$. If the loop transmission at higher mode frequencies is well attenuated (as it will be in a well designed autopilot), it is possible to approximate the deflection function closely with only a small number of terms of the normal mode expansion. For the purposes of this investigation it is assumed that only the first three terms are necessary (see the discussion of this point in the Appendix, p.152). Therefore,

$$u(\xi,t) = \sum_{i=1}^{3} q_i(t)\, \phi_i(\xi) \qquad (2.1)$$

where

$$q_i(t) = i^{th} \text{ generalized bending coordinate} \qquad (2.2)$$

and

$$\phi_i(\xi) = i^{th} \text{ bending mode shape } (i^{th} \text{ eigenfunction of the free-free beam equation}) \qquad (2.3)$$

It is convenient to define the following quantity for use in the later portions of this chapter:

$$\lambda_i(\xi) = \frac{d\phi_i}{d\xi} = \text{slope of the } i^{th} \text{ bending mode eigenfunction} \qquad (2.4)$$

The following notation will be employed when using these mode shapes and mode slopes

$$\phi_{Ai} = \phi_i(\xi_A) \tag{2.5}$$

and

$$\lambda_{Ai} = \lambda_i(\xi_A) \tag{2.6}$$

where

$$\xi_A = \text{value of } \xi \text{ at location A along the axis of the missile} \tag{2.7}$$

The approximate open loop transfer functions of interest are shown in Figure 2.2. A feedback loop has to be closed around the system to provide the capability of following a commanded attitude variation θ_c. Angular position feedback is generally supplied by a position gyro located at coordinate ξ_P, and a rate gyro located at ξ_R is used to provide the prediction necessary to stabilize the system. The complete model of the control system is shown in block diagram form in Figure 2.3. This is the configuration which will be analyzed. The symbols needing definition in the figures are:

$\mu_c = \frac{T}{I}(-\xi_T)$ (note that $\mu_c > 0$)

$\mu_\alpha = \frac{N_\alpha}{I}\xi_{cp}$ ($\mu_\alpha > 0$ if the c.p. is forward of the c.g.

N_α = aerodynamic force acting normal to the undeformed elastic axis per unit angle of attack

ω_i = natural frequency of the i^{th} free-free structural mode

ζ_i = damping of the i^{th} natural mode (for the lightly damped modes of interest it will be assumed that $\zeta_i = 0.005$)

K_D = attitude gain constant

K_R = attitude rate gain constant

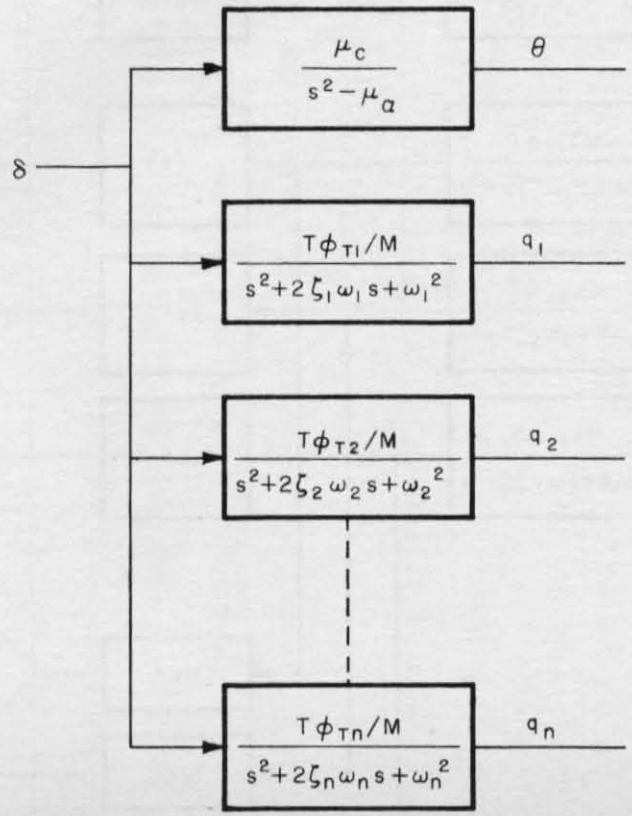


Figure 2.2: Open-loop transfer functions.

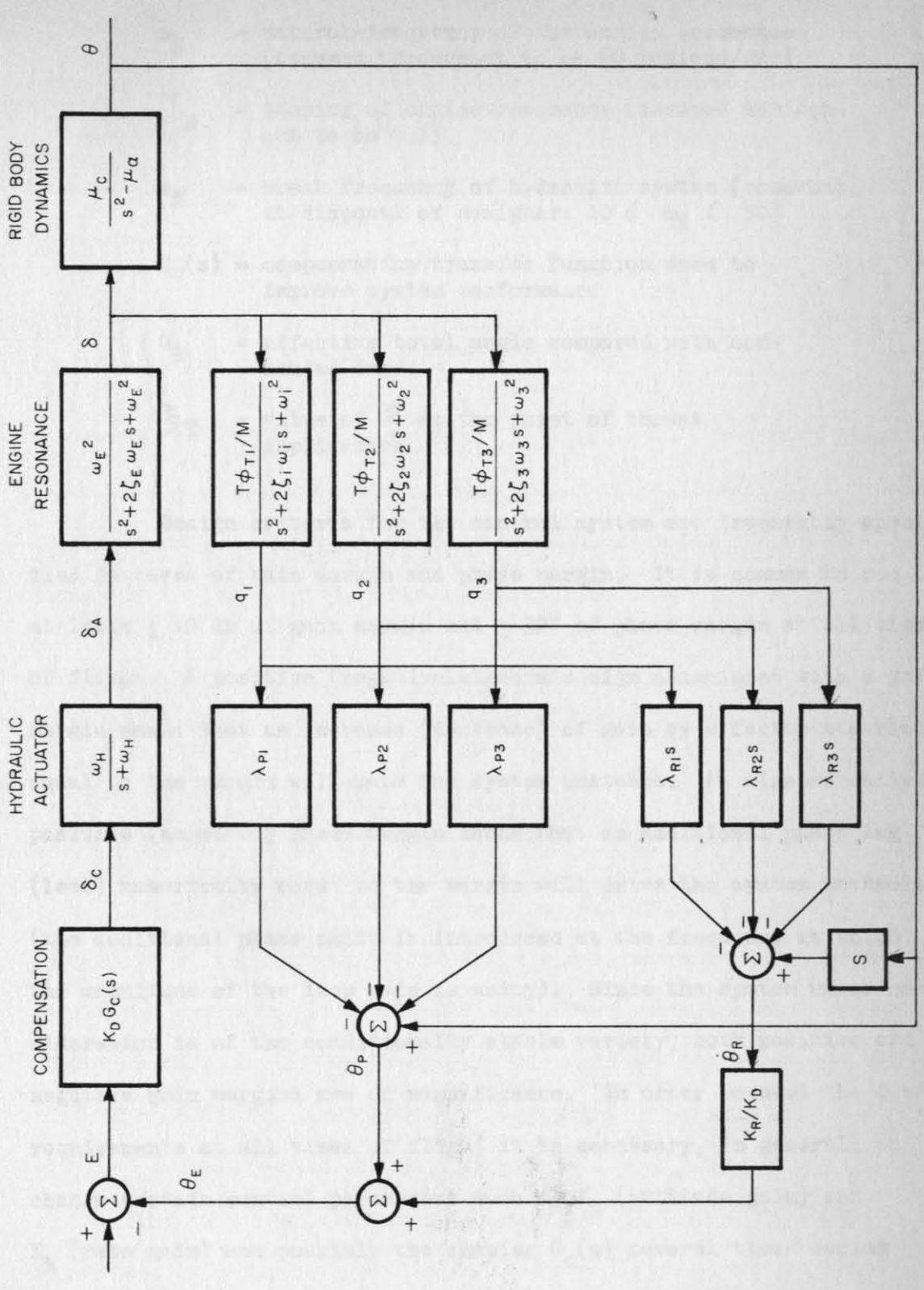


Figure 2.3: Block diagram of multiply resonant missile example.

ω_E = natural frequency of the engine resonance (assumed throughout to be 60 radians/sec)

ζ_E = damping of engine resonance (assumed throughout to be 0.1)

ω_H = break frequency of hydraulic system (somewhat at disposal of designer; $10 \leq \omega_H \leq 50$)

$G_c(s)$ = compensation transfer function used to improve system performance

Θ_E = effective total angle compared with commanded input

ξ_T = value of ξ at the point of thrust application

Design criteria for the control system are frequently specified in terms of gain margin and phase margin. It is common to require at least $\pm$ 10 db of gain margin and $\pm$ 30° of phase margin at all times of flight. A positive (negative) algebraic sign associated with a gain margin means that an increase (decrease) of gain by a factor numerically equal to the margin will make the system unstable. An algebraically positive (negative) phase margin means that an additional phase lag (lead) numerically equal to the margin will drive the system unstable (the additional phase shift is introduced at the frequency at which the magnitude of the loop gain is unity). Since the system under consideration is of the conditionally stable variety, both positive and negative gain margins are of significance. In order to meet the design requirements at all times of flight it is necessary, in general, to change certain control parameters such as K_D (attitude gain) and K_R (rate gain) and possibly the shaping $G_c(s)$ several times during the flight. In order to keep the autopilot reliability high the number of such changes during a flight should be kept as low as possible.

Before examining the effects of the lightly damped feedback paths, it will be instructive to analyze the stability of the system in the absence of bending. To simplify matters further, $G_c(s)$ has been set equal to unity. Figure 2.4 shows this condition. The loop transmission, $\Theta_E/\mathcal{E}$, is

$$G(s) = \frac{\Theta_E}{\mathcal{E}} = \frac{\mu_c (K_D + K_R s) \omega_E^2 \omega_H}{(s^2 - \mu_\alpha)(s + \omega_H)(s^2 + 2\zeta_E \omega_E s + \omega_E^2)} \tag{2.8}$$

The system is stable if the rational function $P(s) = 1 + G(s)$ has no zeros with positive real parts. The number of right-half-plane roots of the equation $P(s) = 0$ can be determined by using the Nyquist stability criterion (16). From Equation 2.8, letting $s = j\omega$,

$$G(j\omega) = \frac{K_D(-\mu_c/\mu_\alpha)\left[1 + j\omega/(K_D/K_R)\right]}{(1 + j\omega/\sqrt{\mu_\alpha})(1 - j\omega/\sqrt{\mu_\alpha})(1 + j\omega/\omega_H)} \cdot \frac{1}{\left[1 + 2\zeta_E\, j\omega/\omega_E + (j\omega/\omega_E)^2\right]} \tag{2.9}$$

The exact shape of the locus of Equation 2.8 depends, of course, upon the particular numerical values chosen for the various parameters appearing in the equation. During operation of a typical control system, μ_c and μ_α will vary strongly with time, ω_E and ω_H will be essentially constant, and the control parameters K_D and K_R will be at the disposal of the autopilot designer. Figure 2.5 shows, in the form of a log magnitude and phase plot, how the transmission might vary as a function of flight time for a typical set of parameters. The parameters which are the same for all the curves are:

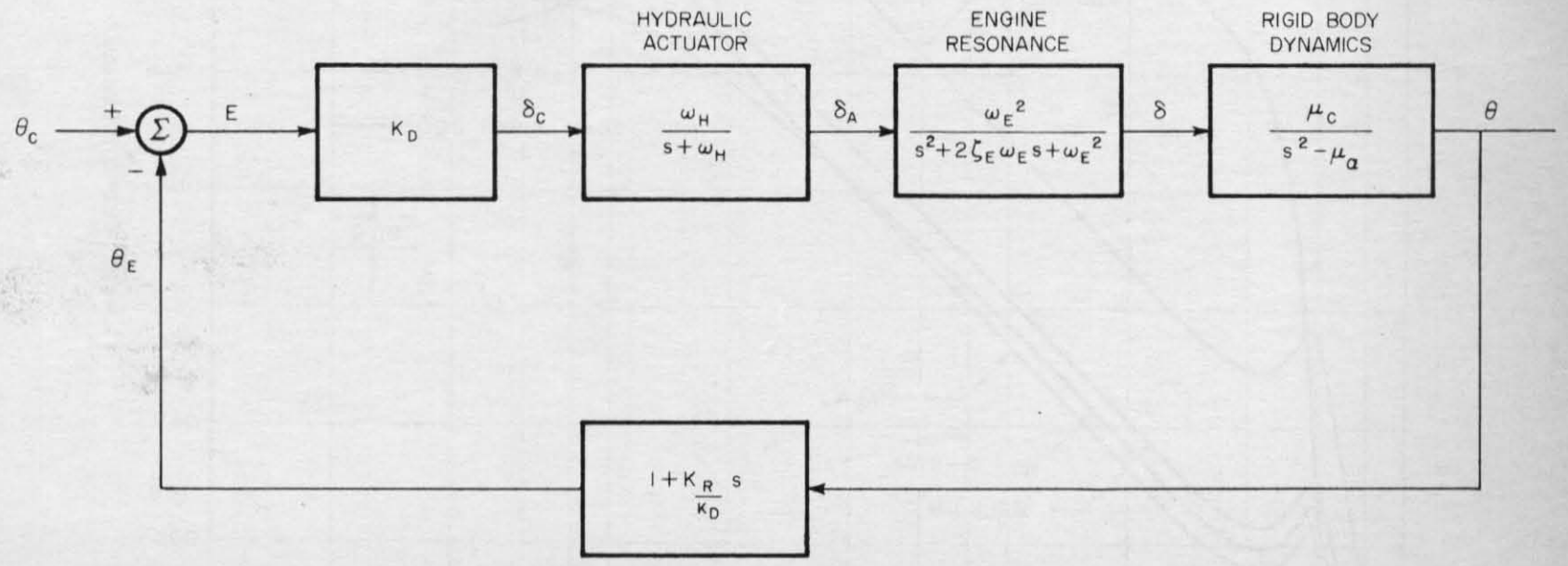


Figure 2.4: Block diagram of missile example when no bending is considered and $G_c(s) = 1$.

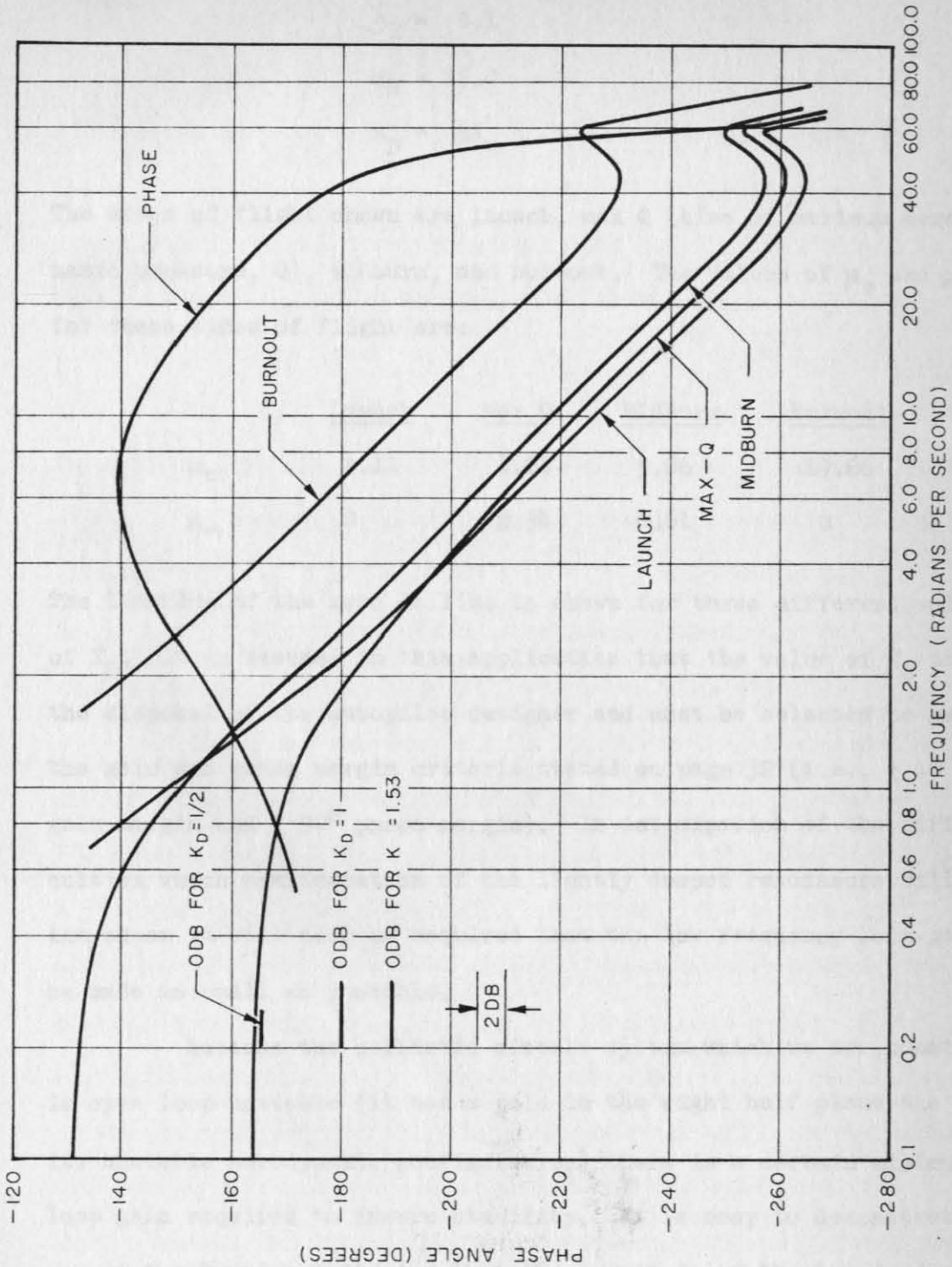


Figure 2.5: Typical Frequency Response Curves for a Rigid Missile.

$$\omega_E = 60.0$$

$$\zeta_E = 0.1$$

$$\omega_H = 15.0$$

$$K_D = 3K_R$$

The times of flight shown are launch, max Q (time of maximum aerodynamic pressure, Q), midburn, and burnout. The values of μ_c and μ_α for these times of flight are:

	Launch	Max Q	Midburn	Burnout
μ_c	4.11	4.84	5.86	19.66
μ_α	0	2.34	0.61	0

The location of the zero db line is shown for three different values of K_D. It is assumed in this application that the value of K_D is at the disposal of the autopilot designer and must be selected to meet the gain and phase margin criteria stated on page 32 (i.e., $\pm$ 10 db gain margin and $\pm$ 30° phase margin). In anticipation of the difficulties which consideration of the lightly damped resonances will introduce it will also be required that the low frequency loop-gain be made as small as possible.

Because the ballistic missile system which we are considering is open loop unstable (it has a pole in the right half plane due to its unstable aerodynamic configuration) there is a certain minimum loop gain required to insure stability. It is easy to demonstrate by use of the Nyquist Criterion that the magnitude of the low frequency loop gain must be greater than unity if the system described by Equation 2.9 is to be stable. But from Equation 2.9, the magnitude of the low frequency gain is $K_D \mu_c/\mu_\alpha$. Therefore the following

inequality must hold just to maintain marginal stability:

$$K_D \mu_c/\mu_\alpha \geqslant 1 \qquad (2.10)$$

In order to meet the specification of at least -10 db gain margin, the right hand side of this inequality must be multiplied by a factor equivalent to 10 db, i.e., by a factor of 3.162. In light of this we rewrite Equation 2.10 in the following form:

$$K_D \geqslant 3.162\, \mu_\alpha /\mu_c \qquad (2.11)$$

In order to keep the gain low, the equality sign is used.

It should be noted from the relation above that the minimum required value of K_D approaches zero as μ_α approaches zero. Thus, in regions of the flight where aerodynamic effects are negligible it appears that the attitude gain K_D can be made as small as we like.

There is, however, a lower limit to the value which we can assign to K_D. This limit arises from a static accuracy requirement. The model presented in the appendix has assumed implicitly that the thrust vector passes through the missile c.m. when $\delta = 0$ and $\mu(\xi,t) = 0$, i.e., for this condition the thrust produces no moment about the c.m. However, the nominal zero of δ is set in practice by optically aligning the center line of the rocket thrust chamber with the center line of the body of the missile. In practice there will be some misalignment error, δ_m, equal to the angle between the line of action of the thrust vector when $\delta = 0$ and the line joining the c.m. of the system and the point of application of the thrust vector. Since in the steady state the thrust vector must pass through the c.m., the steady state

value of δ, δ_{ss}, must satisfy the equation $\delta_{ss} = -\delta_m$. From Figure 2.4 it may be seen that $\delta_{ss} = K_D \varepsilon_{ss}$. The steady state attitude error, ε_{ss}, is therefore

$$\varepsilon_{ss} = -\delta_m / K_D \tag{2.12}$$

A lower limit of 1/2 is frequently specified for K_D, which means that the steady state attitude error due to thrust vector misalignment is at most twice the value of the misalignment. Typically this will keep the attitude error below 1°.

In general, then, the minimum allowable value of K_D at any time of flight is set by the following pair of inequalities:

$$K_D \geq \begin{cases} 3.162\,\mu_\alpha/\mu_c & \text{for } 3.162\,\mu_\alpha/\mu_c > 0.5 \\ 0.5 & \text{for } 3.162\,\mu_\alpha/\mu_c \leq 0.5 \end{cases} \tag{2.13}$$

K_D is usually not programmed to vary continuously. Rather, in the interest of reliability and simplicity it is programmed to change its value at several discrete times. If we limit ourselves to one gain change for the example under consideration, it is reasonable to choose $K_D = 1.53$ from launch to midburn and $K_D = 1$ from midburn to burnout. If a second gain change were allowed it would be desirable to reduce K_D to 1/2 (its lowest allowable value) as soon after midburn as possible consistent with the phase margin requirements. The stability margins can be read directly from Figure 2.5 for the two values of gain chosen and are tabulated below.

		Gain Margins (db)	Phase Margins (degrees)
$K_D = 1.53$	Launch	$-\infty$, + 30	33
	Max Q	-10, + 28.5	31
	Midburn	-23.3, + 26.9	36
$K_D = 1$	Midburn	-19.6, + 30.6	31
	Burnout	$-\infty$, + 20.0	40

In order to get a feeling for how the system should behave when lightly damped resonances are present, the following observations are made from Figure 2.3. Each of the resonant transmissions from δ to Θ_E is summed with the rigid body transmission. The transmission associated with the k^{th} mode will be called $G_k(s)$ and is given by

$$G_k(s) = \frac{-T\,\phi_{Tk}}{M\,\omega_k^2} \; \frac{\omega_k^2}{s^2 + 2\zeta_k\,\omega_k\,s + \omega_k^2} \; (K_D\,\lambda_{Pk} + K_R\,\lambda_{Rk}\,s) \qquad (2.14)$$

To obtain the frequency response one makes the substitution $s = j\omega$, which gives

$$G_k(j\omega) = -\frac{T\phi_{Tk}}{M\omega_k^2} \; (K_D\lambda_{Pk} + j\,K_R\lambda_{Rk}\,\omega) \; \frac{1}{1 - (\omega/\omega_k)^2 + j\,2\zeta_k\,(\omega/\omega_k)} \qquad (2.15)$$

A Nyquist plot of the normalized function

$$F(ju) = \frac{1}{1 - u^2 + j\,2\zeta u} = X + jY \qquad (2.16)$$

is shown in Figure 2.6. It is easy to show that the equation describing this function, when u is eliminated, is

$$(X - 1/4)^2 + (Y + 1/4\zeta)^2 = (1/4\zeta)^2 \quad \text{for } \zeta X/Y \ll 1 \qquad (2.17)$$

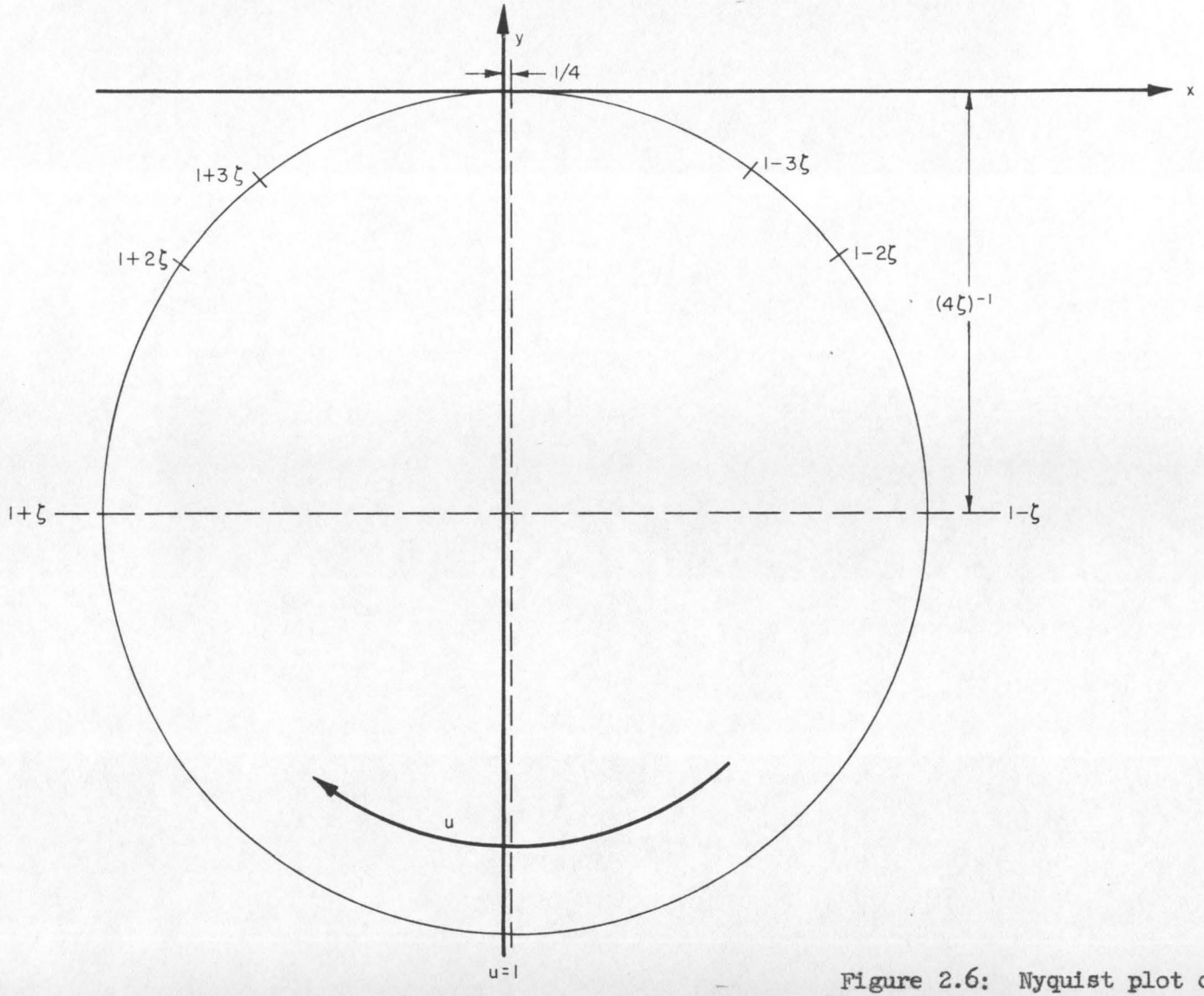


Figure 2.6: Nyquist plot of Equation 2.16.

Furthermore,

$$u \cong 1 + \zeta X/Y \quad \text{for} \quad \zeta X/Y \ll 1 \tag{2.18}$$

For the lightly damped resonances of interest ($\zeta < 0.01$) the figure is a circle of radius $1/(4\zeta)$ centered at $X = 1/4$ and $Y = -1/(4\zeta)$. The peak transmission is very nearly $1/(2\zeta)$ at an angle of -90°.

From Equation 2.15 and 2.16 the phase and magnitude of $G_k(j\omega)$ in the vicinity of resonance is

$$G_k(j\omega) = + \frac{T\ K_R}{M\ \omega_k} (-\phi_{Tk}\lambda_{Rk}) \left(1 - j \frac{K_D \lambda_{Pk}}{K_R \lambda_{RK}\ \omega_k}\right) j\ F(\omega/\omega_k) \tag{2.19}$$

The Nyquist plot of this function for $\omega \approx \omega_k$ is approximately a circle passing through the origin. The diameter of the k^{th} mode circle passing through the origin makes an angle ψ_k with the positive real axis, where

$$\psi_k = \begin{cases} -\tan^{-1} \dfrac{K_D \lambda_{Pk}}{K_R \lambda_{RK}\ \omega_k} \text{ (radians)} & \text{if } \phi_{Tk}\lambda_{Rk} < 0 \\ \pi - \tan^{-1} \dfrac{K_D \lambda_{Pk}}{K_R \lambda_{RK}\ \omega_k} \text{ (radians)} & \text{if } \phi_{Tk}\lambda_{Rk} > 0 \end{cases} \tag{2.20}$$

The total loop transmission through the k^{th} resonance is obtained by multiplying $G_k(j\omega)$ by the contents of the blocks in Figure 2.3 labeled "Compensation," "Hydraulic Actuator," and "Engine Resonance," all evaluated at $\omega = \omega_k$. $\Theta_E/\mathcal{E}$ is then the sum of all the resonant transmissions and the rigid body transmission. Figure 2.7 shows an example of the form of the loop transmission which results when resonances are included for the case where $K_D\ G_c(s) = 1$. Three resonant modes are included at frequencies of 17, 35, and 53 radians per second. A mode

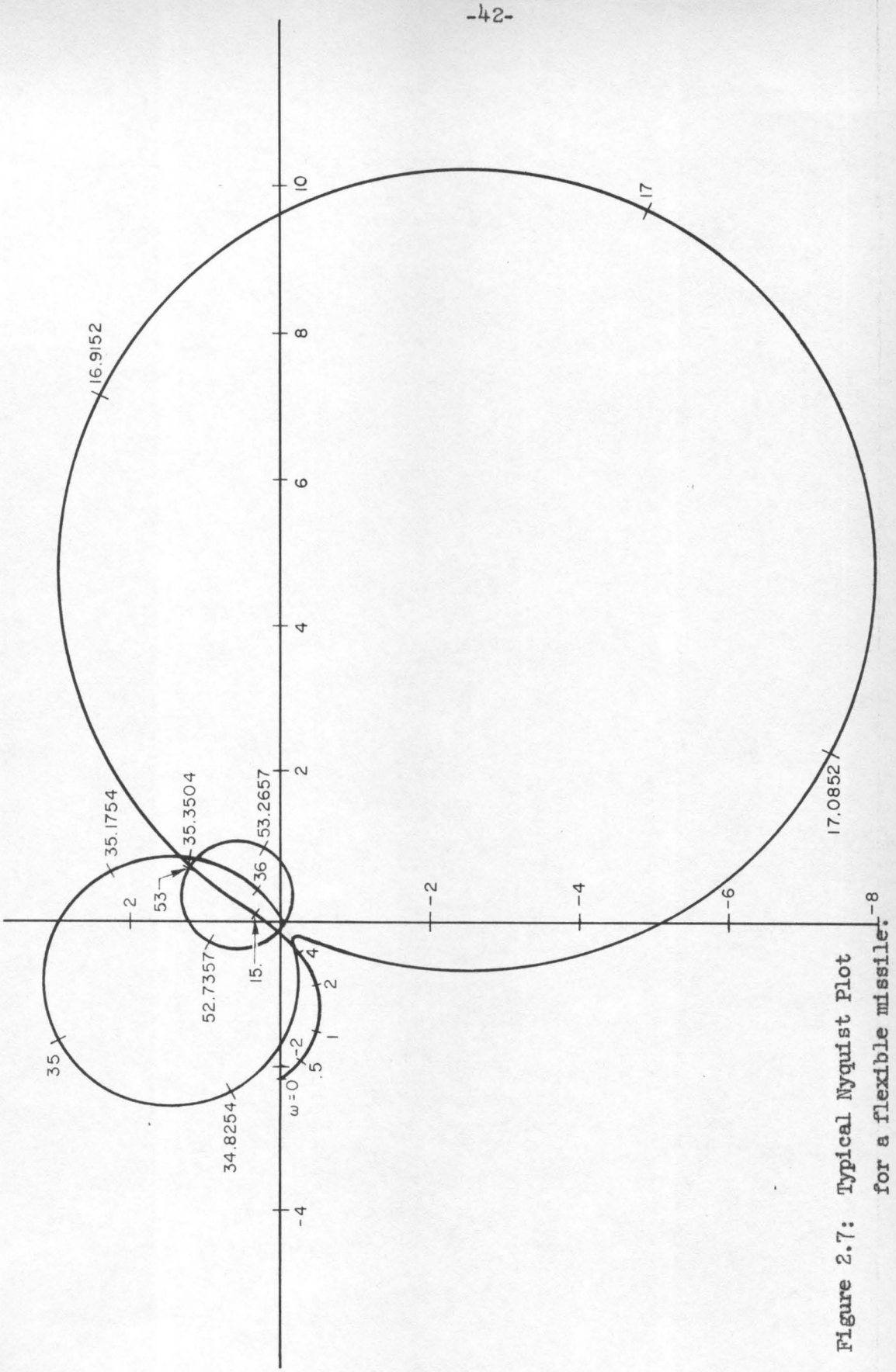


Figure 2.7: Typical Nyquist Plot for a flexible missile.

is unstable if its circle in the Nyquist diagram encircles the -1 point. For the parameters chosen Figure 2.7 shows that the system will have a divergent 2nd mode oscillation. Although the loop transmissions at first and third mode frequencies is greater than unity, these modes are stable due to the orientation of these bending loops.

Equations 2.18 and 2.19 show that the magnitude and orientation of the resonant loops depends upon the bending mode slopes λ_{Pk} and λ_{Rk} at the points where the position and rate sensors are located. In particular, if the slope at either of the sensor locations changes sign the output of the device will be changed in phase by exactly 180^{o}. Figure 2.8 shows how the mode slopes might vary during the course of a flight. This variation means that the phase angle at which the peak resonant transmissions occur can change considerably during the course of a missile flight, as can the magnitude of the resonant transmission.

There are essentially two ways by which a resonant mode may be stabilized. First, the transmission at a resonant frequency may be reduced by appropriate compensation to the point where the loop transmission at resonance is less than unity. In this case the mode is said to be "gain stabilized." A gain stabilized mode will be stable regardless of the phase of the loop transmission at resonance. Second, although a large resonant loop occurs at resonance, a mode will be stable if its orientation on the Nyquist diagram is such that the -1 point is not encircled. In this case the mode is said to be "phase stabilized." A phase stabilized mode will be de-stabilized if the phase shift at resonance changes enough to swing the resonant loop into a position such that it encircles the -1 point. Using these

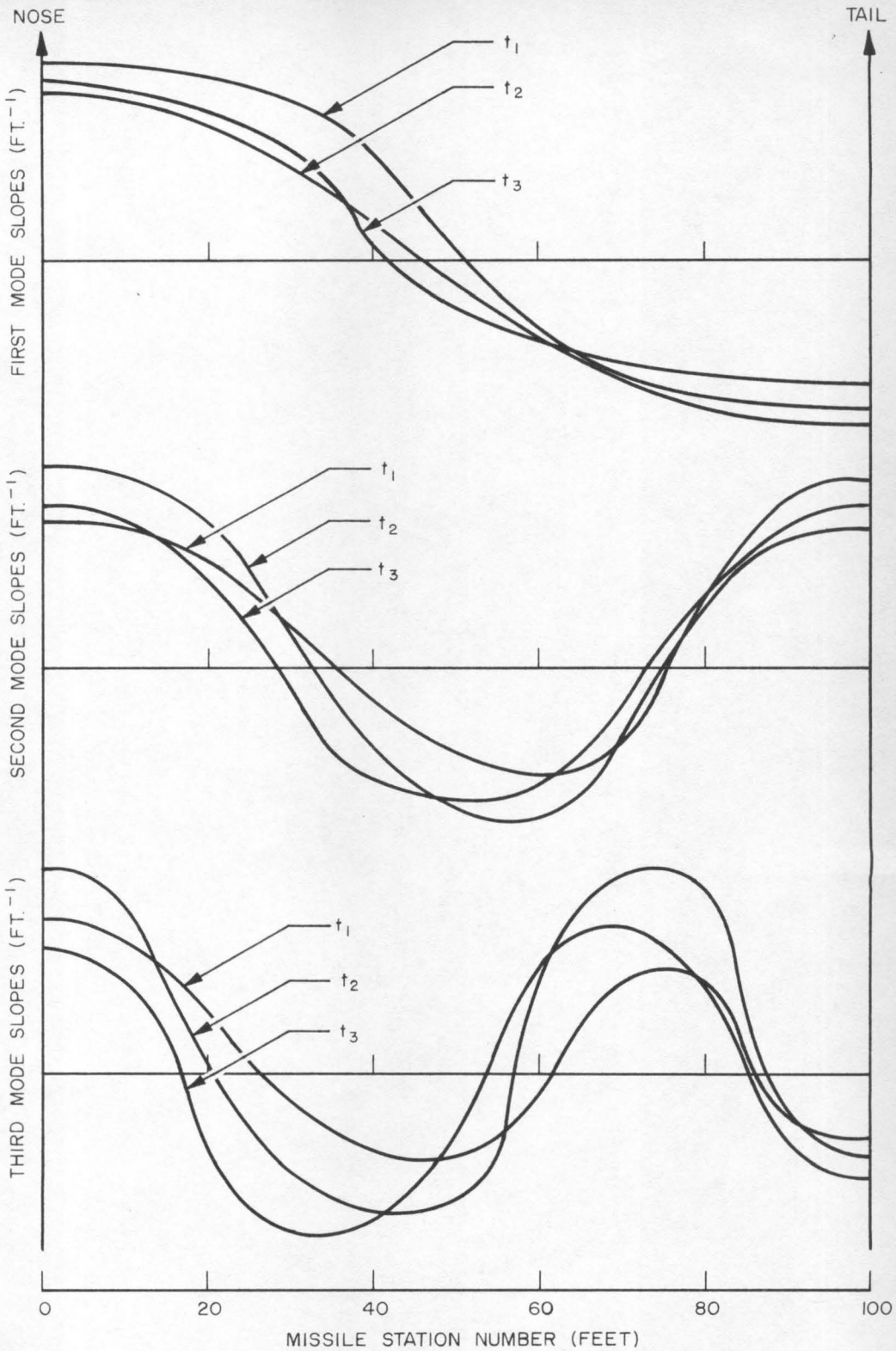


Figure 2.8: Typical normalized bending mode slopes for three times of flight ($t_3 > t_2 > t_1$).

concepts we see that, in Figure 2.7, modes 1 and 3 (at frequencies of 17 and 53 radians per second, respectively) are phase stabilized while mode 2 (at 35 radians per second) is unstable.

In light of the previous discussion concerning the variation with time of the bending mode slopes and the direct effect this has on the phase at resonance, it is evident that great care must be taken to insure that the phase is correct at the frequency of each phase stabilized mode. This is quite difficult to do for several bending modes simultaneously, especially so because the system characteristics are not known exactly and vary considerably with time. In addition, the system characteristics at higher bending mode frequencies are not well known due to inaccuracies at high frequencies in any system model which is used.

The following section describes the approach which is currently followed in the design of autopilots for highly flexible vehicles.

2.3 The Present Design Approach

It should be recognized that in the model described in the preceding section all coupling between the autopilot and the structural resonances occurs via the autopilot sensing devices, i.e., the position and rate gyros. While a more complete representation would indicate other possible methods of coupling, further analysis shows that by far the dominant coupling does indeed occur via the sensors. In addition, because a rate sensor emphasizes high frequency signals, the coupling through the rate gyro is usually far more detrimental than coupling through the position loop. Thus the rate gyro is usually

located in as favorable a location along the length of the missile as is possible.

This favorable position is close to an antinode of the first or second bending mode (a point where $\lambda_i(\xi_R) = 0$ for i = 1 or 2). In this way the gyro effectively does not pick up the mode and thus it is not coupled to the autopilot. However, an accurate knowledge of the missile's mode shapes is required to accomplish this, and the autopilot design, therefore, is highly dependent upon the missile's structural characteristics. Due to the fact that the mode shapes change considerably during the duration of a flight (thus the position of the various antinodes change) it may be necessary to use several rate gyros and use their outputs individually or sum them in pre-selected proportions depending upon the time of flight. It is not possible, in general, to place a single sensor near an antinode of more than one mode.

The properties of the final control system design are highly dependent upon whether the rate gyro is located forward or aft of the first mode antinode. If the rate gyro is located forward of the first mode antinode, $\phi_{T1}\lambda_{R1} > 0$, while if the gyro is located aft of the first mode antinode, $\phi_{T1}\lambda_{R1} < 0$. Referring to Equation 2.19 we see that, for very small λ_{P1},

$$\psi_1 \cong \begin{cases} 0 \text{ (radians)} & \text{for } \phi_{T1}\lambda_{R1} < 0 \text{ (gyro aft)} \\ \pi \text{ (radians)} & \text{for } \phi_{T1}\lambda_{R1} > 0 \text{ (gyro forward)} \end{cases} \tag{2.21}$$

Remember that the angle ψ_1 specifies the phase at which the maximum

first mode transmission occurs when all phase shift in the autopilot is neglected.

When forward mounted gyros are used, an attempt is made to introduce as much phase lag as possible at the first bending frequency by putting lag networks in series in the loop. Since these lag networks also introduce attenuation at higher frequencies, it may be possible to gain stabilize all higher modes when this method is applicable. Unfortunately, when the first mode frequency is too low, it is not possible to introduce enough lag to stabilize the first bending mode without also sacrificing a significant portion of the rigid body stability margins. It is then necessary to place the rate gyro aft of the first mode antinode and minimize control system phase shift at the first mode frequency in order to keep the system stable. Since lag networks in this case cannot be put in at very low frequencies relative to the higher bending frequencies, it is usually not possible to gain stabilize all higher modes. In the system represented by Figure 2.7, for example, the rate gyro is located aft of the first mode antinode. Thus the phase of the first mode transmission through the rate gyro alone (neglecting all phase shift in the autopilot) is 0^{o} (by Equation 2.21). The fact that the first mode resonance as shown actually occurs at an angle of approximately -27^{o} is due to the effect of lag in the autopilot and the contribution of the position sensor (which in the example was located forward of the first mode antinode) to the total loop gain. It is evident from Figure 2.7 that the autopilot does not cut off rapidly enough to gain stabilize either the second or third bending mode.

Aerodynamic and bending problems place two contradictory requirements on the value of the low frequency loop gain. In order to stabilize the rigid body in a region where aerodynamic forces are large (i.e., at max Q) it is desirable to have a high loop gain in order to maintain a tight control loop. However, a low loop gain is desirable at all times to minimize bending transmissions. The resolution of a compromise between these two competing factors depends upon the parameters of the particular system under consideration.

In general a solution is achieved by tailoring the control system parameters at each time of flight to the nominal missile characteristics which apply at that time. Control loop gains and filters are changed either continuously or, more commonly, at several predetermined times during the flight of the missile. Thus an extremely detailed design must be performed if any slight change occurs in missile characteristics. The control system parameters which are the result of the design effort are, in particular, extremely sensitive to changes in bending characteristics. An improvement in design which would result in a system which was relatively insensitive to the magnitude of the bending frequencies and bending mode shapes would be extremely desirable.

The material presented in this chapter was intended to serve as an illustration of how a system with lightly damped resonances might arise. The example of the ballistic missile illustrates the point that these resonances are the most severe obstacle to a straightforward control system design. Any effort to improve control systems for this class of systems should aim directly at ameliorating this

problem. The following chapters describe the principles and operation of an approach to the problem which leads to a system which is no longer highly sensitive to bending characteristics and, in this regard, must be considered an improvement over the conventional design approach.

CHAPTER III

LIGHTLY DAMPED RESONANCES AND ADAPTIVE CONTROL

3.1 Introduction

With the background provided by Chapter II it is possible to examine the desirability of incorporating adaptive features in a controller for systems with lightly damped resonances. Let us first restate the analytic formulation of the adaptive control problem which was presented on page 19.

Given: A dynamic system specified by the following set of equations:

$$dX_i/dt = F_i(X_1, \ldots, X_n, \alpha_1 \ldots, \alpha_m, \beta_1, \ldots, \beta_\ell, t) \qquad (3.1)$$

where

$X_i(t)$ = i^{th} state variable (i=1, ..., n)

$\alpha_i(t)$ = i^{th} control parameter (i=1, ..., m)

$\beta_i(t)$ = i^{th} critical system parameter (l = 1, ..., ℓ)

t = time

Note that the quantities β_i are not known precisely during the design of the system in cases where adaptive control is desirable.

Required: To select the proper values of $\alpha_1, \ldots, \alpha_m$ at all times to insure that the overall system response satisfies some specified design criteria.

The characteristic feature of the hypothetical highly resonant system of concern here is the fact that the frequencies of the resonances are not known with precision. In the ballistic missile illustration it will be recalled that there were an infinite number of

these resonant frequencies corresponding to the natural frequencies of an unsupported (free-free) beam. Fortunately, for all practical purposes, only a finite number of these resonances are of interest. These frequencies correspond to the β_1 in Equation 3.1. It will be demonstrated in Section 3 of this chapter that it is possible to design a particular type of control system which will result in a satisfactory system response if this finite number of critical frequencies is known accurately enough. Certain auxilliary restrictions which must be met are also discussed in Section 3.

Since the system is time-varying and has critical parameters which are not known precisely it will be necessary to measure these parameters in real time as the system is operating. This information will then be used to adjust parameters in the controller. The proposed adaptive system configuration is shown in Figure 3.1. The adaptive feature of this system is the secondary feedback loop. A signal y(t) is processed in a "parameter measurement computer" which provides commands to the controller to adjust internal parameters. In accordance with standard notation, the input signal to the controller is an error signal e(t) and the output of the controller is the "manipulated variable" m(t). m(t) is the signal which activates the dynamic system (it would correspond to the angle δ in the missile example).

As is frequently the case with adaptive systems, it is convenient now to think of the adaptive portion of the control system as performing two relatively distinct functions. The first may be termed the "identification function" and the second termed the "correction function." In any particular system, it may or may not be possible

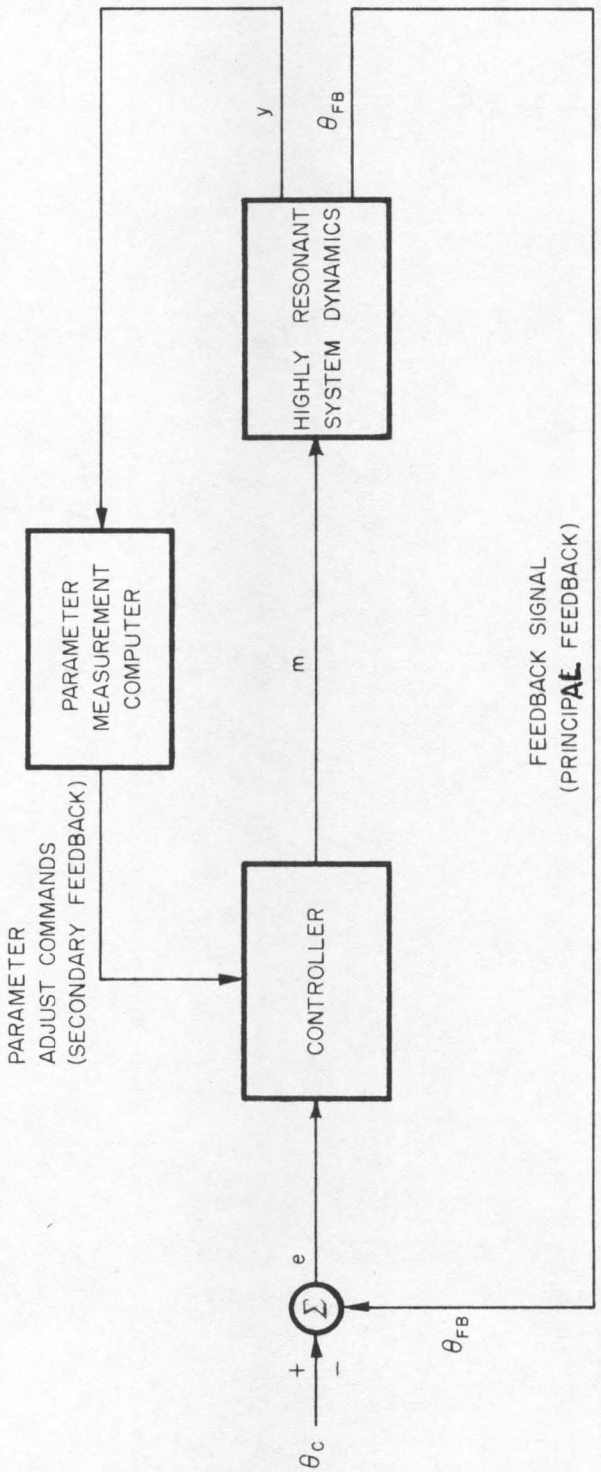


Figure 3.1: Adaptive system schematic.

to identify separate portions of the system performing these functions. Likewise, they may be occurring simultaneously or during different portions of the control cycle. During the first phase of the operation the pre-determined critical system parameters are measured to within the desired degree of accuracy. Action is taken to vary the adjustable system parameters during the second phase of the adapting process. The principle feedback loop is operating throughout the process of adapting.

Information about the resonant frequencies is obtained by processing a signal y(t) (see Fig. 3.1). The signal y(t) can be any arbitrary function of the state of the system and is generated in such a manner as to insure that if any oscillation at a resonant frequency is occurring it will be seen in y(t). In the missile example, y(t) could be the output of a rate-gyro mounted on the missile in a location where all the bending slopes of interest have a non-zero value (the nose and tail of the missile are two such locations). If the system is oscillating at more than one resonant frequency, signal components will be present in y(t) at each of these frequencies.

The adjustable controller parameters will initially be set to values which nominally provide a stable system at $t = 0$. As the characteristics of the dynamic system change with time it is possible that one or more of the resonant modes will become unstable, and an oscillation will begin to build up near the corresponding natural frequency. The Adaptive Frequency Measuring Computer (hereafter referred to as the FMC) will detect a component at this frequency of oscillation in y(t) and via its internally programmed logic will cause parameters

to be changed in the controller to restabilize the system. It is immediately apparent that if this concept is to be successful in a real physical system the adaptation (i.e., the change from an unstable to a stable configuration) must take place rapidly enough to insure that the resonant oscillations do not reach a destructive level. Structural vibrations of a missile, for example, lead to large dynamic loads on the structure, and the missile will break up if the maximum design loads are exceeded.

The nature of the controller itself merits some comment at this point. The controller will be restricted to be a time varying piecewise linear operator which produces an output m(t) through piecewise linear operations performed on the input signal e(t). The physical realization of this controller could be achieved by utilizing active electrical networks to perform the desired operations. In order to modify the transfer characteristics of this continuous controller it would be necessary to change the values of physical components such as resistors and capacitors.

The controller can also be realized within a digital computer by programming the solution of a set of difference equations relating the input and output signals. Such digital systems are considerably more flexible than conventional analogue controllers. In order to modify the transfer characteristics of the digital controller it is only necessary to change the values of certain parameters which are stored in the digital computer. This can be done quite easily.

Control systems which incorporate digital computers directly in the control loop are called "sampled-data" control systems (due to

the fact that digital computers operate only on discrete samples of a continuous signal). The adaptive system which is the subject of this thesis is a sampled-data system. Both the controller and the adaptive parameter measuring computer are instrumented in a digital computer. The sampled-data configuration is shown in Figure 3.2. The two switches represent samplers which sample e(t) and y(t) every T_s and T_y seconds respectively. Two different sampling rates are used because, as will be shown, the signal e(t) need not be sampled at the high rate which is necessary for y(t). Thus some economy of computing capacity is effected by making $T_s \gg T_y$. In accordance with standard notation, starred quantities (e.g., $e^*(t)$) represent sampled signals. Note, in Figure 3.2, that a zero order hold circuit is used to convert the discrete output $m^*(t)$ of the digital computer into a piecewise constant signal m(t) which is used as an input to the highly resonant system dynamics.

A great body of literature is available on the standard techniques which are available for analyzing linear constant coefficient sampled-data systems. It will be assumed that the reader is familiar with the Z-transform as applied to linear sampled-data systems, and this analytic technique will be used freely in subsequent sections of this thesis. Familiarity with the Nyquist stability criterion and the root-locus technique as applied to sampled-data systems is also assumed. A good treatment of these subjects is available elsewhere (1)(4).

The details of an adaptive control technique applicable to highly resonant systems are presented in the following sections.

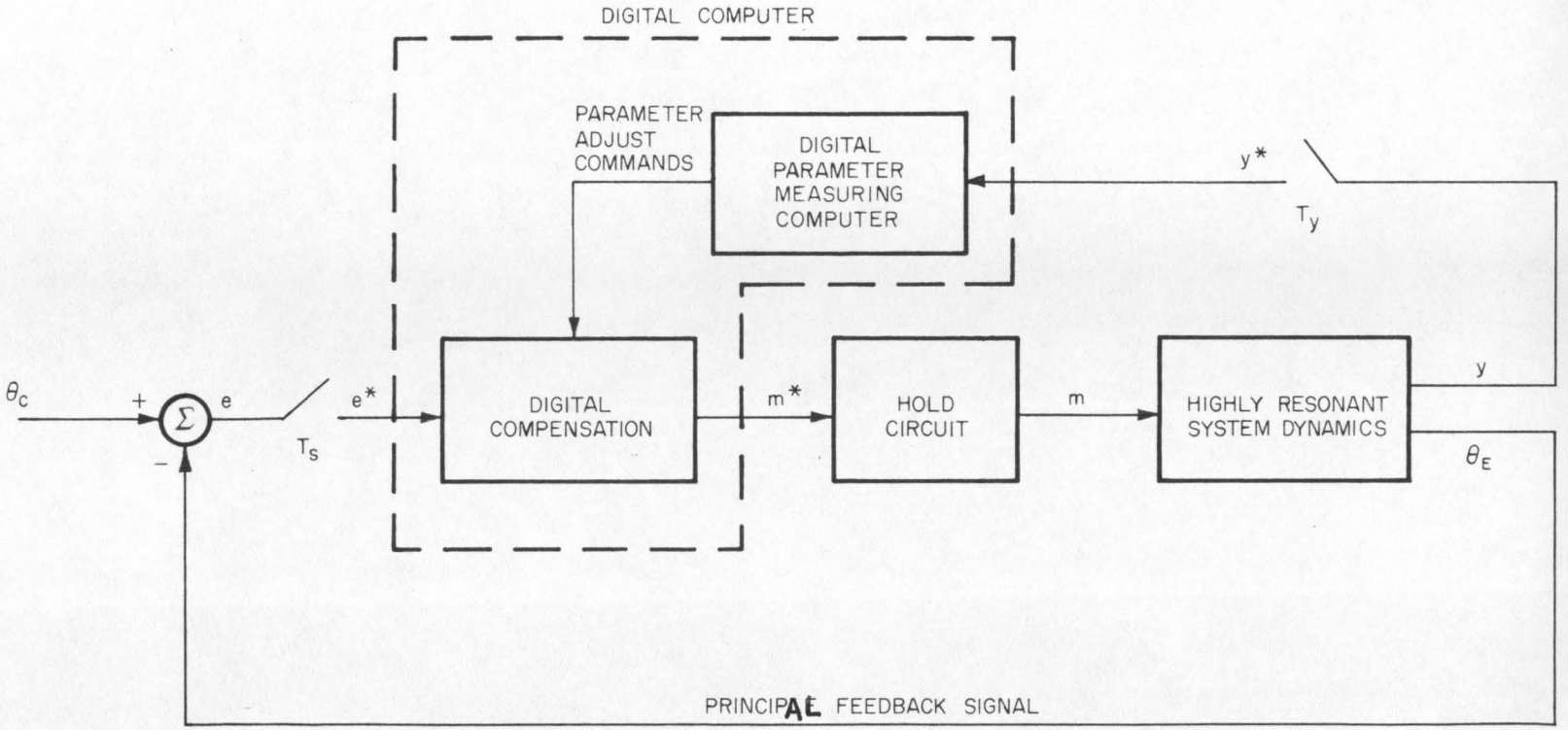


Figure 3.2: Block diagram of digital adaptive control system.

First, the problem of determining the system's natural frequencies will be considered. Following this a means of utilizing this information in a digital controller will be presented. The chapter will close with a summary of the important points developed in the body of the chapter.

3.2 Detection and Frequency Measurement

This section will treat the problem of parameter identification in the system under consideration. The parameters of interest are the frequencies of the system's lightly damped resonances. The numerical values of these frequencies will be obtained by operating on the signal y(t). The basic concept underlying the adaptive portion of the control loop is the following: if the system is unstable due to a particular resonance, an oscillation will begin to build up very near the natural frequency of that resonance. An oscillation at this frequency will appear in y(t), and its presence will be detected and its frequency measured in the adaptive FMC.

Two conflicting requirements are immediately apparent. First, the values of the resonant frequencies must be determined as rapidly as possible to insure that corrective action is taken before the oscillations reach a destructive level. Second, the frequencies must be determined very accurately to insure that effective adaptive adjustments of controller parameters can be made. The accuracy with which one can measure the frequency of a signal, however, is proportional to the length of the signal which one has to examine. Thus in order to locate the frequency with great precision one must wait until an appreciable length of signal is available for processing. On the

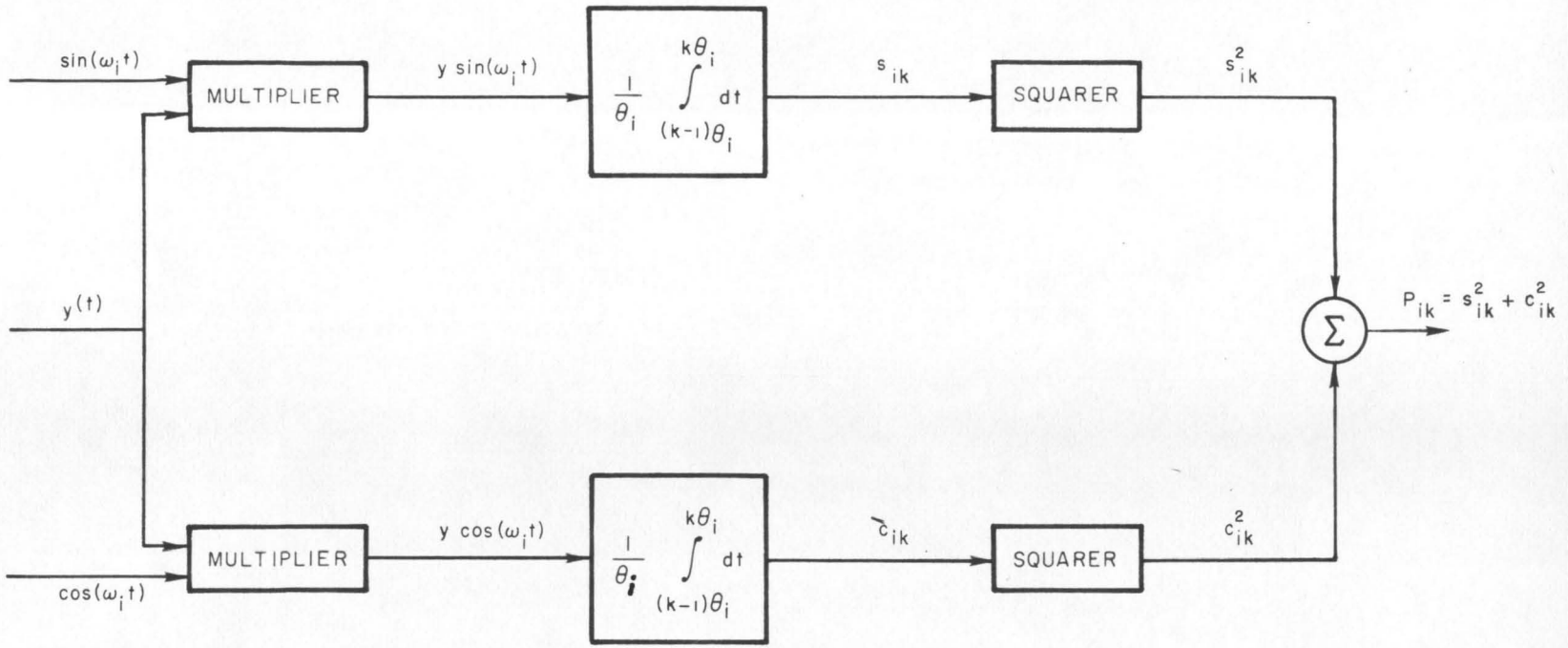
other hand, in order to insure that the system does not destroy itself due to resonant oscillations one must determine the frequencies in the minimum possible time.

The key to rapid determination of the resonant frequencies is parallel operation of many tuned elements. The principle of the method developed in this investigation is shown in Figure 3.3. The signal y(t) is multiplied by a set of sinusoids which spans the frequency range in which the lightly damped resonances are known to occur. Both sines and cosines are necessary to allow for the random phase of the sine wave which may be present in y(t). The process of multiplication produces sum and difference frequency components in the resulting signals. These signals are now averaged by integrations in separate channels for finite lengths of time Θ_i. The integrators have appreciable outputs only when y(t) contains a discrete frequency sinusoidal signal whose frequency is very close to the frequency of the multiplying sinusoid. The output of each integrator is measured every Θ_i seconds. The integrator outputs are then squared and summed in pairs to yield discrete outputs P_{ik} as shown in Figure 3.3, where:

$$P_{ik} = \text{output of } i^{th} \text{ channel during time interval } k\Theta_i \leq t < (k+1)\Theta_i \tag{3.2}$$

P_{ik} will be used to determine whether or not an oscillation of frequency close to ω_i is present at time $t = k\Theta_i$. The precise meaning of "close to" is made explicit in the analysis performed later in this section.

A modification of this technique can be used to simplify the necessary computations. The signal y(t) can be multiplied by two sets



$$s_{ik} = \frac{1}{\theta_i} \int_{(k-1)\theta_i}^{k\theta_i} y \sin(\omega_i t)\, dt$$

$$c_{ik} = \frac{1}{\theta_i} \int_{(k-1)\theta_i}^{k\theta_i} y \cos(\omega_i t)\, dt$$

Figure 3.3: Schematic of a single frequency measuring channel.

of square waves which are 90° out of phase instead of sinusoids 90° out of phase. By this change the relatively complex and time consuming multiplication operation can be replaced by a simple gating operation. Since even harmonics are absent in the Fourier development of a square wave, and the third harmonic is smaller than the fundamental by a factor of three (higher harmonics being even less significant), this modification of the frequency measuring technique should not lead to appreciable error. Care must be taken in the design, however, to insure that a high frequency oscillation does not lead to an appreciable output in a lower frequency (1/3 of the high frequency, for example) channel.

The operation of the FMC is best analyzed by examining the output of a single channel for an assumed input y(t). The only difference between channels is the frequency of the multiplying sinusoidal or square wave signal. This frequency, ω_i, will be called the center frequency of the channel. The analysis of the square wave FMC requires a Fourier expansion of the square wave and a similar analysis for each Fourier component. Since this involves only repeating analyses of the type required for the sinusoidal FMC only the latter will be treated in detail below.

There are several cases of interest. First, and most simple, is the case when y(t) is a pure sine wave of unknown frequency ω and unknown phase ϕ'. Second is the case when y(t) is the sum of two signals, one of which is a pure sine wave of unknown frequency and phase; the second signal is a random noise signal which is uncorrelated with the sine wave and which has specified statistical

characteristics, i.e.:

$$y(t) = s(t) + n(t)$$

where

$$s(t) = A \sin (\omega t + \phi')$$

$n(t)$ = random noise signal

After the analyses of these more or less tractable cases, a few words will be devoted to what one might expect the signal $y(t)$ to be like in actual system operation and how accurately frequency measurements can be performed for this case.

P_{ik} may be expressed analytically in several ways. The fundamental definition of P_{ik} is given by (see Figure 3.3):

$$P_{ik} = s_{ik}^2 + c_{ik}^2 \tag{3.3}$$

where

$$s_{ik} = (1/\Theta_i) \int_{(k-1)\Theta_i}^{k\Theta_i} dt_1 \, y(t_1) \sin \omega_i \left[t_1 - (k-1)\, \Theta_i \right] \tag{3.4}$$

and

$$c_{ik} = (1/\Theta_i) \int_{(k-1)\Theta_i}^{k\Theta_i} dt_1 \, y(t_1) \cos \omega_i \left[t_1 - (k-1)\Theta_i \right] \tag{3.5}$$

In what follows we shall require that Θ_i be an integral number of periods of the frequency ω_i. Therefore

$$\Theta_i = (2\pi N_i/\omega_i) \tag{3.6}$$

where

N_i = number of cycles of frequency ω_i contained in the interval Θ_i

In the remainder of this chapter we shall drop the subscript from the quantity N_i and call it simply N. It should be remembered, however, that N is not necessarily the same for all channels.

Making the change of variables $t = \omega_i[t_1 - (k-1)\theta_i]$ and using Equation 3.6 in Equations 3.4 and 3.5 results in

$$s_{ik} = (1/2\pi N) \int_0^{2\pi N} dt \sin t \; y\left[\frac{t + 2\pi N(k-1)}{\omega_i}\right] \tag{3.7}$$

$$c_{ik} = (1/2\pi N) \int_0^{2\pi N} dt \cos t \; y\left[\frac{t + 2\pi N(k-1)}{\omega_i}\right] \tag{3.8}$$

$$P_{ik} = [1/(2\pi N)^2] \int_0^{2\pi N}\int_0^{2\pi N} dx\,du \cos(x-u) \cdot \left\{ y\left[\frac{x + 2\pi N(k-1)}{\omega_i}\right] y\left[\frac{u + 2\pi N(k-1)}{\omega_i}\right]\right\} \tag{3.9}$$

or

$$P_{ik} = [1/(2\pi N)^2]\left\{\left\langle \int_0^{2\pi N} dt \sin t \; y\left[\frac{t + 2\pi N(k-1)}{\omega_i}\right]\right\rangle^2 + \left\langle \int_0^{2\pi N} dt \cos t \; y\left[\frac{t + 2\pi N(k-1)}{\omega_i}\right]\right\rangle^2\right\} \tag{3.10}$$

In spite of the formidable looking equations defining P_{ik}, there is one particular y(t) for which the integrals yield an extremely simple result. If $y(t) = A \sin(\omega_i t + \phi)$ a direct evaluation of Equation 3.10 leads to the result that $P_{ik} = A^2/4$. Thus the channel output for this case is independent of the random phase of y(t).

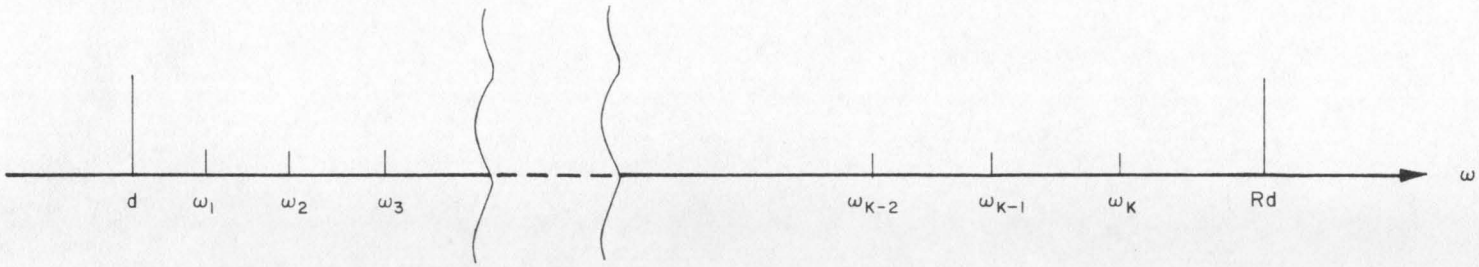
Furthermore, as will be shown later in this chapter, the channel output decreases rapidly as the frequency of the sinusoid present in y(t) departs from ω_i. If a number of channels is used the frequency, ω_m, of the channel which yields the largest output P_{ik} (i.e., $P_{mk} > P_{ik}$ for $i \neq m$) can be used as an estimate of the discrete frequency present in y(t).

The number of frequency measuring channels which must be used is dependent upon the precision with which one wants to measure the signal frequency and upon the amount of a priori information which is available about the signal frequency. Fewer channels are needed, for example, to measure a frequency known to lie within a specified 10 radian band-width than are required to measure the same frequency, with equal accuracy, if it can fall anywhere within a 100 radian interval. It is convenient to define accuracy on a per-unit basis as follows: a frequency estimate ω^* is accurate within a factor ξ (per-unit) if

$$\frac{\omega^* - \omega}{\omega} \leq \xi \qquad (3.11)$$

where ω = the true signal frequency.

We will now compute the minimum number of channels, K, necessary to measure an unknown frequency, ω, to an accuracy $\pm \xi$, assuming that the center frequency of the channel with greatest output, ω_m, is used as the estimate of ω. Thus, $\omega^* = \omega_m$. It is assumed that the inequality $d \leq \omega \leq Rd$ is known to be true (see Figure 3.4). Suppose the center frequencies of the measuring channels are designated by $\omega_i (i = 1, \ldots, K)$ and $\omega_{i+1} > \omega_i$. Then for any frequency ω lying in the interval $\omega_i \leq \omega \leq \omega_{i+1}$ one or the other of the following



$$\omega_1 \leq d(1+\xi)$$

$$\omega_K \geq Rd(1-\xi)$$

$$\omega_i = \left(\frac{1+\xi}{1-\xi}\right)^{i-1} \omega_1$$

Figure 3.4: Placement of frequency measuring channels to span interval $d \leq \omega \leq Rd$ with an accuracy ξ.

relationships must be true if the accuracy requirement is to be satisfied:

$$\frac{\omega - \omega_i}{\omega} \leq \xi \qquad \text{(a)}$$

or (3.12)

$$\frac{\omega_{i+1} - \omega}{\omega} \leq \xi \qquad \text{(b)}$$

The critical case occurs when the equality holds in Equations 3.12 a and b, and then we have, after eliminating ω from the two equations:

$$\frac{\omega_{i+1}}{\omega_i} = \frac{1 + \xi}{1 - \xi} \qquad (3.13)$$

For small ξ this becomes

$$\frac{\omega_{i+1}}{\omega_i} = 1 + 2\xi \qquad (3.14)$$

If the ratio ω_{i+1}/ω_i is any greater than this it will not be possible to satisfy the accuracy requirement for every ω in the interval $\omega_i \leq \omega \leq \omega_{i+1}$. Figure 3.4 shows how the frequency measuring channels will be spaced in the interval $d \leq \omega \leq Rd$. In order to insure an accuracy of ξ for $\omega = d$, it is necessary to select ω_1 as follows:

$$\omega_1 \leq d(1 + \xi) \qquad (3.15)$$

In order to insure an accuracy ξ for $\omega = Rd$ it is necessary to select ω_K as follows:

$$\omega_K \geq Rd(1 - \xi) \qquad (3.16)$$

In order to minimize K it is necessary to use the equality signs in each of the above relations. When the equality signs are used it follows from Equations 3.15 and 3.16 that

$$\frac{\omega_K}{\omega_1} = \frac{R(1 - \xi)}{(1 + \xi)} \tag{3.17}$$

But it follows from Equation 3.13 that

$$\frac{\omega_K}{\omega_1} = \left(\frac{1 + \xi}{1 - \xi}\right)^{K-1} \tag{3.18}$$

Equating the right hand sides of relations 3.17 and 3.18 leads to the following:

$$R = \left(\frac{1 + \xi}{1 - \xi}\right)^{K} \tag{3.19}$$

or

$$\ln R = K \ln \frac{1 + \xi}{1 - \xi} \tag{3.20}$$

For small ξ ($\xi < 0.1$) we may replace $\ln[(1 + \xi)/(1 - \xi)]$ by

$$\ln[(1 + \xi)/(1 - \xi)] = 2\xi + O(\xi^3) \cong 2\xi$$

and retain 3 significant figure accuracy. Thus Equation 3.20 becomes

$$\ln R \cong 2K\xi \tag{3.21}$$

Since K must be an integer, not all values of R and ξ are allowable. A simple example illustrates how Equation 3.21 may be used. Suppose it is specified that a 2:1 range of frequencies must be spanned by the frequency measuring channels and that an accuracy of at least 2% is

required. Thus R = 2 and $\xi = 0.02$. Solving Equation 3.21 for K gives

$$K = (1/2\xi) \ln R = (1/0.04) \ln 2 = (0.693/0.04) = 17.3$$

Since only integer values of K are allowable, it will be necessary to choose K = 18 to get the desired accuracy over the range specified. The final accuracy will be

$$\xi = (1/2K) \ln R = (0.693/36) = 1.925\%$$

If K = 17 were used, the final accuracy would have been

$$\xi = (0.693/34) = 2.038\%$$

It is possible to obtain this accuracy using fewer measuring channels if a more sophisticated technique is used to estimate the frequency which is present. This improved method is based upon an interpolation procedure which allows the use of frequencies other than the ω_i as estimates of the unknown frequency ω. Since the improvement obtainable is dependent upon the form of the P_{ik} outputs as a function of ω, a discussion of this technique will not be presented until the end of the treatment of Case 1 below.

The form of P_{ik} will now be computed for several specific functions y(t).

Case 1: $y(t) = A \sin(\omega t + \phi)$

It is assumed that the phase of y, ϕ, can take on any value with equal probability. Thus we can define a new random variable ϕ_{ik}, where

$$\phi_{ik} = \phi + 2\pi N(k-1) (\omega/\omega_i) \qquad (3.22)$$

which also takes on any value with equal probability. s_{ik} and c_{ik} are therefore functions of the random variable ϕ_{ik}, and are given by (see Equations 3.7 and 3.8)

$$s_{ik} = (A/2\pi N) \int_0^{2\pi N} dt \sin t \sin [(1 + \eta) t + \phi_{ik}] \tag{3.23}$$

and

$$c_{ik} = (A/2\pi N) \int_0^{2\pi N} dt \cos t \sin [(1 + \eta) t + \phi_{ik}] \tag{3.24}$$

where

$$\eta = (\omega/\omega_i) - 1 \geq -1$$

After considerable algebra, the following expression for P_{ik} is obtained (see Equation 3.3):

$$P_{ik} = \frac{A^2}{4} \frac{\sin^2 N\pi\eta}{(N\pi\eta)^2} \left[\frac{1 + 2\eta(1+\eta/2)\, r_{ik}}{(1 + \eta/2)^2} \right] \tag{3.25}$$

where r_{ik} is a random variable defined by

$$r_{ik} = \sin^2 (N\pi\eta + \phi_{ik}) \tag{3.26}$$

and

$$0 \leq r_{ik} \leq 1 \tag{3.27}$$

The bracketed portion of Equation 3.25 is itself a random variable through its functional dependance on r_{ik}. Define

$$F(\eta, r_{ik}) = \frac{1 + 2\eta(1 + \eta/2) r_{ik}}{(1 + \eta/2)^2} \tag{3.28}$$

and

$$P^{o} = P^{o}(N\eta) = \frac{\sin^2 \pi N \eta}{(\pi N \eta)^2} \qquad (3.29)$$

= normalized fundamental output component of any particular frequency measuring channel

Using these definitions it follows that we may write

$$P_{ik} = (A^2/4)\, P^{o}\, F(\eta, r_{ik}) \qquad (3.30)$$

Figure 3.5 is a plot of the normalized fundamental output component, P^{o}, of any particular frequency measuring channel as a function of a normalized frequency error, x, (where $x = N\eta$), between the signal frequency ω and the channel center frequency ω_i.

Note that for a specified value of N the abscissa is directly proportional to the per-unit difference in frequency between ω and ω_i (because $\eta = (\omega - \omega_i)/\omega$). Thus, for example, if N = 10, the output of the channel falls to one half its maximum possible value when $\eta = 0.044$. If N = 5, the output of the channel falls to one-half its maximum possible value when $\eta = 0.088$. In general, P^{o} falls to one-half its maximum possible value for $\eta = 0.44/N$. For the remainder of this section we shall be considering only the channel which produces the largest output for a given sinusoidal input. We can assume that, by proper selection of the interchannel spacing factor ξ and the number, N, of periods over which the integration is carried, there will always be at least one channel for which $|\eta| < 1/2N$. For these channels, which are the ones of interest in the following analysis, it is evident that for sizable N ($N > 5$) it is reasonable to treat η as a small quantity.

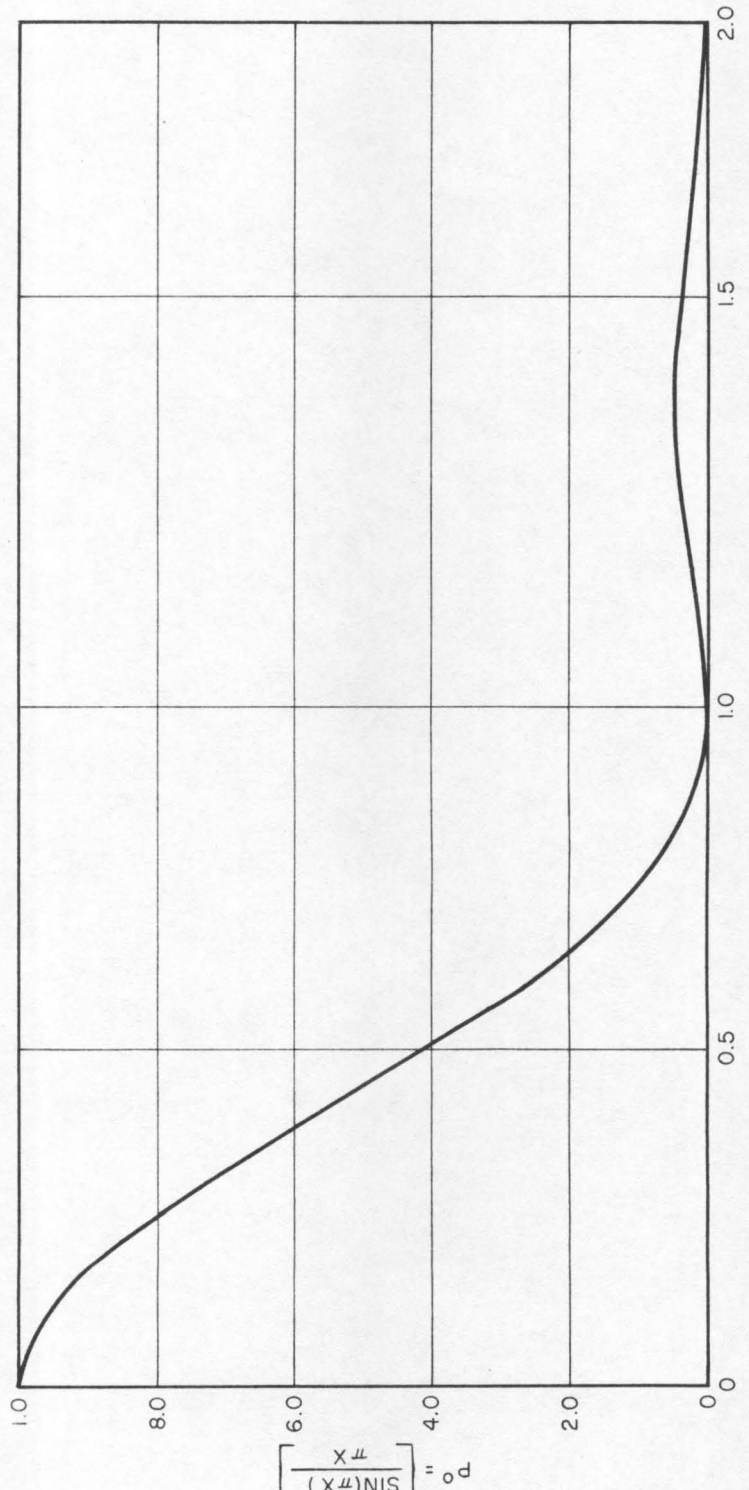


Figure 3.5: Plot of Equation 3.29.

The ordinate in Figure 3.5, i.e., the quantity P^o, is independent of the random phase of the input signal. We should like to consider P^o to be the principle component of the output P_{ik}, of any frequency measuring channel. This is possible if the deviation from P^o due to the random variable r_{ik} is small. To demonstrate that this is indeed the case we shall compute the expected value of P_{ik} and its standard deviation. Since P_{ik} is a random variable only through its dependence on the function $F(\eta, r_{ik})$, we shall first compute the expected value and standard deviation of F. From Equation 3.28 it follows, upon averaging over r_{ik}, that

$$E(F) = 1 + [(\eta/2)/(1 + \eta/2)]^2 = \text{expected value of F} \qquad (3.31)$$

and

$$\sigma_F = \sqrt{2}\,[(\eta/2)/(1 + \eta/2)] = \text{standard deviation of F} \qquad (3.32)$$

For small η these expressions become

$$E(F) = 1 + \eta^2/4$$
$$= 1 \qquad \text{to first order in } \eta \qquad (3.33)$$

and

$$\sigma_F = |\eta|/\sqrt{2} \qquad \text{to first order in } \eta \qquad (3.34)$$

For the channels of interest we have shown that

$$|\eta| < 1/(2N) \qquad (3.35)$$

Therefore we may combine Equation 3.34 and 3.35 to obtain

$$\sigma_F < 1/(2\sqrt{2}\,N) = 0.354/N \qquad (3.36)$$

It is evident that for large N the expected value of F approaches unity very rapidly, and the variance of F approaches zero.

It is possible to obtain expressions for the maximum and minimum values which F may attain by observing that, for a specified value of η, Equation 3.28 is a monotonic function of r_{ik} (monotonic increasing for $\eta > 0$, and monotonic decreasing for $\eta < 0$). Thus F takes on its maximum or minimum value when r_{ik} is at the limit of its range of values. Since, by Equation 3.27, $0 \leq r_{ik} \leq 1$ we conclude that, to first order in η,

$$1 - |\eta| \leq F \leq 1 + |\eta| \tag{3.37}$$

It is now possible to state what the maximum difference is between the true value of P_{ik} and the value $A^2P^o/4$, which is the fundamental component of the output. Using Equation 3.30 we write

$$\begin{aligned} |P_{ik} - A^2P^o/4| &= |(F - 1)| A^2P^o/4 \\ &\leq |\eta| A^2P^o/4 \qquad \text{to first order in } \eta \end{aligned} \tag{3.38}$$

The per unit difference between P_{ik} and the fundamental component of the output is less than $|\eta|$. Recalling the limitation on the possible values of $|\eta|$ we conclude that this per-unit difference is less than 0.5/N. This is a small difference for appreciable values of N.

Since no a priori information is available about the amplitude of the signal, it is not possible to determine the signal frequency from the amplitude of the output of a single channel. However, it is possible to identify the frequency of a signal by operating several frequency measuring channels simultaneously with each channel tuned to a slightly different frequency. One or two of these channels, if

a sinusoidal oscillation is present, will have an output significantly larger than the rest of the outputs. The center frequency of the channel with the largest output is then an estimate of the true frequency of the sine wave.

Two conceptually different functions must be performed by the frequency measuring computer. First, it must determine whether or not a sinusoidal oscillation is present. Second, it must determine the frequency of the oscillation if the oscillation is present. In order to prevent the decision that an oscillation is present when it really is not, a detection threshold level L must be established. If and only if the output P_{ik} of some channel exceeds the threshold L, the measuring computer will decide that an oscillation was present during the time interval $(k-1)\,\Theta_i \lesssim t \lesssim k\,\Theta_i$. The value of L depends upon the response of the measuring computer to random noise inputs (which will be considered later) and upon the amplitude of the smallest oscillation which it is required to detect.

If the oscillation "A sin $(\omega t + \phi)$" occurs at a frequency exactly equal to the center frequency, ω_i, of some channel, the η for that channel will be zero and the corresponding P^o will be unity. The corresponding P_{ik} is $A^2/4$ (see Equation 3.3). It is very unlikely, however, that an oscillation will occur exactly at any channel center frequency.

Suppose the oscillation occurs between ω_i and ω_{i+1} at a frequency $\omega_i(1 + \eta)$. For small values of ξ, two adjacent channels have their center frequencies in the ratio (see Equation 3.14)

$$(\omega_{i+1}/\omega_i) = 1 + 2\,\xi$$

The two channel outputs will be equal when $\eta \cong \xi$, and their magnitudes will be

$$P_{ik} = P_{i+1,\,k} \cong (A^2/4)\, P^o\,(N\xi) \tag{3.39}$$

Since it is necessary to detect the presence of a signal even when its frequency occurs at this worst location, i.e., midway between two channel frequencies, it follows that the minimum input amplitude which can be detected with certainty is

$$A_{min} \cong 2\sqrt{\frac{L}{P^o(N\xi)}} \tag{3.40}$$

Conversely, if it is necessary to detect the presence of oscillations of amplitude A_{min} or greater, the detection threshold level should be

$$L = (A^2_{min}/4)\, P^o(N\xi) \tag{3.41}$$

$P^o(N\xi)$ is simply the ratio of the minimum to the maximum possible output of the channel nearest in frequency to the input frequency (for a constant input amplitude), and it is necessary to specify a numerical value for this quantity. This, in turn, determines the value of $N\xi$ (via either Equation 3.29 or Figure 3.5). Since ξ is determined by the accuracy requirement, we now know N. Thus we have a systematic procedure for selecting all the parameters of the FMC.

A value must now be selected for $P^o(N\xi)$. The system will usually have a requirement to detect all signals above some minimum amplitude. To this end a certain L must be selected. In order to minimize the occurrence of FMC responses to spurious input signals (e.g., random noise which will be discussed later) it is desirable

to set L as high as possible. From Equation 3.41 it is evident that the required L decreases as $P^{o}(N\xi)$ decreases. Therefore it is desirable to keep $P^{o}(N\xi)$ as large as possible. However, if its value is too large it becomes possible for the effects of noise and the random phase of the input sinusoid to cause the output of the wrong channel (i.e., not the channel closest in frequency to the applied frequency) to be a maximum relative to its neighbors. Thus some compromise must be reached. The value of $P^{o}(N\xi) = 1/\sqrt{2}$ seems a convenient compromise. From Figure 3.5 one finds that this requires $N\xi = 0.32$. This is not claimed to be an optimum setting for these parameters but it is a practical one. It is not believed that this setting is extremely critical. The discussion of an interpolation scheme below will shed more light on the effects of changes in this setting.

The conventional quality factor, Q, is defined as $Q = \omega_o/\Delta\omega$, where $\Delta\omega$ is the frequency interval between the lower and upper "half power" points (the frequencies at which the output drops to $1/\sqrt{2}$ of its maximum possible amplitude). But the "half power" frequencies are $\omega_i(1 \pm \eta_{1/2})$, where $\eta_{1/2} = 0.32/N$. Therefore $\Delta\omega = 2\omega_i \eta_{1/2} = 0.64\ \omega_i/N$, and

$$Q = N/0.64 \cong 1.56\ N \qquad (3.42)$$

We will now analyze a simple interpolation technique which leads to a considerable improvement in frequency measuring accuracy. Consider the three frequency measuring channels which lie closest to the signal frequency ω. Call these three channel center frequencies ω_-, ω_o, and ω_+, where for small ξ the defining relations are (see Figure 3.6):

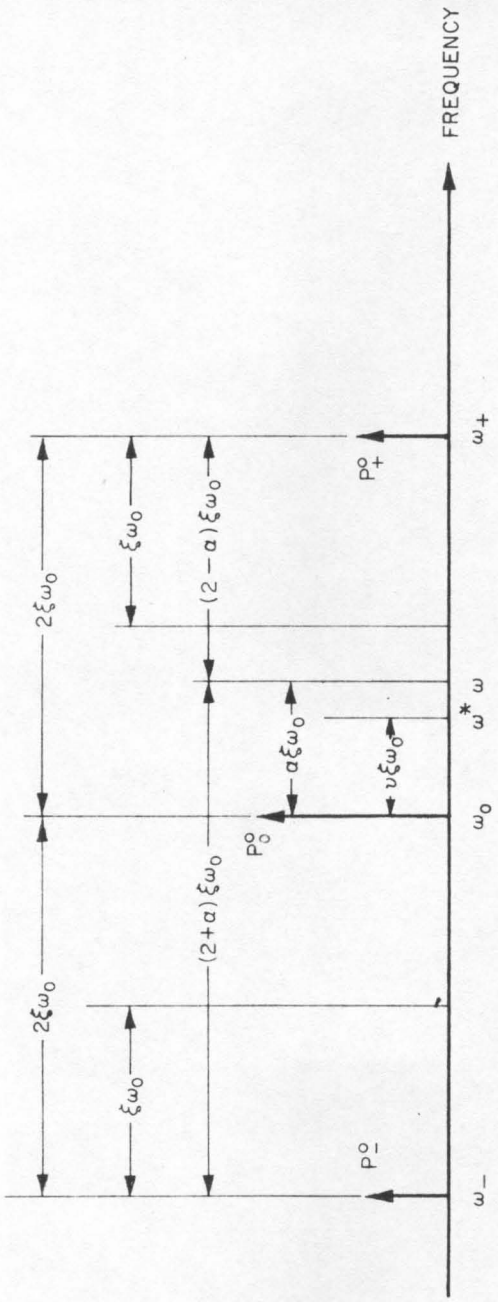


Figure 3.6: Interpolation geometry.

$$\omega_- = \omega_o (1 - 2\xi) \qquad \text{(a)}$$

$$\omega_+ = \omega_o (1 + 2\xi) \qquad \text{(b)} \qquad (3.43)$$

and

$$\omega = \omega_o (1 + \alpha \xi) \qquad \text{(c)}$$

where

$$|\alpha| \leq 1.$$

Define P_-^o, P_o^o, and P_+^o to be the fundamental components of the outputs associated with ω_-, ω_o, and ω_+, respectively.

Define $$\gamma = N \xi \qquad (3.44)$$

Then

$$P_-^o = \frac{\sin^2 \pi\gamma(2 + \alpha)}{[\pi\gamma(2 + \alpha)]^2} \qquad \text{(a)}$$

$$P_o^o = \frac{\sin^2 \pi\gamma\alpha}{(\pi\gamma\alpha)^2} \qquad \text{(b)} \qquad (3.45)$$

$$P_+^o = \frac{\sin^2 \pi\gamma (2-\alpha)}{[\pi\gamma(2-\alpha)]^2}$$

The FMC will supply an estimate, ω^*, of the true input frequency, where

$$\omega^* = \omega_o (1 + \nu \xi) \qquad (3.46)$$

and

$$\nu = \frac{P_+^o - P_-^o}{2P_o^o - P_+^o - P_-^o} = \nu(\gamma, \alpha) \qquad (3.47)$$

This method of estimation is called "parabolic interpolation." In effect a parabola is passed through the three ordinates P_-^o, P_o^o, and P_+^o, and the frequency at which this parabola has its maximum value is ω^*. The resulting accuracy is, by definition

$$A^* = \frac{\omega - \omega^*}{\omega}$$

$$= \frac{(\alpha - \nu)\,\xi}{1 + \alpha\,\xi} \qquad (3.48)$$

$$\simeq (\alpha - \nu)\,\xi$$

If no interpolation had been performed, the frequency estimate would have been ω_o and the accuracy

$$A' = \frac{\omega - \omega_o}{\omega}$$

$$= \frac{\alpha\,\xi}{1 + \alpha\,\xi} \qquad (3.49)$$

$$\simeq \alpha\,\xi$$

The ratio of the accuracy with interpolation to the accuracy with no interpolation is

$$\frac{A^*}{A'} = \frac{\alpha - \nu}{\alpha} \qquad (3.50)$$

Figures 3.7 and 3.8 display Equations 3.48 and 3.50 in graphical form. The abscissa, α, in each of these figures is directly proportional to the input frequency (Equation 3.43c). The ordinate of Figure 3.7 is the accuracy with which the input frequency is determined as a fraction of the channel spacing ξ. For example, if we choose $\gamma = 0.3$, we find that the poorest accuracy will be achieved when $\alpha \simeq 0.65$. This accuracy will be $A^* \simeq 0.206\,\xi$. Thus if ξ had been 0.05 (implying $N = \gamma/\xi = 6$), $A^* \lesssim 0.206\,\xi \simeq 0.01$. Thus the frequency of the input signal would be determined to better than 1%. Without interpolation

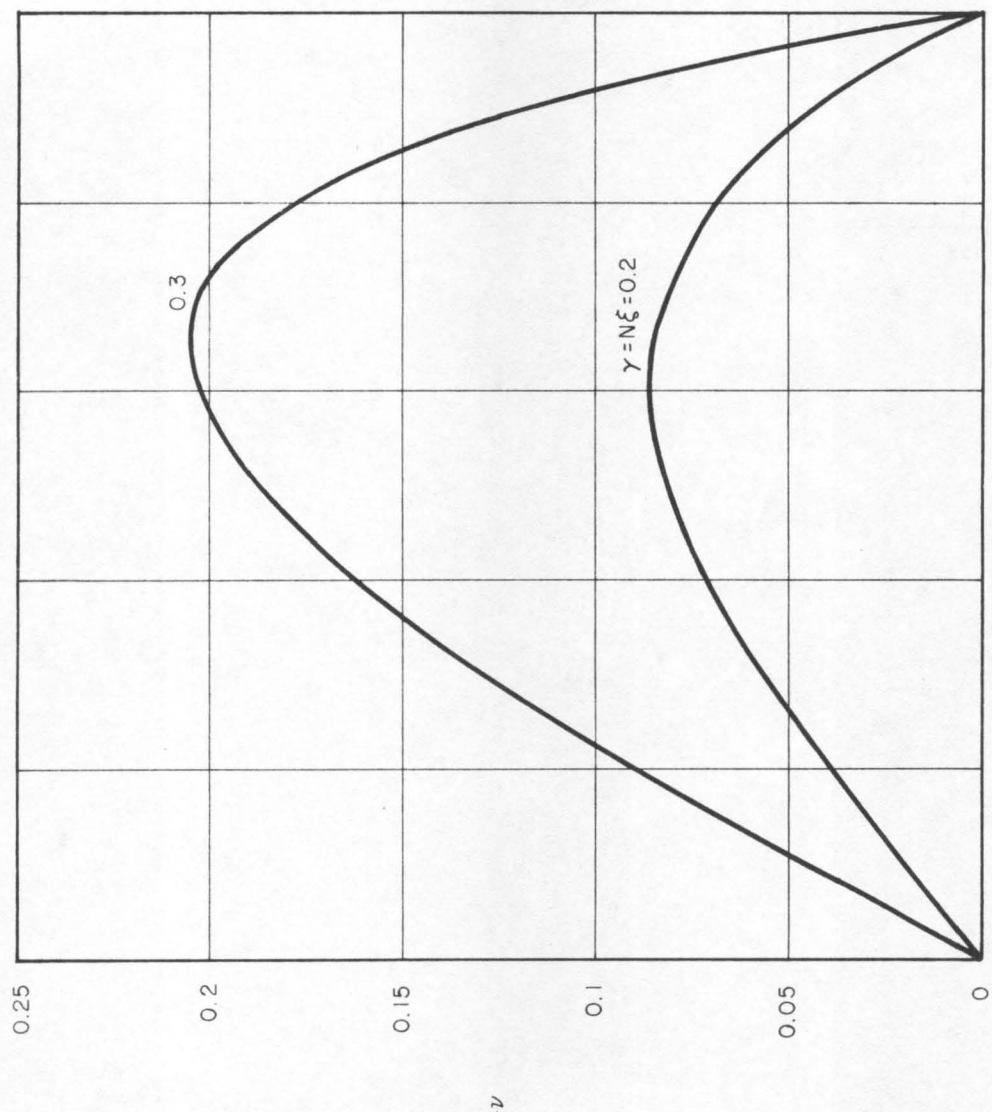


Figure 3.7: Normalized accuracy obtainable using parabolic interpolation.

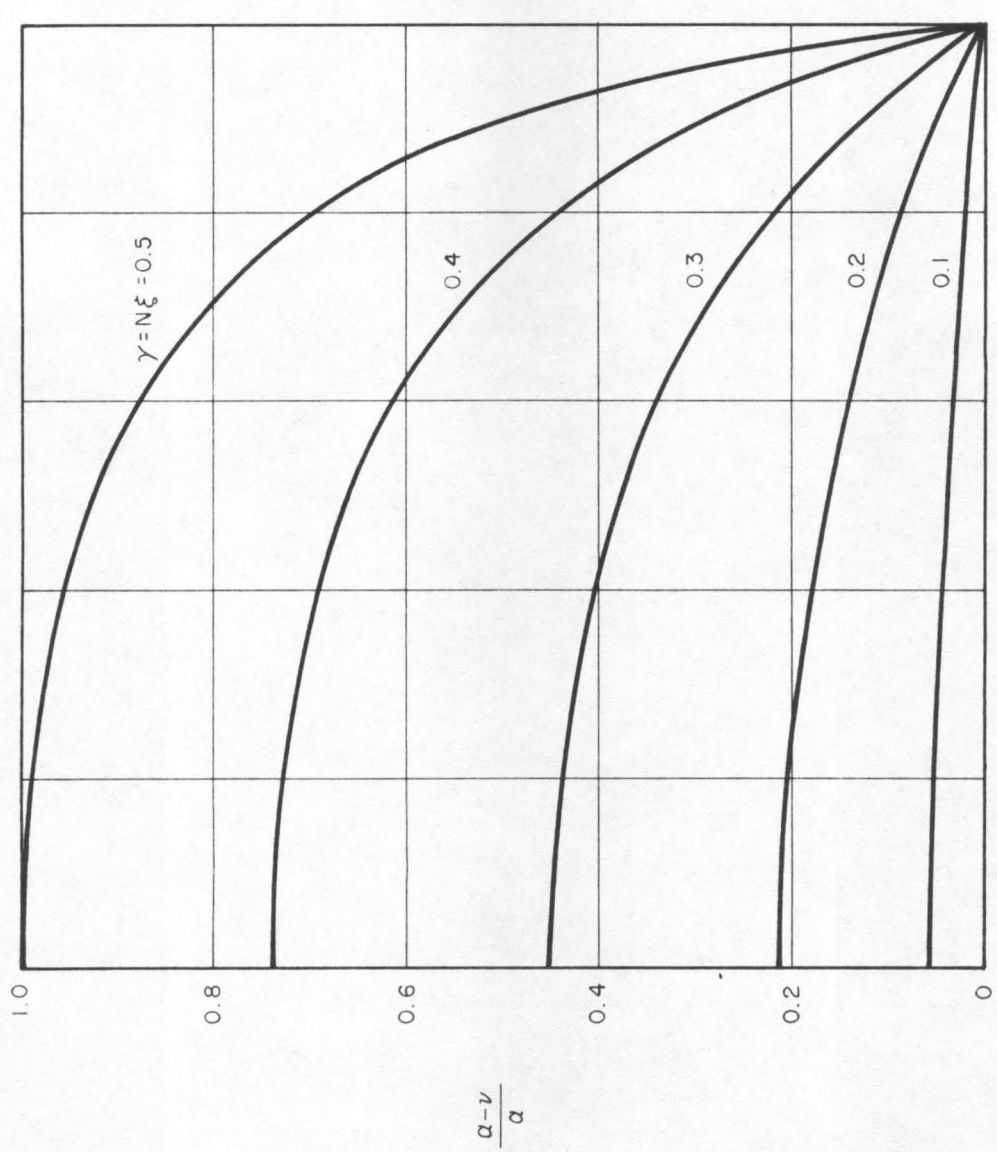


Figure 3.8: Plot of equation 3.50.

it is possible to have an accuracy as poor as ξ. For the example above $\xi = 0.05$, so the input signal is determined to better than 5%. The ordinate in Figure 3.8 is equal to the ratio of the accuracy with interpolation to the accuracy without interpolation.

Figures 3.7 and 3.8 are based upon calculations made with P^o alone. In reality the output of any frequency tracker channel will not be P^o, but will be P_{ik}. A discussion of how much improvement in accuracy interpolation provides in the more realistic case will be presented at the close of the following section, which describes the response of the FMC when "noise" is present.

It should be observed in Figure 3.7 that the final accuracy provided by the interpolation technique is strongly dependent upon the value of the quantity $N\xi$. This parabolic fitting technique would yield precisely the correct answer (under the assumption that the fundamental component is the complete output of any channel) if $P^o(N\xi)$ were exactly a parabolic function of its argument. The range over which P^o must be parabolic for the solution to be exact in form extends from $-3N\xi \leq N\eta \leq 3N\xi$. For small values of $N\xi$, P^o is quite close to parabolic in shape over this range. For larger values of $N\xi$ the precision of the parabolic approximation deteriorates and, therefore, so does the accuracy obtainable from this method of interpolation. In order to have an interpolation result which provides an accuracy five times better than the inter-channel spacing factor, ξ, it is necessary to select a value for $N\xi$ of approximately 0.3. This leads roughly to the value $P^o(N\xi) = 1/\sqrt{2}$ and agrees with our earlier selection of this parameter.

Case 2: $y(t) = n'(t) + A \sin(\omega t + \phi)$

The question of how accurately the frequency of a pure sinusoid can be measured becomes more complex when the signal is obscured by the presence of noise. Following the nomenclature of detection theory we shall call y(t) the "stimulus" of the FMC. Each of the outputs, P_{ik}, of the FMC will be called a test "statistic." The detection problem is that of determining whether the "signal," $A \sin(\omega t + \phi)$, is or is not present during a certain finite length sample of the stimulus. Each of the FMC channels is looking for a signal close to its own center frequency, ω_i. If the test statistic is a good one, only the channels close in frequency to ω will produce sizable outputs, and the channel having the greatest output should be closest in frequency to ω.

Before it is possible to approach this problem analytically, it is necessary to know something of the statistical characteristics of the noise, n'(t). (It is assumed that the reader is familiar with the theory of random processes and the concepts of correlation functions, spectral densities, stationary processes, etc. Definitions of the terms used in the following analysis agree with those of Reference 17. For further background material, see References 17 and 18.) The problem of characterizing the noise in y(t) is formidable for several reasons. A detailed study of all the possible noise sources which contribute to n'(t) would be an ambitious project. In a missile, the primary sources of random noise are the winds and gusts through which the missile flies during its trajectory in the atmosphere. Since missile parameters and environmental conditions all change with time,

it is evident that n'(t) does not have statistical characteristics which are independent of time. Furthermore, different missiles will have different statistics at n'(t) because their transfer functions between the input source and the stimulus are significantly different. Even if an accurate statistical characterization of n'(t) were obtainable, the computations involved in its application would be prohibitive. Therefore, rather than attempt to determine the statistical characteristics of n'(t) from the basic underlying random processes, we shall assume that certain properties of the noise are known a priori. Before stating what these properties are it is necessary to justify one simplifying approximation which will be made.

It should be recalled from the previous paragraph that n'(t) is not stationary noise (i.e., noise whose statistical properties are independent of time). However, in computing the effect of noise on any particular frequency measuring channel one only considers a short sample of the noise function, specifically a finite length sample whose duration is equal to the integration time Θ_i pertinent to the channel under consideration. In order for the overall adaptive system to function properly it is necessary that the system change only slightly during this finite time interval. Thus, over the length of any single sample of n'(t) which is used in the FMC it is safe to say that the statistical properties of the noise do not change, i.e., that it is stationary. This stationary short sample is now assumed to have come from a stationary random process with the same statistical characteristics. We now state the information which we assume to know about the random noise:

1. $n'(t)$ is a sample function of a stationary process
2. $n'(t)$ and ϕ are statistically independent random variables
3. The autocorrelation function $R'(t_1, t_2)$ is known, where

$$R'(t_1,t_2) = E[n'(t_1)\, n'(t_2)] \tag{3.51}$$

= the mean value of $n'(t_1)\, n'(t_2)$ averaged over the ensemble of sample functions $n'(t)$

The autocorrelation function is one of the simplest statistical characterizations of a random signal. For a stationary process, knowledge of the autocorrelation function implies knowledge only of the way in which the signal energy is distributed as a function of frequency. The autocorrelation function will arise naturally in the mathematical development of the problem and, in order to get numerical results it will be necessary to assume a mathematical form for it.

Assumption 1 implies that the autocorrelation function is a function only of the difference between t_1 and t_2, i.e.,

$$R'(t_1,t_2) = R'(t_1 - t_2) \tag{3.52}$$

Defining $\tau = t_1 - t_2$

and $t = t_1$

we have

$$R'(\tau) = E[n'(t)\, n'(t + \tau)] \tag{3.53}$$

which is the fundamental definition of the autocorrelation function of a stationary process.

Assumption 2 is a very reasonable one. It states that the phase of the sinusoidal signal is statistically independent of the noise n'(t). This assumption is stated explicitly because in the analysis which follows the sinusoidal signal is treated as a deterministic signal, i.e., it is completely known. Averages will be taken over the ensemble of noise signals n'(t) treating ϕ as a known quantity. This method is possible only if n'(t) and ϕ are statistically independent.

We must now demonstrate that P_{ik} is a useful test statistic when the stimulus consists of a discrete frequency sinusoid and additive stationary random noise. To accomplish this goal we shall compute the expected value and the variance of P_{ik}. We shall show that the expected value, under reasonable assumptions concerning the input signal-to-noise ratio in y(t), is approximately equal to P^o whenever a discrete frequency is present whose frequency is close to the center frequency of the channel. The output of all other channels will be very much smaller. Furthermore, the variance will be quite small. Thus, it will be shown, P_{ik} is a suitable statistic for this type of input signal.

In what follows, P_{ik} is assumed to be a random variable only through its dependence upon n'(t). In order to make the problem manageable it will be necessary to assume, at an early stage of the analysis, that n'(t) has a Gaussian distribution. Numerical results will then be computed for an assumed form of the autocorrelation function of the noise. It is convenient to define the following functions:

$$S(t) = A \sin \left[(\omega/\omega_i)\, t + \phi + 2\pi N (k-1)\, (\omega/\omega_i) \right] \tag{3.54}$$

$$n(t) = n' \left[\frac{t + 2\pi N(k-1)}{\omega_i} \right] \tag{3.55}$$

Since, for this case,

$$y(t) = n'(t) + A \sin (\omega t + \phi) \tag{3.56}$$

it follows by combining Equations 3.54, 3.55, and 3.56 that

$$y \left[\frac{t + 2\pi N(k-1)}{\omega_i} \right] = n(t) + S(t) \tag{3.57}$$

Thus, using Equation 3.57 in Equation 3.9, it follows that

$$P_{ik} = \left[1/(2\pi N)^2 \right] \int_0^{2\pi N} \int_0^{2\pi N} dx\,du \cos (x-u) \cdot \left[\bar{n}(x) + S(x) \right] \left[\bar{n}(u) + S(u) \right] \tag{3.58}$$

and

$$P_{ik}^2 = \left[1/(2\pi N)^4 \right] \int_0^{2\pi N} \int_0^{2\pi N} \int_0^{2\pi N} \int_0^{2\pi N} dx\,du\,dv\,dz \cos (x-u) \cos (v-z) \cdot \left[\bar{n}(x) + S(x) \right] \left[\bar{n}(u) + S(u) \right] \left[\bar{n}(v) + S(v) \right] \left[\bar{n}(z) + S(z) \right] \tag{3.59}$$

From Equation 3.58, it follows that (recalling that the process of taking an ensemble average commutes with the integral operator):

$$E\left[\bar{P}_{ik} \right] = \left[1/(2\pi N^2) \right] \int_0^{2\pi N} \int_0^{2\pi N} dx\,du \cos (x-u) \cdot \left\{ E\left[\bar{n}(x)\, n(u) \right] + S(x)\, S(u) \right\} \tag{3.60}$$

Terms involving the mean value of the noise, $E[n(x)]$, do not appear in Equation 3.60 because of the limits of integration which have been chosen, i.e., because

$$\int_0^{2\pi N} dx \cos(x - u) = 0 \tag{3.61}$$

The variance of P_{ik}, σ_P^2, is a measure of the spread of the random variable P_{ik} about its mean value. By definition,

$$\sigma_P^2 = E[P_{ik}^2] - [E(P_{ik})]^2 \tag{3.62}$$

The variance is computed in a straightforward manner by taking the ensemble average of both sides of Equation 3.59 and subtracting the square of Equation 3.60 from both sides of the equality. After much algebra and repeated use of Equation 3.61 and symmetry properties of the autocorrelation functions, the following expression for the variance is obtained:

$$\sigma_P^2 = \int_0^{2\pi N}\int_0^{2\pi N}\int_0^{2\pi N}\int_0^{2\pi N} dx\,du\,dv\,dz \cos(x-u)\cos(v-z) \cdot \left\{ E[n(x)\, n(u)\, n(v)\, n(z)] + 4\, S(x)\, S(v)\, E[n(u)\, n(z)] - E[n(x)\, n(u)]\, E[n(v)\, n(z)] \right\} \tag{3.63}$$

The expression is particularly formidable because it contains an ensemble average of the product of four random variables. In general, knowledge of the noise autocorrelation function alone is not sufficient to evaluate expressions of this type. For the case when n(t) is a

Gaussian random variable with zero mean, however, the following relationship holds (see Reference 18, Equation 4.3 - 5):

$$E[\bar{n}(x)\, n(u)\, n(v)\, n(z)] = E[\bar{n}(x)\, n(u)]\, E[\bar{n}(v)\, n(z)] + E[\bar{n}(x)\, n(v)]\, E[\bar{n}(u)\, n(z)] + E[\bar{n}(x)\, n(z)]\, E[\bar{n}(u)\, n(v)] \tag{3.64}$$

It is possible to analyze the case when n(t) does not have zero mean, but the algebra involved is considerably more messy. For the remainder of this analysis it will therefore be assumed that

$$E[\bar{n}(t)] = 0 \tag{3.65}$$

Assuming that n(t) is a Gaussian signal and using Equation 3.64 in Equation 3.63 we find that

$$\sigma_P^2 = [2/(2N\pi)^4] \int_0^{2N\pi}\int_0^{2N\pi}\int_0^{2N\pi}\int_0^{2N\pi} dx\,du\,dv\,dz \cos(x-u)\cos(v-z) \cdot\left\{E[\bar{n}(x)\, n(v)]\, E[\bar{n}(u)\, n(z)] + 2\, S(x)\, S(v)\, E[\bar{n}(u)\, n(z)]\right\} \tag{3.66}$$

Equations 3.60 and 3.66 are the expressions which result when a direct calculation is made of the expected value and variance of P_{ik}. These expressions are still too complicated to yield much information from a qualitative examination. It is convenient now to define several new variables in terms of which the significance of these expressions becomes more apparent. First we define a normalized autocorrelation function $\rho(t_1 - t_2)$ by the relation

$$E[\bar{n}(t_1)\, n(t_2)] = W_N\, \rho(t_1 - t_2) \tag{3.67}$$

where

$$W_N = E[n^2(t)] \tag{3.68}$$

= the average power in the noise

and, therefore

$$\rho(0) = 1 \tag{3.69}$$

It should be recalled that the original noise signal in the stimulus was defined to be $n'(t)$, and the function $n(t)$ was defined by Equation 3.55. By use of Equations 3.53, 3.55 and 3.67 it follows that

$$\rho(\tau) = (1/W_N)\, R'(\tau/\omega_i) \tag{3.70}$$

Now define the following quantities:

$$I_s = [1/(2\pi N)^2] \int_0^{2\pi N} \int_0^{2\pi N} dx\,du \sin x \sin u\, \rho(x-u) \tag{3.71}$$

$$I_c = [1/(2\pi N)^2] \int_0^{2\pi N} \int_0^{2\pi N} dx\,du \cos x \cos u\, \rho(x-u) \tag{3.72}$$

$$\bar{P} = [1/(2\pi N)^2] \left\{ \left[\int_0^{2\pi N} dx \sin x\, S(x) \right]^2 + \left[\int_0^{2\pi N} dx \cos x\, S(x) \right]^2 \right\} \tag{3.73}$$

$= E(P_{ik})$ when $n'(t) = 0$

Note that by the definition of $\bar{P}$ we imply that $\bar{P}$ is equal to the quantity which was called P_{ik} in the preceding section where noise was not considered (see Equation 3.25).

By using the quantities defined above it is possible to reduce

Equations 3.60 and 3.66 to the following forms:

$$E(P_{ik}) = \bar{P} + W_N(I_s + I_c) \tag{3.74}$$

$$\sigma_P^2 = 2\left\{W_N^2(I_s^2 + I_c^2) + \left[2W_N/(2\pi N)\right]\left\langle I_s\left[\int_0^{2\pi N} dx \sin x\, S(x)\right]^2 + I_c\left[\int_0^{2\pi N} dx \cos x\, S(x)\right]^2\right\rangle\right\} \tag{3.75}$$

It should be emphasized that these two expressions are the exact solution for the expected value and variance of the output of any FMC channel when the stimulus is of the form (see Equation 3.56):

$$y(t) = n'(t) + A \sin(\omega t + \phi)$$

provided that $n'(t)$ is a stationary Gaussian random variable with zero mean.

Several special cases will now be discussed. First, observe that when no noise is present the variance becomes zero and the expected value reduces to Equation 3.73 (which is the same as Equation 3.25) as we would expect. For this case the analysis is precisely the one which was discussed in the previous section which dealt with the no-noise case.

Next we examine the output when there is no discrete frequency sinusoid present in the stimulus. For this case application of Equations 3.74 and 3.75 yields

$$E(P_{ik}) = W_N(I_c + I_s) \tag{3.76}$$

$$\sigma_P = W_N \sqrt{2(I_s^2 + I_c^2)} \tag{3.77}$$

Thus the expected value and variance of P_{ik} are both proportional to the average power contained in the noise. The magnitude of the "constant" of proportionality depends upon the form of the noise autocorrelation function through Equations 3.71 and 3.72. For most cases of interest it will be a number considerably smaller than unity.

Finally we consider the case where the stimulus contains both a discrete frequency sinusoid and additive random noise. For this case σ_P is a complicated function to estimate due to the form of its dependence on S(t) (see Equation 3.75). It is possible, however, to obtain an upper bound for σ_P in a relatively convenient form. To this end we define the variable I to be

$$I = \max (I_s, I_c) \tag{3.78}$$

where the definition implies that I equals the larger of the two variables I_s and I_c. Using this definition and Equation 3.75 we find that

$$\sigma_P \leq 2\sqrt{W_N I \bar{P}}\ \sqrt{(1 + W_N I/\bar{P})} \tag{3.79}$$

It follows from Equation 3.74 that

$$E(P_{ik}) \leq \bar{P}(1 + 2W_N I/\bar{P}) \tag{3.80}$$

$\bar{P}$ can be related in a direct manner to the average power in the signal component of the stimulus. The average power in a signal of the form A sin $(\omega t + \phi)$ is just $A^2/2$. But $\bar{P}$ for this same input is $A^2/4$ (see Equation 3.25) if ω is equal to the center frequency of the channel under consideration. If the input frequency is not exactly

at one of the channel center frequencies, the output of the channel nearest this frequency will be somewhat less than $A^2/4$. By selection of the quantity $P^o(N\xi) = 1/\sqrt{2}$ (see page 74) we have insured that, for the channel closest in frequency to ω, the following inequality holds:

$$A^2/4 \gtrsim \bar{P} \gtrsim A^2/(4\sqrt{2})$$

If we now define the input average power by

$$W_S = A^2/2 \tag{3.81}$$

we may write

$$W_S/2 \gtrsim \bar{P} \gtrsim W_S/(2\sqrt{2}) \tag{3.82}$$

It is evident that the poorest output signal to noise ratio will occur when the right hand equality holds above, i.e., when

$$\bar{P} = W_S/2\sqrt{2} \tag{3.83}$$

Finally we define the input signal to noise ratio, D, by

$$D = W_S/W_N \tag{3.84}$$

If we use Equation 3.83 to define the relation between $\bar{P}$ and the input power level and use Equation 3.84 in expressions 3.79 and 3.80 we find that

$$\sigma_P \lesssim \bar{P}\sqrt{8\sqrt{2}\ I/D}\ \sqrt{1 + 2\sqrt{2}\ I/D} \tag{3.85}$$

and

$$E(P_{ik}) \lesssim \bar{P}(1 + 4\sqrt{2}\ I/D) \tag{3.86}$$

Once the autocorrelation function of the noise is specified, I_c and I_s can be evaluated by integration. The algebra, unfortunately,

becomes extremely involved unless the autocorrelation function has a simple form. Numerical methods can always be used and the functions evaluated by digital computation if the need arises. The calculation has been performed analytically for the case of first order low pass noise. By definition, this noise has an autocorrelation function

$$R'(\tau) = W_N e^{-\alpha\omega_i|\tau|} \tag{3.87}$$

The power spectral density of this noise is just the Fourier cosine transform of the autocorrelation function and is given by

$$G(\omega) = \frac{2W_N/\alpha\omega_i}{1 + (\omega/\alpha\omega_i)^2} \tag{3.88}$$

and

$\alpha\omega_i$ = the half-power frequency of the noise

From Equation 3.70

$$\rho(x - u) = e^{-\alpha|x - u|} \tag{3.89}$$

Direct integration of Equations 3.71 and 3.72 gives the results

$$I_s = \frac{\alpha}{2\pi N(\alpha^2 + 1)}\left[1 + \frac{1 - e^{-2\pi N\alpha}}{\pi N\alpha(\alpha^2 + 1)}\right] \tag{3.90}$$

$$I_c = \frac{\alpha}{2\pi N(\alpha^2 + 1)}\left[1 - \alpha^2\,\frac{1 - e^{-2\pi N\alpha}}{\pi N\alpha(\alpha^2 + 1)}\right] \tag{3.91}$$

Figures 3.9 and 3.10 are graphs of the functions I_s and I_c as a function of α for several different values of N. It is evident that, for a given value of α, I_s is always larger than I_c. Therefore, in Equations 3.85 and 3.86 we may set $I = I_s$. Furthermore, for a given

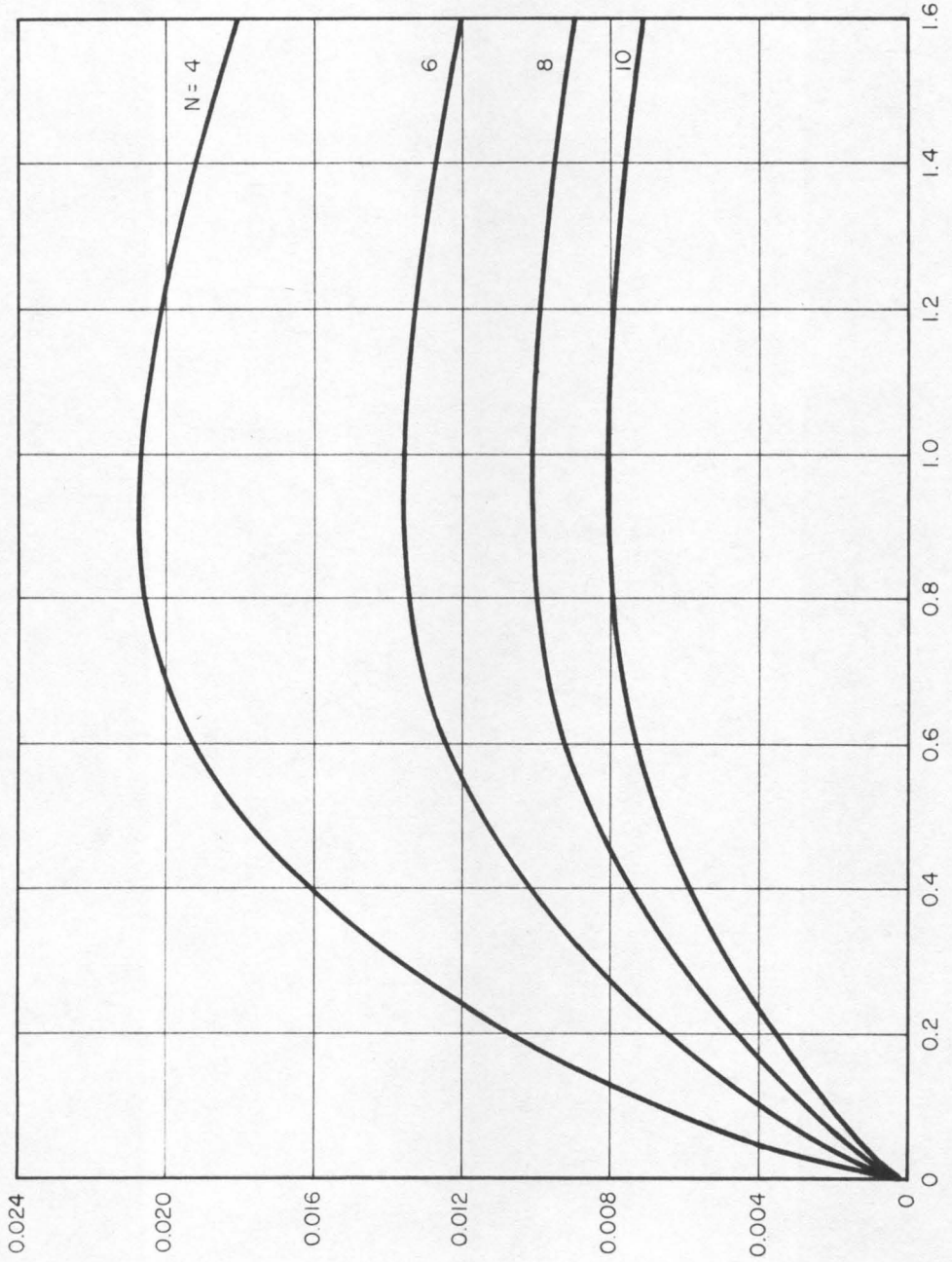


Figure 3.9: Plot of equation 3.90.

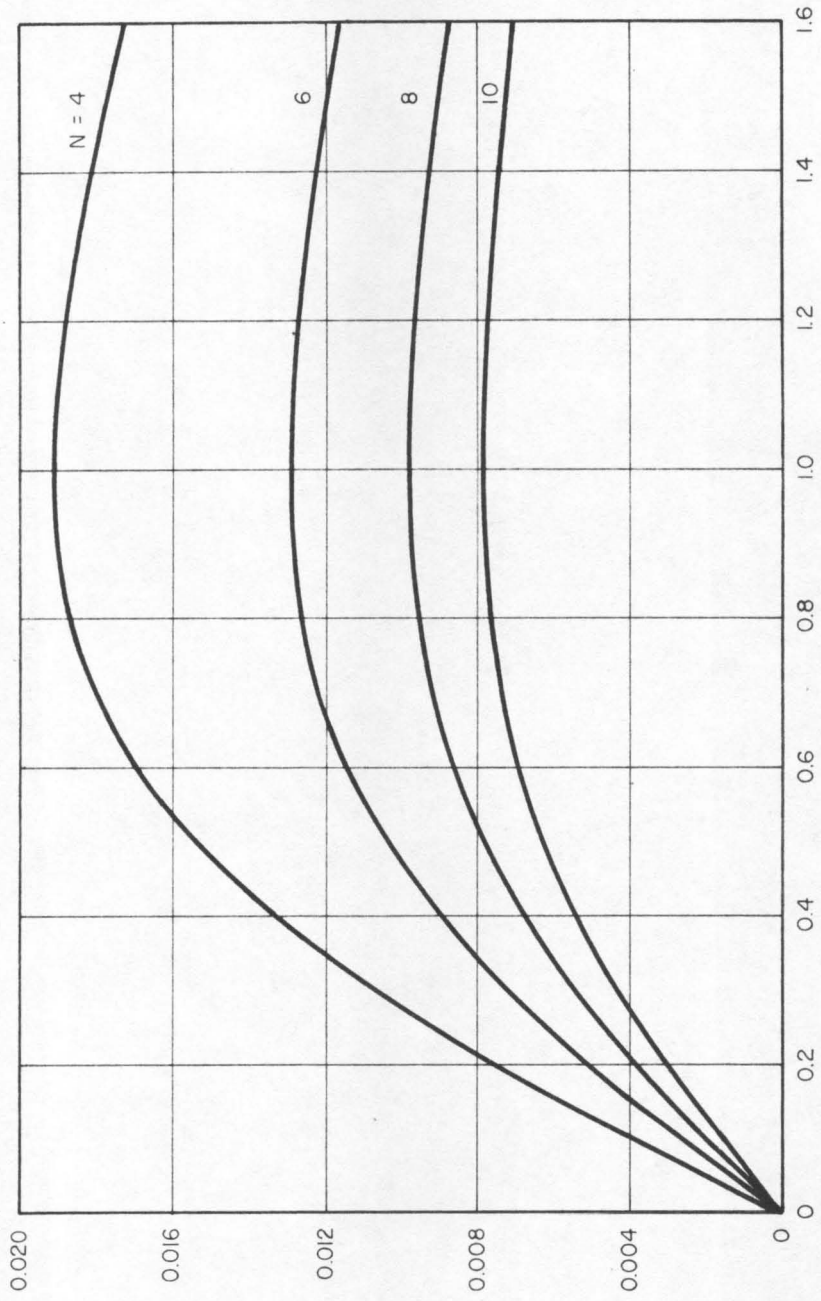


Figure 3.10: Plot of equation 3.91.

value of N these curves attain their maximum values for $\alpha \cong 1$, at which point they both are approximately equal to $1/4\pi N$ (for $N > 2$). Substituting

$$I = 1/4\pi N \tag{3.92}$$

into Equations 3.85 and 3.86 gives

$$\sigma_P = \bar{P}\sqrt{0.9/ND}\ \sqrt{1 + 0.225/ND} \tag{3.93}$$

$$E(P_{ik}) = \bar{P}(1 + 0.45/ND) \tag{3.94}$$

These results apply for the worst possible case of low pass noise, i.e., the case where $\alpha \cong 1$. For any other value of α, the value of I is smaller, which in turn insures that the variance of the output is smaller and the deviation from $\bar{P}$ is smaller. Note that the case $\alpha = 1$ implies that the half-power frequency of the noise is exactly equal to the center frequency of the FMC channel which produces the largest output. For this worst case the requirement $\sigma_P \ll \bar{P}$ implies

$$\sqrt{0.9/ND} \ll 1 \tag{3.95}$$

If we assume that "much less than" should mean at least an order of magnitude, we arrive at the requirement (approximately),

$$D > 90/N \tag{3.96}$$

Since other considerations restrict N to moderate values, roughly ten or less, we require an input signal to noise power ratio of 9 or greater to insure that noise does not affect our FMC output significantly. This result applies for the worst case of low pass noise, i.e., when the bandwidth of the noise is approximately equal to the center

frequency of the FMC channel under consideration.

Further investigation of this problem would require more information about $R'(\tau)$. As was indicated at the beginning of this section, this is a large problem in itself. It is obvious that if the signal to noise ratio in the stimulus is large enough, random additive noise will not cause appreciable errors in frequency measurement. What "large enough" implies must be computed in detail for the type of noise that is present in any particular situation. The results of this section indicate that the problem requires additional consideration. In the missile control system example, it is felt that noise will not be a serious problem because the noise in the stimulus is primarily the missile's transient response to winds, gusts and guidance inputs. These wave forms have a frequency content which usually falls off well below the bending frequencies. Thus α will be considerably less than unity if this noise is approximated as low pass noise. In addition, the amplitude of this noise is expected to be quite small.

Only two types of signals have so far been considered as possible stimuli for the FMC. The first of these, a pure sine wave, is an idealization which rarely, if ever, will be met in practice. Even for this ideal case, however, it was observed that a small amount of noise is present in the FMC output due to the random phase of the sinusoid. The second assumed form for the stimulus consisted of a pure sinusoid and additive stationary noise. An undesirable noise output is produced due to the input noise and its interaction with the signal. Finally, it should be noted that the real stimulus will not, in general, precisely resemble either of these cases. It will

resemble more closely additive noise combined with a sinusoid whose phase, frequency and amplitude are slowly changing with time. Obviously, if the frequency measuring scheme proposed here is to be effective neither phase, frequency, nor amplitude can be allowed to change appreciably during the integration time of any channel intended to measure that frequency.

The rate at which the amplitude of the sine wave grows (or decreases) is related to the damping factor of the sinusoid. Since this rate must be small the damping factor is small and, as a first approximation the growing sinusoidal signal will be

$$S(t) = A(1 + \zeta\omega t)\sin(\omega t + \phi) \qquad \text{for } |\zeta\omega t| \ll 1 \qquad (3.97)$$

where the origin of the time scale is assumed to be exactly at the start of an integration interval. The integration will extend from $t = 0$ to $t = 2\pi N/\omega_i$. At the beginning of this interval, the amplitude of the signal envelope is A, and at the end of the interval the envelope has grown to an amplitude $A(1 + 2\pi N\zeta\omega/\omega_i)$. Since for all channels having a significant output the relation $\omega \cong \omega_i$ must be true, the final envelope amplitude is approximately $A(1 + 2\pi N\zeta)$. For this case the channel output is approximately the same value as would be obtained if the channel were excited by a constant amplitude sinusoid of amplitude $A(1 + N\pi\zeta)$. This amplitude is just equal to the envelope amplitude at the center of the integration interval.

The system with lightly damped resonances which is of concern has already been constrained to involve parameters which change slowly with time. The frequency of the sinusoid in the stimulus changes

at a rate which is very nearly equal to the rate at which the natural frequencies of the system vary. The performance of the FMC will be satisfactory provided that, during any integration interval associated with the measurement of a particular frequency, that frequency changes by an amount smaller than the bandwidth of the channel which is making the measurement.

It will now be assumed that the system to which the adaptive frequency measuring technique is to be applied satisfies the constraints discussed above. The question now arises as to what to do with the information which the FMC is obtaining. This matter is treated in the following section.

3.3 Digital Compensation

Once the critical frequencies are determined, it is necessary to synthesize compensation which will stabilize the system. It was shown in Chapter II that the frequency response curves of systems with lightly damped resonances are characterized by large amplitude peaks at the resonant frequencies. Figure 2.7 showed the type of Nyquist plot which one expects. If some way could be found to reduce the amplitudes of the large bending "loops" in this figure so that their diameters were always less than unity, all of these resonant modes would be "gain stabilized." (See p. 43).

One possible way to reduce the amplitude response at the resonant frequencies is to place zeros of transmission at these frequencies. Since the Frequency Measuring Computer is continually determining the frequencies of all troublesome modes, we know quite accurately what the desired zero locations are. Digital compensation

is the most convenient way to instrument such shaping because of the ease with which poles and zeros may be moved about in the Z-plane. Associated with the pole-zero configuration which is introduced to stabilize the system will be a particular frequency response (magnitude and phase) characteristic. Other than providing attenuation at the frequencies specified by the FMC, the digital shaping should have as little effect as possible. In particular, the compensation selected should have the following properties:

1. It should introduce transmission zeros at the frequencies specified by FMC.
2. It should not introduce appreciable low frequency phase lag.
3. It should have a low frequency gain of unity.
4. It should have as low a gain as possible at high frequencies in order to prevent amplification of higher bending mode transmission.

The objective of the adaptive control system ideally is to insure that all resonant modes of importance are gain stabilized, i.e., the loop transmission at and near all resonant frequencies is less than unity. If this aim is achieved instability cannot occur due to the lightly damped resonances.

Highly complex, high order filters could be designed in an attempt to satisfy all the requirements stated above. Several poles and zeros could be used to compensate for each resonant instability. The purpose of this investigation, however, is to demonstrate the feasibility of a particular concept of an adaptive control system, and we should like to do it as simply as possible. No quantitative attempt will be made at an optimization of the filter configuration

postulated below; a qualitative argument will serve to demonstrate its reasonableness.

The simplest possible way to satisfy requirement 1 above is to introduce conjugate pairs of zeros on the unit circle in the Z-plane. The angular position is chosen in accordance with the following condition (see Figure 3.11):

$$\beta_i = \omega_i^* \, T_s \tag{3.98}$$

where

ω_i^* = i^{th} mode frequency as specified by the FMC (radians per second)

T_s = control sampling period (seconds)

β_i = angular position around unit circle of zeros introduced to compensate for oscillation at i^{th} resonant frequency (radians)

i = subscript denoting which mode is being accounted for by a particular set of zeros ($i = 1, 2, .., n$)

For the compensation to be physically realizable, it is necessary to have at least as many poles as zeros in the Z-plane. As far as requirement 1 is concerned, it does not matter where these poles are located (provided they do not coincide with the zeros). The closer the poles are to the zeros, however, the sharper will be the filter (i.e., the filter will have less and less effect at frequencies far removed from ω_i^* as the pole is moved toward the zero). It is desirable to locate the poles inside the unit circle, for poles outside the unit circle would very likely lead to unstable closed loop roots. The compensation poles must be kept an appreciable distance inside the unit circle or they will lead to a peaked frequency response similar to the response due to the resonances.

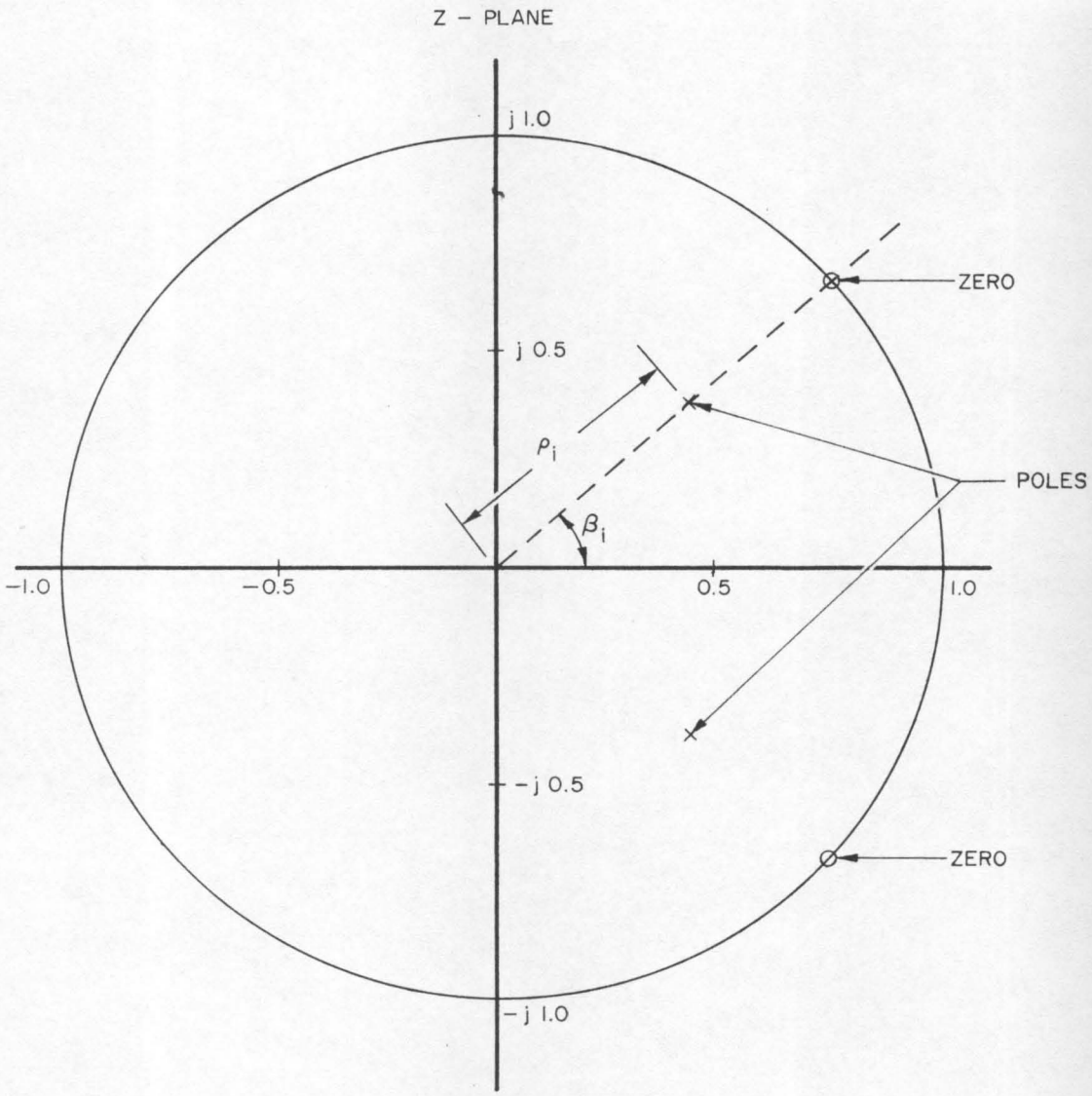


Figure 3.11: Pole-zero configuration used to stabilize the i^{th} resonant frequency.

The configuration selected in this investigation is shown in Figure 3.11. A pole is located on the radius from the origin of the Z-plane out to each zero. Its distance from the origin is ρ_i, where $0 \leq \rho_i \leq 1$. The form of the compensation introduced to stabilize the i^{th} mode is therefore

$$D_i(z) = A_i \frac{z^2 - 2z \cos \beta_i + 1}{z^2 - 2\rho_i z \cos \beta_i + \rho_i^2} \tag{3.99}$$

where A_i is selected to make the zero frequency gain unity. Thus,

$$A_i = \frac{1 + \rho_i^2 - 2\rho_i \cos \beta_i}{2(1 - \cos \beta_i)} \tag{3.100}$$

In order to obtain the frequency response of this transfer function, it is only necessary to make the substitution

$$z = \exp(j \omega T_s) = \exp(j \psi) \tag{3.101}$$

where

$$\psi = \omega T_s \text{ (radians)} \tag{3.102}$$

and

$$\omega = \text{frequency at which response is being calculated (radians per second)} \tag{3.103}$$

It is possible to plot a normalized frequency response using ψ as a normalized frequency variable. This normalization eliminates T_s as a significant parameter, and the resulting curves can be used to obtain results for any specified value of T_s. Note that ψ represents the angle around the unit circle at which a particular frequency, ω, is located. Figures 3.12 a to 3.14 b show the gain and phase characteristics of Equation 3.99 as a function of ψ (measured in degrees) for three different values of β_i (measured in degrees).

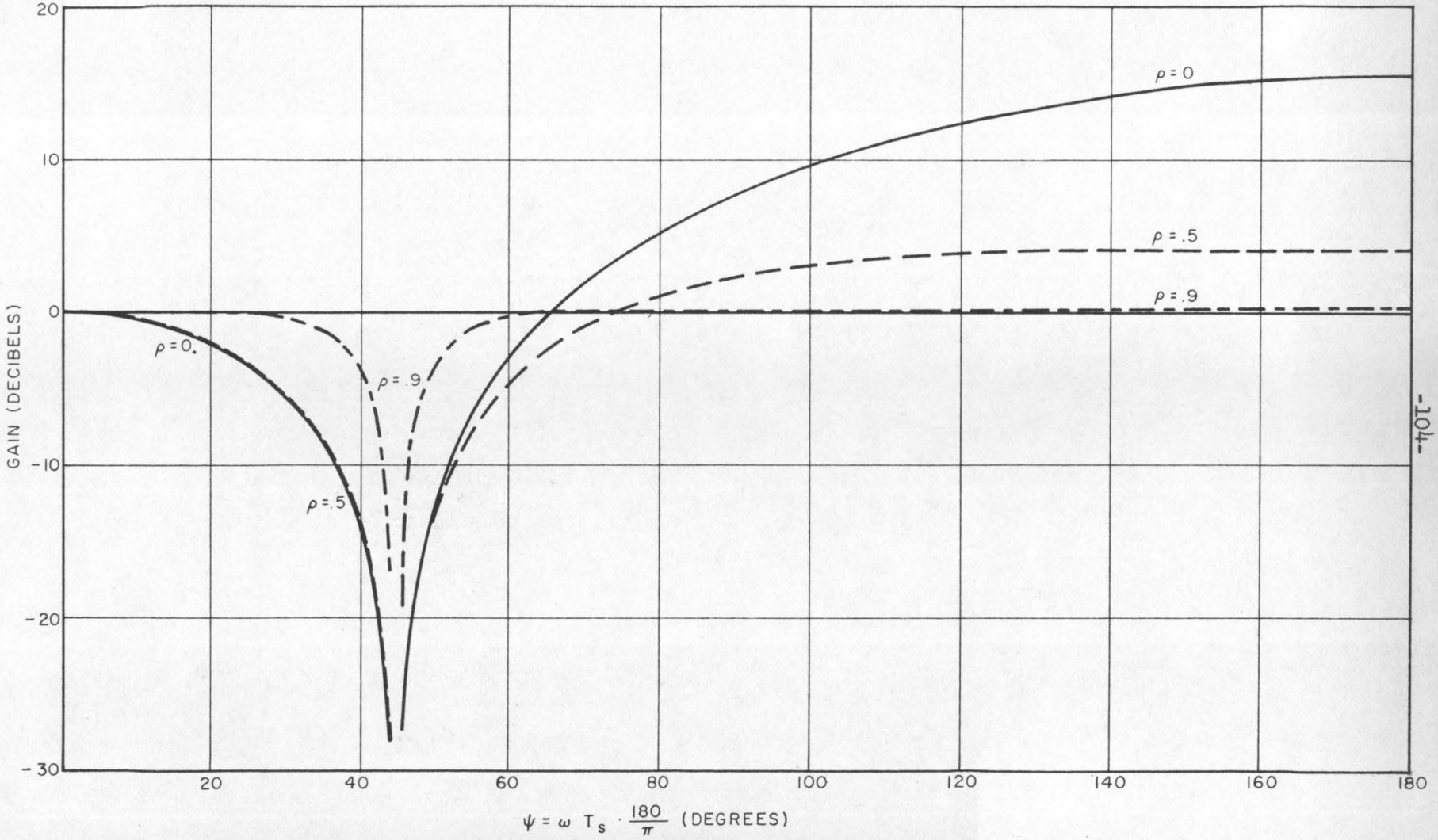


Figure 3.12a: Gain response of sampled compensation for $\beta_1 = 45^\circ$.

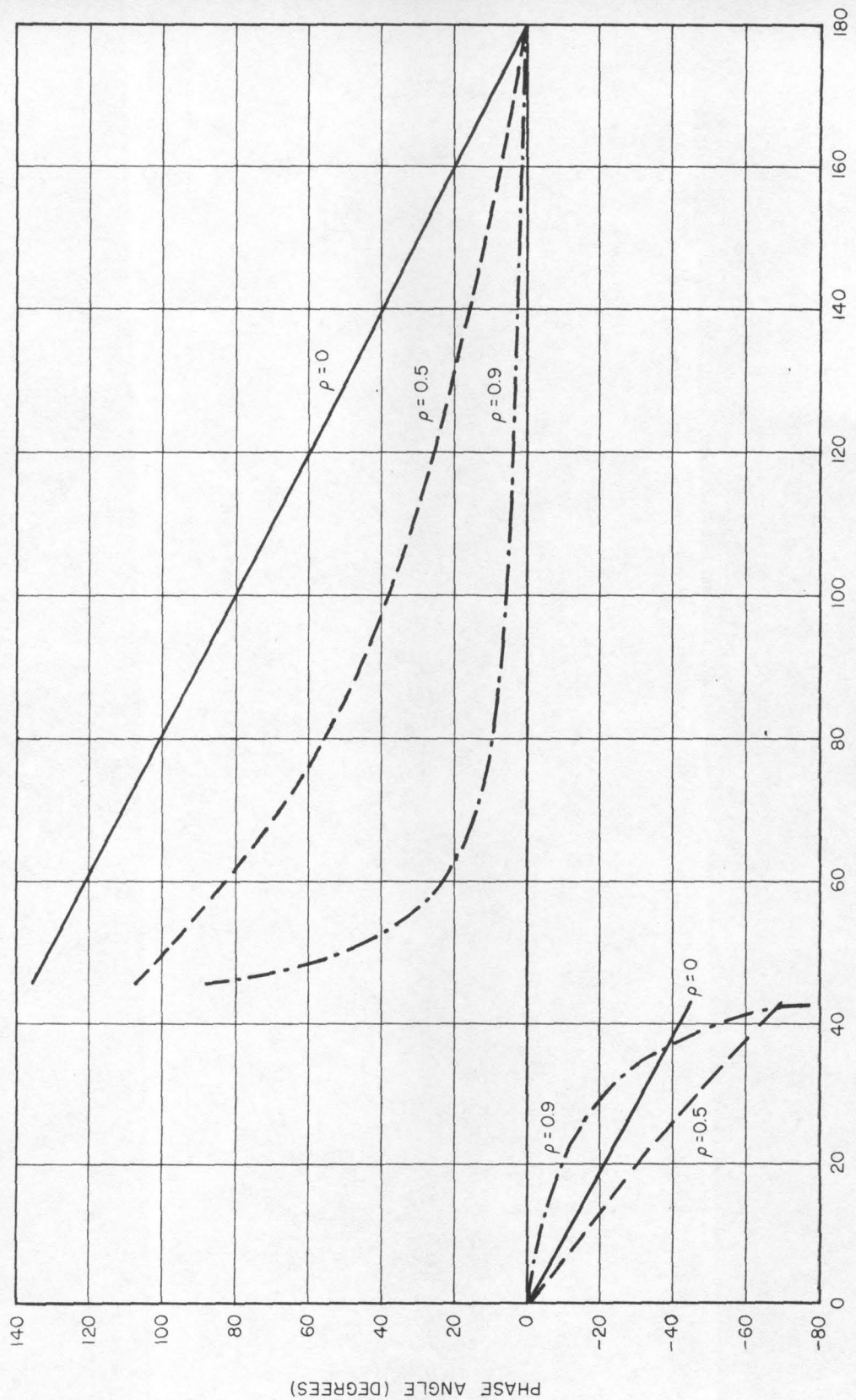


Figure 3.12b: Phase response of sampled compensation for $\beta_1 = 45^0$.

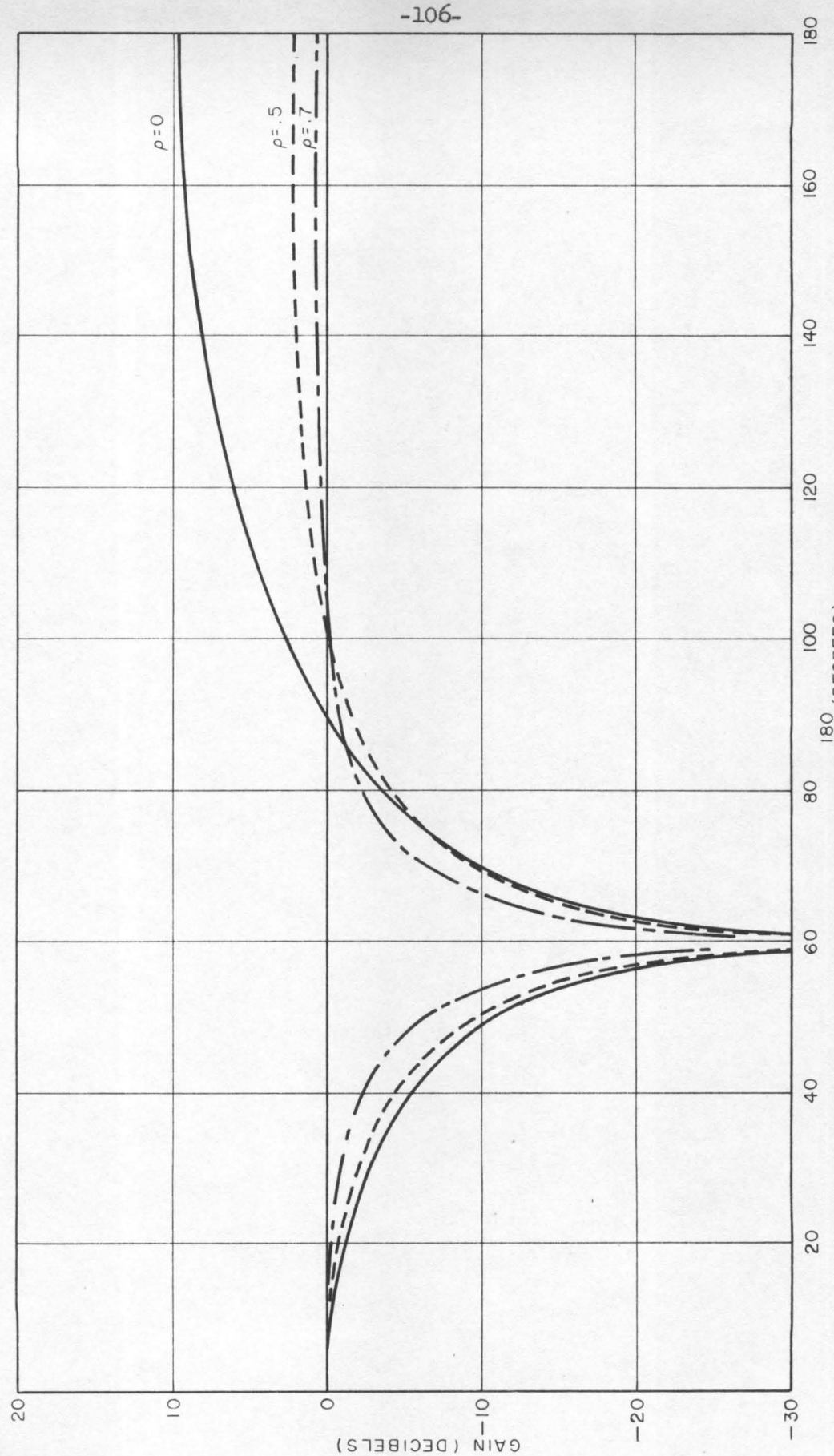


Figure 3.13a: Gain response of sampled compensation for $\beta_1 = 60^\circ$.

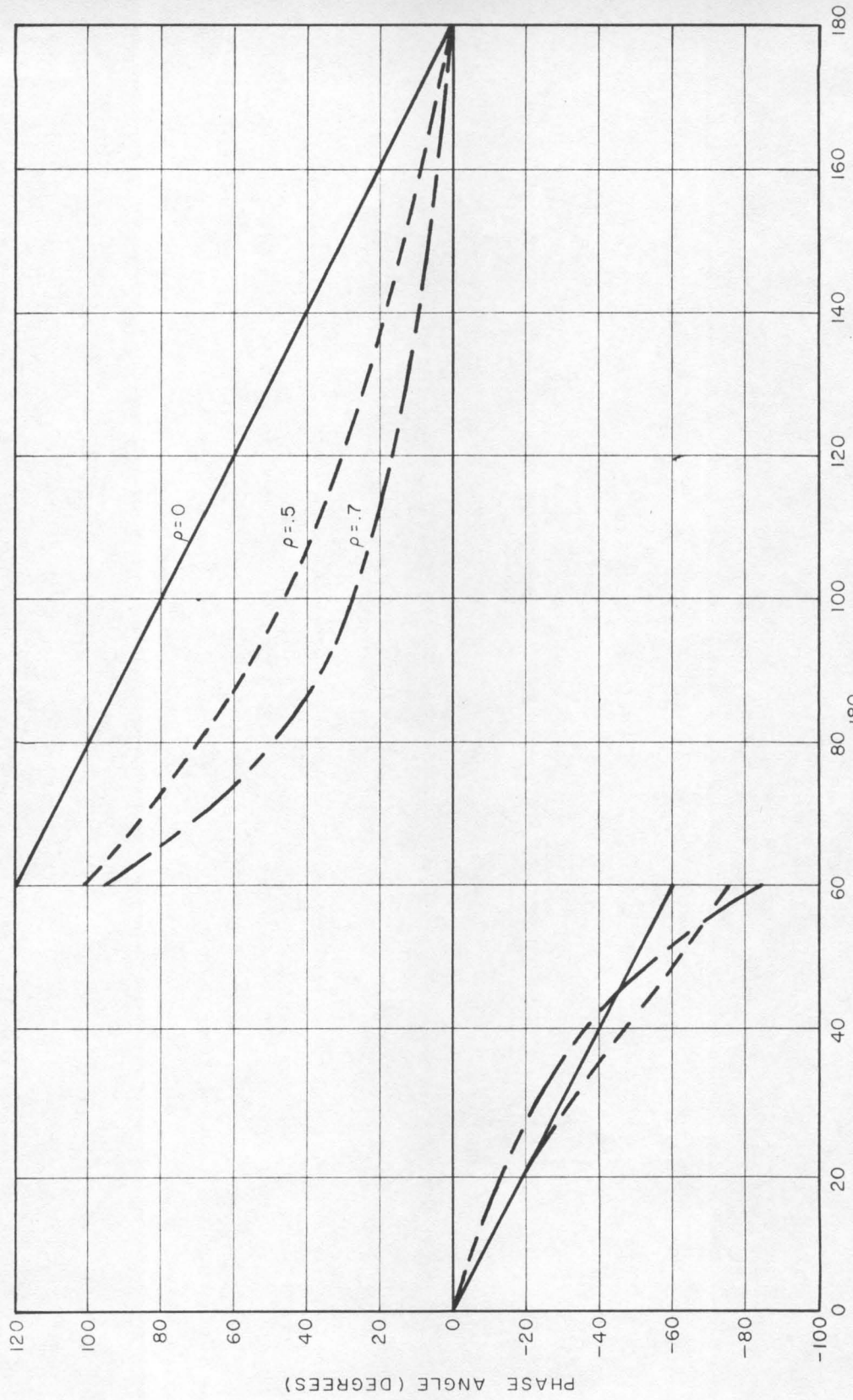


Figure 3.13b: Phase response of sampled compensation for $\beta_1 = 60^\circ$.

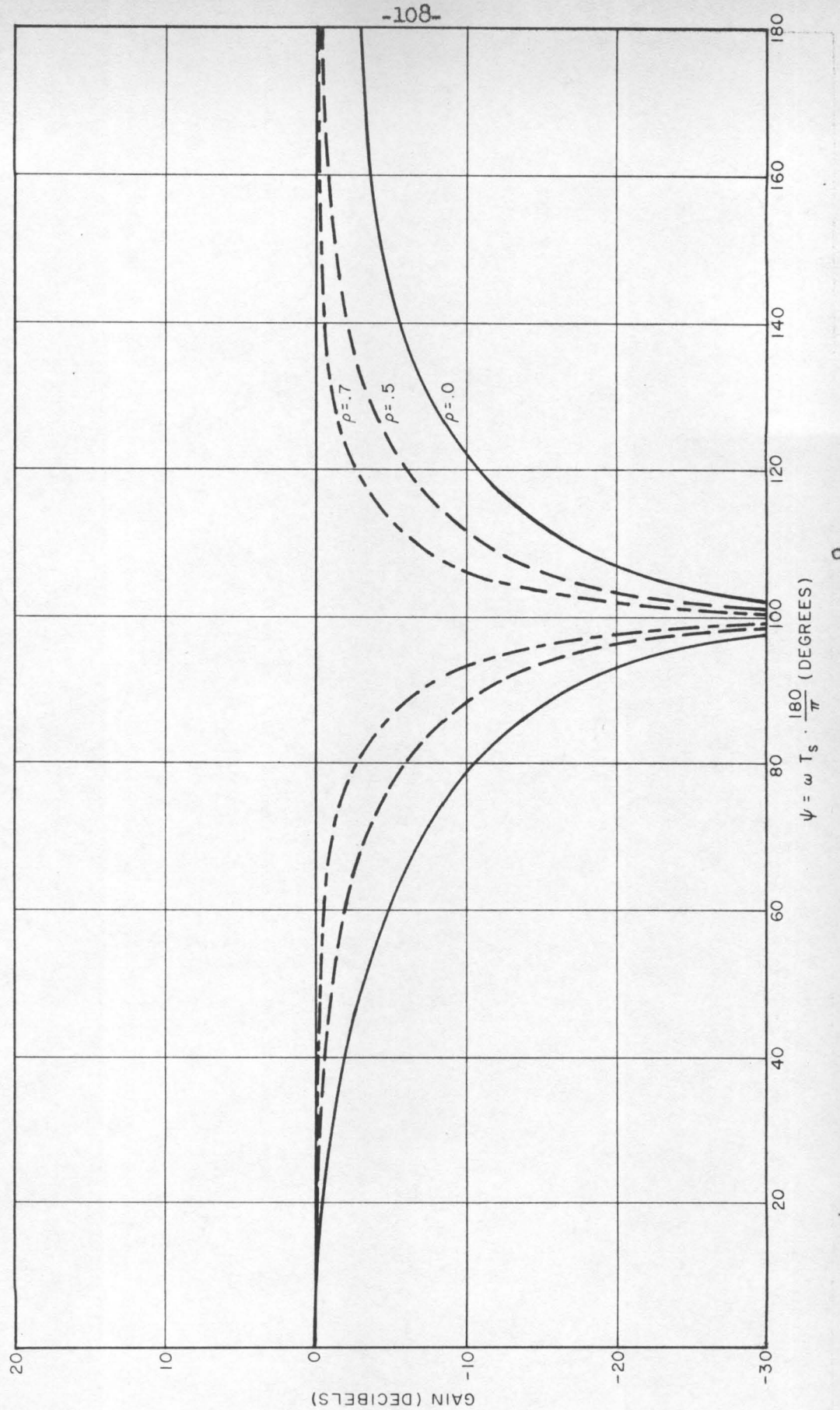


Figure 3.14a: Gain response of sampled compensation for $\beta_1 = 100^\circ$.

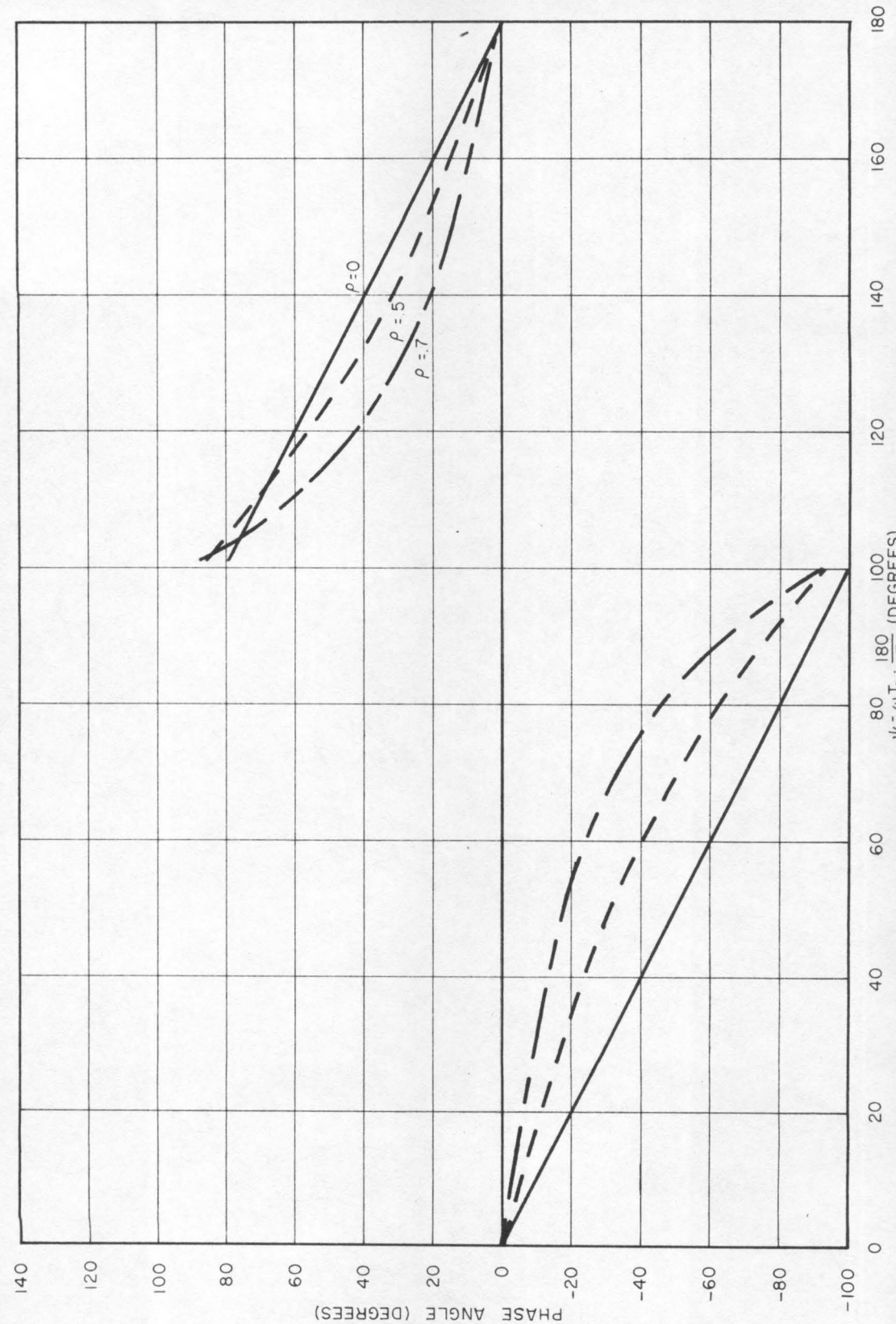
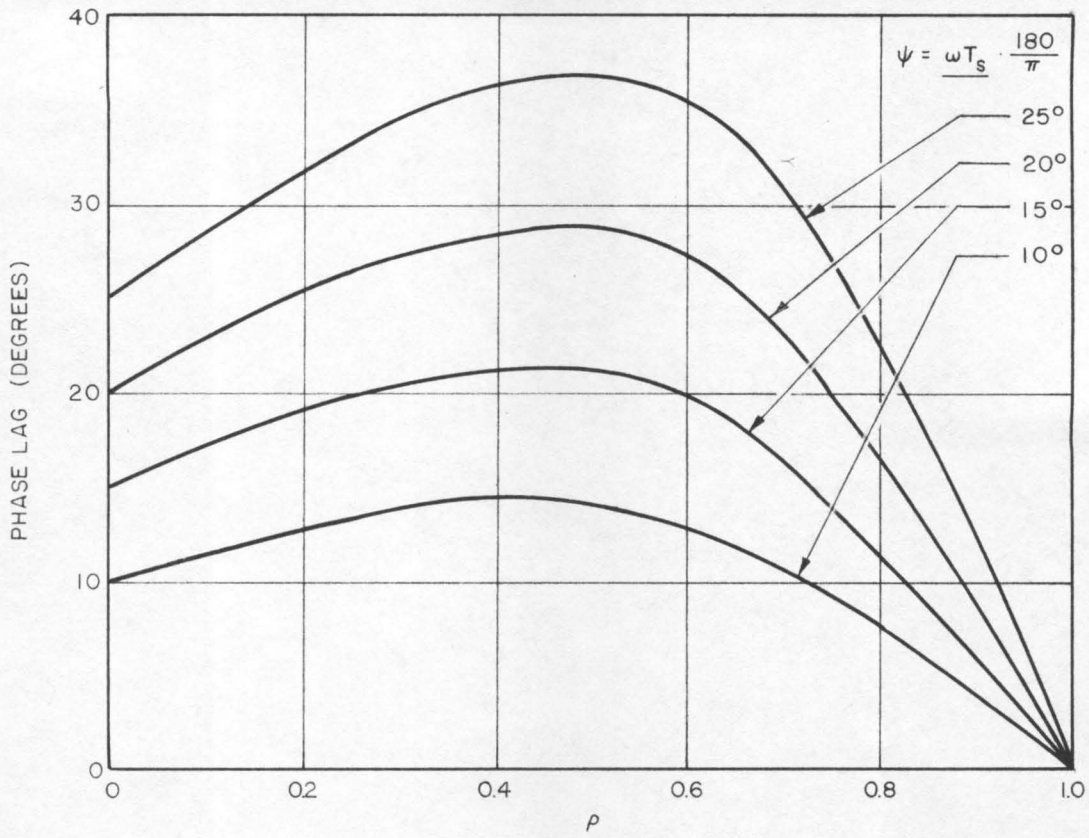


Figure 3.14b: Phase response of sampled compensation for $\beta_i = 100^o$.

Note that the bandwidth of the filter decreases rapidly as ρ_i approaches unity. Also, the low frequency phase lag for a given β_i decreases significantly as ρ_i approaches unity. Finally, the high frequency gain (gain for $\omega T_s \to \pi$) decreases as ρ_i approaches one.

First, a small ρ_i is desirable to insure that there is adequate attenuation at the true bending frequency even if the ω_i^* supplied by the FMC is slightly in error. The width of the frequency band which is attenuated appreciably is the factor which sets the FMC accuracy requirement. Second, a large ρ_i is desirable to minimize low frequency phase shift and to keep the high frequency gain as low as possible for this filter configuration. Figures 3.15a, b and c are approximate plots of the data in Figures 3.12 to 3.14 in a manner which shows the effect of ρ_i directly. (The subscript i is not indicated in the curves since they are applicable for any i.) The use of these curves is best demonstrated by an example such as the following: find the maximum allowable value for ρ_i if it is known that the FMC is accurate to $\pm$ 2% and that 20 db or more of attenuation must be introduced in a mode with $\beta_i = 100^\circ$. An examination of Figure 10c (because $\beta_i = 100^\circ$) shows that ρ_i must be less than 0.7 (the 98° curve corresponds to - 2%).

The compensation will initially be chosen with zeros located at their nominal position at t = 0. As time progresses, the resonant frequencies will slowly change. When an oscillation begins to build up because one or more resonances become unstable, the FMC will detect the oscillation, measure its frequency and adjust the appropriate compensation zeros to lie at the measured frequency. For each



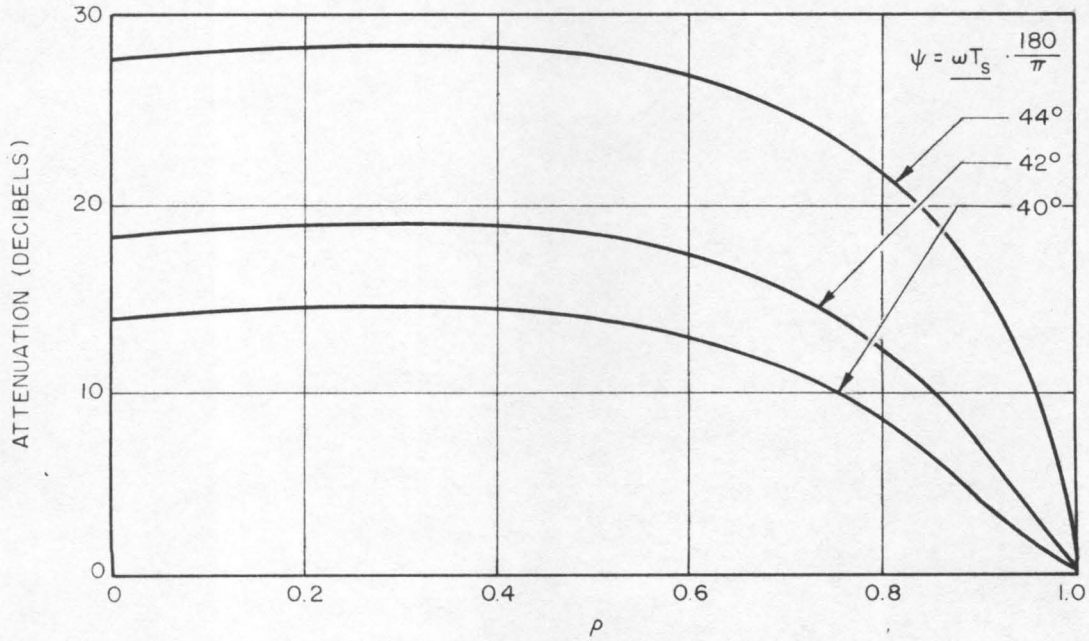


Figure 3.15a:Gain and phase characteristics for $\beta_i = 45^o$.

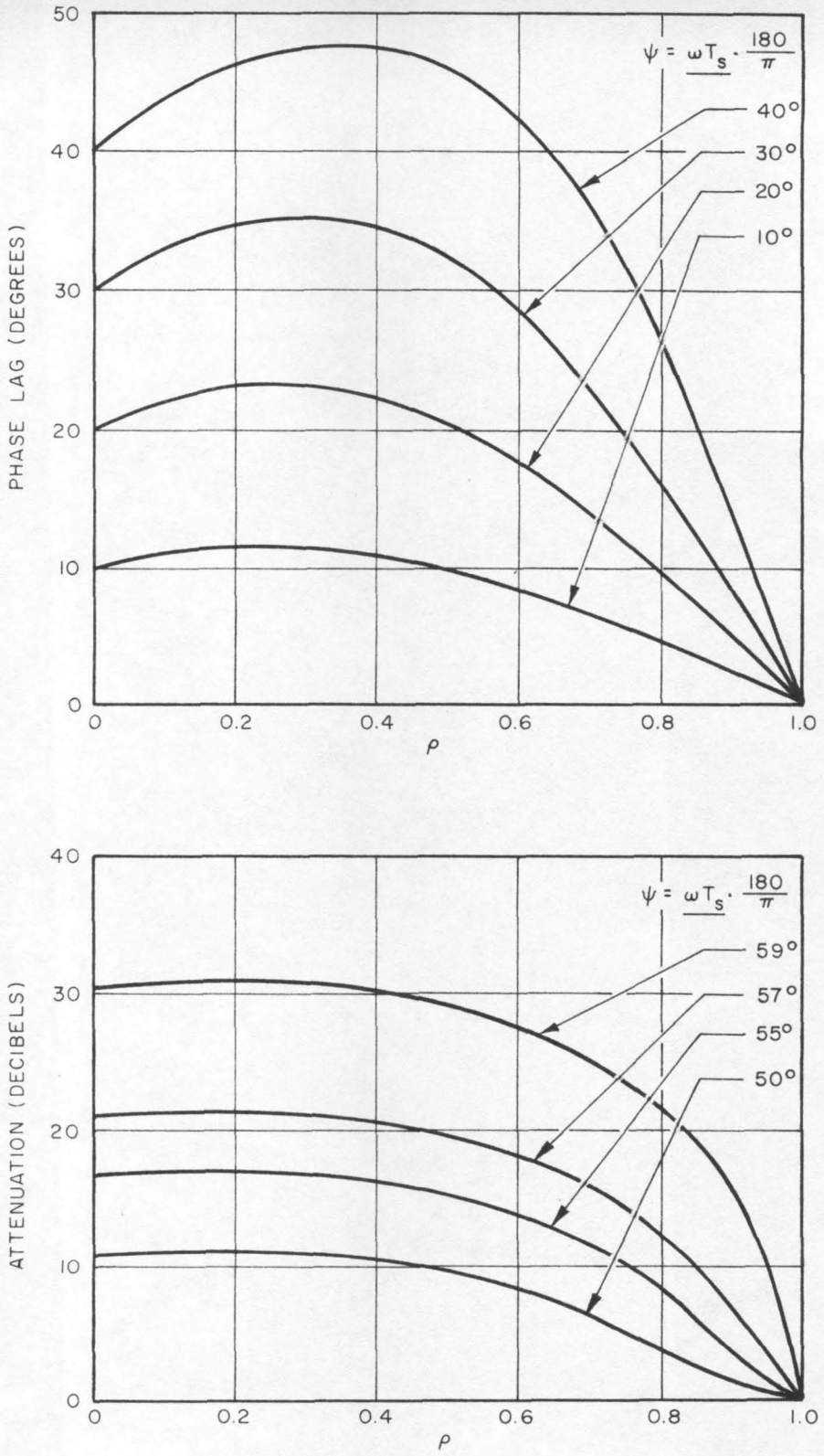


Figure 3.15b: Gain and phase characteristics for $\beta_i = 60^\circ$.

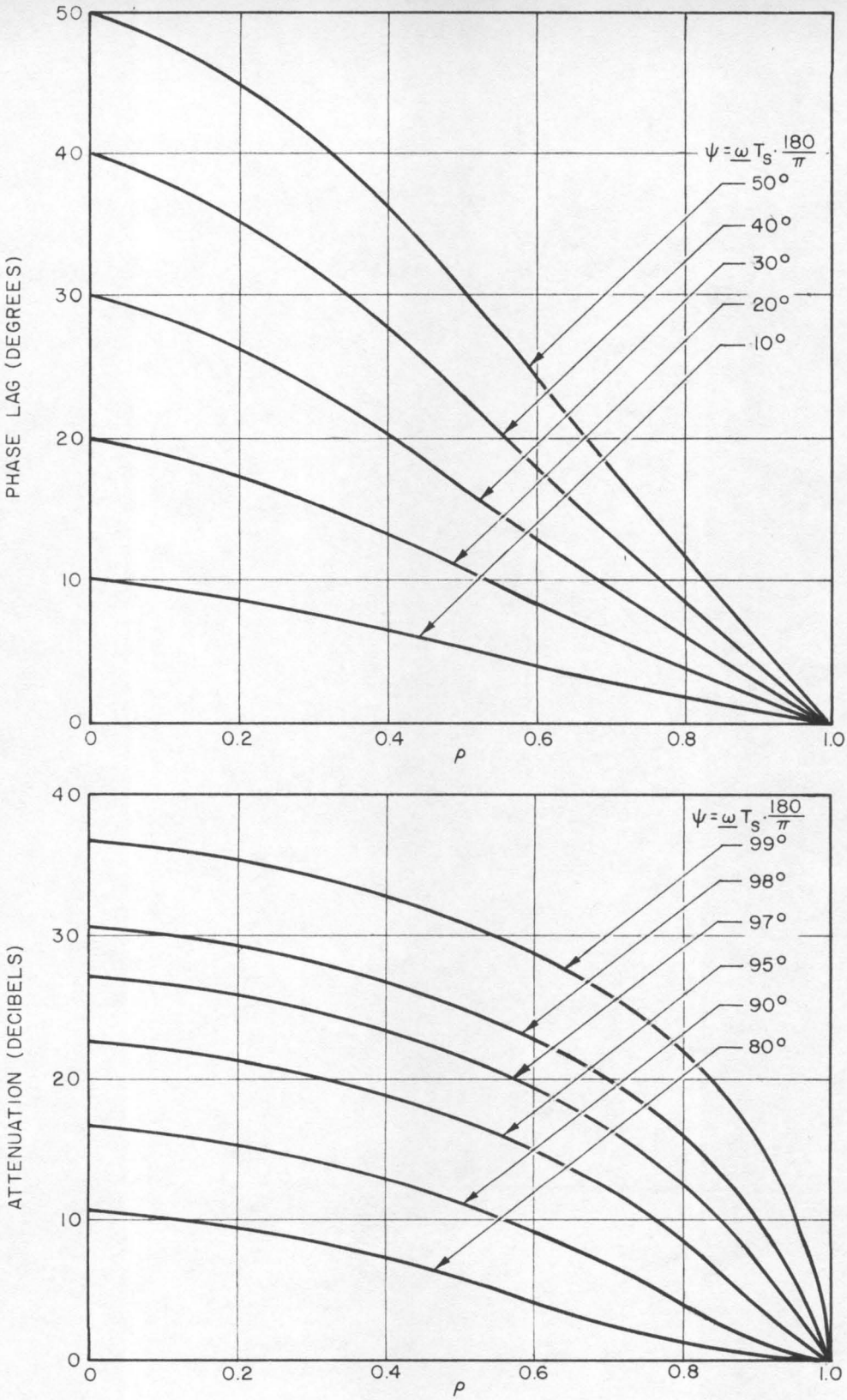


Figure 3.15c: Gain and phase characteristics for $\beta_i = 100^\circ$.

resonance which is to be cancelled there will be a corresponding pair of cancellation zeros.

For frequencies ω which lie very close to the frequency ω_i^*, it is possible to express the magnitude of the function $D_i(z)$ in the following form (see Equations 3.98, 3.99, 3.100, and 3.102):

$$\left|D_i(e^{j\omega T_s})\right| \tilde{=} 2\gamma \left|\psi - \beta_i\right|/\beta_i \quad \text{for} \quad \left|\psi - \beta_i\right| \ll 1 \qquad (3.104)$$

where

$$\gamma = \frac{\beta_i \cot(\beta_i/2)\,(1 + \rho^2 - 2\rho\cos\beta_i)}{2(1-\rho)\left[(1-\rho)^2 + 2\rho(1-\rho)\sin^2\beta_i\right]^{1/2}} \qquad (3.105)$$

The constant γ is simply related to the geometrical location of the Z-plane poles and zeros of $D_i(z)$. Typically it is of the order of one to three so long as the poles associated with the compensation function do not lie too close to the unit circle. Evaluated for the parameters $\rho = 0.7$ and $\beta_i = 60^\circ$, for example, we find that $\gamma = 1.74$.

Necessary and sufficient conditions (Equations 3.112 and 3.125) will now be derived which place bounds on the allowable FMC measurement error as a function of the uncompensated loop transmission at any resonant frequency we wish to gain stabilize. If the necessary condition is <u>not</u> satisfied it is in general impossible to gain stabilize the mode being considered. If the sufficiency condition <u>is</u> satisfied, it is certain that the adaptive technique will gain stabilize the mode being considered. There is a region of operation where the necessary condition is fulfilled but the sufficiency condition is not met. Under these circumstances it is possible, but not certain, that the system will work.

It is convenient to define the following quantities:

ω_i = open loop natural frequency of the i^{th} resonant mode

ω_c = closed loop frequency of oscillation of the i^{th} mode

ω_c^* = FMC estimate of the frequency ω_c

$\beta_i = \omega_i T_s$

$\beta_c = \omega_c T_s$

$\beta_c^* = \omega_c^* T_s$

M = the magnitude of the loop gain at the open loop resonant frequency ω_i, excluding the effects of $D_i(z)$

$K = \left| D_i(e^{j\omega_i T_s}) \right|$

The compensation, $D_i(z)$, which is introduced to stabilize the i^{th} mode must satisfy the following inequality if this mode is to be gain stabilized:

$$\left| D_i(z = e^{j\omega_i T_s}) \right| M = KM < 1 \qquad (3.106)$$

Since we know that the compensating zeros are placed at angles β_c^* while the angle at which the open loop resonant frequency occurs is β_i, we may evaluate K approximately using relation 3.104. Thus

$$K \cong 2\gamma |\beta_i - \beta_c^*| / \beta_c^* \cong 2\gamma |\beta_i - \beta_c^*| / \beta_i \qquad (3.107)$$

where it has been assumed that $|\beta_i - \beta_c^*| \ll 1$. Combining Equations 3.106 and 3.107 gives

$$|\beta_i - \beta_c^*| / \beta_i < 1/(2\gamma M) \qquad (3.108)$$

The left hand portion of the above inequality is just the per unit

error that is made in the attempt to position compensation zeros at the open loop resonant frequency. This error is attributable to two distinct causes. First, even if the FMC were perfect there would be a per unit error ε because the FMC measures the closed loop frequency of oscillation rather than the open loop natural frequency of the system. Second, there is a per unit error Δ because the FMC is not perfect and gives an erroneous estimate of what the closed loop frequency is. ε and Δ are, by definition,

$$\varepsilon = (\beta_c - \beta_i)/\beta_i \tag{3.109}$$

$$\Delta = (\beta_c^* - \beta_c)/\beta_c \tilde{=} (\beta_c^* - \beta_c)/\beta_i \tag{3.110}$$

Combining Equations 3.108, 3.109 and 3.110, we require that the inequality

$$|\varepsilon + \Delta| < 1/(2\gamma M) \tag{3.111}$$

must be satisfied if the mode under consideration is to be gain stabilized. The necessary condition that must be satisfied by the FMC is obtained by setting $\varepsilon = 0$ in Equation 3.111, i.e., even if the closed loop and open loop frequencies are equal (the best condition so far as the FMC is concerned) it is necessary for the FMC to produce estimates of the resonant frequency such that

$$|\Delta| < 1/(2\gamma M) \tag{3.112}$$

Similarly, if we assume that the FMC is perfect (i.e., $\Delta = 0$) we obtain from Equation 3.111 an upper bound on the tolerable difference between open loop and closed loop natural frequencies:

$$|\varepsilon| < 1/(2\gamma M) \tag{3.113}$$

Finally we recognize that inequality 3.111 will certainly be satisfied if the necessary condition 3.113 is satisfied and if

$$|\Delta| < 1/(2\gamma M) - |\varepsilon| \tag{3.114}$$

When relation 3.114 holds for any particular resonant mode we are certain that the adaptive control technique which has been suggested will gain stabilize that mode, i.e., this condition is sufficient to insure success.

We now obtain an upper bound for $|\varepsilon|$ by examining the locus of the closed loop root which originates at the Z-plane open loop pole location p_i, where

$$p_i = e^{(-\zeta\omega_i + j\omega_i)T_s} = e^{(-\zeta + j)\beta_i} \tag{3.115}$$

The loop transmission may be written as

$$G(z)/(z - p_i) \tag{3.116}$$

The pole originating at p_i will be found at the point

$$z' = p_i + \mu \tag{3.117}$$

By application of the usual Root Locus technique we find that

$$\mu = -G(z') \tag{3.118}$$

If we assume that the closed loop pole at z' is much closer to p_i than are any of the singularities of $G(z)$ we may write

$$G(z') \cong G(e^{j\beta_i}) \tag{3.119}$$

Thus, the magnitude of μ may be written

$$|\mu| \cong |G(e^{j\beta_i})| \tag{3.120}$$

But $|G(e^{j\beta_i})|$ is related to the peak loop gain at resonance, M, by (see Equations 3.115 and 3.116):

$$\begin{aligned} M &= |G(e^{j\beta_i})| / (1 - e^{-\zeta\beta_i}) \\ &\cong |G(e^{j\beta_i})| / (\zeta\beta_i) \end{aligned} \tag{3.121}$$

Combining Equations 3.120 and 3.121 leads to

$$|\mu| \cong \zeta M\beta_i \tag{3.122}$$

= magnitude of the distance between the open and closed loop pole locations for the i^{th} resonant mode

Since both the closed and open loop poles lie very close to the unit circle it follows that the difference between the angular locations of the closed loop pole, β_c, and the open loop pole, β_i, are related by the following inequality:

$$|\beta_c - \beta_i| \leq |\mu| \tag{3.123}$$

Combining relations 3.109, 3.122 and 3.123 we obtain

$$|\varepsilon| \leq \zeta M \tag{3.124}$$

Finally, combining Equations 3.114 and 3.124 yields

$$|\Delta| < 1/(2\gamma M) - \zeta M \qquad (3.125)$$

Relation 3.125 is a sufficiency condition which, if satisfied, guarantees the success of the proposed adaptive system.

As an example of the application of sufficiency condition 3.125, consider the following hypothetical lightly damped resonant mode and compensation:

$$\zeta = 0.005$$

$$\gamma = 2.0$$

The largest allowable loop gain at the resonant frequency assuming a perfect FMC (i.e., $\Delta = 0$) is

$$M = 1/\sqrt{2\gamma\zeta} = 7.08 = 17 \text{ db.}$$

If $M = 6.0$ (15.56 db), the accuracy must be greater than

$$100\left[1/(2\gamma M) - \zeta M\right] = 1.17\%$$

On the other hand, the necessary condition 3.112 gives, for this latter condition

$$|\Delta| < 4.17\%$$

3.4 Summary

Section 3.2 contained an analysis of the sine-wave-correlation Frequency Measuring Computer (FMC) for several different assumptions concerning the form of the input stimulus y(t). The analysis examined the output of a single FMC channel and showed that such a channel would have a sizable output only when the stimulus contained a sinusoidal signal at a frequency close to the center frequency of

the channel. By utilizing a group of such channels tuned to a set of frequencies which spans the frequency interval of interest it is possible to detect the presence of a discrete oscillation and to measure its frequency with considerable precision in a short length of time.

A modification of the sine-wave-correlation technique was mentioned which reduces the amount of computing necessary to measure a frequency. This modification is the square-wave-correlation FMC, in which the stimulus is correlated with a set of square waves of different frequencies. This device can be analyzed by making a Fourier Series expansion of the square waves and performing a calculation identical to the one for the sine wave FMC for each Fourier component. Since the amplitude of the fundamental frequency component of the square wave is a factor of three greater than the amplitude of the third harmonic and all other harmonics are even smaller, only the first harmonic need be considered in the first approximation. By using the square-wave FMC the complicated and relatively lengthy multiplication operation can be replaced by a simple and rapid gating operation, thus reducing considerably the necessary computing capability.

It is possible, in principle, to measure the frequency of a discrete frequency sinusoid imbedded in additive noise with great precision using the method described in this chapter. This great accuracy is achieved at the expense of integrating over a long time interval and using a very large number of very closely spaced FMC channels. This is not possible in practice for several reasons: first, no real oscillation is going to continue indefinitely with constant amplitude and frequency, i.e., an absolutely pure infinite length sinusoid is a

mathematical fiction; second, an answer is needed in a relatively short length of time if adaptive adjustments of controller parameters are going to be made on the basis of these measurements; finally, hardware requirements dictate the use of only a relatively few channels if the system is to be at all practical. Thus compromises must be made between the accuracy which is desired and the length of time over which the correlations are to be performed.

One possible form which might be selected for the adaptive digital compensation was presented in Section 3.3. It was shown that, if certain conditions are satisfied, the proposed adaptive technique can gain stabilize the resonant modes of lightly damped systems. The adaptive system will work even though very little a priori information is available concerning the frequencies and amplitudes of the various resonant modes.

The next chapter describes a detailed simulation study which demonstrates the feasibility of this type of adaptive controller for systems with lightly damped resonances.

CHAPTER IV

SIMULATION OF THE SYSTEM

4.1 Introduction

In the preceding chapter an adaptive control system was proposed to stabilize a system with multiple lightly damped resonances. It is assumed that the frequencies of the resonances are imprecisely known and can vary slowly with time. Under these circumstances the application of an adaptive system which can measure the conditions of the "plant" that it is controlling and adjust its own parameters on the basis of this information seems desirable. The operation of the overall system was discussed in the preceding chapter, and the adaptive Frequency Measuring Computer (FMC) was analyzed in detail. The complete system is a high order multiple loop non-linear feedback control system in which the non-linearity is a complicated function of a system variable y(t). At present there are no adequate analytic techniques for evaluating exactly the performance of a system of this type. For this reason a detailed simulation of the complete adaptive system was implemented to investigate its performance characteristics. The specific example of a highly flexible missile was chosen because the author is familiar with systems in this category and was able to obtain the pertinent parameter data from Space Technology Laboratories. The simulation was programmed for the Burroughs-220 digital computer which is available for research use at the California Institute of Technology.

The missile dynamic equations which were used in the simulation are presented in the Appendix. These equations were integrated

step by step by the standard fourth order Runga-Kutta integration technique (19). Difference equations were programmed to represent the sampled-data portion of the control loop. Certain coefficients in these difference equations were adjusted in accordance with logic programmed into the simulation of the FMC. The computer output consists of the following information:

1. Transient response points for any desired system variables.
2. The outputs of the FMC channels.
3. The coefficients of the difference equation compensation (these are printed whenever they are changed by command of the FMC).

Due to the fact that the missile dynamic equations are of high order, the integration of the differential equations takes considerable time. The program ran roughly with a ratio of 300 to 1 between computer time and real time, i.e., to obtain one second of transient response data took 300 seconds (five minutes) of machine time. Each run took from 45 minutes to 1 hour, since roughly ten seconds of real time were needed to observe whether or not the adaptive action of the system caused it to converge to a stable operating configuration.

Because of the lengthy running time it was not possible to conduct a very extensive investigation of the effects of variations of a great number of parameters. Only the square wave FMC was simulated for there is no doubt that the sine wave FMC will perform at least as well. Parameters (e.g., the interchannel spacing and integration times) were selected for the FMC logic after a bit of trial and error experimentation. It was found that for these FMC parameters

the system did effectively stabilize responses for several different sets of bending characteristics. The transient responses all started with the system unstable in one or more bending modes and with no sampled-data compensation in the loop. The FMC then measured the frequency (or frequencies) of the unstable response which began to build up and introduced digital compensation which placed zeros of transmission at the measured bending frequencies. The simulation results give one confidence that the technique can succeed when applied to other lightly damped systems.

The following section of this chapter presents the results of the computer simulation study. The final section describes some of the details of the simulation which was performed.

4.2 Results of the Simulation

The results presented in this section are almost entirely in graphical form. The transient response curves shown are plots of θ_p, the output of the position sensing device, as a function of time. All initial conditions were assumed to be zero and the commanded attitude, θ_c, in each case was a step input of amplitude 1° applied at time $t = 0$. The graphs are presented in pairs, the first showing the transient response of the system when the adaptive loop is inoperative (therefore no digital compensation is present in the control loop), and the second showing the response when the adaptive logic is functioning. θ_p was selected as the output variable because both the rigid body attitude, θ, and the bending oscillations show up clearly in this function. It is not necessary, therefore, to plot each of the generalized bending coordinates, q_i, and the rigid body attitude separately.

In the original tabulated computer outputs all of this information is available.

All the runs were made with aerodynamic parameters which correspond to a time of flight where aerodynamic effects are very important (i.e., near max Q). This, in general, is the most difficult time of flight to design for (due to the competing interests of bending and aerodynamics with regard to optimum gain settings). It was assumed that if the adaptive technique could be successful for this flight condition it would be possible to insure that its operation would be successful at any other time of flight. Bending parameters (i.e., the mode slopes at the rate and position sensing locations, and the mode frequencies) differ significantly in the following examples.

The transient response runs are presented in Figures 4.1a through 4.5b. The curves demonstrate how the action of the adaptive FMC stabilizes an otherwise unstable dynamic system. The parameters pertaining to the missile dynamic equations for each of these graphs are listed in Section 4.3, as are the parameters pertaining to the FMC portion of the simulation. (See Tables 4.1 and 4.2 at the end of this chapter.)

Note that in each of these examples the bending frequency oscillations are not completely eliminated by the adaptive loop, i.e., although the closed loop bending poles are forced into a stable region they do not have appreciable damping. This is not necessarily detrimental provided that these oscillations remain at tolerable levels. That the oscillations do not have appreciable damping is closely connected with the fact that a cancellation technique is being used

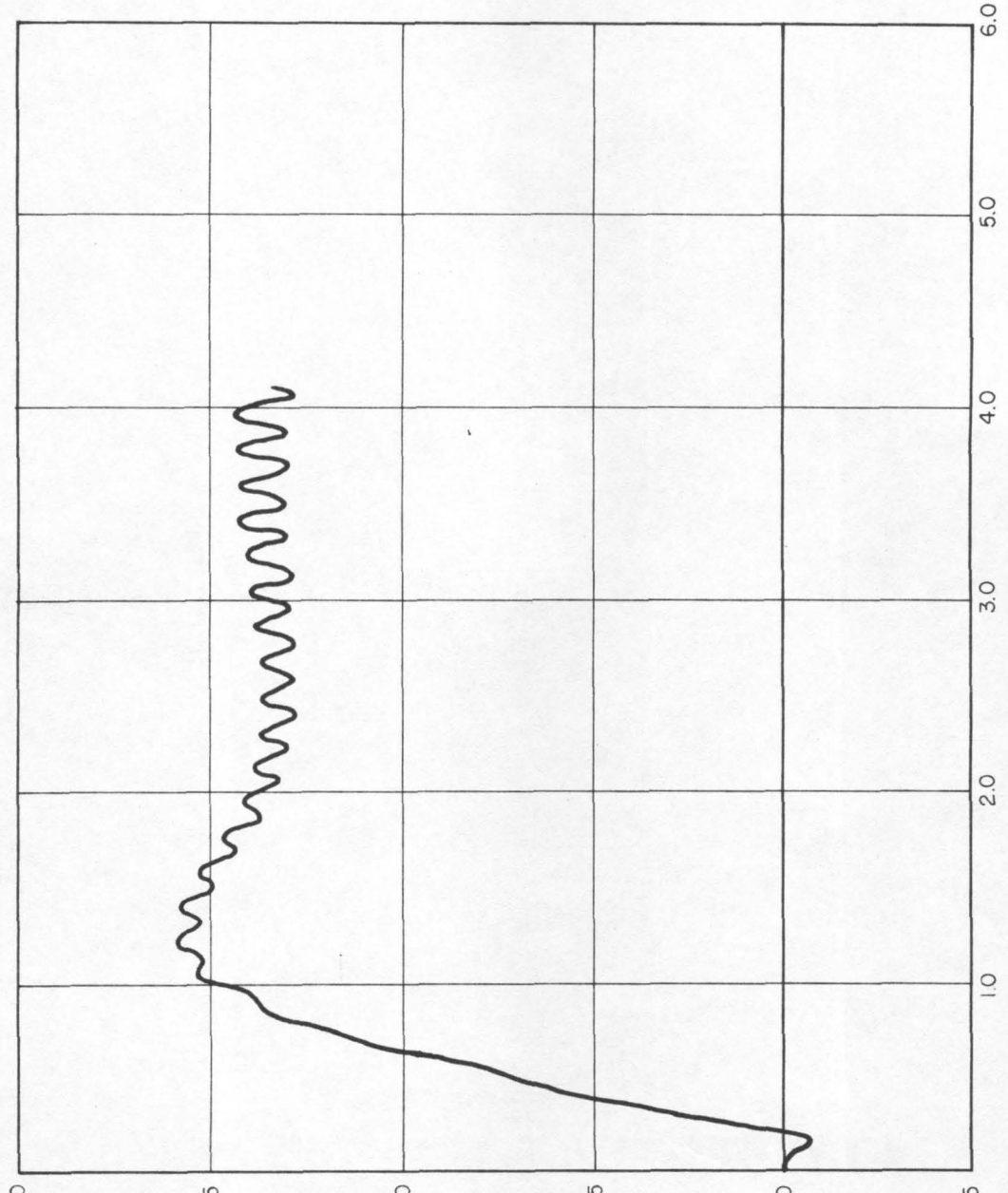


Figure 4.1a: No adaptation.

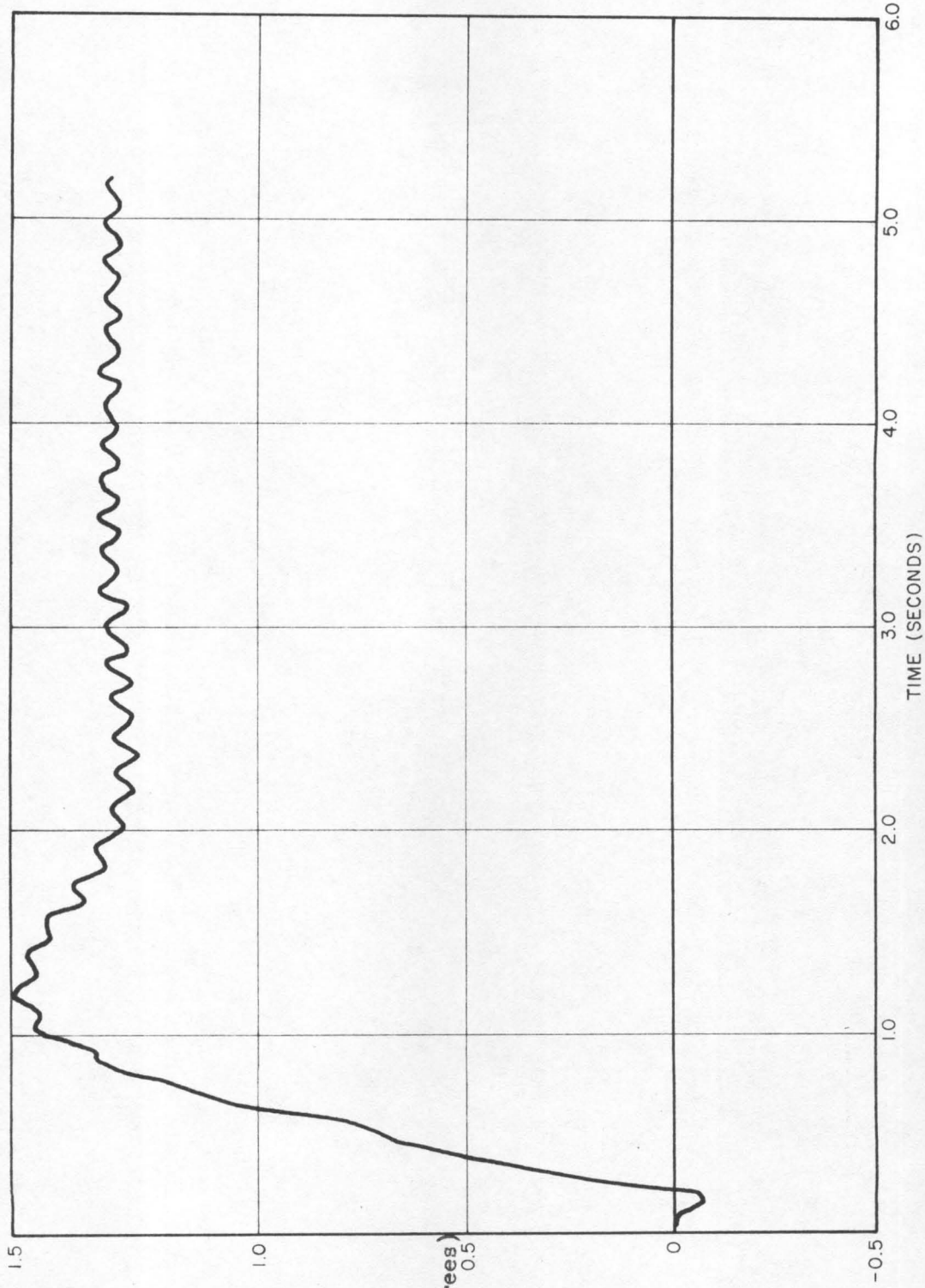


Figure 4.1b: Adaptive.

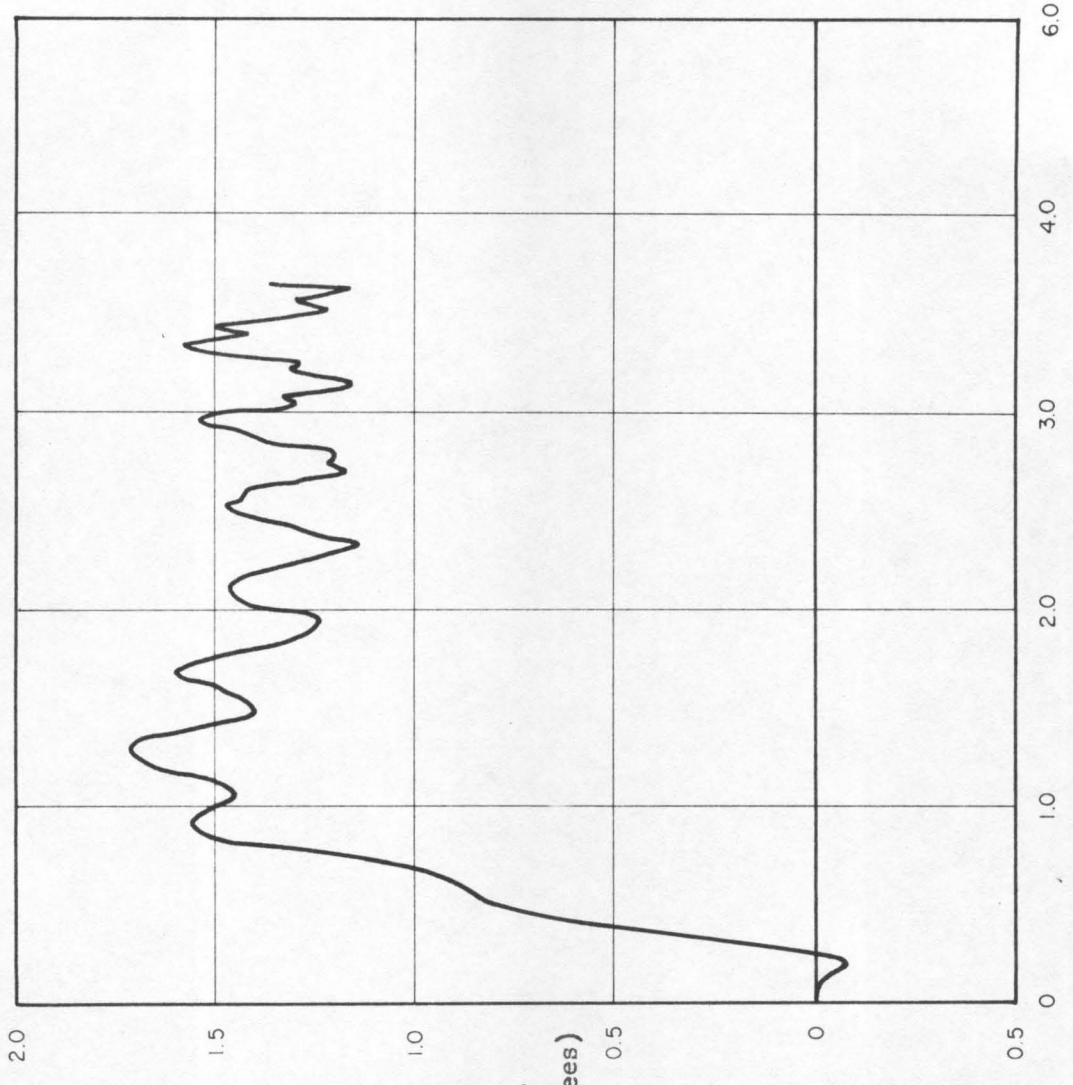


Figure 4.2a: No adaptation.

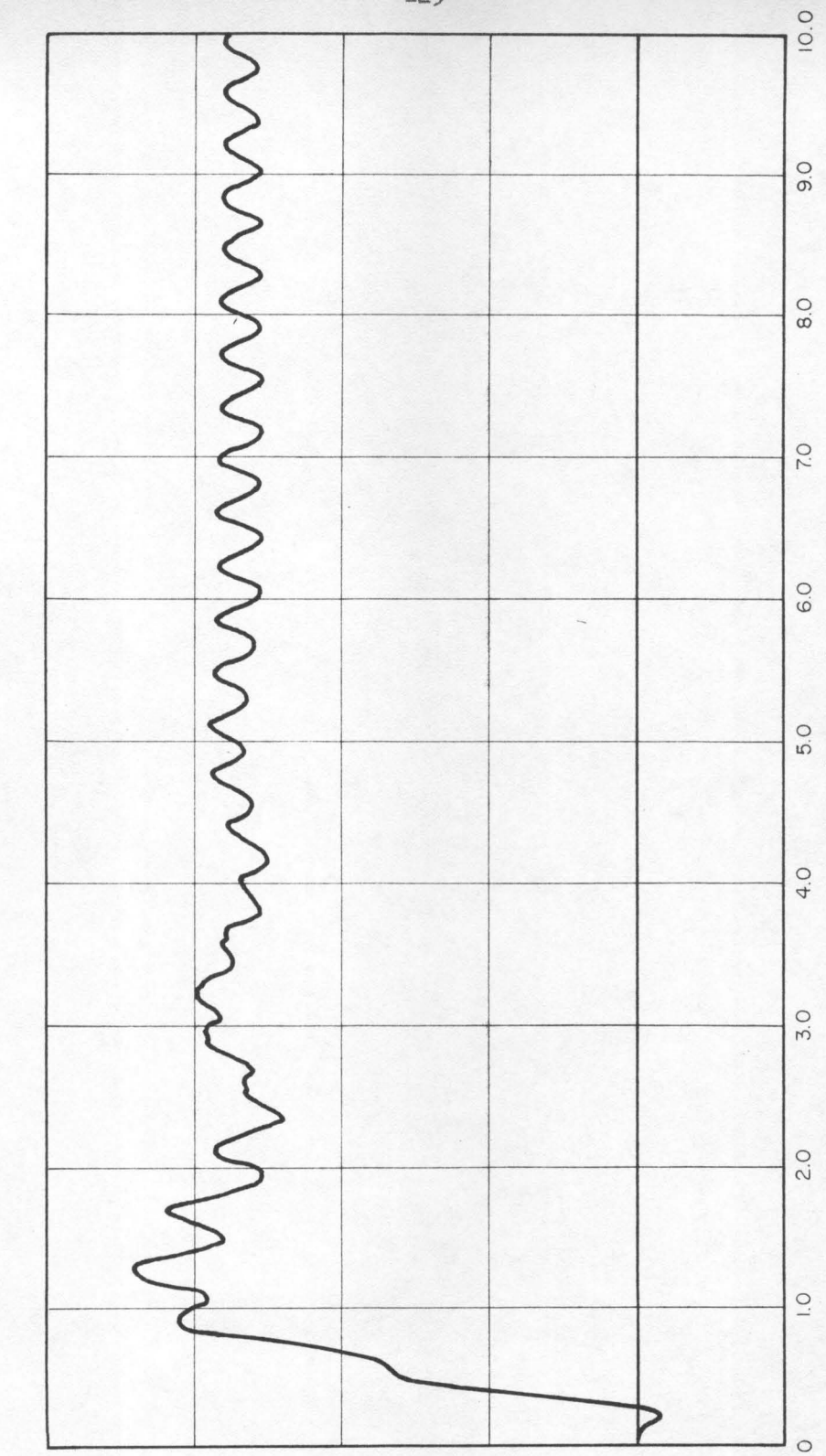


Figure 4.2b: Adaptive.

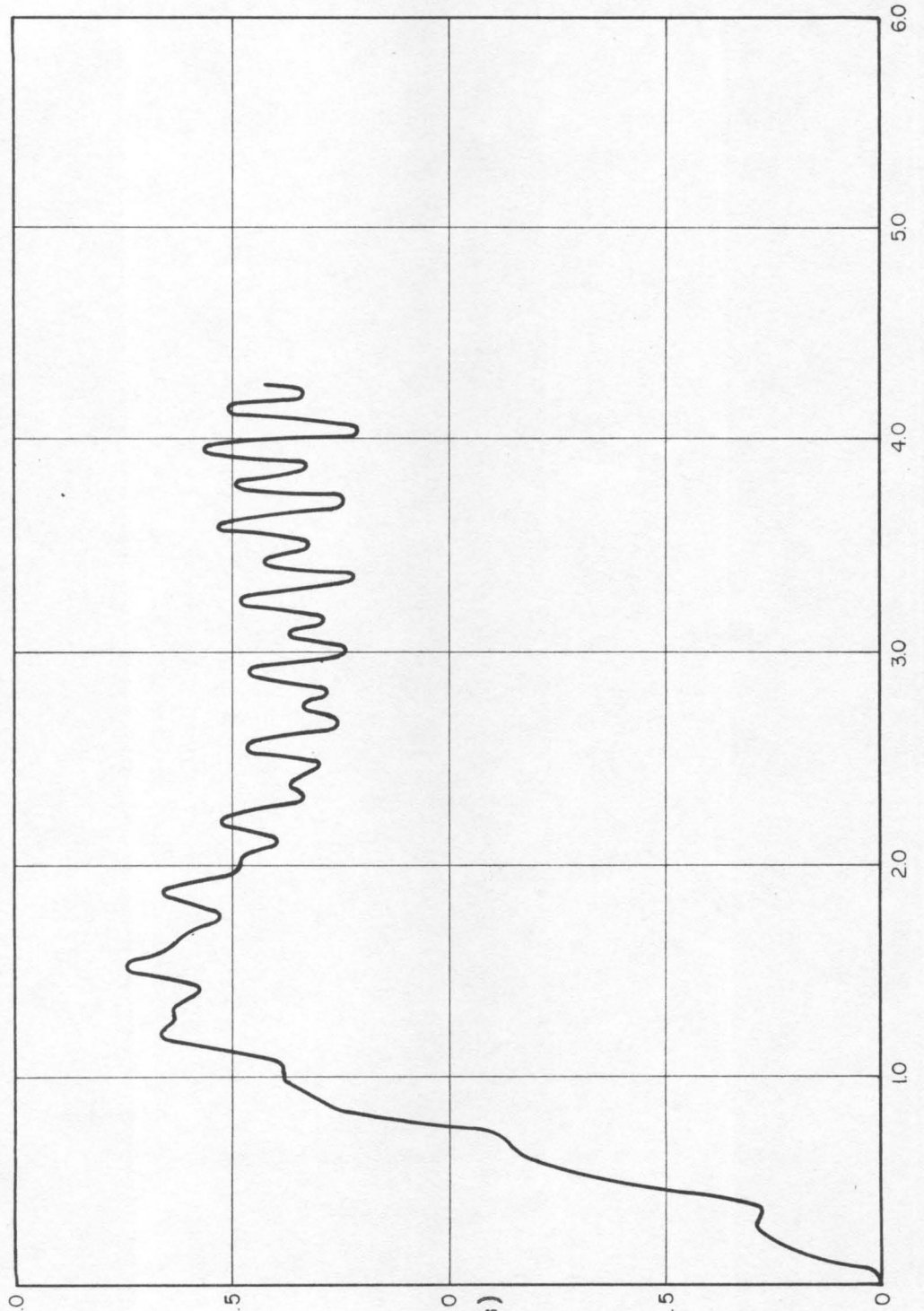


Figure 4.3a: No adaptation.

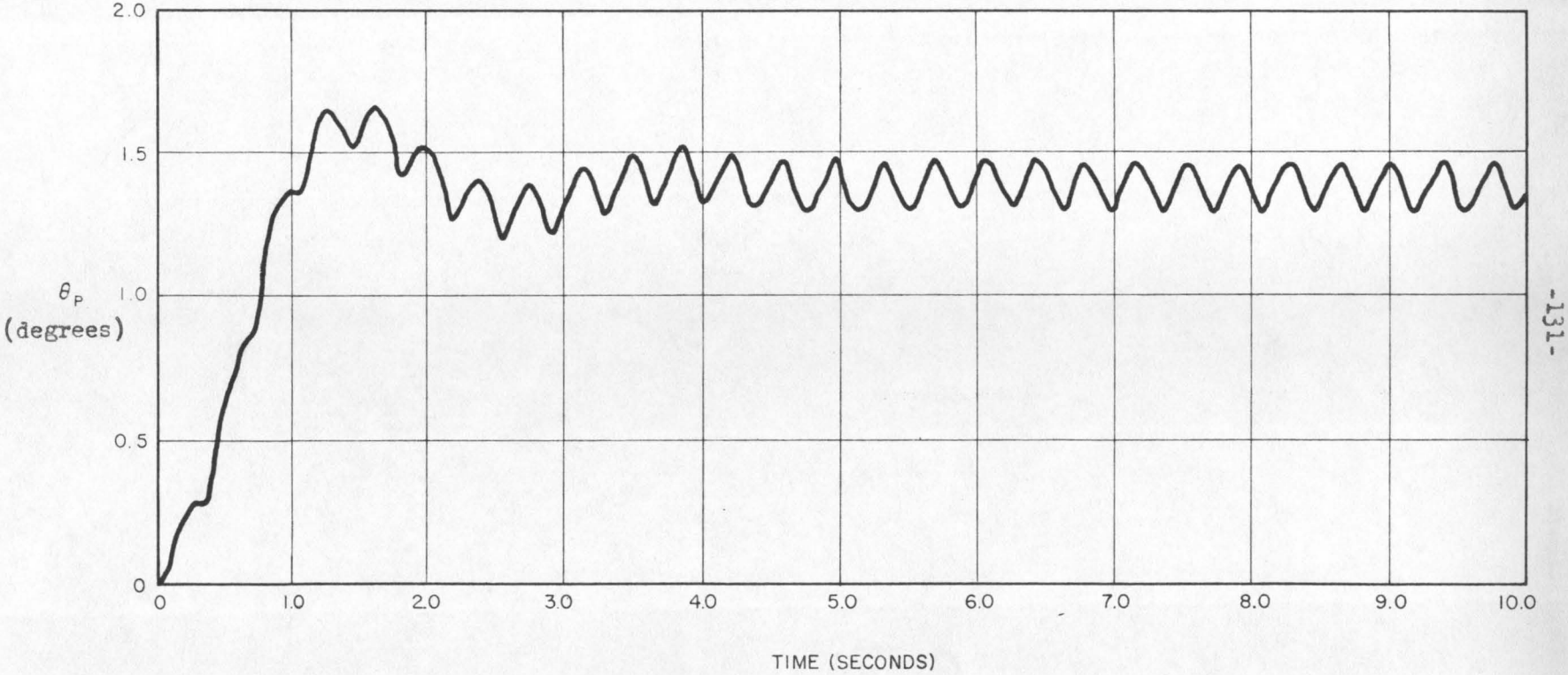


Figure 4.3b: Adaptive.

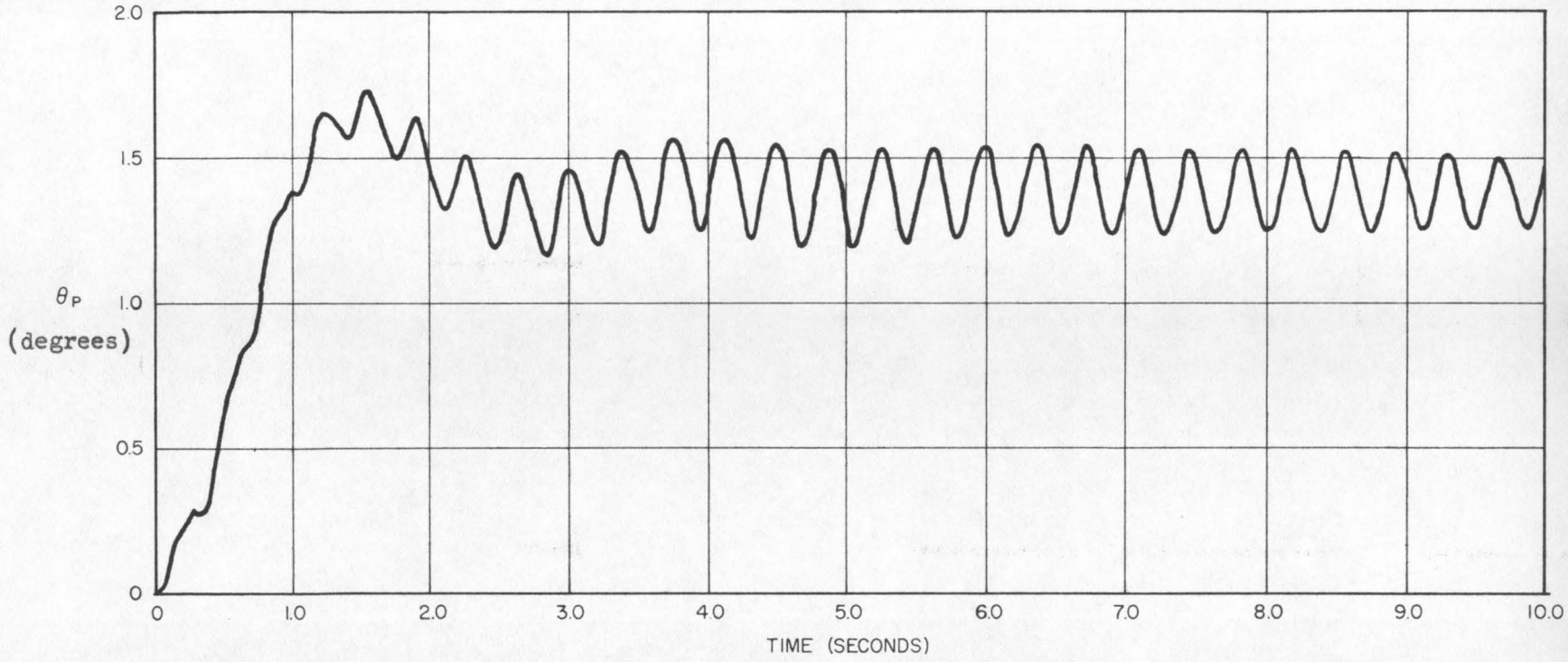


Figure 4.3c: Adaptive.

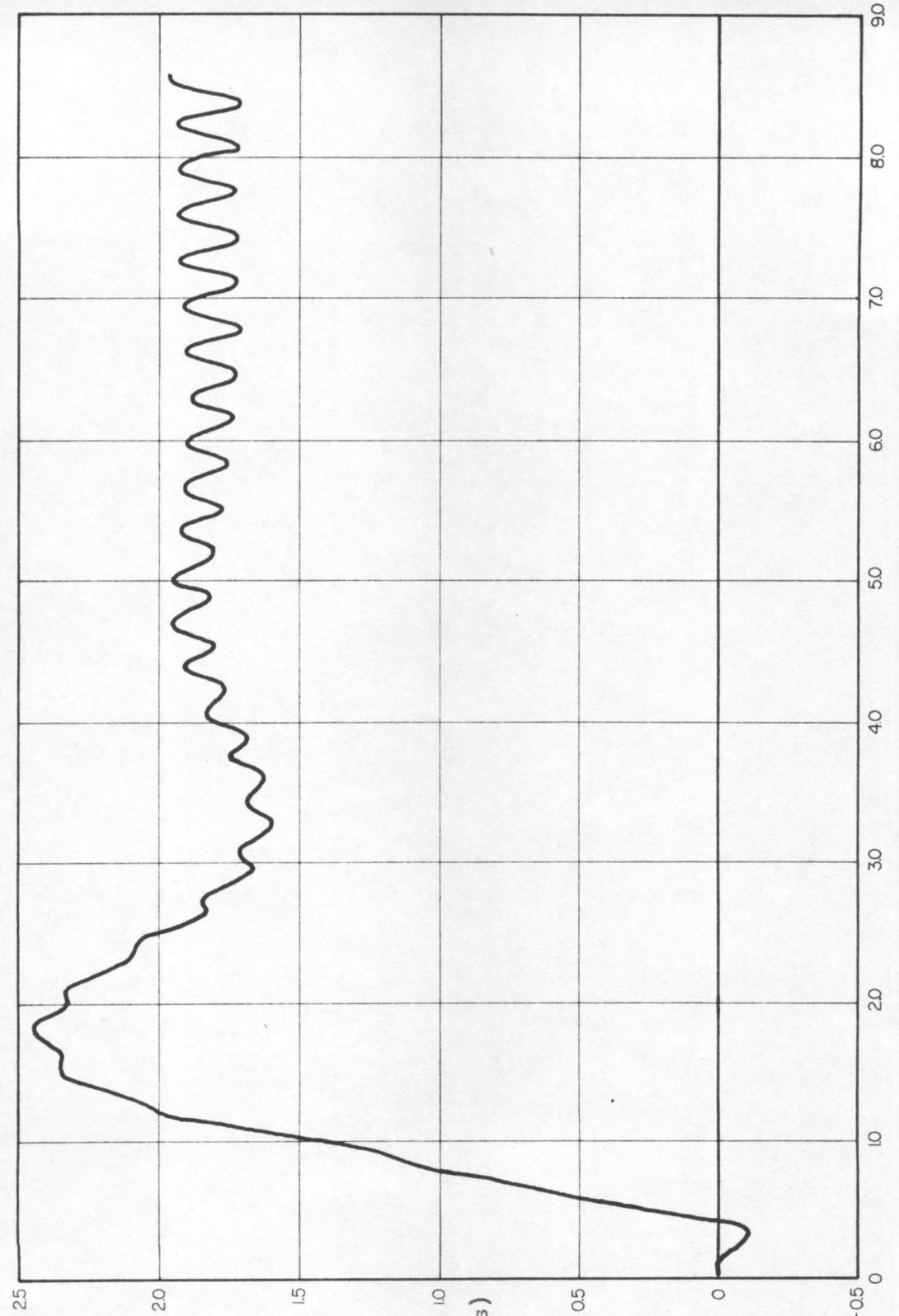


Figure 4.4a: No adaptation.

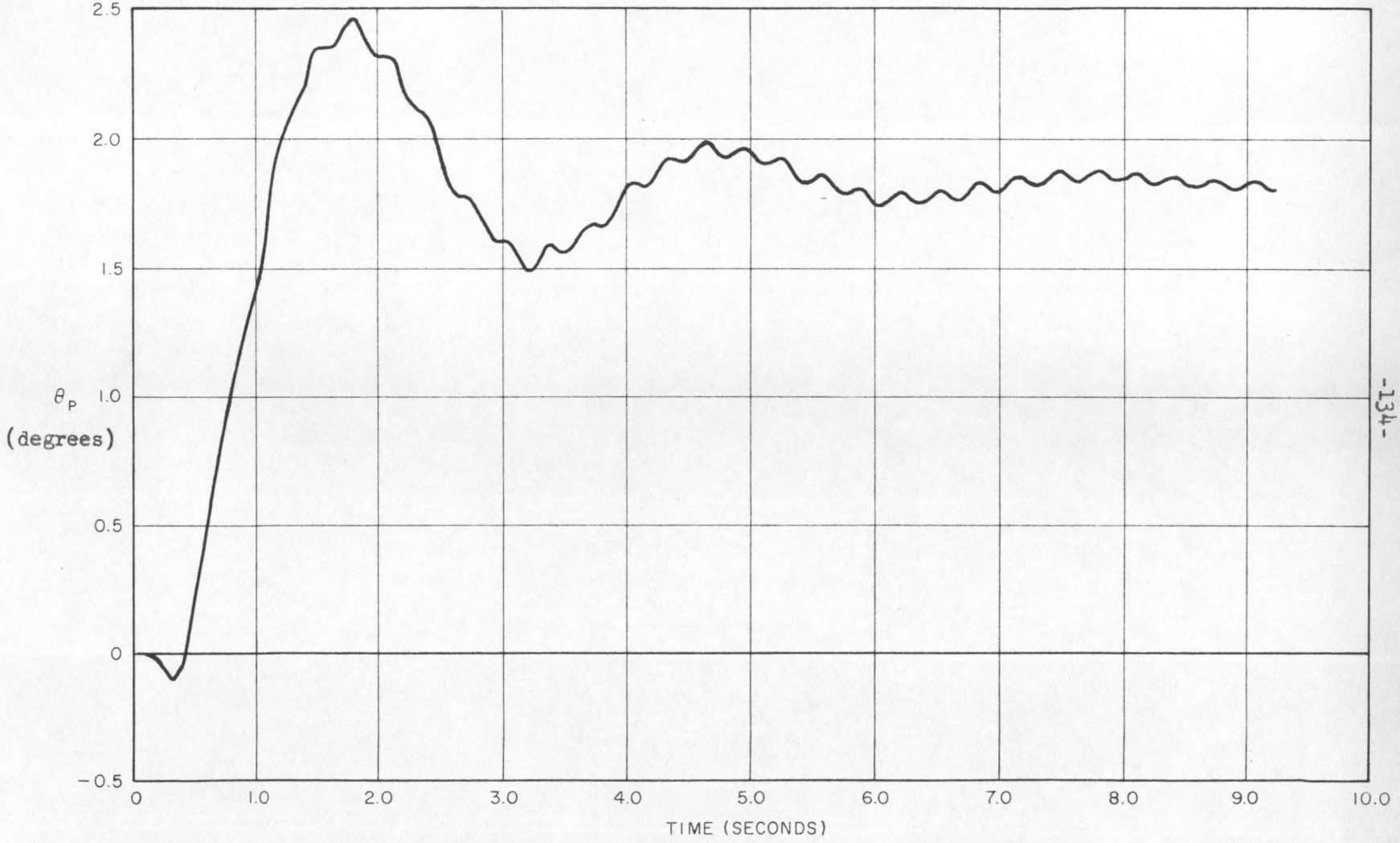


Figure 4.4b: Adaptive.

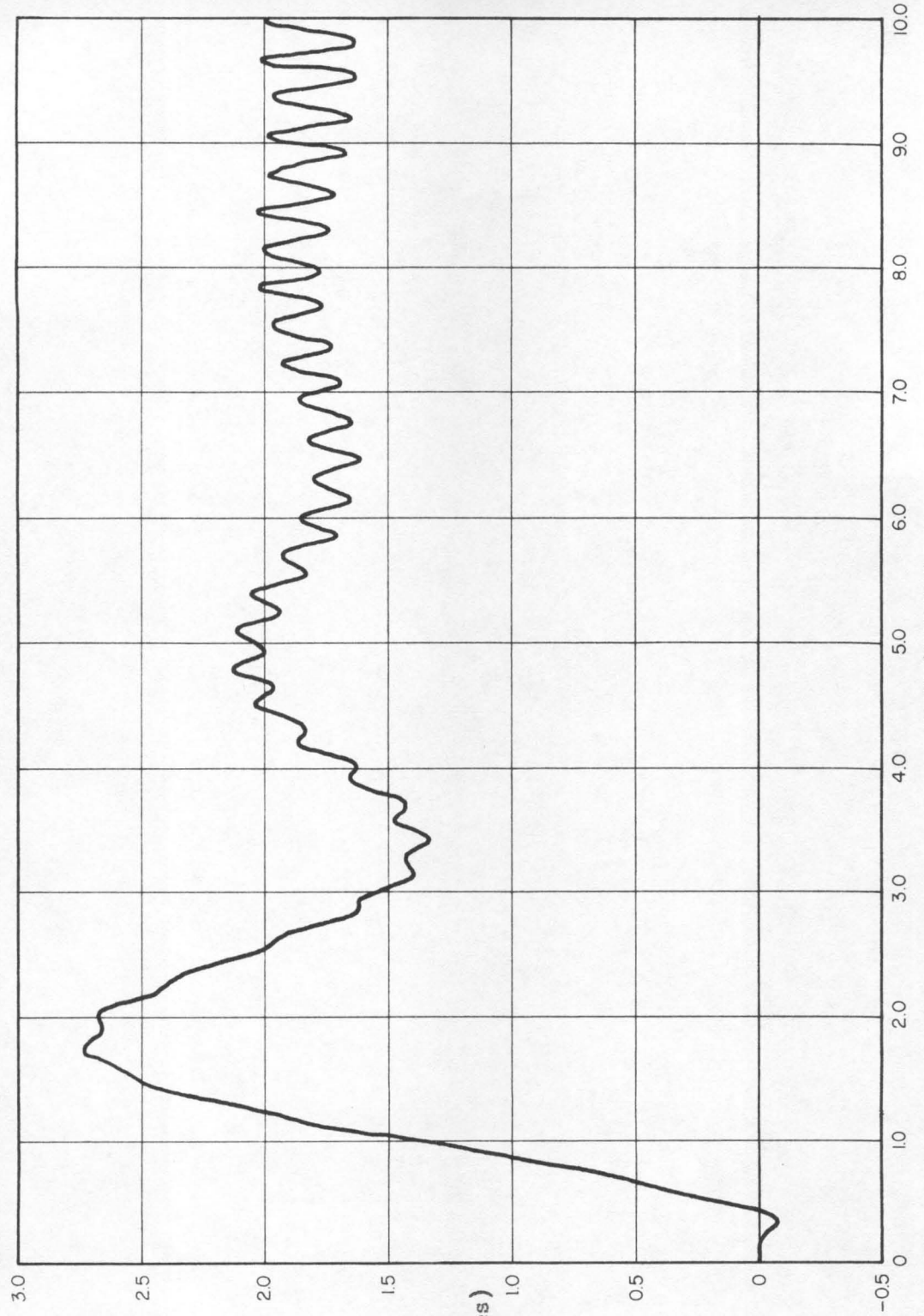


Figure 4.5a: No adaptation.

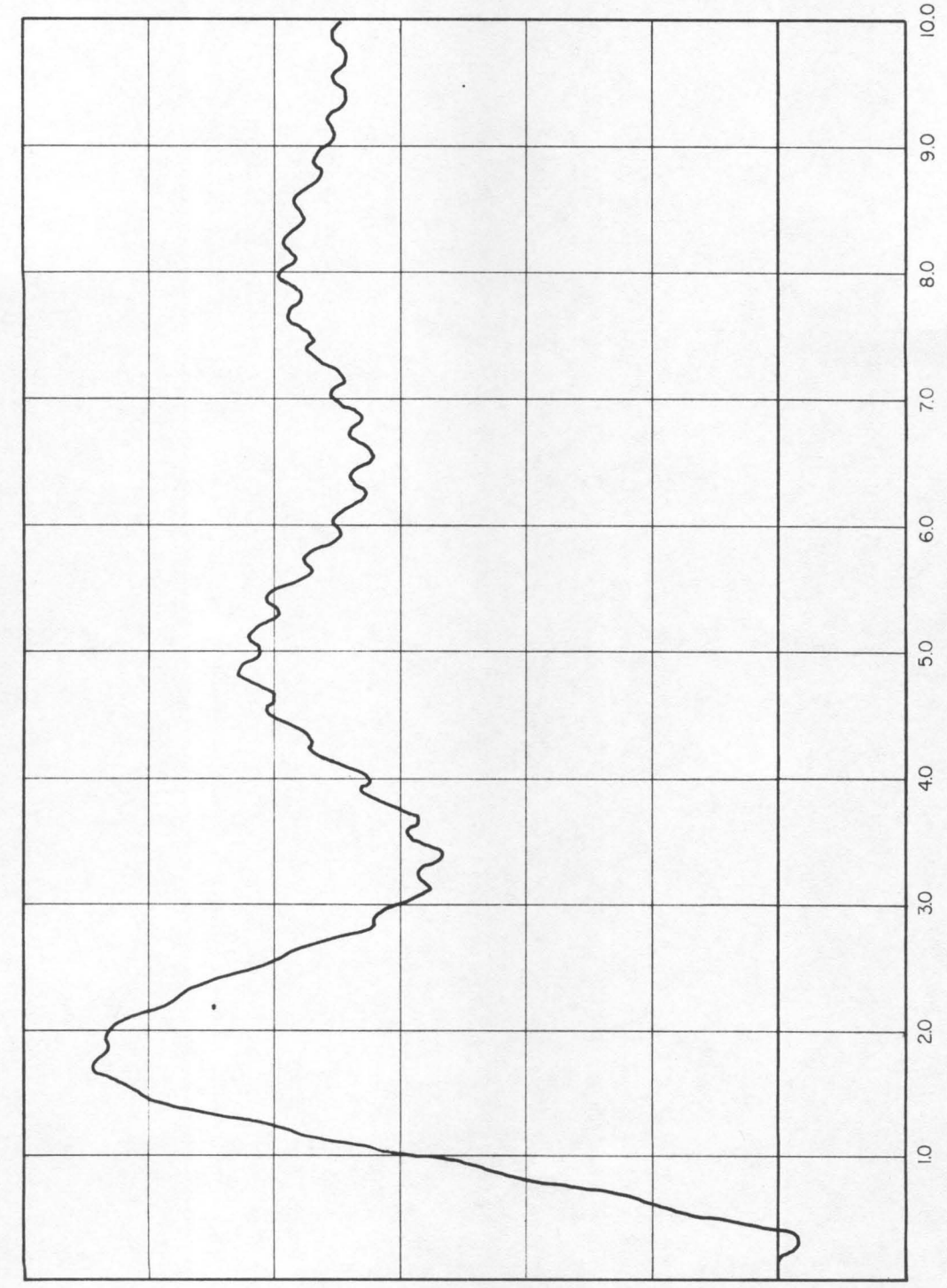


Figure 4.5b: Adaptive.

(i.e., the lightly damped system poles are being cancelled by compensation zeros). When cancellation is inexact (as it always is) a closed loop pole will exist very close to the location of the open loop pole and the cancellation zero. This closed loop pole is responsible for the oscillations under discussion. Since the amplitude of the oscillation is almost constant it is the type of signal whose frequency the FMC can measure with greatest accuracy. As the missile parameters change, the frequency of the oscillation changes, and the compensation zeros can follow this motion.

The fact that unstable oscillations initially build up quite rapidly results from the lack of any compensation at $t = 0$. Because no cancellation zeros are located close to the open loop bending poles these poles can move an appreciable distance outside the unit circle (in the Z-plane), which leads to a rapid build up of the oscillation. In practice the system would begin operating with compensation zeros which would hopefully be quite close to the proper locations. Only as system parameters slowly changed would the system gradually drift into an unstable condition, and the FMC would then move the zeros to better locations.

Figure 4.1a shows a system which is initially unstable at its second mode frequency (approximately 35 radians/second). Figure 4.1b shows that this instability is removed when the adaptive loop is in operation.

Figure 4.2a shows a system which is unstable at several bending frequencies. Figure 4.2b shows how the response is stable when the adaptive loop is in operation.

Figure 4.3a shows a system which has an unstable second mode and possibly an unstable first mode oscillation. Figures 4.3b and c show how the response is changed for two different sets of FMC parameters. From Table 4.2 we find that the integration intervals used in measuring the first mode frequency were three and six periods of the multiplying square wave for Figures 4.3b and c, respectively. The parameter ξ was equal to 0.06667 in Figure 4.3b and 0.03333 in Figure 4.3c. It follows from the analysis presented in Chapter III that the FMC is supplying first mode frequency estimates twice as often in Figure 4.3b as in Figure 4.3c. On the other hand, the accuracy of these estimates in Figure 4.3c is roughly twice as good as the accuracy in Figure 4.3b. Note that in Figure 4.3b it is not apparent that the oscillations are diminishing in amplitude. Evidently the FMC has not located the natural frequency of the first mode with sufficient accuracy to stabilize it. In Figure 4.3c the amplitude of the oscillation is definitely decreasing, although very slowly, and its initial magnitude is approximately twice as great as in Figure 4.3b. The larger amplitude is due to the fact that the unstable oscillation had an opportunity to grow for twice as long in the former case as in the latter.

Figures 4.4a and b, and Figures 4.5a and b illustrate once again how the adaptive action of the FMC leads to a stable response in an otherwise unstable configuration.

4.3 Description of the Simulation

A block diagram of the system under consideration is shown in the Appendix in Figure A.1. The simulation can be conveniently

divided into three parts which are discussed independently below. Each part requires an input from, and generates an output to, one or both of the remaining parts. The three parts are:

1. The missile dynamic equations.
2. The compensation difference equations.
3. The FMC and its associated logic.

Each of these building blocks of the simulation will now be discussed.

Integration of the Missile Dynamic Equations

The equations which describe the dynamic behavior of the open-loop missile system are presented in the Appendix as Equations A.2 to A.11. These equations constitute a twelfth order system of linear ordinary differential equations with constant coefficients. These equations can be written in the canonical form

$$\frac{du_i}{dt} = F_i(u_1, \ldots, u_{12}, \delta_c)(i = 1, \ldots, 12) \qquad (4.1)$$

Note that the F functions are dependent on the quantity $\delta_c(t)$ (through Equation A.6), which is the output of the difference equation portion of the simulation. These equations were integrated by the standard fourth order Runga-Kutta integration technique.

Compensation Simulation

The equations describing the digital compensation which was used are presented in the Appendix as Equations A.13 to A.19. The difference equations are sixth order. Certain coefficients in the difference equations are under the control of the FMC. The compensation block requires an input, $\varepsilon_o(nT_s)$ from the simulation of the missile dynamic equations, and an input from the FMC which controls

the compensation parameters. An output, $\delta_c(t)$, is generated which is used in the missile dynamic equation simulation.

FMC Simulation

The FMC requires an input, $y(t)$, from the missile dynamic equation simulation. This signal is not available in continuous form but is sampled every T_y seconds. Each channel of the FMC performs in exactly the way described in Chapter III. Because we must now consider a large number of frequency measuring channels it is convenient to rewrite Equations 3.3, 3.4 and 3.5 in different notation. Since three different frequencies are being measured we define (for i=1,2,3)

ξ_i = the interchannel separation factor of the channels used to measure the i^{th} mode frequency

$\bar{\omega}_i$ = an a priori estimate of the i^{th} mode angular frequency (radians/second)

K_i = the number of channels being used to measure the i^{th} mode frequency (K_i must be odd)

N_i = the number of cycles of frequency ω_{ij} contained in the time interval θ_{ij}

where

$$\omega_{ij} = \bar{\omega}_i(1 + 2\xi_i)^{j-(K_i+1)/2} \quad (i=1,2,3) \quad (j=1,2,\ldots,K_i) \tag{4.2}$$

= the angular frequency of the j^{th} channel that is being used to measure the i^{th} mode frequency (radians/second)

and

$$\theta_{ij} = 2\pi N_i/\omega_{ij} \quad (i=1,2,3) \quad (j=1,2,\ldots,K_i) \tag{4.3}$$

= the time interval for the ij^{th} channel over which the stimulus is correlated with the local signal (seconds)

Since a square wave FMC is being used, the local multiplier signals are defined to be

$$f_{ij}(t) = \mathrm{sgn}(\sin \omega_{ij} t) \tag{4.4}$$

and

$$g_{ij}(t) = \mathrm{sgn}(\cos \omega_{ij} t) \tag{4.5}$$

where

$$\mathrm{sgn}(x) = \begin{cases} +1 & \text{if} \quad x \geqslant 0 \\ -1 & \text{if} \quad x < 0 \end{cases} \tag{4.6}$$

The ij^{th} channel produces an output every Θ_{ij} seconds. This output at time $t = k\Theta_{ij}$ (where k is an integer) is just

$$P_{ij}(k\Theta_{ij}) = (1/\Theta_{ij})^2 \left\{ \left\langle \int_{(k-1)\Theta_{ij}}^{k\Theta_{ij}} dt\, y(t)\ f_{ij}\lfloor t - (k-1)\Theta_{ij} \rfloor \right\rangle^2 + \left\langle \int_{(k-1)\Theta_{ij}}^{k\Theta_{ij}} dt\, y(t)\ g_{ij}\lfloor t - (k-1)\Theta_{ij} \rfloor \right\rangle^2 \right\} \tag{4.7}$$

Memory space is available for storage of all of the P_{ij} quantities (i.e., there are $\sum_{i=1}^{3} K_i$ storage locations used for this purpose). Each time an output P_{ij} becomes available it is stored in its allotted location (the previous P_{ij} value is therefore lost) and compared with a preselected threshold level L (it would be possible to have different threshold levels, L_i, for each of the modes being measured but this was not done in the simulation). If $P_{ij} < L$, no adaptive action takes place. If $P_{ij} \geqslant L$, the presence of a discrete frequency oscillation is indicated and its frequency is determined by a parabolic interpolation scheme as described below.

Suppose that the test statistic which has exceeded the threshold is P_{ks}. The quantity P_{km} is now determined by sorting through the P_{ks} (for s=1,..., K_s), where

$$P_{km} = \max_s P_{ks} \qquad (s=1,\ldots,K_s) \tag{4.8}$$

Since P_{km} is the largest of all the P_{ks} it follows that the discrete frequency, ω_k, which is present in the stimulus is closer to ω_{km} than to any of the other ω_{ij}. An estimate, ω_k^*, of ω_k is made by the FMC according to the following parabolic interpolation algorithm (see p. 77):

$$\omega_k^* = \omega_{km}(1 + \nu_k \xi_k) \tag{4.9}$$

where

$$\nu_k = (P_{k,m+1} - P_{k,m-1})/(2P_{km} - P_{k,m-1} - P_{k,m+1}) \tag{4.10}$$

It is now necessary to use the estimate of the k^{th} mode frequency to modify the digital compensation which is being used to gain stabilize that mode. The compensation difference equations are presented in the Appendix in Equations A.14 to A.16. In Z-transform notation these become

$$\varepsilon_k^*(z)/\varepsilon_{k-1}^*(z) = D_k(z) = (z^2 + a_{k1}z + a_{k2})/(z^2 + b_{k1}z + b_{k2}) \qquad (k=1,2,3) \tag{4.11}$$

A comparison with Equation 3.99 leads to the following expressions for the coefficients:

$$\begin{aligned} a_{k1} &= -2\cos\omega_k^* T_s && \text{(a)} \\ a_{k2} &= 1 && \text{(b)} \\ b_{k1} &= \rho_k a_{k1} && \text{(c)} \\ b_{k2} &= \rho_k^2 && \text{(d)} \end{aligned} \tag{4.12}$$

Since T_s and ρ_k are known fixed quantities, the value of ω_k^* supplied by Equation 4.9 is all that is needed to change the compensation coefficients. Actually, only a_{k1} and b_{k1} are altered. The compensation gain, A_o, is now recomputed using Equation A.17 which is repeated below for convenience:

$$A_o = \prod_{i=1}^{3} [(1 + b_{k1} + b_{k2})/(1 + a_{k1} + a_{k2})] \qquad (4.13)$$

In all the simulation runs made, the system was started with $a_{k1} = a_{k2} = b_{k1} = b_{k2} = 0$, and $A_o = 1$. Thus no compensation was in the loop. As the FMC measured the frequencies of the oscillations present in y(t) the necessary coefficients were modified as required by Equations 4.12 and 4.13. Thus the system adapted on the basis of real time measurements of its performance. Tables 4.1 and 4.2 present lists of the many parameters which are necessary to perform the system simulation described in this chapter.

TABLE 4.1

MISSILE EQUATION PARAMETERS

Parameter	Figure Number 4.1a,b	4.2a,b	4.3a,b,c	4.4a,b	4.5a,b
μ_α	2.34	2.34	2.34	2.34	2.34
μ_c	4.84	4.84	4.84	4.84	4.84
$\zeta_1, \zeta_2, \zeta_3$	0.005	0.005	0.005	0.005	0.005
$T\phi_{T1}/M$	145.0	145.0	145.0	145.0	145.0
$T\phi_{T2}/M$	-65.5	-65.5	-65.5	-65.5	-65.5
$T\phi_{T3}/M$	65.5	65.5	65.5	65.5	65.5
λ_{y1}	0.14	0.14	0.14	0.14	0.14
λ_{y2}	0.31	0.31	0.31	0.31	0.31
λ_{y3}	0.14	0.14	0.14	0.14	0.14
T_s	0.03	0.03	0.03	0.04	0.04
ω_E	∞	60.0	60.0	60.0	60.0
ζ_E	-	0.1	0.1	0.1	0.1
ω_H	35.0	15.0	15.0	10.0	10.0
ω_y	5.0	5.0	5.0	5.0	5.0

TABLE 4.1 (continued)

MISSILE EQUATION PARAMETERS

Parameter	Figure Number 4.1a,b	4.2a,b	4.3a,b,c	4.4a,b	4.5a,b
K_R/K_D	0.333	0.333	0.333	0.4	0.4
K_D	1.8	1.8	1.8	1.0	1.0
ω_1	17.0	17.0	17.0	10.0	10.0
ω_2	35.0	35.0	35.0	20.0	20.0
ω_3	53.0	53.0	53.0	30.0	30.0
λ_{P1}	0.125	0.125	0.125	0.125	0.125
λ_{P2}	0.2	0.2	0.2	0.3	0.3
λ_{P3}	0.045	0.045	0.045	0.1	0.1
λ_{R1}	-0.05	0.05	0.05	-0.05	-0.125
λ_{R2}	-0.05	0.05	0.05	-0.05	0.075
λ_{R3}	0.09	-0.09	-0.09	-0.1	-0.125

TABLE 4.2

FMC PARAMETERS

Parameter	Figure Number 4.1b	4.2b	4.3b	4.3c	4.4b	4.5b
N_1	3.0	6.0	3.0	6.0	6.0	6.0
N_2	6.0	6.0	6.0	6.0	6.0	6.0
N_3	6.0	6.0	6.0	6.0	6.0	6.0
ζ_1	0.08333	0.03333	0.6667×10^{-1}	0.3333×10^{-1}	0.3333×10^{-1}	0.3333×10^{-1}
ζ_2	0.04167	0.03333	0.4167×10^{-1}	0.4167×10^{-1}	0.3333×10^{-1}	0.3333×10^{-1}
ζ_3	0.04167	0.03333	0.4167×10^{-1}	0.4167×10^{-1}	0.3333×10^{-1}	0.3333×10^{-1}
$\bar{\omega}_1$	18.0	17.5	18.0	18.0	9.5	9.5
$\bar{\omega}_2$	37.0	36.0	37.0	37.0	19.5	19.5
$\bar{\omega}_3$	56.0	56.0	56.0	56.0	31.0	31.0
L	0.2	0.2	0.2	0.2	0.2	0.2
ρ	0.7	0.7	0.7	0.7	0.7	0.7
T_y	0.01	0.005	0.01	0.01	0.01	0.01

CHAPTER V

FINIS

5.1 Summary and Conclusions

A particular type of Adaptive Controller was synthesized which is capable of stabilizing dynamic systems which contain multiple lightly damped resonances. The controller acts to gain stabilize (see p. 43) the dynamic system by inserting into the control loop compensation which has zeros of transmission very close to the critical resonant frequencies. The zeros must be close enough to the open loop resonant frequency to insure that the loop gain at and near these frequencies is less than unity. Very little a priori knowledge is needed about the frequencies at which the resonances occur because the Adaptive controller itself measures these frequencies while the system is operating. It then adjusts its internal parameters on the basis of the measurements to insure that the overall system performance is satisfactory. Since the measurement process can be performed continually this adaptive control technique is applicable to systems whose resonant frequencies change slowly with time.

Both the measurement and compensation functions are performed by a digital computer. The resonant frequencies are measured by cross-correlating a signal generated by the dynamic system with a set of periodic signals whose frequencies span the intervals in which the resonant frequencies are known to occur. Thus the only information needed to allow successful operation of the Frequency Measuring Computer (FMC) is the frequency range of the resonant frequencies which are present. The necessary compensation is instrumented in a set of

difference equations which are stored in the digital computer. Certain coefficients which appear in these difference equations are adjusted according to logic which is programmed into the FMC.

The circumstances for which the proposed system can be successful are stated by the necessary and sufficient conditions derived in Section 3.3. The fact that the system actually performs as expected was demonstrated by a detailed simulation of a complete adaptive control system. The particular example studied in the simulation involved the design of an autopilot for a highly flexible ballistic missile where lightly damped structural resonances are a serious problem.

It is possible that certain resonant modes cannot be gain stabilized by the adaptive technique proposed in this investigation. This will be the case if the loop transmission at a resonant frequency is so large that the FMC errors prevent the insertion of adequate attenuation at this frequency. In the ballistic missile example, for instance, it is usually the case that the first mode resonance is too great to be gain stabilized by the proposed adaptive technique. It is necessary, therefore, to phase stabilize the first mode and apply the adaptive technique to higher mode resonances only.

A characteristic feature of the cancellation compensation which is employed is that it leads to very lightly damped (but stable) oscillations in the resulting closed loop system. These oscillations will not be detrimental provided that their amplitudes are not excessive. In fact, the presence of these oscillations with relatively constant amplitudes for long periods of time makes possible more accurate operation of the FMC than would otherwise be the case.

Although the proposed technique insures stability of the resonant modes at almost all times (there will be short intervals in which unstable oscillations can grow slowly until they are detected by the FMC), care must be taken in the overall system design to insure that the compensation that is introduced does not tend to destabilize the low frequency roots of the system. Thus quasi-stationary stability analyses must be performed for different times of flight, in the missile example, to insure that the low frequency (rigid body) performance of the system is satisfactory for the various possible compensation configurations which can occur.

5.2 Suggestions for Further Study

There are several possible areas in which additional effort could lead to interesting results. Further investigation of the properties of the FMC when noise is present in the stimulus is one such area. One would like to be able to make a precise statement regarding the capability of the FMC to measure the frequency of a sinusoidal component in the stimulus within a specified accuracy. The precise mathematical statement would involve the probability with which the accuracy could be obtained as a function of certain input parameters (e.g., signal to noise ratio, and parameters of the FMC such as ξ and N). It would then be possible to arrive at design criteria which would allow an optimization of this portion of the system.

The actual design and instrumentation of a special purpose computer to perform the frequency measuring function in an efficient manner would be of interest. Before this should be attempted, however,

it would be desirable to have the answers to the problems stated in the paragraph above.

A more detailed examination of the forms of digital compensation which can be used to stabilize the system once the resonant frequencies have been measured would be of interest. The compensation chosen was selected only because of its simplicity and is not optimum in any known sense. For example, it might be desirable to locate the zeros somewhat inside the unit circle instead of exactly on the unit circle as was done in this investigation. Or, if some information were available concerning the phase of the loop transmission at the resonant frequencies, it might be possible to increase the damping of the closed loop resonant poles by deliberately placing the compensation zeros slightly higher (or lower) in frequency than that which is measured.

Finally, a great deal more effort can be applied to develop a better understanding of the role adaptive systems should play in the control area. Any approach would be welcomed which can give one a quantitative measure of the effectiveness of an "adaptive" system compared with one of more conventional design. Attention should be directed toward the basic limitations of the parameter measurement function in terms of accuracies obtainable and time required to make the measurements.

APPENDIX

APPROXIMATE EQUATIONS OF MOTION FOR A FLEXIBLE BALLISTIC MISSILE

Large rocket vehicles are excellent examples of physical systems which exhibit multiple lightly damped resonances. Equations of motion useful for analyzing the stability of a control system for such a vehicle are presented in this appendix. Equations of motion based on a more exact model of the real physical system are available elsewhere (20).

The equations presented here describe an idealized model of the true physical system since several effects of importance have been neglected in the interest of simplicity. The characteristic feature of the flexible missile control problem, i.e., the presence of multiple lightly damped resonances, is included in the equations. Had the equations of motion been derived for a much more realistic model, their solution would have taken a prohibitive amount of time on the digital computer which was available to the author (a Burroughs 220). Among the more important approximations implicit in the form of the equations are the following:

1. The effects of propellent sloshing are neglected.
2. Only three bending modes are considered.
3. The generalized force corresponding to each bending mode consists solely of the engine thrust component acting normal to the centerline of the missile multiplied by the corresponding mode deflection at the point of thrust application.
4. The mass and inertia of the rocket engine do not produce forces or moments on the missile due to inertial reaction effects (i.e., the "tail wags dog" effect is neglected).

In addition, the following approximations were made:

5. The hydraulic power servo transfer function is represented by a lag and a quadratic pole pair due to the dynamic load of the rocket engine.

6. Aerodynamic effects are represented by a normal force acting at the center of pressure of the missile.

7. The effects of the equation stating the balance of forces normal to the missile's nominal trajectory are neglected.

8. The dynamics of all sensing devices are neglected.

9. Non-linear and second order effects are neglected.

10. System parameters vary slowly compared to the response time of the control system and are considered to be constants at any particular time of flight.

As stated in Assumption 2, the normal mode expansion of the deflection function, $u(\xi, t)$, is truncated after three terms. Thus we write (see list of Definition of Symbols)

$$u(\xi, t) = \sum_{i=1}^{3} q_i(t)\, \phi_i(\xi) \qquad (A.1)$$

For many cases of interest this is a reasonable approximation. In certain applications, however, the approximation is not valid. Nonetheless, the resulting system of equations describes a model of the true physical system which is adequate for our simulation study.

The geometry of the system to be analyzed was shown in Figure 2.1 and most of the pertinent parameters were discussed there (see pp. 25ff). Since the derivation of the equations of motion is a relatively routine application of Newton's laws and Lagrange's equations, the details will not be spelled out here. The resulting

linearized equations which describe the complete closed loop control system are presented below in the form which was used in the simulation. A block diagram representation of these equations is shown in Figure A.1.

Missile Dynamic Equations

Rigid Body Dynamics

$$\ddot{\theta} = \mu_c \delta + \mu_\alpha \theta \tag{A.2}$$

Bending Resonances

$$\ddot{q}_i = T\phi_{Ti} \delta/M - 2\zeta_i \omega_i \dot{q}_i - \omega_i^2 q_i \qquad (i=1,2,3) \tag{A.3,A.4,A.5}$$

Hydraulic Actuator and Engine Resonance

$$\dot{\delta}_A = \omega_H(\delta_c - \delta_A) \tag{A.6}$$

$$\ddot{\delta} = \omega_E^2 \delta_A - 2\zeta_E \omega_E \dot{\delta} - \omega_E^2 \delta \tag{A.7}$$

Sensor Outputs

$$\theta_P = \theta - \sum_{i=1}^{3} \lambda_{Pi} q_i \tag{A.8}$$

$$\dot{\theta}_R = \dot{\theta} - \sum_{i=1}^{3} \lambda_{Ri} \dot{q}_i \tag{A.9}$$

FMC Stimulus

$$\dot{\theta}_y = \dot{\theta} - \sum_{i=1}^{3} \lambda_{yi} \dot{q}_i \tag{A.10}$$

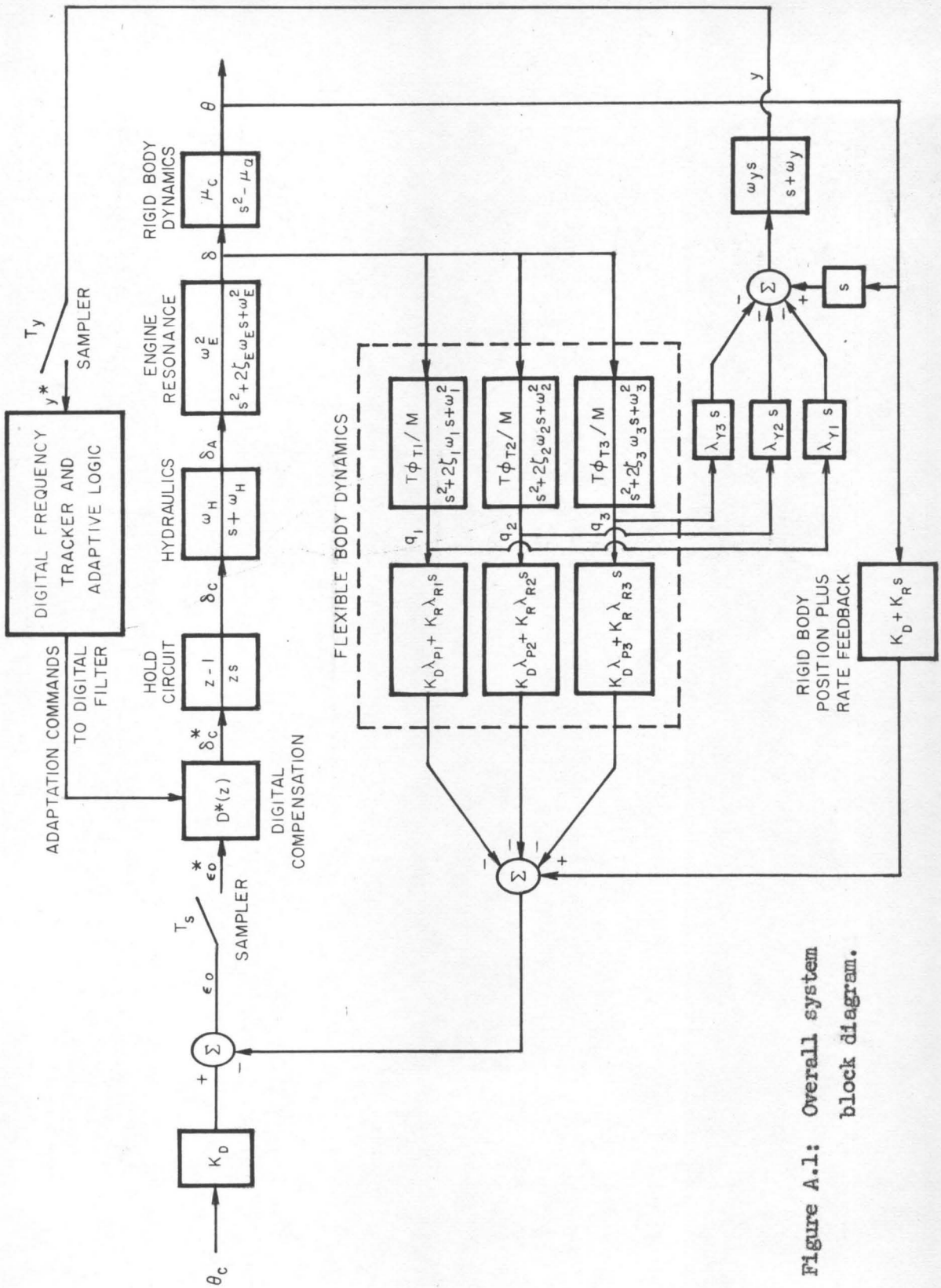


Figure A.1: Overall system block diagram.

$$\dot{y} = \ddot{\theta}_y - \omega_y y \tag{A.11}$$

Autopilot Error Signal

$$\varepsilon_o = K_D (\theta_c - \theta_P) - K_R \dot{\theta}_R \tag{A.12}$$

Digital Compensation Equations

$$\varepsilon^*_{o,n} = \begin{cases} \varepsilon_o(n T_s) & \text{for } t = n T_s \\ 0 & \text{elsewhere} \end{cases} \tag{A.13}$$

$$\varepsilon^*_{k,n} = \varepsilon^*_{k-1,n} - \sum_{j=1}^{2} (b_{kj} \varepsilon^*_{k,n-j} - a_{kj} \varepsilon^*_{k-1,n-j})$$

$$(k=1,2,3) \tag{A.14,A.15,A.16}$$

$$A_o = \prod_{i=1}^{3} [(1 + b_{k1} + b_{k2})/(1 + a_{k1} + a_{k2})] \tag{A.17}$$

$$\delta^*_{c,n} = A_o \varepsilon^*_{3,n} \tag{A.18}$$

$$\delta_c(t) = \delta^*_{c,n} \quad \text{for } nT_s \leq t < (n + 1)T_s \tag{A.19}$$

PARTIAL LIST OF SYMBOLS

A	amplitude of discrete frequency sinusoid present in $y(t)$
A'	FMC accuracy when interpolation scheme is not used
A^*	FMC accuracy when parabolic interpolation scheme is used
A_i	low frequency gain of $D_i(z)$ (defined by Equation 3.100)
A_o	low frequency gain of the adaptive digital compensation = $A_1A_2A_3$
a_{kj}, b_{kj}	difference equation coefficients of the adaptive filter
c_{ik}	cosine component of the output of the i^{th} FMC channel at time $2\pi N_i k/\omega_i$ (defined by Equation 3.8)
D	signal to noise ratio in $y(t)$ (defined by Equation 3.84)
$D_i(z)$	compensation introduced to stabilize the i^{th} resonant mode (defined by Equation 3.99)
d	lowest possible value of an unknown frequency ω (see p. 63) (radians per second)
$e(t)$	error signal in general control system
$\vec{e}_x$, $\vec{e}_y$	defined on p. 25
$\vec{e}_X$, $\vec{e}_Y$	defined on p. 25
FMC	Frequency Measuring Computer
$F(\eta, r_{ik})$	random variable defined by Equation 3.28
I	1. moment of inertia of the missile (defined on p. 25) (slug ft.^{2}) 2. the larger of the two quantities I_s and I_c (defined by Equation 3.78)

I_s — defined by Equation 3.71

I_c — defined by Equation 3.72

K —
1. the number of FMC channels necessary to measure an unknown frequency to within a specified accuracy (see p. 63)
2. the magnitude of the digital compensation introduced to stabilize the i^{th} resonant mode, evaluated at the true open loop natural frequency, ω_i, of that mode (see p. 115)

K_D — attitude position gain constant

K_R — attitude rate gain constant

L — FMC threshold detection level

M —
1. mass of the missile (defined on p. 25) (slugs)
2. magnitude of the loop gain at the open loop resonant frequency ω_i, excluding the effects of $D_i(z)$ (defined on p. 115)

N — number of cycles of local multiplier frequency over which any particular FMC channel integrates

N_i — number of cycles of frequency ω_i contained in the time interval Θ_i

$n(t)$ — normalized random variable defined by Equation 3.55

$n'(t)$ — stationary additive noise present in $y(t)$

N_α — aerodynamic normal force per unit angle of attack (lb./radian)

P_{ik} — output of i^{th} FMC channel during time interval $k\Theta_i \leq t < (k+1)\Theta_i$. Used as test statistic to detect the presence of a discrete frequency oscillation in $y(t)$; also used to estimate the frequency of the signal present in $y(t)$

P^o	normalized fundamental output component of any frequency measuring channel (defined by Equation 3.29)	
$\bar{P}$	Output of any frequency measuring channel when y(t) contains no additive noise (defined by Equation 3.73)	
$q_i(t)$	i^{th} generalized bending coordinate	(ft.)
R	ratio of highest to lowest possible values of an unknown frequency ω (see p. 63)	
$R'(\tau)$	autocorrelation function of the noise in y(t) (defined by Equation 3.53)	
r_{ik}	random variable defined by Equation 3.26	
S(t)	normalized signal component of y(t) defined by Equation 3.54	
s	Laplace transform complex variable	
s(t)	signal component of y(t)	
s_{ik}	sine component of the output of the i^{th} FMC channel at time $2\pi N_i k/\omega_i$ (defined by Equation 3.7)	
T	rocket engine thrust	(lb.)
T_s	sampling interval at input to digital compensation	(seconds)
T_y	sampling interval at input to FMC	(seconds)
$u(\xi,t)$	bending deflection function (defined on page 27)	(ft.)
$\vec{V}$	velocity of the missile's center of mass	(ft./second)
W_N	average power of the noise component of y(t) (defined by Equation 3.68)	

W_S — average power of the signal component of y(t) (defined by Equation 3.81)

x — normalized frequency variable = $N\eta$ (see p. 69)

$y(t)$ — input "stimulus" to the FMC

z — Z-transform variable

α — 1. angle of attack (radians)
2. normalized frequency variable defined by Equation 3.43c

α_i — i^{th} adjustable control parameter in general adaptive system (see Equations 1.15 and 3.1)

β_c — angular position around unit circle (in Z-plane) of i^{th} closed loop frequency of oscillation (see p. 115) (radians)

β_c^* — angular position around unit circle (in Z-plane) at which FMC places compensation zeros to stabilize i^{th} resonant mode (see p. 115) (radians)

β_i — 1. i^{th} critical system parameter in general adaptive system (see Equations 1.15 and 3.1)
2. angular position around unit circle (in Z-plane) of zeros introduced to compensate for oscillations at the i^{th} resonant frequency (see p. 101) (radians)
3. angular position around unit circle (in Z-plane) of open loop resonant frequency of the i^{th} resonant mode (see p. 115) (radians)

γ — normalized variable = $N\xi$ (defined by Equation 3.44)

Δ — per unit measurement error due to FMC (see p. 116)

δ — angle between line of action of T and the deflected elastic axis at the point of thrust application (defined on p. 27) (radians)

δ_c — engine deflection commanded by autopilot (radians)

δ_A equivalent angular displacement at the base of the hydraulic actuator lever arm (radians)

ε per unit measurement error due to difference between open loop natural frequency and closed loop frequency of oscillation (see p. 116)

$\varepsilon^*_{k,n}$ sampled signal in digital computer at time $t = nT_s$

ζ_i damping of i^{th} resonance

ζ_E damping of engine resonance

η per unit difference between the frequency of the sinusoidal signal present in y(t) and the center frequency of a particular FMC channel (see p. 68)

$\eta_{1/2}$ normalized "half power" frequency of an FMC channel (see p. 75)

θ angle between inertial reference line and undeformed elastic axis of a missile (see Figure 2.1) (radians)

θ_P angular position measured by a position gyro located at position $\xi = \xi_P$ (see Equation A.8) (radians)

$\dot{\theta}_R$ angular rate measured by a rate gyro located at position $\xi = \xi_R$ (see Equation A.9) (radians/second)

θ_c attitude commanded by guidance loop (radians)

θ_i integration time of i^{th} FMC channel (see p. 58) (seconds)

θ_{ij} integration time of ij^{th} FMC channel (see p. 140) (seconds)

$\dot{\theta}_y$ angular rate measured by a rate gyro located at position $\xi = \xi_y$ (see Equation A.10)

$\lambda_i(\xi)$ bending mode slope ($= d\phi_i(\xi)/d\xi$) for i^{th} mode ($ft.^{-1}$)

μ	vector (in the Z-plane) between the open and closed loop i^{th} mode pole locations	
ν , ν_k	normalized correction in frequency estimated by the FMC by parabolic interpolation method (see Equations 3.46 and 4.10)	
ξ	1. distance from the center of mass of the missile to any point on the undeflected elastic axis (see p. 25) 2. per unit accuracy of FMC when interpolation scheme is not used (see p. 63) 3. interchannel spacing factor of FMC channels (see p. 65)	(ft.)
ξ_i	interchannel spacing factor for FMC channels being used to measure the i^{th} resonant frequency (see p. 140)	
$\rho(\tau)$	normalized autocorrelation function defined by Equation 3.70	
σ_P	standard deviation of P_{ik} (see p. 87)	
ϕ, ϕ'	random phase of sinusoidal component of y(t)	(radians)
$\phi_i(\xi)$	bending mode deflection for i^{th} mode	(dimensionless)
ϕ_{ik}	random variable related to ϕ; defined by Equation 3.22	(radians)
ψ	normalized variable ($= \omega T_s$)	(radians)
ω	frequency of sinusoidal signal present in y(t)	(radians/second)
ω^*	FMC estimate of the frequency ω	(radians/second)
ω_E	natural frequency of engine resonance	(radians/second)
ω_H	hydraulics break frequency	(radians/second)
ω_i	1. i^{th} mode natural frequency 2. center frequency of i^{th} FMC channel	(radians/second) (radians/second)

$\bar{\omega}_i$	a priori estimate of i^{th} mode resonant frequency (see p. 140)	(radians/second)
ω_c	closed loop frequency of oscillation (see p. 115)	(radians/second)
ω_c^*	FMC estimate of the frequency ω_c (see p. 115)	(radians/second)
ω_{ij}	center frequency of the ij^{th} FMC channel (see p. 140)	(radians/second)

REFERENCES

1. Truxal, J. G.: Automatic Feedback Control System Synthesis, McGraw-Hill Book Company, Inc., New York, 1955.

2. Newton, G. C., Jr., Gould, L.A., and Kaiser, J. F.: Analytical Design of Linear Feedback Controls, John Wiley and Sons, Inc., New York, 1957.

3. Jury, E. I.: Sampled-Data Control Systems, John Wiley and Sons, Inc., New York, 1958.

4. Ragazzini, J. R., and Franklin, G. F.: Sampled-Data Control Systems, McGraw-Hill Book Company, Inc., New York, 1958.

5. Kalman, R. E., and Bertram, J. E.: "Control System Analysis and Design Via the 'Second Method' of Lyapunov. I. Continous-Time Systems," Journal of Basic Engineering, Transactions of the American Society of Mechanical Engineers, Vol. 82, Series D, No. 2, June 1960, pp. 371-393.

6. Kalman, R. E., and Bertram, J. E.: "Control System Analysis and Design Via the 'Second Method' of Lyapunov. II. Discrete-Time Systems," Journal of Basic Engineering, Transactions of the American Society of Mechanical Engineers, Vol. 82, Series D, No. 2, June 1960, pp. 394-400.

7. Aseltine, J. A., Mancini, A. R., and Sarture, C. W.: "A Survey of Adaptive Control Systems," Institute of Radio Engineers Transactions on Automatic Control, PGAC-6, Dec. 1958, pp. 102-108.

8. Gregory, P. C. (Editor): Proceedings of the Self-Adaptive Flight Control Symposium, W.A.D.C. Technical Report 59-49, March 1959, p. 199 and p. 348.

9. Ashby, W. R.: Design for a Brain, 2nd Ed., John Wiley and Sons, Inc., New York, 1960.

10. Goodman, T. P., and Hillsley, R. H.: "Continuous Measurement of Characteristics of Systems with Random Inputs: A Step Toward Self-Optimalizing Control," American Society of Mechanical Engineers Transactions, Vol. 80, November 1958, pp. 1839-1848.

11. Shinbrot, M.: "On the Analysis of Linear and Non-Linear Systems," American Society of Mechanical Engineers Transactions, Vol. 79, April 1957, pp. 547-552.

12. Kalman, R. E.: "Design of a Self-Optimizing Control System," American Society of Mechanical Engineers Transactions, Vol. 80, February 1958, pp. 468-478.

13. Anderson, G. W., Aseltine, J. A., Mancini, A. R., and Sarture, C. W.: "A Self-Adjusting System for Optimum Dynamic Performance," Institute of Radio Engineers National Convention Record, Part 4, 1958, pp. 182-190

14. Staffin, R.: Executive-Controlled Adaptive Systems, Doctoral Dissertation at the Polytechnic Institute of Brooklyn, 1959.

15. Staffin, R.: "Executive-Controlled Adaptive Systems," American Institute of Electrical Engineers Transactions, Part II, Vol. 78, January 1960, pp. 523-530.

16. Wilts, C. H.: Principles of Feedback Control, Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1960.

17. Davenport, W. B., Jr., and Root, W. L.: An Introduction to the Theory of Random Signals and Noise, McGraw-Hill Book Company, Inc., New York, 1958.

18. Laning, J. H., Jr., and Battin, R. H.: Random Processes in Automatic Control, McGraw-Hill Book Company, Inc., New York, 1956.

19. Hildebrand, F. B.: Introduction to Numerical Analysis, McGraw-Hill Book Company, Inc., New York, 1956, p. 237.

20. Young, D.: "Generalized Missile Dynamics Analysis II. Equations of Motion," Space Technology Laboratories Technical Report, GM-TR-0165-00359, Los Angeles, April 7, 1958.