

PERFORMANCE LIMITATIONS OF VIBRATORY GYROSCOPES

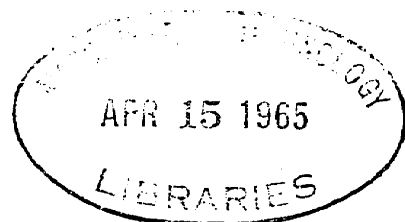
by

RICHARD WRIGHT BUSH

B.S., Purdue University
(1959)

S.M., Massachusetts Institute of Technology
(1961)

E.E., Massachusetts Institute of Technology
(1962)



SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
January, 1965

Signature of Author _____

Department of Electrical Engineering, January 22, 1965

Certified by _____

Thesis Supervisor

Accepted by _____

Chairman, Departmental Committee on Graduate Students

PERFORMANCE LIMITATIONS OF VIBRATORY GYROSCOPES

by

Richard Wright Bush

Submitted to the Department of Electrical Engineering on January 25, 1965
in partial fulfillment of the requirements for the degree of Doctor of Science.

ABSTRACT

Vibratory gyroscopes are angular motion sensors in which vibratory forces produce the output signals. Potentially, these devices have important advantages over conventional rotary type gyroscopes; however, their present day precision is relatively low and, as a result, they have not received appreciable use. The basic types of vibratory gyroscopes are reviewed first in this report and then the theoretical ultimate and practical performance limitations of rate type instruments are examined in detail. It is shown that the most important factor which is presently limiting the performances of vibratory rate gyroscopes is unwanted cross coupling between the basic drive member and the sensing system. Therefore, to improve the performances of these instruments it is necessary to reduce the magnitude of the cross coupling or the errors associated with it. Although both of these methods are examined in this report, the major emphasis is on the latter. The two ways of reducing the cross-coupled errors in vibratory rate gyroscopes which are examined are called "double modulation" and "optimum processing of suppressed carrier signals."

Double modulation is a process by which the driving and sensing frequencies in vibratory gyroscopes are separated. It is achieved by rotating or vibrating the basic drive member of a more conventional vibratory gyroscope about a rate insensitive axis. When the frequencies are separated in this manner, the errors associated with cross coupling between the drive member and the sensing system are reduced. Theoretical and experimental evidence in support of the double modulation concept are presented in this report.

Optimum processing of suppressed carrier signals is an analytical design procedure for synchronous suppressed carrier systems, such as vibratory gyroscopes. The techniques which are developed in this report treat the problem of minimizing the mean square error when the signal processor is a linear time-varying filter, and the harmonic errors are weighted either equally with the other errors or not at all. The sources of error which are examined in the suppressed carrier problem are quadrature signals, unmodulated additive noise and drifts of the carrier frequency, the parameters of the fixed elements and the phase of the demodulator synchronizing signal.

Thesis Supervisor: George C. Newton, Jr.

Title: Professor of Electrical Engineering

ACKNOWLEDGEMENT

The author gratefully acknowledges the support of this research by the National Aeronautics and Space Administration under NASA Research Grant NsG 149-61, and also the supervision and guidance of this research by Professors George C. Newton, Jr., Wilbur B. Davenport, Jr., and William W. Seifert.

The author wishes to thank his wife, Suzanne, for her patience and understanding during the writing and rewriting of this document and also for her financial support for the last four years.

Finally, the author wishes to thank the members of the Electronic Systems Laboratory who typed the document--Mrs. Ionia Lewis, Mrs. Laurel Retajczyk and Mrs. Susan L. Lackner--and also those who did the drafting--Mr. Harold Tonsing, Mr. Norman Darling, Mr. Arthur Giordani and Miss Janet McGonnigal--with special thanks to Mrs. Laurel Retajczyk.

CONTENTS

CHAPTER I	INTRODUCTION	<u>page</u>	1
CHAPTER II	SUMMARY AND CONCLUSIONS		5
	2.1 BACKGROUND		5
	2.2 SUMMARY AND CONCLUSIONS		6
CHAPTER III	BRIEF SURVEY OF THE BASIC TYPES OF VIBRATORY GYROSCOPES		13
	3.1 INTRODUCTION		13
	3.2 CORIOLIS FORCE		13
	3.3 BASIC DRIVING SCHEMES		14
	3.4 VIBRATORY RATE INTEGRATING GYROSCOPES		15
	3.5 BASIC TYPES OF VIBRATORY RATE GYROSCOPES		19
	3.5.1 Introduction		19
	3.5.2 Rotary Drive Instruments		20
	3.5.3 Vibratory Drive Instruments		25
	3.5.4 Instruments Employing Double Modulation		36
	3.6 SUMMARY		40
CHAPTER IV	ULTIMATE PERFORMANCE LIMITATIONS IN VIBRATORY GYROSCOPES		44
	4.1 INTRODUCTION		44
	4.2 SIGNAL PROCESSING IN VIBRATORY GYROSCOPES		44
	4.3 SENSING SYSTEM IN VIBRATORY RATE GYROSCOPES		48
	4.4 THERMAL NOISE IN VIBRATORY GYROSCOPES		52
	4.5 NORMALIZED CORIOLIS TORQUES (FORCES) IN VIBRATORY RATE GYROSCOPES		54
	4.6 STRAIN CONSTRAINTS IN VIBRATORY GYROSCOPES		59
	4.7 THEORETICAL ULTIMATE PERFORMANCE OF VIBRATORY RATE GYROSCOPES		64

CONTENTS (Continued)

	<u>page</u>
CHAPTER V PRACTICAL PERFORMANCE LIMITATIONS IN VIBRATORY RATE GYROSCOPES	72
5.1 INTRODUCTION	72
5.2 TUNING FORK GYROSCOPE	73
5.2.1 History of Tuning Fork Gyroscopes	73
5.2.2 Cross Coupling	74
5.2.3 Time-Varying Cross Coupling	80
5.2.4 Differential Frequency Drifts	86
5.2.5 Differential Phase Shift Drift	90
5.2.6 Tuning Fork Gyroscope Summary	94
5.2.7 Comments	98
5.3 CROSS COUPLING IN VIBRATORY RATE GYROSCOPES	98
5.4 CONCLUSIONS	104
CHAPTER VI REDUCTION OF ERRORS IN VIBRATORY GYROSCOPES BY DOUBLE MODULATION	107
6.1 INTRODUCTION	107
6.2 DOUBLE-MODULATED TUNING FORK GYROSCOPE	108
6.3 SOME ASPECTS OF DOUBLE MODULATION	111
6.4 SOURCES OF ERROR IN VRG's EMPLOYING DOUBLE MODULATION	112
6.5 EXPERIMENTAL APPARATUS	117
6.6 EXPERIMENTAL DATA	124
6.7 CONCLUSIONS	132
CHAPTER VII OPTIMUM PROCESSING OF SUPPRESSED CARRIER SIGNALS	139
7.1 INTRODUCTION	139
7.2 STATEMENT OF THE PROBLEM	141
7.3 FREE CONFIGURATION PROBLEM WITH IDEAL COMPONENTS	145
7.3.1 Review of Linear Time-Varying Filter Theory	145
7.3.2 Free Configuration Problem Integral Equation	149
7.3.3 Special Case When $\phi_{mm}(\tau) = \phi_{qq}(\tau)$	150

CONTENTS (Continued)

7.3.4	General Case When $\phi_{mm}(\tau) \neq \phi_{qq}(\tau)$	153
7.3.5	Upper Bound on the Minimum Mean Square Error	161
7.3.6	Lower Bound on the Minimum Mean Square Error	165
7.3.7	Examples of the Developed Methods	167
7.3.8	Summary	173
7.4	PROBLEM WITH FIXED ELEMENTS LOCATED BEFORE THE SIGNAL PROCESSOR	175
7.4.1	Unrestricted Optimum Filter Problem	175
7.4.2	Restricted Optimum Filter Problem	176
7.4.3	Upper Bound on the Minimum MSE in the Restricted Problem	183
7.4.4	Lower Bound on the Minimum MSE in the Restricted Problem	185
7.4.5	Summary	188
7.5	PROBLEM WITH FIXED ELEMENTS LOCATED AFTER THE SIGNAL PROCESSOR	189
7.6	SATURATION CONSTRAINTS ON SIGNALS FOLLOWING THE OPTIMUM LTV FILTER	194
7.7	RANDOM DRIFT OF THE CARRIER FREQUENCY	199
7.7.1	Optimization without Fixed Elements	199
7.7.2	Optimization with Fixed Elements	203
7.7.3	Conclusions	206
7.8	RANDOM DRIFT OF THE FIXED ELEMENTS	207
7.9	RANDOM PHASE SHIFTS ON THE REFERENCE SIGNAL	208
7.10	SUMMARY OF OPTIMUM PROCESSING OF SUPPRESSED CARRIER SIGNALS	211

CONTENTS (Continued)

CHAPTER VIII	EXAMPLE OF AN OPTIMUM FILTER FOR A VIBRATORY RATE GYROSCOPE	212
8.1	INTRODUCTION	212
8.2	VIBRATORY DRIVE, VIBRATORY RATE GYROSCOPE EXAMPLE	214
8.3	SUMMARY	221
APPENDIX A	THEORY OF OPERATION OF TUNING-FORK GYROSCOPE EMPLOYING DOUBLE MODULATION	222
APPENDIX B		227
APPENDIX C	CROSS COUPLING COMPARISON	234
APPENDIX D	OPTIMIZATION PROBLEM WITH FIXED ELEMENTS FOLLOWING THE OPTIMUM FILTER	235
REFERENCES		239
BIOGRAPHICAL SKETCH		243

LIST OF FIGURES

		<u>page</u>
3.1	Basic Drive Schemes	17
3.2	Vibrating String Gyroscope	17
3.3	Rotary Drive, Drive Frequency Output Device	24
3.4	Rotary Drive, Double Frequency Output Device	24
3.5	Acoustic Gyroscope	30
3.6	Single Axis, Vibratory Drive, Drive Frequency Output Device	30
3.7	Two Axis, Vibratory Drive, Drive Frequency Output Device	33
3.8	Single Axis, Vibratory Drive, Double Drive Frequency Output Device	39
3.9	Single Axis, Vibratory Drive, Drive Frequency Odd Harmonic Output Device	39
3.10	Double Modulated Single Axis Device	39
3.11	Vibratory Gyroscopes	41
4.1	Approximation for a Suppressed Carrier System	47
4.2	VRG Sensing System Model	51
4.3	VRG Sensing System Employing Feedback	51
4.4	Approximately Equivalent Thermal Noise Representation	56
4.5	Model of a Vibratory Rate Gyroscope	56
4.6a	Rotary Drive, Drive Frequency Output Device	63
4.6b	Vibratory Drive, Drive Frequency Output Device	63
5.1	Idealized Double Tuning Fork	76
5.2	Simple Phase Shifter	92
6.1	Double Modulation of Tuning Fork	109
6.2	Tuning-Fork Unit	118
6.3	Experimental Double Tuning Fork	120
6.4	Double Modulation Carriage	122
6.5	Block Diagram of the Experimental Electronics	123
6.6	Experimental Data Without Double Modulation	126
6.7	Experimental Data With Double Modulation	127
6.8	Drift Data Without Double Modulation and Without Temperature Control	131
6.9	Drift Data Without Double Modulation and With Temperature Control	131
6.10	Drift Data With Double Modulation and Without Temperature Control	133
7.1	Suppressed Carrier Problem	152
7.2	Optimum Filter in the Special Case	152

LIST OF FIGURES (Continued)

7.3	Alternate Form for the Optimum Filter in the Special Case	152
7.4	Graphical Spectrum Factorization	157
7.5	Series Form of the Optimum Filter	159
7.6	Sampling Demodulator and Optimum Discrete Filter	166
7.7	Ideal Demodulator and Optimum Linear Filter	166
7.8	Fixed Element Before the Signal Processor	178
7.9	Fixed Element Before a Restricted Signal Processor	178
7.10	Practical Sensitivity Functions	178
7.11	Fixed Element Following the Signal Processor	192
7.12	Periodic Multipliers	192
7.13	Saturating Signal Following the Signal Processor	197
8.1	Vibratory Gyroscope Model	213
8.2	Two Components of the Mean Square Error in the Vibratory Gyroscope Example	219
A.1	Instrument Coordinate Systems	223
B.1	Tuning Fork Drive	228
B.2	Vibration Detector	229
B.3	Error Angle Detector	230
B.4	SCR Motor Control	231
B.5	First Demodulator	232
B.6	Second Demodulator	233

CHAPTER I

INTRODUCTION

Broadly speaking, vibratory gyroscopes are angular motion sensors in which vibratory forces or torques produce the output signals. This is in contrast to conventional rotating type gyroscopes in which the forces or torques producing the output signals are steady for constant applied angular rates. The potential advantages of vibratory gyroscopes over conventional gyroscopes are primarily low power consumption, low production costs, long life and high reliability. Up to the present time, the precision of vibratory gyroscopes has been relatively low and no instruments have successfully been marketed.* The major factors which limit the performance are common to all vibratory rate gyroscope configurations employing macroscopic phenomena and are the subject of this thesis. Instruments employing microscopic phenomena, such as nuclear magnetic resonance or electromagnetic waves, for angular rate measurements are not considered because they are not generally classified as vibratory gyroscopes.**

Vibratory type angular rate sensors are not conceptually new. They are found on house flies in the form of vibratory rods located behind the wings^{1***} and Foucault's pendulum,² which was demonstrated in 1851, was a vibratory rotation sensor. Nevertheless, vibratory type gyroscopes are relatively unknown compared to the conventional rotary type gyroscopes presently used for stabilization and navigation. The main reason for this is that vibratory gyroscopes have received relatively little attention in the past few decades

* This comment refers to November 1964. The author realizes that several vibratory gyroscopes will probably be marketed in the next 5 years.

** They are presently classified as exotic gyroscopes.

*** Superscripts refer to number items in the Bibliography.

because conventional gyroscopes have always been considered better suited for the inertial guidance systems of airplanes, missiles and submarines and it has been these applications toward which the majority of the gyroscope research and development effort has been directed. As a result, the drift of present day conventional gyroscopes are phenomenally low compared to the drift of conventional gyroscopes of a decade ago, whereas the drift of vibratory gyroscopes have changed very little over the past two decades and presently are orders of magnitude larger than the drift of conventional gyroscopes of similar size.

In the past few years, interest in vibratory and exotic types of gyroscopes has been increasing. The major reasons for this are the increased emphasis being placed on reliability in military equipment and the advent of new applications for gyroscopes in which the specifications differ greatly from those of the guidance systems for which most present day rotary type gyroscopes have been developed. Many of these new applications are in space where the environment differs from that on earth and great emphasis is placed on reliability over a long* period of time and low operating power. In other new applications emphasis is on one or more of the following characteristics; low cost, ruggedness, wide bandwidth, large dynamic range, ability to start and stop numerous times, and often extremely high precision is not one of the requirements. In many of these applications vibratory gyroscopes probably satisfy most of the specifications better than conventional gyroscopes; however, the relatively low precision of present day vibratory gyroscopes has discouraged people from even considering them.

* This is in contrast with the emphasis placed on reliability over a short period of time for the guidance systems of some missiles.

At the present time it is difficult to estimate what the precision of these two types of gyroscopes will be in the future. In view of the large discrepancy in the engineering efforts expended to date on the two types, however, it is apparent that present day precision* of vibratory and conventional gyroscopes may not necessarily be the best criterion upon which to estimate future performance. A criterion which may be more useful for this purpose is the standard deviation of the output drift which is caused by thermal fluctuations in the mechanical components of the instruments because this noise is the ultimate performance limitation in all instruments. Although the drift rates of the best conventional gyroscopes are well above the thermal fluctuation limit, this fundamental barrier has been approached in other instruments³ (galvanometers, electronic circuits, etc.) and conceivably may be approached in gyroscopic instruments eventually.

When the thermal fluctuation noise in vibratory and conventional gyroscopes of similar size are compared, rotary type gyroscopes generally are approximately an order of magnitude⁴ superior. This is not discouraging though, because vibratory gyroscopes with drift rates anywhere near the thermal fluctuation limit would be extremely useful because of their potential ruggedness, low power requirements and longevity. It is the unfortunate fact, however, that the drift rates of present day vibratory gyroscopes are generally more than three orders of magnitude larger than the ultimate threshold

* All interesting numbers for conventional and vibratory gyroscopes are classified so the reader must apply a suitable classification factor to the following numbers which are typical of those found in the open literature.

Conventional gyroscopes - 10^{-2} °/hr

Vibratory gyroscopes - 1 °/hr (averaged over a day or more)

predicted on the basis of thermal fluctuations and strain limitations in materials. This indicates that factors other than thermal fluctuations are presently limiting the performances of vibratory rate gyroscopes.

On the basis of existing data on various types of vibratory rate gyroscopes, it appears that the most important of these "other factors" are practical problems associated with the stability of materials and the limited tolerances to which mechanical parts can be made. In view of the many orders of magnitude improvement obtained in past years on conventional gyroscopes as a result of concentrated engineering efforts on similar "practical problems," it is certainly worthwhile examining all the factors that limit the performances of vibratory rate gyroscopes to determine what can be done to improve their performances. These factors are examined in this thesis after a brief survey of the basic types of vibratory gyroscopes is presented. Then two methods for improving the performances of vibratory gyroscopes are examined in detail. The first is called double modulation and the second is called optimum processing of suppressed carrier signals. A by product of this latter research is a filter theory for synchronous suppressed carrier systems which treats the practical problems of carrier frequency drift, drift of component parameters and ~~phase~~ of the carrier frequency reference signal available to the demodulator.

CHAPTER II

SUMMARY AND CONCLUSIONS

2.1 BACKGROUND

Vibratory gyroscopes are angular motion sensors in which vibratory forces or torques produce the output signals. The numerous configurations that have been proposed for these instruments can all be classified into a few basic types. The classification criteria used for this purpose in Chapter III are: first, whether an instrument is basically a rate or rate integrating device; second, whether the basic drive scheme is rotary or vibratory; and third, whether the carrier frequency of the suppressed carrier output signals containing the angular position or velocity information is equal to the drive frequency, a harmonic of the drive frequency or a linear combination of the harmonics of two drive frequencies. Although certain types of rate integrating instruments have been studied by others, the present research is restricted to vibratory rate gyroscopes.

Several important types of vibratory rate gyroscopes are briefly examined in Chapter III; however, no performance figures are quoted. Then in Chapter IV, the ultimate performance limitations in these instruments are examined. These limitations are that the maximum signal power out of a vibratory rate gyroscope for a constant angular rate input is determined by strain in the drive member and the minimum noise power is determined by the thermal noise in the sensing system components. These two factors together determine the theoretical ultimate threshold of a vibratory rate gyroscope. Equations showing the effects of the material properties and the

instrument size on this threshold are derived and briefly discussed in Chapter IV. Then the equations are applied to an example and numerical values for the theoretical ultimate thresholds of a rotary and a vibratory drive device are determined. It is apparent from this example that if instruments could be built with thresholds even two orders of magnitude larger than the theoretical ultimate, they would be useful in many applications. Unfortunately, this precision has been achieved to date because of "practical" problems.

2.2 SUMMARY AND CONCLUSIONS

The "practical" performance limitations in vibratory rate gyroscopes are examined in Chapter V. This is accomplished by first examining the tuning fork gyroscope in detail and then generalizing the results. It is concluded that the major factor that is presently limiting the performances of vibratory rate gyroscopes is unwanted cross coupling between the drive member and the sensing system. This cross coupling is caused by inhomogenities, nonlinearities and instabilities of materials and the limited tolerances to which mechanical parts can be made. Cross coupling is troublesome because the signals produced by it have components at the same frequencies as the suppressed carrier angular rate information and can be distinguished from the angular rate signals only to the extent that they remain constant and can be bucked or calibrated out or else can be discriminated against by phase sensitive demodulation. Unfortunately, the cross coupling is not constant but varies in a stochastic manner with time, temperature, linear accelerations and numerous other factors and also quadrature rejection is not perfect in practice. As a result, large

cross-coupled, zero-rate errors occur at the outputs of vibratory rate gyroscopes. Therefore, it appears that there are only two basic ways to improve the performances of vibratory gyroscopes. They are:

1. reduce the magnitude of the stochastic cross coupling (assuming it is normalized in terms of equivalent input angular rates).
2. reduce the errors that occur at the output of an instrument because of the cross coupling.

The cross-coupling problem in rotary and vibratory drive, vibratory rate gyroscopes is thoroughly examined in Chapter V. It is concluded that in vibratory drive instruments the magnitude of the cross coupling can be minimized by operating large, low frequency, well balanced drive members made out of stable materials at low strain levels in a temperature controlled environment. Unfortunately, an instrument with all these characteristics is expensive (because of the precision balancing operation required), is not a low power device (because of the power required for a temperature control) and may require a long warm up time and may be very acceleration sensitive. Furthermore, the precision of the instrument may not be any better than that of more conventional gyroscopes of comparable cost. Therefore, it appears doubtful that high precision, vibratory drive instruments have sufficient advantages over more conventional gyroscopes to justify additional research on them. However, this conclusion is not necessarily true for medium precision instruments ($10 - 100^{\circ}/\text{hr}$ long term drift), in which the balance need not be as good and the temperature need not be controlled. At the present time it appears possible to build low cost, medium precision, vibratory drive instruments having wide bandwidths and short warm up times and requiring very little operating power. Also, the short term noise or drift in these instruments will be lower than the long term drift which makes

them useful in applications in which a good but noisy angular position reference is available. Therefore, it appears that future research on vibratory drive instruments should be directed toward these instruments. It also appears that serious consideration should be given to using stable materials, such as quartz, for the drive members of these instruments because theoretically there is sufficient signal power in some configurations so that the attenuation caused by the lower maximum allowable strains will not cause transducer noise problems while the cross coupling stability for a specific degree of balance may be significantly improved.

The cross-coupling problem in rotary drive instruments is also examined in Chapter V. This examination is brief because few data are available in the open literature on these instruments. It is ~~concluded~~ that in rotary drive instruments the cross coupling problem is primarily caused by rotor unbalances and non-ideal bearings and bearing supports. However, the machining operations required to reduce the magnitude of the cross coupling in rotary drive instruments are relatively simple compared to the operations required for vibratory drive instruments. On the basis of some crude estimates made in Chapter V on these instruments, it appears that medium precision ($0.1 - 10^0$ /hr long term drift) rotary drive instruments are feasible. However, these instruments will not be low power devices, like vibratory drive instruments, and will have some type of bearings.

Reducing the magnitude of the cross coupling is just one of the two basic ways listed to improve the performances of vibratory rate gyroscopes. Although it is very important, a more detailed analysis requires considerable experimental work with materials and machining techniques and is beyond the scope of the present research. The present research is mainly concerned with reducing the errors that occur at the outputs of vibratory gyroscopes because of cross coupling. Two methods of doing this which are examined are called

"double modulation" and "optimum processing of suppressed carrier signals". The double modulation research is reported in Chapter VI and the research on optimum signal processing is reported in Chapters VII and VIII.

Double modulation is a process by which the angular rate information in vibratory gyroscopes is modulated twice so that the suppressed carrier rate information does not occur at the same frequencies as the cross coupling between the basic drive member and the sensing system. This additional modulation can be obtained by either rotating or vibrating the basic drive member of a more conventional vibratory gyroscope about a rate insensitive axis. When the frequencies of the cross coupling and angular rate information are separated in this manner, the zero-rate errors that result from cross coupling between the drive member and the sensing system are greatly reduced. However, double modulation also introduces other types of cross coupling which have components at the same frequencies as the angular rate information. In order to investigate all of these aspects of double modulation, an experimental rotary doubly-modulated tuning fork gyroscope was built. The results of the experimental and theoretical research conducted to date on this instrument are reported in Chapter VI and summarized at the end of that chapter. The most important conclusions are:

1. It appears that a rotary doubly modulated gyroscope using a ring configuration for the basic drive member could be built which would have a long term drift of the same order of magnitude as a super precision, vibratory drive instrument without double modulation but with temperature control. An order of magnitude estimate for the drift of these instruments is 10^{-1} °/hr.

The doubly modulated instrument would undoubtedly be cheaper, less sensitive to linear accelerations and require much shorter warm up times than the super precision instrument but probably would be less reliable. Neither instrument would be a low power device.

2. In order to achieve high precision in a rotary double modulated instrument, additional research on bearings may be necessary. This research should be conducted in such a manner that the results are applicable to both rotary double modulated instruments and rotary drive instruments because one troublesome cross coupling problem is common to both types of instruments.
3. In vibratory double modulated instruments, the second harmonic terms must be used rather than the first because of cross coupling between the double modulation motion and the sensing system. This is undesirable because it means the signal power into the basic configuration being double modulated is severely attenuated. Therefore these instruments do not appear to be promising.

The second method which is examined for reducing the errors that occur at the output of a vibratory rate gyroscope because of cross coupling is called optimum processing of suppressed carrier signals. In Chapter VII, a filter theory for synchronous suppressed carrier systems, such as vibratory gyroscopes, is developed and then in Chapter VIII it is applied to one aspect of the vibratory gyroscope signal-processing problem. This work is concerned with determining

the optimum linear time-varying filter to minimize the mean square error at the output of a suppressed carrier system when:

1. there are quadrature signals present.
2. there is unmodulated noise added to the modulated signals.
3. there are fixed elements in the system, such as the resonant sensing system in a vibratory rate gyroscope.
4. there are saturation constraints on signals following the optimum signal processor.
5. the carrier frequency drifts slowly in a random manner.
6. the parameters of the fixed elements are not constant but drift slowly in a random manner.
7. the phase of the synchronizing signal available to the synchronous demodulator drifts slowly in a random manner.

By using this theory one can determine the minimum mean square error in a suppressed carrier system and also the optimum signal processor when the harmonic errors either are weighted equally with the other errors, or are neglected. The form of the optimum signal processor may be completely unrestricted, which results in the absolute minimum mean square error, or it may be restricted in some manner so that the errors are larger but the signal processor is easier to build.

The most important conclusions of the research on the optimum processing of suppressed carrier signals are:

1. an optimum or approximately optimum "practical" signal processor can be determined in suppressed carrier problems by using the theory developed in Chapter VII.
2. processing signals in vibratory gyroscopes in an optimum manner reduces the errors due to cross coupling but the

percentage improvement is small if the instruments must have a flat response to zero frequency. The reason for this conclusion is that the cross-coupled, zero-rate errors that result from the stochastic inphase cross coupling and the quadrature rejection errors caused by differential phase shift drifts cannot be filtered if the instrument must have a response to zero frequency because the bandwidth of the zero-rate errors from these sources is generally much smaller than the bandwidth of the input angular rates.

3. a significant improvement can be obtained in the mean square error at the output of a vibratory rate gyroscope by designing the signal processing electronics in an optimum manner if the instrument need not have a flat response to zero frequency.

CHAPTER III

BRIEF SURVEY OF THE BASIC TYPES OF VIBRATORY GYROSCOPES

3.1 INTRODUCTION

Although numerous, greatly different, vibratory gyroscope configurations have been proposed, all known vibratory gyroscopes can be classified into a few basic types. The classification criteria used in this chapter for that purpose are: first, whether an instrument is basically* a rate** or rate integrating gyroscope; second, the type of drive scheme or schemes used in an instrument; and third, the relationship between the carrier frequency of the suppressed carrier signals containing the rate information and the drive frequency of an instrument. In addition to this classification, the differences between one and two axis instruments of the basic types and some of the characteristics of the basic types are pointed out in this chapter. This is accomplished by examining extremely simple configurations which are representative of each basic type of instrument and a few examples of more practical configurations. A convenient place to start such a survey, is with the concept of the Coriolis force which is used extensively in the following approximate analyses.

3.2 CORIOLIS FORCE

The vibratory torques or forces that produce the output signals in vibratory gyroscopes are the result of Coriolis effects. Although the exact

* Basic here refers to the characteristics of an instrument when it is operated without feedback or external integrators.

** Rate gyroscopes are angular motion sensors in which the output is constant when a constant input angular rate is applied. Rate integrating gyroscopes are angular motion sensors in which the output is a ramp for a constant input angular rate. In both cases, the output (amplitude or ramp slope) ideally is proportional to the input angular rate.

equations describing the operation of a particular configuration must be obtained by Lagrange's or some other method, approximate* equations can be determined by considering the Coriolis forces on moving masses in the instrument. Recall that the Coriolis force \vec{F}_c apparently exerted on a mass m moving at a velocity \vec{v}_d in the instrument frame of reference is

$$\vec{F}_c = 2m(\vec{v}_d \times \vec{\Omega}) \quad (3.1)$$

where $\vec{\Omega}$ is the inertial angular rate of the instrument frame of reference.

The component (s) of $\vec{\Omega}$ along a particular instrument axis (axes) is the quantity that is sensed in a vibratory gyroscope by measuring an output produced by a component (s) of the total Coriolis force on all the moving mass particles or the total Coriolis torque resulting from the Coriolis forces. This force (torque) per unit mass per applied angular rate is proportional to the velocity, \vec{v}_d , which is often called the drive velocity. Since there are only two basic methods recognized for driving the mass elements, the drive scheme provides one convenient method for the classification of vibratory gyroscopes.

3.3 BASIC DRIVING SCHEMES

The two basic schemes which are recognized for driving the mass elements in vibratory gyroscopes are called rotary drive and vibratory drive. In rotary drive instruments the mass elements move in paths that are approximately circular

* The approximate differential equations which are found when this method is used do not contain terms in Ω , do not contain all the time varying coefficients of some derivatives, do not contain interaxial cross-coupling terms and do not contain the terms showing the source of the output energy. Therefore the approximation is best when Ω is small and varies slowly with time, and the driven masses are a small percentage of the total mass of the instrument. Since none of these conditions is unreasonable, the approximations are quite good for most instruments and applications of these instruments.

whereas in vibratory drive instruments the paths are arcuate or linear.

Idealized examples of these two types of drive are shown in Fig. 3.1.*

When no angular rates are applied to these devices**, the point mass m_1 in the rotary drive example rotates at a radius R and angular velocity ω_d about the instrument y axis and the point mass m_1 in the vibratory drive example oscillates sinusoidally with amplitude r_0 and frequency ω_d parallel the instrument x axis. When angular rates are applied, Coriolis forces are exerted on the moving masses and the trajectories are altered. When a spring type restoring force acts on the particles in the opposite direction of the Coriolis force, the resulting change in the trajectories is relatively small and when the angular rate is removed the trajectories relax to their initial states. These instruments are called rate gyroscopes. When the restoring force is absent, the particles do not return to their initial trajectories when angular rates are removed but instead remain in new trajectories which are a function of the input angular rate history of the instrument. These instruments are called rate integrating gyroscopes. The basic types of rate and rate integrating gyroscopes are examined in the following two sections.

3.4 VIBRATORY RATE INTEGRATING GYROSCOPES

Foucault's¹ pendulum, which was demonstrated in 1851, was basically an integrating angular rate sensor. With it, Foucault showed that the earth was spinning because the plane of oscillation of the pendulum rotated at a constant angular rate which was equal to the vertical component of the earth's sidereal spin. Thus an observer on earth could measure the time integral of the vertical component of earth rate by observing the angle through which the plane of the pendulum oscillation traversed in a period of time.

* The dotted masses in these figures should be ignored at the present time. They are referred to in other sections of this chapter.

** The word *device* is used for lack of a better word and also because practical configuration will be evolved from these simple configurations in this chapter.

The same principle is used in practical vibratory rate integrating gyroscopes. In these instruments, mass particles are made to oscillate in a uniform central force field and the plane of oscillation of the particles in the instrument is measured to provide an output. An illustrative* example of such an instrument is shown in Fig. 3.2. In this instrument² a string or ribbon oscillates in the radial direction at a constant amplitude and frequency. When an angular rate is applied about the y axis, ideally the plane of the string oscillation remains fixed in inertial space and therefore precesses in the instrument coordinate system. Angular rates about the instrument x and z axes have a negligible effect on the motion of the string in the instrument because the Coriolis forces associated with them are in the y direction. When the plane of the string oscillation in this instrument is initially the xy plane, ideally the integral of the subsequent input angular rate can be determined by measuring the x coordinate (or its derivative) of an element of the string.

$$x(t, y=0) = \left[\sin \int_0^t dt \Omega_y(t) \right] r_0 \sin \omega_d t = r_0 \sin \theta_y(t) \sin \omega_d(t) \quad (3.2)$$

When $\theta(t)$ is small, so that the approximation $\sin \theta = \theta$ is valid, a signal proportional to $x(t)$ is a suppressed carrier representation of the integral of the input angular rate. Therefore the low frequency component of $\theta(t)$ can be obtained by demodulating this signal. For larger angles, it is necessary to find the inverse sine of the output or else process the signals differently; perhaps by measuring both the x and z positions of the string. The alternative is to prevent $\theta(t)$ from becoming large by using some type of feedback around the instrument. However, when this is done the instrument output is no longer the integral of the input angular rate and angular information must be obtained elsewhere in the loop.

* The means by which the string or ribbon is driven or the position of the plane of oscillation sensed are not shown. Energy can be supplied to the string drive either by applying a double drive frequency force in the y direction or a drive frequency force in the radial direction. Sensing can be accomplished in numerous ways but must not torque the string. The major difficulty with instruments of this type is building them so that they are circularly symmetric.

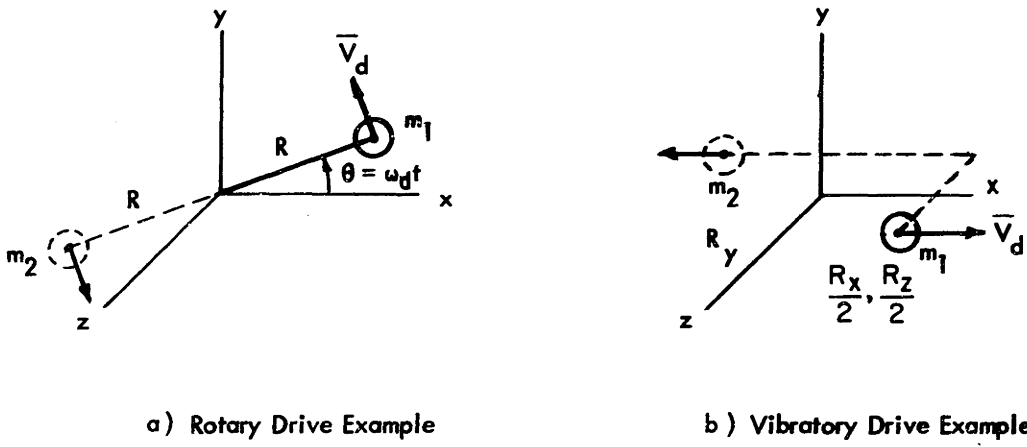


Fig. 3.1 Basic Drive Schemes

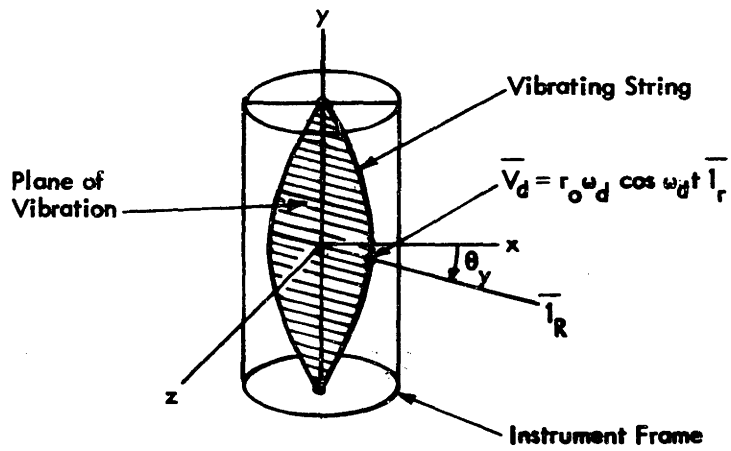


Fig. 3.2 Vibrating String Gyroscope

Vibratory rate integrating gyroscopes of the type shown in Fig. 3.2 are classified as vibratory drive, drive frequency output instruments because the carrier frequency of the output is equal to the drive frequency of the mass elements. These instruments are single axis sensors in which, ideally, the plane of the drive oscillation remains inertially stationary as the instrument case is rotated about the input axis and remains approximately unchanged in the instrument for angular rates about orthogonal axes. Two axis instruments of this type are theoretically possible; however, the practical problems associated with driving the mass elements and processing the signals are difficult and the resulting instruments have few if any advantages over free³ gyroscopes. Therefore instruments of this type have received very little attention.

Rotary drive, drive frequency output instruments operating in the same manner are ~~theoretically~~ also a possible type of vibratory rate integrating gyroscope. In practice, however, these instruments are difficult to build and have few if any advantages over other, more conventional, gyroscopes. The reason for this is that it is difficult to drive an unbalanced drive member, such as m_1 shown in Fig. 3.1a, without the existence of spring type restoring forces to constrain the motion caused by the Coriolis force. The easiest way in which this difficulty can be overcome is by using a balanced drive member (rotor such as the m_1m_2 combination in Fig. 3.1a); however, when this is done the resulting instrument is more properly classified as a "free" gyro with an asymmetrical rotor than a vibratory gyro. Under the broad definition of vibratory gyroscopes in Section I.1, however, these instruments are still vibratory gyroscopes unless they are ruled out because of the steady component of Coriolis torque which is exerted on the drive member for constant input angular rates. Since this escape route does exist, I choose to use it and thereby not classify free gyros with balanced asymmetrical rotors as vibratory gyroscopes.

The classification difficulty which has just been encountered, arises primarily* because the definition of vibratory gyroscopes given in Section 1.1 was quite broad in order to make it as simple as possible. The definitions could be lengthened to clarify the distinction between free and vibratory gyroscopes; however, it is not necessary at this point because I have chosen to restrict the following work to vibratory rate gyroscopes. The primary reasons for this restriction are; first, it was necessary to restrict the scope of the work because only a limited amount of time was available for this research; second, the difference between a rate and rate integrating instrument is well defined; and third there are several important differences between vibratory rate and rate integrating gyroscopes which make the generalization of the performance limitations of all these instruments more difficult than necessary**. Therefore, unless otherwise stated, all references to vibratory gyroscopes in the following chapters refer to vibratory rate gyroscopes (abbreviated VRG) and not rate integrating gyroscopes.

3.5 BASIC TYPES OF VIBRATORY RATE GYROSCOPES

3.5.1 Introduction

In vibratory rate gyroscopes, spring type restoring forces restrict the motion of the driven mass elements when the instruments are rotated. As a result, the change in the trajectories of the driven mass elements is relatively small and, as an excellent first approximation, can be neglected when computing the Coriolis forces on the driven mass elements. Approximate equations describing the operation of the basic types of vibratory rate gyroscopes are obtained in this section by using this approximation. The penalty paid for this approximation is that

* It also arises because the Coriolis force was used in the analyses. The concept of Coriolis forces is much more useful in analyzing the behavior of rate gyroscopes than rate integrating gyroscopes.

** Nevertheless, much of the work done in the following chapters is applicable to vibratory rate integrating gyroscopes. The places where it is not, however, are not pointed out

terms in $\dot{\Omega}$ and some of the time-varying coefficients in the resulting differential equations are omitted along with inter-axial cross coupling and nonlinear terms. In most instruments these omitted terms are not of significant importance in understanding the principles of operation but only in determining second order effects. Therefore this technique is considered adequate for the following sections of the chapter.

3.5.2 Rotary Drive Instruments

A convenient place to start a brief survey of the basic types of rotary drive, vibratory rate gyroscopes is with the simple instrument shown in Fig. 3.1a. By examining the Coriolis forces on the single point mass m_1 or the Coriolis torque on the two point masses m_1 and m_2 , the manner in which most instruments of this type operate can be deduced.* When this instrument is rotated in inertial space, the Coriolis force apparently exerted on m_1 is given by Eq. 3.4.

$$\bar{v}_{m_1} = -R\omega_d(\bar{I}_z \cos \omega_d t + \bar{I}_x \sin \omega_d t) \quad (3.3)$$

$$\bar{F}_c = 2mR\omega_d \left\{ \begin{array}{l} \bar{I}_x \Omega_y \cos \omega_d t - \bar{I}_z \Omega_y \sin \omega_d t \\ + \bar{I}_y (\Omega_z \sin \omega_d t - \Omega_x \cos \omega_d t) \end{array} \right\} \quad (3.4)$$

where Ω_i is the angular rate of the instrument frame of reference about the i axis ($i = x, y, z$)

\bar{I}_i is a unit vector along the i axis ($i = x, y, z$)

The components of the Coriolis force are suppressed carrier representations of the components of the applied angular rate in which the carrier frequency is the drive frequency ω_d . Therefore, in theory, by measuring the components of the Coriolis force along the x and y or y and z axes and demodulating the resulting

* This is true because the total torque or force that is exerted on a drive member when an instrument is rotated is simply an integral over all the point masses of the drive member.

signals with a drive frequency reference signal of the proper phase, the low frequency* content of the three components of the applied angular rate can be sensed.

In most practical rotary drive, vibratory gyroscopes the rotation axis passes through the centroid of the spinning member to simplify the support and drive problems and a differential arrangement is used to produce the output signals by vectorially subtracting the Coriolis forces on diametrically opposed masses. An illustrative example of an instrument employing a solid rotor which operates in this manner is shown in Fig. 3.1a. Two equal masses m_1 and m_2 at equal radii R rotate at constant angular velocity about a spin axis passing through their centroid. The Coriolis forces acting on these masses are opposite in sign and cause Coriolis torques to be exerted on the rotating assembly. These torques produce the output signals; however, there are two distinct ways in which this is accomplished.

In one type of rotary drive instrument the Coriolis torque in a coordinate frame fixed in the rotor (labeled $x'y'z'$ in Fig. 3.1) produces the output. This torque is confined to the z' axis and is expressed as

$$\bar{M}'_c = I'_z 4mR^2 \omega_d (\Omega_z \sin \omega_d t - \Omega_x \cos \omega_d t) \quad (3.5)$$

The torque M'_c is a suppressed carrier representation of the angular rates applied about the x and z axes in which the carrier frequency is the frequency of the drive rotation. The Ω_x and Ω_z components are 90° out of phase in time so the output signal produced by M'_c can be processed by phase sensitive demodulation and filtering to yield both of the x and z axis rate components. Instruments of this type are classified as two axis, rotary drive, drive frequency output, vibratory rate gyroscopes. Several such instruments^{4, 5, 6} have been built and others are currently being investigated. An illustrative example of

* Low frequency here implies frequencies much lower than the drive frequency.

these instruments is shown in Fig. 3.3.* The mechanical output of this instrument is the torsional oscillation of the rotating bar about the z' axis. This motion is caused by a Coriolis torque** similar to that of Eq. 3.5. The displacement or velocity of the bar with respect to the rotating support is sensed to provide a suppressed carrier electrical signal which is a known function of the angular rates about the x and z axes.

In the other type of rotary drive instrument, the Coriolis torque in the xyz (instrument) axes produces the output signal and an interesting change results. This can be seen in the following equation for the torque \bar{M}_C .

$$\bar{M}_C = 2mR^2\omega_d \{ \bar{I}_x [-\Omega_x \sin 2\omega_d t + \Omega_z (1 - \cos 2\omega_d t)] + \bar{I}_z [\Omega_x (1 + \cos 2\omega_d t) - \Omega_z \sin 2\omega_d t] \} \quad (3.7)$$

The x and z torque components are suppressed carrier representations of the applied rates about the x and z axes plus the steady conventional gyroscopic terms. The carrier frequency of the vibratory torques is twice the frequency of the drive rotation in contrast to the previous example in which it was equal to it. It is sufficient to use either the x or z component of the Coriolis torque to produce the output because both contain the same information about the applied angular rates. The torques about either axis corresponding to the x and z axis rates are 90° out of phase in time so the output signal produced by either of the two torques can be processed by phase sensitive demodulation to yield the x and z axis rate components. Instruments of this type are classified as two axis, rotary

* A more practical configuration for the drive member is a spoked wheel or cylinder as shown in Fig. 4.6a.

** The total Coriolis torque in this and other configurations of this type can be written

$$M'_C = 2I_{r0}\omega_d (\Omega_z \sin \omega_d t - \Omega_x \cos \omega_d t) \quad (3.6)$$

where I_{r0} is the planar moment of inertia of the output member (the bar in Fig. 3.3) from the plane defined by the axis of rotation (x) and the output axis (z').

drive, double drive frequency output, vibratory rate gyroscopes. Several such instruments⁷ have been built and others are currently being investigated. An illustrative example of these instruments is shown in Fig. 3.4. The mechanical output of this instrument is the torsional oscillation of the gimbal about the x axis. This motion is caused by the x component of a Coriolis torque* similar to that of Eq. 3.7 which is exerted on the asymmetrical rotor. When this motion is sensed, it provides a suppressed carrier electrical signal which is a known function of the x and z components of the angular rate applied to the instrument.

All presently known rotary drive, vibratory rate gyroscopes fit into one of these two classifications, i.e., the output frequency is either equal to the drive frequency or twice the drive frequency. The instruments in both classifications are two axis devices and the angular rates about the two input axes, which are orthogonal to the nominal plane of the drive oscillation, must be separated by phase sensitive demodulation of the suppressed carrier signals from the output transducer. The precision of instruments of this type which employ solid rotors, such as those shown in Figs. 3.3 and 3.4, depends to a large extent on the quality of the rotor bearings and the balance of the rotating member.

Although the bearing problems differ from those of conventional gyroscopes, bearings are still one of the major sources of error and failure in these instruments. Because of these shortcomings and also the relatively large power required in rotary drive instruments for the drive mechanism, much of the past research on vibratory gyroscopes has been concentrated on instruments employing the second type of drive--vibratory drive. These instruments generally require considerably less power and have longer life expectancies;

* The total Coriolis torque in this and other configurations of this type can be written as

$$M_c = (I_{rr_{max}} - I_{rr_{min}}) \omega_d (-Q_x \sin 2\omega_d t) \quad (3.8)$$

where I_{rr} is the planar moment of inertia of the rotor about a plane containing the axis of rotation.

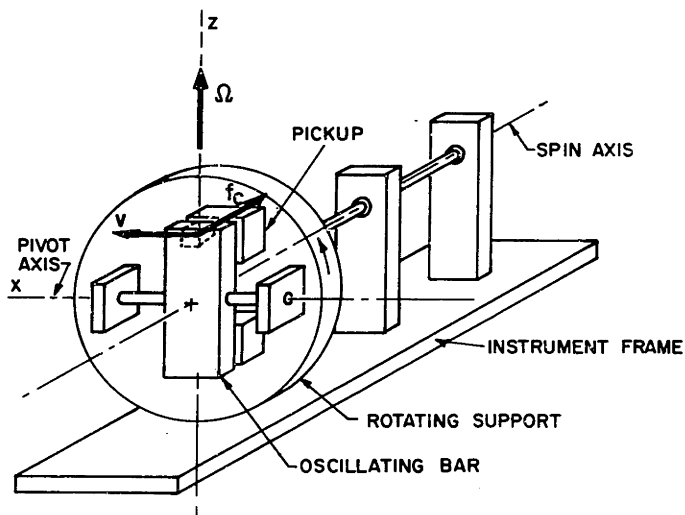


Fig. 3.3 Rotary Drive, Drive Frequency Output Device

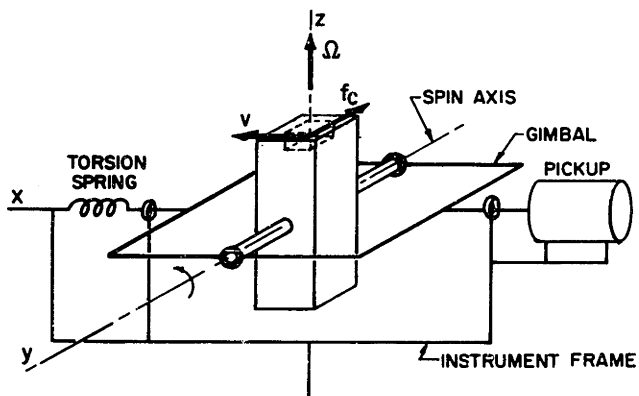


Fig. 3.4 Rotary Drive, Double Drive Frequency Output Device

however, the thermal fluctuation thresholds* are generally larger than those of rotary drive instruments of equal size and a severe cross-coupling problem generally exists.

3.5.3 Vibratory Drive Instruments

A convenient place to start a survey of the basic types of vibratory drive, vibratory rate gyroscopes is with the instrument shown in Fig. 3.1b. By examining the Coriolis force on the single point mass or the Coriolis torque on two point masses m_1 and m_2 , the manner in which most instruments of this type operate can be deduced. When this instrument is rotated in inertial space, the Coriolis force apparently exerted on the point mass m_1 with a velocity \bar{v}_{m_1} is given by Eq. 3.10.

$$\bar{v}_{m_1} = \bar{I}_x r_o \omega_d \cos \omega_d t \quad (3.9)$$

$$\bar{f}_o = 2m_1 r_o \omega_d \cos \omega_d t [-\bar{I}_y \Omega_z + \bar{I}_z \Omega_y] \quad (3.10)$$

The components of the Coriolis force are suppressed carrier representations of the y and z components of the instrument angular rate in which the carrier frequency is the drive frequency of the mass element m_1 . The Coriolis forces corresponding to the y and z axis rates are spatially orthogonal rather than orthogonal in time so the output signals must be produced by either the y or z force components but not both if the rate components are to be distinguished. Therefore two transducers are required for a two axis instrument.

Instruments of this type are classified as two axis,** vibratory drive, drive frequency output, vibratory rate gyroscopes. An illustrative example is shown in Fig. 3.5. In this instrument a longitudinal standing wave in a fluid provides the drive motion to the particles and pressure pickups located on the walls of the

* The differences in the thermal fluctuation thresholds is examined in Chapter IV and the cross coupling problem is examined in Chapter V.

** These instruments are potential two axis devices. The designer may restrict or not measure one component of the output motion and therefore only have a single axis instrument.

of the tube along the y and z axes, detect the Coriolis pressures. These pressures are given by Eq. 3.12

$$\bar{v}_d^* = \bar{I}_y v_o \sin \omega_d t \cos \frac{2\pi x}{l} \quad (3.11)$$

$$\bar{F}_c = 2d\rho v_o \sin \omega_d t (\bar{I}_x \Omega_z - \bar{I}_z \Omega_x) \quad (3.12)$$

where ρ is the density of the fluid, d is the diameter of the tube, v_o is maximum velocity of the drive particles in the tube and l is the length of the tube. The output of each pressure pickup is an electrical suppressed carrier signal which is a known function of the y or z component of the instrument angular rate.

It is interesting** to examine this simple configuration further since it leads to other configurations for instruments of the basic type which are not immediately obvious. First, if the tube in Fig. 3.5 is lengthened so that it is n wavelengths long rather than $l/2$ and the pressure pickups are removed and instead a straight flexural beam is located in the center of the tube for sensing purposes, it follows that the Coriolis forces on the beam in the y and z directions are periodic in both time and space. The same condition exists if a longitudinal standing wave is established in the beam rather than in the fluid surrounding the beam. In either case, if the ends of the beam are properly terminated and a constant angular rate applied about the x or z axis, a flexural standing wave

* This equation assumes the tube is lossless. In actual tubes the pressure wave and velocity are not orthogonal because energy must be supplied to all the oscillating mass particles.

** Configurations of vibratory drive vibratory rate gyroscopes are examined in more detail than configurations of the other basic types because these instruments appear to be superior to others with respect to long life expectancy and low power requirements. Also, it has been these instruments that have been the subject of the majority of vibratory gyroscope research.

spatially synchronized with the Coriolis forces is established in the beam in the z direction for y axis rates and in the y direction for z axis rates. One of the most convenient methods of terminating the beam is to form the tube and beam into a ring and make n an even number so that the drive wave can exist either in the tube or beam. When this is done, the Coriolis pressure apparently exerted on the beam is given by Eq. 3.13.

$$\bar{P}_c = kv_o \sin \omega_d t \cos (n\theta) [\bar{I}_r \Omega_y - \bar{I}_y (\Omega_z \cos \theta + \Omega_x \sin \theta)] \quad (3.13)$$

where \bar{I}_r is a unit vector in the radial direction, θ is the angle about the y axis and k is a constant which depends on whether the Coriolis forces result from a standing wave in the fluid surrounding the ring or a longitudinal wave in the ring itself.

Several interesting observations can now be made from Eq. 3.13. First, it follows from the \bar{I}_r component of \bar{P}_c that an n wavelength radial* flexural standing wave** of frequency ω_d will be established in the ring for constant applied rates about the y axis. Second, and much more important, it can be shown^{12, 13} from the \bar{I}_x component of \bar{P}_c that flexural standing waves*** in the y direction with n+1 and n-1 wavelengths*** and frequency ω_d will be established in the ring for constant angular rates about either the y or z axes. The (n+1) or (n-1) wavelength standing waves containing the y and z axis rate information

*

A radial flexural standing wave with n wavelengths is described by the \bar{I}_r component of \bar{v}_d in Eq. 3.12.

**

This wave is spatially synchronized with the appropriate Coriolis force component in Eq. 3.11 and the phase relationship in time between the Coriolis force component and the output wave depends on the drive frequency and the dimensions, material, etc. of the ring which determine its resonant frequencies.

This becomes apparent upon application of two trigonometric identities to Eq. 3.11

$$\begin{aligned} 2 \cos n\theta \cos \theta &= \cos(n+1)\theta + \cos(n-1)\theta \\ 2 \cos n\theta \sin \theta &= \sin(n+1)\theta - \sin(n-1)\theta \end{aligned}$$

are spatially orthogonal and any of these standing waves can be eliminated by constraining the motion of various elements of the ring. Third, when the ring is used as the drive member rather than a surrounding fluid, the Coriolis forces can be increased by a factor of $n+1$ by using a radial flexural standing wave, such as

$$\bar{v}_d = \bar{I}_r v_o \cos n\theta \cos \omega_d t + \bar{I}_\theta \frac{v_o}{n} \sin n\theta \cos \omega_d t \quad (3.14)$$

to provide the drive velocity to the mass elements. The resulting instrument is extremely simple because it contains only a ring in which one flexural standing wave* is used to provide the drive velocity to the mass elements and another to detect the instruments rotation about one or two axes in the plane of the ring.

The vibratory drive instruments studied thus far in this section have all been potential two axis configurations in which the carrier frequency of the output was the drive frequency. These instruments evolved from the single point mass instrument shown in Fig. 3.1b. In order to examine other types of vibratory drive, vibratory rate gyroscopes, it is convenient to add the dotted mass m_2 to Fig. 3.1b because in these configurations the Coriolis forces acting on symmetrical drive masses generate a Coriolis torque which produces the output.

When the two masses m_1 and m_2 in Fig. 3.1b, oscillate in phase opposition along axes parallel to the instrument x axis, the Coriolis torque apparently exerted on a drive member containing the two masses is

$$\bar{M}_c = -4m r_o \omega_d \left\{ \begin{array}{l} R_x \left[\cos \omega_d t + \frac{r_o}{2R_x} \sin 2\omega_d t \right] [\bar{I}_y \Omega_y + \bar{I}_z \Omega_z] \\ R_z \left[\bar{I}_x \Omega_x \cos \omega_d t + \frac{r_o}{2R_z} \sin 2\omega_d t (\bar{I}_y \Omega_y + \bar{I}_z \Omega_z) \right] \end{array} \right\} \quad (3.15)$$

* It can be shown that the two flexural waves are interchangeable.

The components of the Coriolis torque are suppressed carrier representations of the angular rates about the y and z instrument axes. The carrier frequencies of these torques are the fundamental and the second harmonic of the drive frequency. In practice, only the torques at one frequency are used to produce the output so instruments of this type are classified as vibratory drive, drive frequency or double drive frequency output, vibratory rate gyroscopes.

The majority of vibratory gyroscope configurations of this type use the fundamental component(s) of the Coriolis torque to produce the output because it is considerably larger than the second harmonic in instruments of equal size.* These instruments can be evolved from the simple example of Fig. 3.6 by letting either R_x or R_z equal zero. These two conditions are now examined separately even though the resulting instruments are all classified as vibratory drive, drive frequency output devices, because these configurations have always been the most popular for vibratory gyroscope research.

An illustrative example of the type of instrument that results when $R_z = 0$ is the tuning fork¹⁴⁻¹⁷ shown in Fig. 3.6. The mechanical output of this single axis instrument is the torsional oscillation of the tuning fork about the y axis. This motion, which is caused by the fundamental of the y component of the Coriolis torque [Eq. 3.15 when $R_z = 0$], is sensed to provide a suppressed carrier electrical signal which is a known function of the y axis angular rate. The tuning fork is rigidly**supported except in the torsional mode about the y axis in order to minimize the motion caused by the z axis Coriolis torque.

This configuration has been the most popular over the past 20 years for both experimental and theoretical work on vibratory gyroscopes and it is examined in more detail in Chapter V. The configurations which have evolved

* The ratio of the magnitudes of the torques in Eq. 3.13 is $\frac{2(R_x \text{ or } R_y)}{r_0}$ for the example of Fig. 3.1b.

** Not in Fig. 3.6 but in practice.

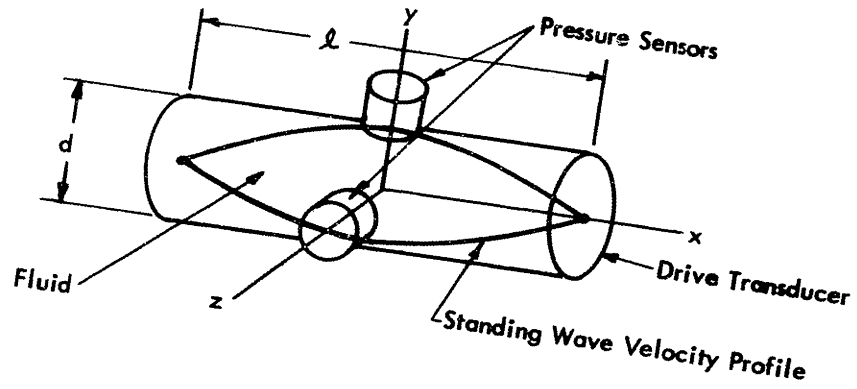


Fig. 3.5 Acoustic Gyroscope

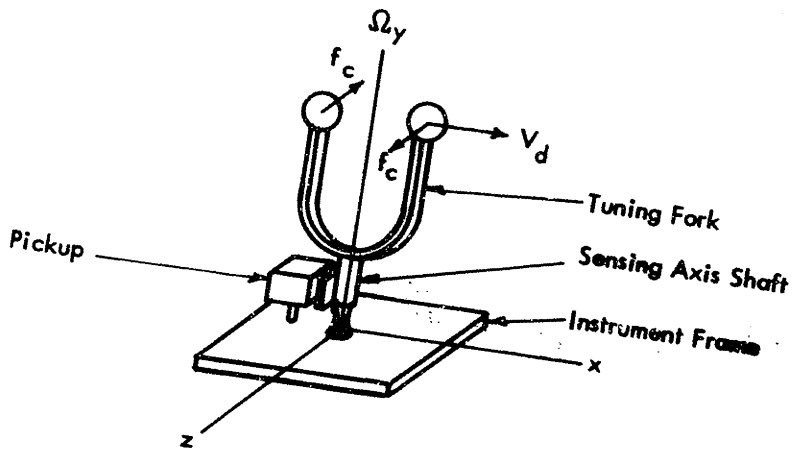


Fig. 3.6 Single Axis, Vibratory Drive, Drive Frequency Output Device

from it are usually symmetric with respect to the xz plane--a double tuning fork-- and also sometimes circularly symmetric with respect to the y axis. One such instrument is a cylinder in which the top and bottom halves oscillate extensionally in the radial direction in phase opposition to provide the drive velocity to the symmetrically located mass elements. The output mechanical motion when rates are applied about the axis of the cylinder is an oscillatory twisting of the two halves of the cylinder in opposite directions because the Coriolis torques are opposite in sign for the two halves of the cylinder.

Instruments of this type, which can be evolved from the example of Fig. 3.1b by letting $R_z = 0$, are generally single axis devices. The moment of inertia of the drive member (or a part of it in the case of the cylinder mentioned above) about the input axis in these instruments varies sinusoidally with time and can be written as

$$I_y = I_0 (1 + a_{y1} \sin \omega_d t) \quad (3.16)$$

Therefore when constant angular rates are applied about this axis, a torque equal to $\Omega_y \dot{I}_y$ is exerted on the drive member about the y axis. This torque is equal to the sum of the Coriolis torques on all the driven mass elements.

Most potential* two axis vibratory drive, drive frequency output instruments can be evolved from the simple example of Fig. 3.1b by making $R_z = 0$ rather than R_x . An illustrative example of the type of instrument that results is shown in Fig. 3.8. In this instrument a drive member consisting of four point masses symmetrically located along the x and z axes is twisted sinusoidally about the y axis to provide the drive velocity. The mechanical output of this two axis instrument is a torsional oscillation of the drive member about the $x(z)$ axis for angular rates about the $z(x)$ axis. The Coriolis torque which causes the twisting about the x axis can be obtained from Eq. 3.15. Since the output motions caused

* These instruments are potentially two axis sensors although the designer may choose to restrict the output motion about one of the axes and therefore make a single axis sensor.

by the x and z axes torques are spatially orthogonal, two transducers must be used to sense the twisting and provide suppressed carrier electrical signals which are known functions of the x and z components of the applied angular rate. A more sophisticated¹⁸ version of this instrument uses two, counter oscillating wheels as the drive members and the outputs are the angles between the wheels about the x and z axes. The torques that cause this motion can be derived by integrating the Coriolis torques on the mass elements or by using the equation which is used in conventional gyroscopes

$$\bar{M} = \bar{\Omega} \times \bar{H} \quad (3.17)$$

where

$$\bar{H} = I_y \dot{\theta}_0 \omega_d \cos \omega_d t \quad (3.18)$$

and θ_0 is the amplitude of the drive oscillation of the wheel of inertia I_z about the z axis.

In all of the vibratory drive instruments that have been examined thus far, the output frequency was equal to the drive frequency. This is true of most practical configurations; however, other configurations have been proposed in which the output frequency is a harmonic of the drive frequency. A double drive frequency Coriolis torque exists in all the examples of Fig. 3.7 but it is much smaller than the fundamental and therefore is not used. Configurations which use a double frequency Coriolis torque to produce the output are generally designed so that the fundamental torque component does not exist. This can be done in the example of Fig. 3.1b by making $R_x = 0$ and sensing an output caused by the y and/or z axis component of the Coriolis torque in Eq. 3.15. An example of a single axis instrument of this type is shown in Fig. 3.8. The output is the torsional oscillation of the bar about the y axis. The double drive frequency torque which causes this motion can be calculated by integrating the Coriolis torques

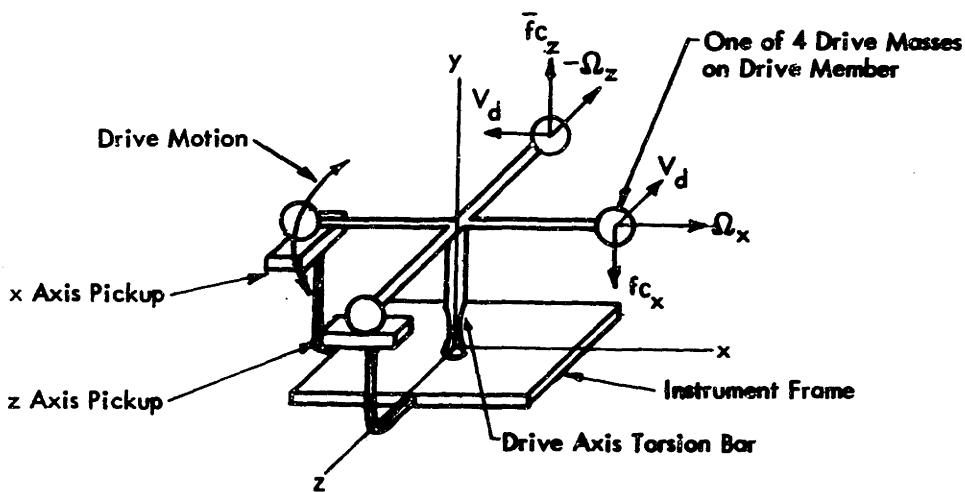


Fig. 3.7 Two Axis, Vibratory Drive, Drive Frequency Output Device

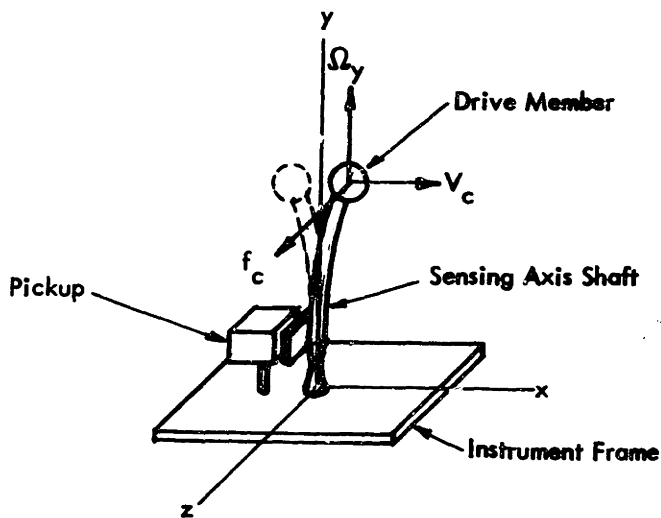


Fig. 3.8 Single Axis, Vibratory Drive, Double Drive Frequency Output Device

on all the masses in the drive member or by using the expression for the moment of inertia of the drive member about the y axis

$$I_y = I_0 (1 - a_{y2} \cos 2 \omega_d t) \quad (3.19)$$

where the position of the vibrating mass elements from the rest position is sinusoidal in time.

Instruments of this type, which are classified as vibratory drive, double drive frequency output, vibratory rate gyroscopes, have received relatively little attention compared to that given vibratory drive, drive frequency output instruments over the past years because in two instruments of equal size the double frequency Coriolis torques in the one instrument are much smaller than fundamental torques in the other while the magnitudes of the noise in the two instruments are about the same. The same reasoning is valid for explaining why instruments have not been built with higher harmonics of the drive frequency as the output frequency. In theory, such instruments could be built because Coriolis torques at these frequencies can be generated. An example is the simple configuration shown in Fig. 3.9. In this instrument the angle θ , which is about the z axis, is defined as

$$\theta = \theta_0 \sin \omega_d t \quad (3.20)$$

and the Coriolis torque* about the y axis on symmetric mass elements in the bar at a radius R from the axis of oscillation is

$$\bar{I}_y \cdot \bar{M}_c = 2mR^2 \theta_0 \omega_d \cos \omega_d t [\sin 2 \theta \sin \omega_d t] \Omega y \quad (3.21)$$

* The same expression can be obtained by computing the Coriolis forces on the masses and then the torques but this method ($\bar{I}_y \cdot \bar{M}_c = \Omega \bar{i}_y$) is obviously much more efficient in the present example.

since the moment of inertia of these mass elements about the y axes is

$$I_y = 2m[R \cos(\theta_0 \sin \omega_d t)]^2 = mR^2 [1 + \cos 2\theta_0 \sin \omega_d t] \quad (3.22)$$

The higher harmonic Coriolis torque expressions can be obtained from Eq. 3.21 by using the identity

$$\sin 2\theta \sin \omega_d t = 2 \sum_{k=0}^{\infty} J_{2k+1}(2\theta_0) \sin(2k+1)\omega_d t \quad (3.23)$$

where J_n is a Bessel function of order n. It is now apparent that the amplitude of the higher harmonic torques fall off quite rapidly.

All presently known, practical, * vibratory drive, vibratory rate gyroscopes can be classified as either drive frequency or a double drive frequency output instruments. These instruments can all be evolved from the simple examples examined in this section. The two axis gyroscopes of both these types require two transducers because the two components of the Coriolis forces which cause the outputs are in phase in time but spatially orthogonal.

The drive member or media in vibratory drive, vibratory rate instruments may be either a solid, liquid, or gas or a combination of these and the drive motion may be imparted by either a flexural, longitudinal or torsional oscillation. The drive power required is generally quite low because of the large quality factors that can be obtained, particularly in solids. This is one of the major advantages of instruments of this type over other types of gyroscopes. Some other advantages are that there are no moving parts to wear out, the electronics are extremely simple, the instruments can be turned on and off with no difficulty and theoretically require very short warm up times. Also the instruments

* Practical is inserted here to rule out instruments similar to that shown in Fig. 3.9 in which higher harmonic torques exist and may be used to produce the output signals.

are rugged, simple and therefore cheap. Unfortunately, vibratory drive instruments also have important disadvantages. The most important of these are examined in the following chapters and are simply stated here. They are that unwanted cross coupling exists between the drive and sensing systems, and the maximum drive velocity of the mass elements in vibratory drive instruments is lower than in rotary drive instruments of the same size.

3.5.4 Instruments Employing Double Modulation

In order to make the present survey of the basic types of vibratory rate gyroscopes complete, it is necessary to include instruments which employ double modulation.* In these instruments there are two separate drive motions and the frequency of the output is a linear combination of harmonics of the two drive frequencies. Although these instruments are more accurately classified as either rotary or vibratory drive, vibratory rate gyroscopes employing double modulation, it is more convenient in a survey such as this to consider them separately. The reason for this is that a vibratory rate gyroscope employing double modulation is basically a rate gyroscope which fits into one of the classifications discussed in Sections 3.5.2 and 3.5.3, in which the angular rate about its input axis (axes) is a modulated representation of two or more components of the overall instrument angular rate; and the effect of double modulation on the output of an ideal instrument is approximately independent of the basic configuration that is double modulated. The primary purpose for introducing this additional modulation is to separate the carrier frequency of the Coriolis forces from the drive frequency or harmonics of the drive frequency of the basic vibratory rate gyroscope. In this way some of the errors resulting from unwanted cross coupling can be removed; however, other difficulties are encountered. The whole concept of double modulation is discussed in Chapter 6 so it is unnecessary to elaborate more upon it here.

* The only known vibratory double modulated rate gyroscope was built by the author and is examined in detail in Chapter 6.

The second modulation* of the Coriolis forces or torques in a double modulated instrument can be introduced by either rotating or vibrating a basic VRG about a rate insensitive axis. This is a planar double modulation and it turns a basic single axis instrument into a two axis instrument. A simple example** of such an instrument is shown in Fig. 3.10. In this device

$$\Omega_{I.A.} = \Omega_y \cos \theta + \Omega_z \sin \theta \quad (3.24)$$

where $\Omega_{I.A.}$ is the angular rate about the input axis of the basic single axis instrument and θ is the angle between the y axis and the input axis.

When rotary double modulation is employed the angle θ can be written as

$$\theta = \omega_m t \quad (3.25)$$

where ω_m is the frequency of rotation. Therefore $\Omega_{I.A.}$ and also the output of the basic single axis instrument are suppressed carrier representations of Ω_y and Ω_z in which the carrier frequency is ω_m . The two rate components, Ω_y and Ω_z , can be obtained from the instrument output by using two phase sensitive demodulators. The Coriolis forces (torques) inside the instrument which cause

* The first modulation occurs in the basic vibratory rate gyroscope and it causes the carrier frequency of the output signals to be a harmonic of the drive frequency.

** This simplified analysis assumes the characteristics of the single axis instrument are unchanged by the double modulation motion. In reality, the complexity of the equations describing the operation of the overall instrument is noticeably increased by double modulation but the approximations are very good and certainly consistent with the approximations made in prior sections of this chapter.

the output are modulated an additional time by the double modulation process so that they now occur at $n\omega_d \pm \omega_m$ instead of $n\omega_d$. Therefore the signal processing can also be accomplished by a single demodulator for each component of rate if the frequency of the demodulator reference is $n\omega_d \pm \omega_m$. In this case the energy of one sideband is discarded.

When vibratory double modulation rather than rotary double modulation is employed, the angle θ in Fig. 3.10 can be written as

$$\theta = \theta_o \sin \omega_m t \quad (3.26)$$

where θ_o is the amplitude of the torsional oscillation. The angular rates about the input axis of the single axis instrument now occur at all harmonics of the double modulation frequency ω_m .

$$\begin{aligned} \Omega_{I.A.} = \Omega_y [J_o(\theta_o) + 2 \sum_{k=1}^{\infty} J_{2k}(\theta_o) \cos 2k\omega_m t] + \\ \Omega_z [2 \sum_{k=0}^{\infty} J_{2k+1}(\theta_o) \sin (2k+1)\omega_m t] \end{aligned} \quad (3.27)$$

where J_k is a Bessel function of order k . Because of the nature of Bessel functions, the magnitudes of the higher harmonic terms decrease quite rapidly. The Coriolis forces (torques) inside the instrument which cause the output are modulated an additional time by the process and now occur at $n\omega_d \pm 2k\omega_m$ for y axis rates and $n\omega_d \pm (2k+1)\omega_m$ for z axis rates. Therefore different demodulation frequencies must be used to obtain the y and z axis rates from the suppressed carrier output signals when a single axis instrument is double modulated. If a two axis instrument, with both input axes in the plane of the double modulation motion, is used, this difficulty is avoided because the output of the other channel of the simple instrument would be similar to that of Eq. 3.27 with Ω_z replaced by Ω_y and Ω_y by $-\Omega_z$.

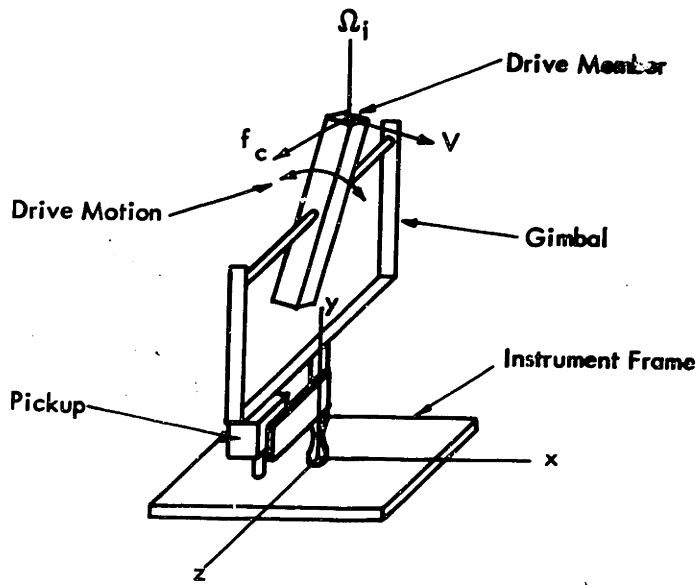


Fig. 3.9 Single Axis, Vibratory Drive, Drive Frequency Odd Harmonic Output Device

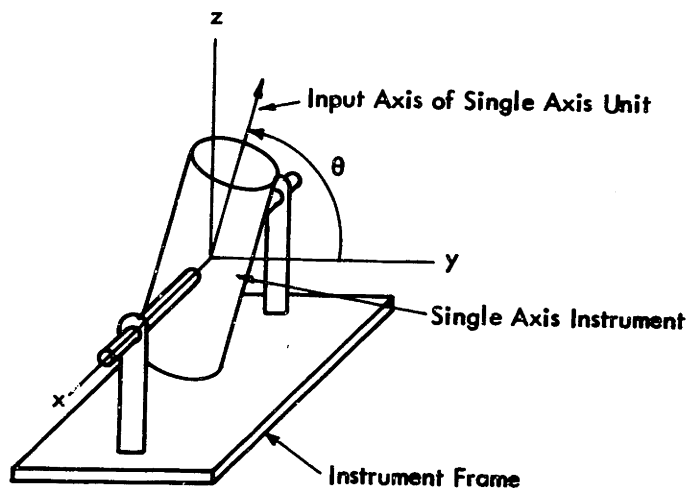


Fig. 3.10 Double Modulated Single Axis Device

In theory, double modulation can also be non-planar but this is generally impractical because it greatly complicates the instrument and the signal processing. Therefore, the rotary and vibratory double modulation examples which have just been presented are representative of the two basic types of double modulated vibratory gyroscopes.

3.6 SUMMARY

Any known vibratory gyroscope can be classified into one of the 8 basic types (numbered 1, 5, 6, 9, 10, 11, 14 and 15) shown in solid lines in Fig. 3.11. No configurations have yet been proposed which fit into the 7 other basic types (numbered 2, 3, 4, 7, 8, 12, 13) shown in dotted lines but instruments of these types are not completely unreasonable. Types which appear to have no possible merits other than the remote possibility of poorly detecting extremely large angular rates have not been included in Fig. 3.11. Obviously, most of the types presently shown in dotted lines belong in the latter-omitted-category but, since they were briefly discussed in this chapter, they are included in this summary section.

The first classification criterion used in the chart of Fig. 3.11 is whether a vibratory gyroscope is a rate or rate integrating instrument. Although certain instruments in classification No. 1 presently appear promising, the work in the following chapters is restricted to vibratory rate gyroscopes. These instruments are conveniently classified by the second criterion in Fig. 3.11 as either rotary or vibratory drive instruments depending on the drive scheme used.

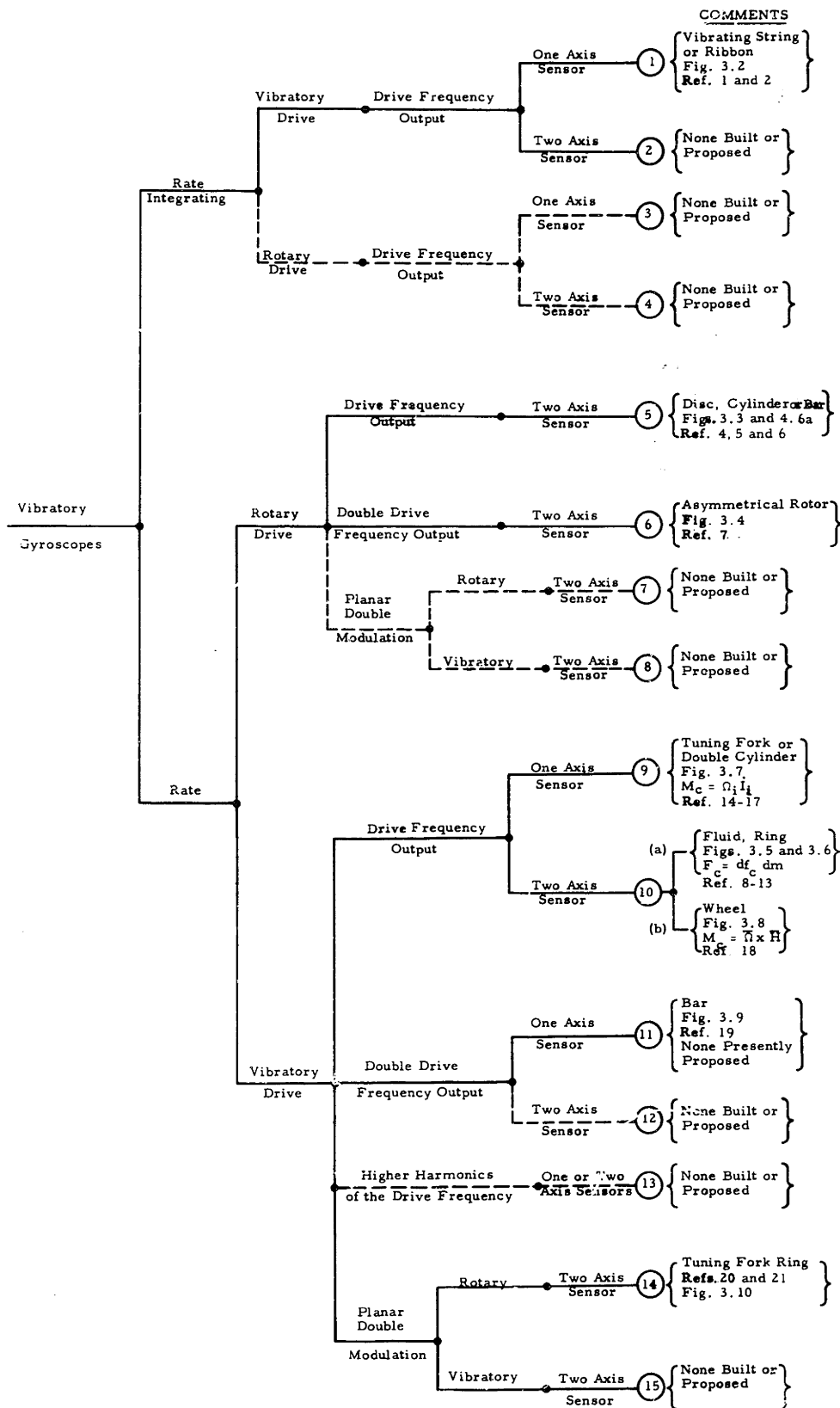


FIG. 3.11 VIBRATORY GYROSCOPES

All practical rotary drive vibratory rate gyroscopes are two axis sensors in which a single transducer is used to detect the suppressed carrier rate information about the two input axes. The two axial rate components can be obtained from the transducer output by phase sensitive demodulation because the carriers of the two components are orthogonal in time. The carrier frequency in these instruments is either the frequency of rotation or twice the frequency of rotation.* The major sources of error in these instruments are the rotor bearings and unbalanced rotors.

Practical vibratory drive, vibratory rate gyroscopes may be either one or two axis sensors, but in instruments not employing double modulation, a separate transducer is needed to detect the suppressed carrier rate information about each input axis. This is because the components of the Coriolis force containing the axial rate information are not orthogonal in time as in rotary drive instruments but orthogonal in space. The carrier frequency in these instruments is almost always equal to the frequency of the drive vibration although it is possible to have it equal to harmonics of the drive frequency. The fundamental frequency is generally preferred because the fundamental Coriolis forces (torques) are always much larger than the harmonics while the noise in the instruments is the same order of magnitude at both frequencies. The major sources of error in these instruments is unwanted cross coupling between the driving system and the sensing system.

One possible way to reduce the errors associated with unwanted cross coupling in vibratory drive rate gyroscopes is by double modulation. In this manner, the carrier frequency of the angular rate information is no longer equal to the drive frequency but is a linear combination of the two drive

* This is not true when double modulation is used but it is doubtful if instruments of this type have any advantages over the other two types whereas they certainly are more complex. Therefore they are not mentioned in the above summary.

frequencies used in the instrument. This concept is examined more thoroughly in Chapter 6.

The primary purpose of this chapter was to acquaint the reader who is unfamiliar with vibratory gyroscopes with the basic types of these instruments and to define the most important of these basic types so that they can be referred to in the following work. Therefore many approximations and omissions were made and the performance characteristics of instruments of the basic types were not presented.* In all of the examples the outputs of these instruments were simply said to be a suppressed carrier representation of the angular rates about one or more axes but the noise also present and the exact relationship between the angular rates and the modulation envelope were not examined. These factors must all be known before one can determine the characteristics and performance limitations of these instruments; therefore, they are the subject of the following chapters.

* Another reason for not presenting performance information is that much of it is classified--usually company confidential.

CHAPTER IV

ULTIMATE PERFORMANCE LIMITATIONS IN VIBRATORY GYROSCOPES

4.1 INTRODUCTION

In vibratory gyroscopes, the output is produced by vibratory forces or torques that result from Coriolis effects on moving masses. These forces or torques cause an oscillatory displacement of the output member or media and this displacement or a derivative of it or a strain associated with it is sensed by an electromechanical transducer. The output of the transducer ideally is a suppressed carrier signal in which the modulation is a function of the applied angular rate about one or more axes and the carrier frequency is the output frequency of the instrument. In practice, the signal from the transducer also contains noise and the designer attempts to maximize the signal to noise ratio. The constraint which is imposed on the maximum signal power by the strains in the drive member and the constraint on the minimum noise power which is imposed by thermal noise in the sensing system components are examined in this chapter. Also, the constraint imposed on the bandwidth of the applied rates by the suppressed carrier modulation is reviewed. These three factors - thermal noise, strain constraints in materials, and suppressed carrier signal processing - are the ultimate performance limitations in vibratory gyroscopes and a knowledge of them allows us to determine the theoretical ultimate threshold of an instrument. This threshold is a useful figure of merit for comparing different configurations and types of gyroscopes even though it is unlikely that the thresholds of actual VRG's will approach this limiting value.

4.2 SIGNAL PROCESSING IN VIBRATORY GYROSCOPES

In vibratory gyroscopes, the output motion which is sensed by an electromechanical transducer to obtain a suppressed carrier electrical signal

which is a function of the applied angular rate, is caused by Coriolis effects. As a good first approximation,* the total force or torque which causes this motion is the integral of the Coriolis forces or torques on all the driven mass elements and the structure or media on which it acts is a linear time-invariant system. Therefore, in the linear range of the output transducer, a good approximation of the suppressed carrier signal out of the electro-mechanical transducer can be obtained by convolving the suppressed carrier Coriolis torque (force) with the impulse response of the output linear sensing system - transducer combination. This signal is then amplified and filtered and demodulated to provide the instrument output which ideally is only a function of the past input angular rate.

For constant input angular rates, the output mechanical motion associated with the Coriolis torque is a sinusoidal oscillation of constant frequency and amplitude so the low frequency output of the demodulator is proportional to the angular rate. When the input angular rate is not constant but varies slowly with time, the demodulator output is no longer directly proportional to the angular rate because of the transfer characteristics of the sensing system and the filter characteristics of the demodulator. However, in most VRG's the relationship between the demodulator output and the input angular rate can be approximated by a linear time-invariant differential equation (and therefore a transfer function) as long as the carrier frequency of the suppressed carrier signals is much larger than the bandwidth of the input angular rate. This transfer function can be determined from the transfer functions of the sensing system-transducer combination and any linear filters which follow the transducer but precede the demodulator. The relationship between the

* This approximation was discussed in Ch. 3 and for most practical configurations it is extremely good. The actual differential equations which describe the instrument can best be described as horrendous because of the numerous time-varying coefficients. An example of these equations, for the tuning fork are in Section and References .

approximate transfer function and the transfer function of the suppressed carrier filters is familiar to most engineers and is just reviewed here.

Figure 4.1a represents a suppressed carrier synchronous communication link and is a good approximation of an ideal, noise free vibratory gyroscope. The transfer function of the linear system which filters the suppressed carrier signal is denoted as $G_{s.c.}(s)$. The demodulator is a multiplier with a sinusoidal sensitivity function which is the modulating signal phase shifted by an angle ψ , followed by an ideal low pass filter with a cutoff frequency equal to the carrier frequency.

The signal out of the multiplier, $f_m(t)$ in Fig. 4.1a, can be written as

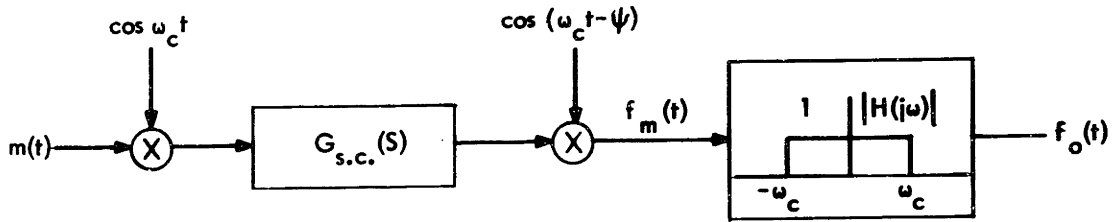
$$f_m(t) = \frac{1}{2} \int_0^t d\tau_1 g_{s.c.}(\tau_1) m(t - \tau_1) \{ \cos(\omega_c \tau_1 - \psi) + \cos[\omega_c(2t - \tau_1) - \psi] \} \quad (4.1)$$

where $m(t)$ is the input to the suppressed carrier system and $g_{s.c.}(\tau)$ is the impulse response of the suppressed carrier linear filter with a transfer function $G_{s.c.}(s)$. When the input signal is bandlimited and the bandwidth is much smaller than the carrier frequency ω_c , the spectrum of the first term of $f_m(t)$ does not overlap that of the second and the output of the ideal band pass filter is simply the first term in Eq. 4.1. By transforming this term, the transfer function $G_e(s)$ can be obtained.

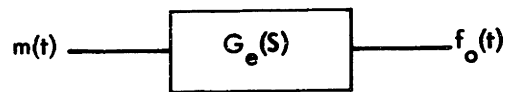
$$F_o(s) = \frac{1}{4} [e^{-j\psi} G_{s.c.}(s - j\omega_c) + e^{j\psi} G_{s.c.}(s + j\omega_c)] M(s) = G_e(s) M(s) \quad (4.2)$$

$$G_e(s) = \frac{1}{4} [e^{-j\psi} G_{s.c.}(s - j\omega_c) + e^{j\psi} G_{s.c.}(s + j\omega_c)] \quad (4.3)$$

Therefore, the effect of the suppressed carrier transmission link shown in Fig. 4.1a on a bandlimited message is approximately the same as a linear time invariant filter with a transfer function $G_e(s)$. This equivalent system



a) Suppressed Carrier Linear System



b) Approximate Equivalent Time-Invariant System

Fig 4.1 Approximation for a Suppressed Carrier System

is shown in Fig. 4. 1b.

In most vibratory rate gyroscopes and applications, the bandwidths of the input angular rates are much smaller than the output frequencies. Therefore, the approximation of Fig. 4. 1b and Eq. 4.3 is applicable for estimating the instrument transfer function if the transfer characteristics of the sensing system in the frequency range around the carrier are known. These transfer characteristics are examined in the following section.

4.3 SENSING SYSTEMS IN VIBRATORY RATE GYROSCOPES

The sensing system of a vibratory rate gyroscope consists of the output member or media which oscillates when the instrument is rotated about an input axis, the spring type element or media which provides the restoring force on the output member or media, and the output transducer which detects the output motion or a derivative thereof.* In most instruments, the transfer² characteristics of the sensing system at frequencies around the carrier frequency are similar to those of a lumped parameter second order system. Therefore, since it has been shown that this is the frequency range of major importance for estimating the transfer characteristics of a vibratory gyroscope, it is convenient to postulate** such a model for the sensing systems of vibratory gyroscopes and use it in the following work.

An electrical analog of the model chosen*** for the sensing system of a vibratory rate gyroscope is shown in Fig. 4.2. where M_c is the input Coriolis torque, I_o is the moment of inertia of the output member about the output axis, ω_n is the natural frequency of the sensing system without the transducer, Q_m

* These three components may all be the same in pizeoelectric and similar devices.

** It has been shown - Reference 4.2- that this approximation is good for distributed parameter systems.

*** A torsional model was chosen because the majority of known VRG's operate in this manner. The analogy between torsional and other types of sensing systems should be apparent to the reader.

is the quality factor of the sensing system without the transducer, and Z_i is the mechanical input impedance of the transducer. Usually, Z_i is purely resistive (in the model of Fig. 4.2) in the frequency range of interest because the stability of electrical reactive components is much poorer than that of mechanical components and stability is extremely important in VRG's. Then the sensing system-transducer combination is also a second order linear system.

When the transducer is a velocity measuring instrument, such as a variable reluctance pickup, and the normalized input impedance in Fig. 4.2 is purely resistive (which it is assumed to be in all the following work) and equal to

$$Z_i(s) = \frac{Q_t}{I_o \omega_m} \quad (4.4)$$

the transfer function of the sensing system is

$$G_{s.c.}(s) = \frac{s\theta_z(s)}{M_c(s)} = \frac{s}{I_o \left[s^2 + \frac{\omega_n}{Q_s} s + \omega_n^2 \right]} \quad (4.5)$$

where

$$Q_s = \frac{Q_m Q_t}{Q_m + Q_t} \quad (4.6)$$

In order to maximize the power into the output transducer** for slowly varying angular rates, the natural frequency ω_n should be made equal to ω_d . When this is done and $\psi = 0$ to maximize the demodulator output, the approximate transfer function for the sensing system is

* This will be shown in Chapter V.

** The impedances should also be matched.

$$G_e(s) = \frac{s Q(s)}{M_c(s)} \approx \frac{1}{4I_0} \left(\frac{\tau}{\tau s + 1} \right), \quad s \ll \omega_c \quad (4.7)$$

where

$$\tau = \frac{2Q_s}{\omega_n} \quad (4.8)$$

If feedback is used around the sensing system, as shown in Fig. 4.3, to increase or decrease the bandwidth, the value of τ in Eq. 4.7 is changed to τ' .

$$\tau' = \frac{2}{\omega_n} \left(\frac{Q_s + H}{Q_s H} \right) \quad (4.9)$$

where $\frac{I_0 \omega_n}{H}$ is the gain of the feedback loop. Therefore, the gain bandwidth product of the sensing system is unchanged by feedback.

When a position rather than a velocity type transducer is used, the approximate transfer function at resonance, $G_e(s)$, has an addition factor ω_n^{-1} . Although at first this appears to indicate that velocity type transducer are better, a closer examination reveals that the power delivered into the two types of output transducers are the same if their mechanical input impedances are the same at ω_c . When either of these systems operate off resonance, the approximate transfer function for the instrument differs greatly from Eq. 4.7 and must be computed from Eq. 4.3. This condition is not examined further because most practical vibratory gyroscopes operate at or near resonance with or without feedback. This is done so that the effect of noise appearing after the transducer is minimized because the signal is preamplified before the noise enters. Obviously, this preamplification does not occur for the noise which originates in the sensing system. The irremovable portion of this noise is caused by thermal fluctuations of the components in the sensing system and is examined in the following sections.

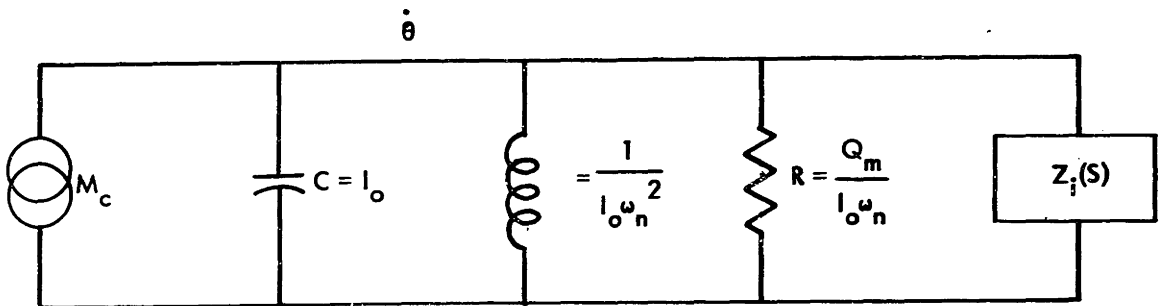


Fig. 4.2 VRG Sensing System Model

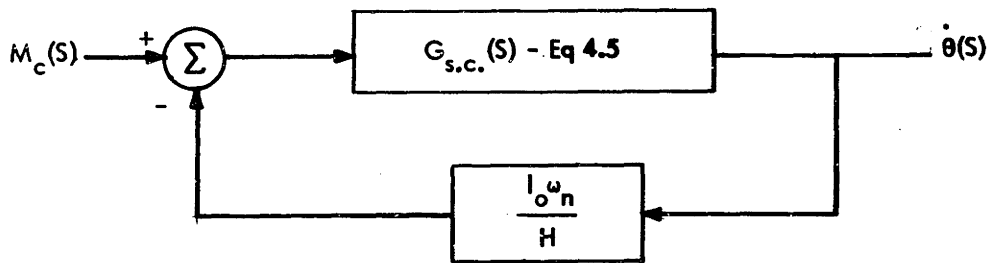


Fig. 4.3 VRG Sensing System Employing Feedback

4.4 THERMAL NOISE IN VIBRATORY GYROSCOPES

The ultimate barrier to high precision in most devices, including vibratory gyroscopes, is thermal noise. A knowledge of spectrum of this noise in an instrument is useful because it allows the theoretical threshold of the instrument to be computed. In vibratory rate gyroscopes, thermal noise results from thermal fluctuations of the components in the sensing system. The noise which is at the same frequencies as the suppressed carrier rate information cannot be distinguished from the rate signals and therefore causes zero-rate errors at the instrument output. From our knowledge of vibratory gyroscopes and statistical physics it is possible to determine the spectrum of this noise and add an appropriate noise generator to a generalized VRG model. This model is then used in Section 4.6 to determine the ultimate thresholds of the basic types of vibratory rate gyroscopes.

It is well known³ that any system which is in thermal equilibrium with its environment will exhibit thermal fluctuations of the state variables which determine the energy of the system. When the component of the systems energy that is associated with a state variable is a quadratic function of that variable, the variable must have a mean square value such that the average value of the energy is $\frac{kT}{2}$ where k is Boltzman's constant and T is the absolute temperature. Therefore, in a lumped parameter second order system, such as the model of the sensing systems of vibratory gyroscopes in Fig. 4.2a, the following relationship can be written

$$\overline{I_0 (\dot{\theta}_\tau)^2} = kT \quad (4.10)$$

where $\dot{\theta}_\tau$ is the angular velocity of the inertia I_0 which is caused by thermal noise. In this or an analogous manner the standard deviation of $\dot{\theta}_\tau$ can be determined for most vibratory rate gyroscopes. What we desire to know, however, is not the mean square value of $\dot{\theta}_\tau$ but the power density spectrum of $\dot{\theta}_\tau$ so that

the thermal angular rate threshold of a vibratory gyroscope can be computed.

Since vibratory rate gyroscopes operate at relatively low frequencies, it is safe to assume that the power density spectrum of the thermal torques* acting on the sensing system is flat and proportional to the loss term in the resonant system without the transducer. Then by using Parseval's theorem it is possible to derive the expression for the magnitude, N^2 , of the power density spectrum of the torques.

$$(\dot{\theta}_T)^2 = \frac{kT}{I_0} = \frac{1}{2\pi} \int_{-\infty}^{\infty} N^2 G_{s.c.}(j\omega) G_{s.c.}(-j\omega) d\omega = \frac{N^2 Q_s}{2I_0 \omega_n} \quad (4.11)$$

$$N^2 = 2kT \left(\frac{I_0 \omega_n}{Q_s} \right) = \frac{4kTI_0}{\tau_s}, \quad Q_s = Q_m \quad (4.12)$$

where $\frac{I_0 \omega_n}{Q_s}$ is the loss term in the resonant system. This noise torque source can now be added to the generalized VRG model in Fig. 4.2a and the power density spectrum of $\dot{\theta}_T$ can be represented as

$$\dot{\theta}_T \dot{\theta}_T(s) = N^2 G_{s.c.}(s) G_{s.c.}(-s) \quad (4.13)$$

The output transducer has been neglected in this analysis because the noise characteristics of transducers and the associated electronics vary so much from one type to another. As a reasonable approximation, let us assume** that the noise temperature of the transducer is the same as the temperature of the sensing system components. Then Eq. 4.12 is valid for any relationship

* This is analogous to the assumption made when analyzing the noise in RLC circuits. See Reference 4.

** This assumption is discussed in Chapter V and is not as bad as one would first think.

between Q_m and Q_t and we can simply measure the time constant of the transducer output to determine Q_s as long as feedback is not used around the sensing system. When feedback is used, the noise torques ideally are unchanged but the spectrum of $\dot{\theta}_T$ changes because $G_{s.c.}(s)$ changes.

Although it is useful to add a noise torque generator to the VRG sensing system model of Fig. 4.2, to represent thermal noise, it is much more useful to add an approximately equivalent noise source to a linear model such as is in Fig. 4.1b. This can be done as long as the bandwidth of the applied angular rate is much smaller than the carrier frequency by using the signal processing scheme of Fig. 4.1a and neglecting overlapping spectra or by approximating the actual noise torque by the sum of statistically independent sinusoidally and cosinusoidally modulated band limited white noise torques as shown in Fig. 4.4. The noise $n(t)$ in Fig. 4.4 is equivalent to the actual thermal noise in the frequency range of interest around the carrier frequency. Then, if the sinusoidally modulated noise is discriminated against in the demodulator and the cosinusoidally modulated noise is not, the thermal noise torque $n_c(t)$ can be considered as an input to a simplified model of a VRG in which $G_e(s)$ in Eq. 4.7 replaces the suppressed carrier instrument. This model is shown in Fig. 4.5 where $a_y \omega_c I_o$ is a normalized representation of the amplitude of the torque which is derived in the following section.

4.5 NORMALIZED CORIOLIS TORQUES (FORCES) IN VIBRATORY RATE GYROSCOPES

The total Coriolis torque (force) that causes the output in a vibratory rate gyroscope is the integral of the Coriolis torque (force) over all the driven mass elements. This torque can usually be written in a normalized manner as

$$M_c = \Omega_1 (a I_o \omega_c) \cos \omega_c t \quad (4.14)$$

where I_o is the moment of inertia of the output member in Fig. 4.2 about the output axis, α is a constant which is a function of the configuration and the material, and ω_c is the carrier frequency of the output signals. This normalized representation is useful because it means that only four parameters, α , I_o , Q_s , and ω_c , are necessary to completely describe the VRG model in Fig. 4.5.

The terms I_o and ω_c in the normalized Coriolis torque are self explanatory and have been examined previously. The remaining term, α , is now examined for the four major types of vibratory rate gyroscopes. The maximum values of α can be obtained in rotary and vibratory drive, drive and double drive frequency output instruments from the equations in Section 3.3 for the instruments shown in Fig. 3.1. The exact equations for α can also be determined from the equations presented in Chapter 3 for the more practical configurations.

In the simple rotary drive, drive frequency output instrument shown in Fig. 3.1a, the Coriolis torque can be written as

$$M_c = \Omega_1(2I_o \omega_c) \cos \omega_c t \quad (4.15)$$

from Eq. 3.5. Therefore, the value of α is 2. In more practical configurations, α , which is given the subscript R_1 for this type of instrument, is less* and can be written as

$$\alpha_{R_1} = \frac{2I_{ro}}{I_o}, \quad 0 < \alpha_{R_1} < 2 \quad (4.16)$$

where I_{ro} is the planar moment of inertia of the driven and sensed member from a plane containing the axis of rotation and the output axis; and I_o is the

* In a practical configuration of this type α_{R_1} is probably nearer to unity than 2.

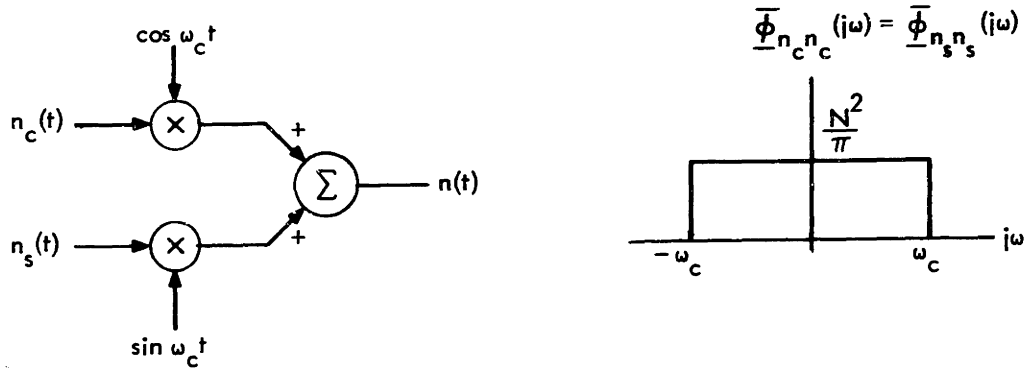


Fig. 4.4 Approximately Equivalent Thermal Noise Representation

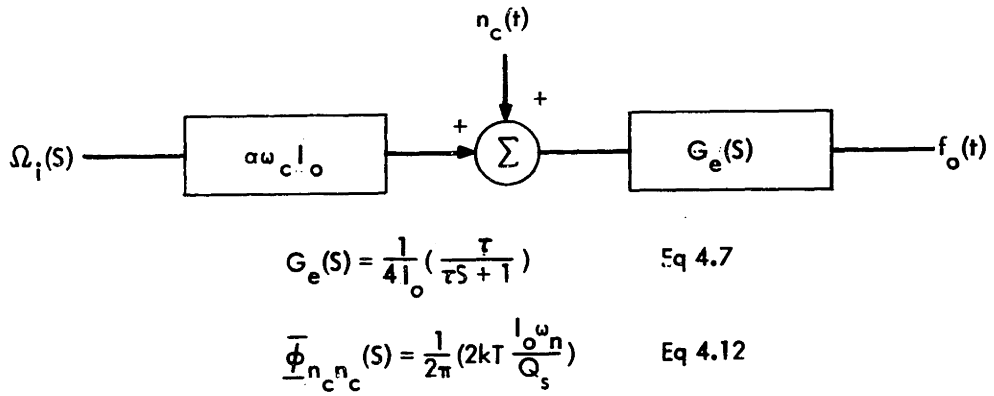


Fig. 4.5 Model of a Vibratory Rate Gyroscope

moment of inertia of the same member about the output axis.

In the simple rotary drive, double drive frequency output instrument shown in Fig. 3.1a, the Coriolis torque can be written as

$$M_c' = \left(\frac{I_o \omega_c}{2} \right) \Omega_1 \cos \omega_c t \quad (4.17)$$

from Eq. 3.7. Therefore, the value of a is $\frac{1}{2}$. In more practical configurations a , which is given the subscript R_2 for this type of instrument, is less* and can be written as

$$a_{R_2} = \frac{I_{rr \max} - I_{rr \min}}{[I_{r \max} + I_{r \min} + 2I_{g_o}]}, \quad 0 < a_{R_2} < \frac{1}{2} \quad (4.18)$$

where I_{rr} is the planar moment of inertia of the rotating member from a plane containing the axis of rotation, I_r is the moment of inertia of the rotating member about the output axis, and I_{g_o} is the moment of inertia of the gimbal about the output axis.

In both types of rotary drive instrument, the value of a_R is strictly a function of the configuration of the rotating member. This configuration is generally selected so that the strain in the rotating member is approximately uniform. When the size of an instrument is changed by linearly scaling all the dimensions, the value of a_R remains constant and independent of size. Therefore, in order to increase the Coriolis torques in a rotary drive instrument, the product $I_o \omega_c$ must be increased until the strain in the drive member reaches its maximum allowable value. This constraint is examined in Section 4.6.

In the simple vibratory drive, drive frequency output instrument shown

* In practical configurations of this type a_{R_2} is probably nearer to 0.2 than 0.5.

in Fig. 3.1b, the Coriolis torque can be written as

$$M_c = \Omega_1 \left(\frac{2I_o \omega_c r_o}{R^*} \right) \Omega_1 \cos \omega_c t \quad (4.19)$$

from Eq. 3.15. Therefore, the value of a is $\left(\frac{2r_o}{R} \right)$, which has a maximum of 2 when $r_o = R_o$. In more practical configurations a , which is given the subscript V_1 for this type of instrument, is less** and can be determined from the equation for the moment of inertia of the drive member about the output axis

$$I_o = I_o (1 + a_{V_1} \cos \omega_d t) = I_o (1 + a_{V_1} \cos \omega_c t) \quad (4.20)$$

or for instruments such as in Fig. 3.7 it is simply

$$a_{V_1} = \frac{I_d \theta_d}{I_o} \quad (4.21)$$

where I_d is the moment of inertia of the driven and sensed member about the drive axis and θ_d is the maximum angle of the drive oscillation.

In the simple vibratory drive, double drive frequency output instrument shown in Fig. 3.1b, the Coriolis torque can be written as

$$M_c' = \Omega_1 \left(\frac{I_o \omega_c r_o^2}{2R^2} \right) \cos \omega_c t \quad (4.22)$$

from Eq. 3.15. Therefore, the value of a is $\frac{1}{2} \left(\frac{r_o}{R} \right)^2$ which has a maximum value of $\frac{1}{2}$ when $r_o = R$. In more practical configurations, a , which is given

* This R is either R_x or R_y in the model of Fig. 3.1b.

** It is difficult to make a general statement about a for this type of instrument. In practice, however, it rarely exceeds 0.1 and often is orders of magnitude lower.

the subscript V_2 for this type of instrument is considerably less* and can be determined from the equation for the moment of inertia of the drive member about the output axis

$$I_o = I_o(1 + a_{V_1} \cos \omega_d t + a_{V_2} \cos 2\omega_d t) \quad (4.23)$$

In both types of vibratory drive instruments, the values of a_V is a function of the configuration of the drive member and the maximum allowable strain in the drive member. It can easily be shown that the strain in the drive member is proportional to $\left(\frac{r_o}{R}\right)$ in these instruments so that when the size of an instrument is changed by linearly scaling all the dimensions, the value of $a_{V_{max}}$ remains constant and independent of size. It can also be shown that the configuration of a vibratory drive instrument is extremely important and the product $a_V I_o \omega_d$ must be examined more closely before conclusions can be made about materials and operating frequencies. This is done in the following two sections.

4.6 STRAIN CONSTRAINTS IN VIBRATORY GYROSCOPES

The Coriolis force that is exerted on a driven mass element in a vibratory gyroscope is proportional to the drive velocity of the mass element. In order to maximize the sensitivity of an ideal fixed configuration and size VRG, the drive velocity of the mass elements must be maximized. For the model of Fig. 4.5 this is equivalent to maximizing the factor ω_c since I_o is fixed. In rotary drive instruments the velocity is maximized by increasing the drive frequency until the strain in the rotor reaches the maximum safe value - ϵ_{max} . In vibratory drive instruments the velocity is maximized by increasing the amplitude of the drive oscillation until the strain in the energy storage member

* It is impossible to make a general comment other than that the values of a_V for these instruments are much lower than for the other three types which have been considered.

reaches a maximum safe value. By examining normalized equations for the strain in these two types of gyroscope, the effect of size and material properties on the maximum velocities can be determined. For this analysis it is useful to consider only instruments in which all the dimensions are uniformly and linearly scaled when the size of the instrument changes.

The strain in the rotor of a linearly scaled rotary drive device can be written in a normalized manner as

$$\xi = \beta_{\xi} \left(\frac{\omega_d R}{c} \right)^2 \quad (4.24)$$

where R is the maximum radius of the rotor and therefore a dimension which is representative of the size of the instrument, β_{ξ} is a constant which is only a function of the rotor configuration, and c is the velocity of sound in the material. Recall that

$$c^2 = \frac{E}{\rho} \quad (4.25)$$

where E is Young's modulus for the material and ρ is the density. Therefore, both the configuration constant β_{ξ} and the maximum drive velocity

$$V_{d_{\max}} = R\omega_{d_{\max}} = c \sqrt{\frac{\xi_{\max}}{\beta_{\xi}}} \quad (4.26)$$

in a linearly scaled device are independent of the size; and the maximum drive velocity is proportional to the product $c \xi_{\max}^{1/2}$, where both are properties of the material of the drive member.

The strain in the energy storage member in a linearly scaled vibratory drive device can be written in a normalized manner as

$$\xi = \gamma^{-1} \left(\frac{r}{R} \right) \quad (4.27)$$

where γ is a constant which is only a function of the configuration, R is the average radius of the device and therefore a dimension which is comparable to R in a rotary drive device, and r is the maximum oscillatory displacement of the driven mass elements with the largest drive velocities. Another equation for the strain can be determined by equating the maximum potential energy stored in the strained member to the maximum kinetic energy.* This equation is

$$\xi = \beta_{\xi} \left(\frac{\omega_d r}{c} \right) \quad (4.28)$$

where β_{ξ} is a constant which is only a function of the configuration and $c = \sqrt{\frac{E}{\rho'}}$ where E refers to the energy storage member of ρ' refers to both the energy storage and drive member. Therefore, the maximum drive velocity in a linearly scaled vibratory drive instrument

$$V_{d_{\max}} = r_o \omega_d = \frac{c \xi_{\max}}{\beta_{\xi}} = \gamma R \omega_d \xi_{\max} \quad (4.29)$$

is independent of the size and proportional to $c \xi_{\max}$. The configuration constants a_{V_1} and a_{V_2} are also independent of the size and can be written as

$$a_{V_1} = a'_{V_1} \left(\frac{2r_o}{R} \right) = 2a'_{V_1} \gamma \xi \quad (4.30)$$

$$a_{V_2} = a'_{V_2} \left(\frac{r_o^2}{2R^2} \right) = \frac{a'_{V_2}}{2} \gamma^2 \xi^2 \quad (4.31)$$

where a'_{V_1} and a'_{V_2} are functions of the configuration which have maximum values of unity.

* At constant r or amplitude of oscillation.

Upon comparison of Eqs. 4.26 and 4.29, it is apparent that the equation for the maximum drive velocity for vibratory drive instruments has an additional factor of $\xi_{\max}^{1/2}$ in the numerator. Since ξ_{\max} is usually a number less than unity, it appears that the maximum drive velocities in vibratory drive instrument may be lower than in rotary drive instruments of equal size. This is usually true and is best illustrated by examining the two instruments shown in Fig. 4.6. The instrument shown in Fig. 4.6a is a rotary drive, drive frequency output device and the one in Fig. 4.6b is a vibratory drive, drive frequency output device. Both instruments employ a ring of radius R and cross sectional area S as a drive member which is why these particular configurations were chosen for this illustration. The maximum drive velocity for the ring elements in the rotary drive device is

$$V_{d_{\max}} = R\omega_{d_{\max}} = c \sqrt{\frac{\xi_{\max}}{2}} \quad (4.32)$$

while in the vibratory drive device it is

$$V_{d_{\max}} = r_o \omega_d = c \xi_{\max} \quad (4.33)$$

Also the drive frequency in the vibratory device is

$$\omega_d = \frac{c}{R} \quad (4.34)$$

Therefore, the ratio of the maximum drive velocities is $\sqrt{\frac{2}{\xi_{\max}}}$ which for steel is approximately 50. This means that the Coriolis forces in the rotary drive, drive frequency output device are a factor of $\sqrt{\frac{2}{\xi_{\max}}}$ greater than the fundamental Coriolis forces in the vibratory drive device and a factor of $\frac{2}{\xi_{\max}}$ greater than the second harmonic Coriolis forces in the vibratory drive device. This large difference arises because in the rotary drive device the kinetic

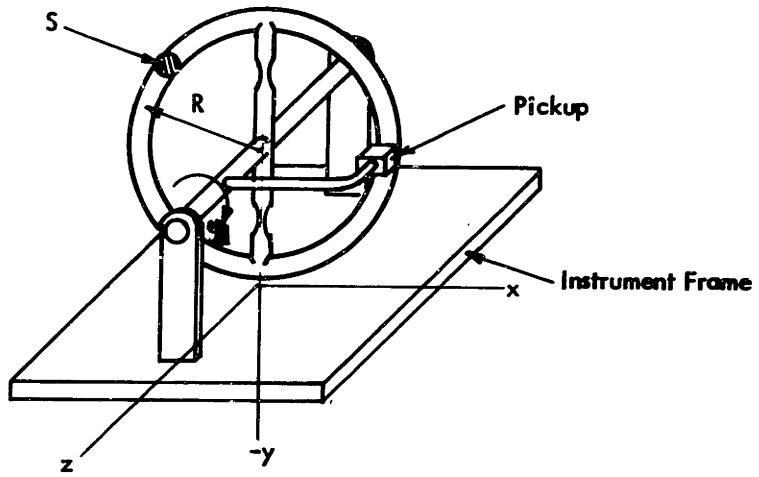


Fig. 4.6a Rotary Drive, Drive Frequency Output Device

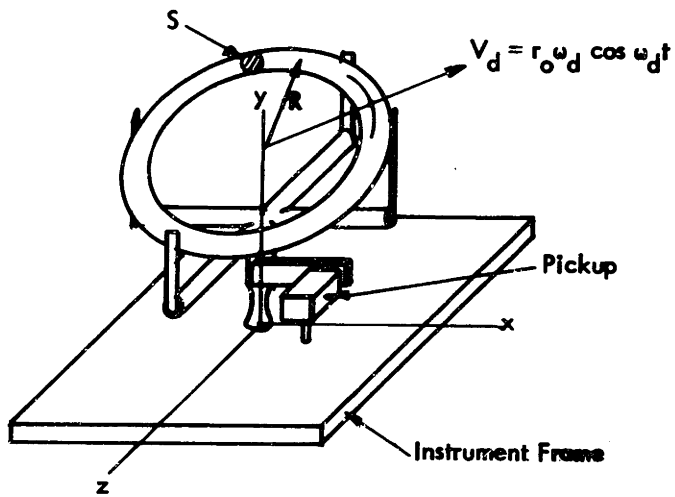


Fig. 4.6b Vibratory Drive, Drive Frequency Output Device

energies of the driven mass elements is much greater than the potential energy stored in the strained rotor whereas in the vibratory drive device the two must be equal.

It appears from the examples in Fig. 4.6 and Eqs. 4.26 and 4.29 that rotary drive devices may be more sensitive than vibratory drive devices of the same size. Although this is generally true, it cannot be determined from the magnitude of the drive velocities alone because vibratory drive configurations can probably be found in which the maximum drive velocities are equal to those of a rotary drive instrument of the same size and material. Instead, it is necessary to know the relative magnitudes of the signals and noise in both types of instruments since this is a meaningful and useful criterion for comparing the sensitivities of instruments. In the following section the signal to thermal noise ratios for the basic types of vibratory rate gyroscopes are determined and the relative sensitivities of the basic types compared.

4.7 THEORETICAL ULTIMATE PERFORMANCE OF VIBRATORY RATE GYROSCOPES

The theoretical ultimate performance of a vibratory rate gyroscope is limited by the thermal noise in the sensing system, the maximum allowable strain in the drive member and the ratio of the bandwidth of applied angular rates to the drive frequency. In an instrument in which the bandwidth of the applied angular rates is much smaller than the output carrier frequency, the noise temperature of the transducer is equal to the temperature of the instrument, and the signal processing electronics are noise free; the VRG model shown in Fig. 4.5 and the strain relations derived in Section 4.6 can be used to determine the theoretical ultimate performance. It is convenient to express the ultimate performance in terms of an angular rate threshold - Ω_{θ} - which is defined as the constant angular rate which causes the signal to noise ratio at the output of the VRG model in Fig. 4.5 to be unity. Because the noise in the model

of Fig. 4.5 is white, Ω_T is proportional to the square root of the bandwidth of $G_e(s)$. Therefore, in order to make Ω_T a figure of merit which is useful for comparing the basic types of vibratory rate gyroscopes, it is necessary to make the bandwidths of the instruments identical by using an ideal noise free feedback loop around the sensing system or else an external filter. In either case the following equation can be obtained from the model of Fig. 4.5 for a VRG.

$$\Omega_T = \left(\frac{a \omega_c I_o}{kT \frac{\omega_n}{Q_s} \text{ B. W.}} \right)^{-1} \quad (4.35)$$

where all the terms have been defined previously except B. W. which is the instrument bandwidth in radians per second.

It is convenient to rewrite the equations for Ω_T as

$$\Omega_T^{-1} = \left(a \sqrt{I_o \omega_c' Q_s} \right) \left(\frac{1}{\sqrt{kT}} \right) \left(\frac{1}{\sqrt{\text{B.W.}}} \right) \quad (4.36)$$

where $\omega_c' = \frac{\Delta}{\omega_n} \omega_c \left(\frac{\omega_c}{\omega_n} \right)$ (4.37)

because the first term is a function of the configuration and the materials, the second is a function of the instrument temperature and the third is a function of the instrument and noise bandwidth. Therefore, the first factor, which hereafter is called F_i where the subscript i denotes the type of instrument,

$$F_i = a_i \sqrt{I_o \omega_c' Q_s} \quad , \quad i = R_1, R_2, V_1, V_2 \quad (4.38)$$

is the only term of interest for comparing configurations and types of instruments and it should be maximized in order to minimize Ω_T .

When the equations from Sections 4.5 and 4.6 are used to define a and ω_c and the inertia of the output member is written in a normalized manner as

$$I_o = \beta_1 m R^2 \quad (4.39)$$

where β_1 is a function of the configuration and m is the mass of the output member which is proportional to ρR^3 , the following equations can be determined for F . In these equations the assumptions have been made that Q_s is not a function of ω_c and also $\omega_c = \omega_n$. This frequency relationship has been used to simplify the equations and also because it is the relationship usually employed in VRG's to maximize the amplification of the rate signals before the noise in the signal processing electronics is added. If another frequency relationship is used, * the following equations for F must be multiplied by $\sqrt{\frac{\omega_c}{\omega_n}}$.

$$F_{R_1} = a_{R_1} \sqrt{\beta_1 m R Q_s (R \omega_c)_R}, \quad \omega_c = \omega_n = \omega_d \quad (4.40)$$

$$F_{R_2} = a_{R_2} \sqrt{\beta_1 m R Q_s (R \omega_c)_R}, \quad \omega_c = \omega_n = 2\omega_d \quad (4.41)$$

where

$$(R \omega_c)_R \triangleq c \sqrt{\frac{\xi_{\max}}{\beta_{\xi}}}$$

$$F_{V_1} = 2a'_{V_1} \left(\frac{r_o}{R}\right) \sqrt{\beta_1 m R Q_s (R \omega_c)_V}, \quad \omega_c = \omega_n = \omega_d \quad (4.42)$$

$$F_{V_2} = \frac{a'_{V_2}}{2} \left(\frac{r_o}{R}\right)^2 \sqrt{\beta_1 m R Q_s (R \omega_c)_V}, \quad \omega_c = \omega_n = 2\omega_d \quad (4.43)$$

where

$$\left(\frac{r_o}{R}\right) \triangleq \gamma \xi_{\max}$$

* Other frequency relationships may be used to reduce errors caused by a temperature shift of ω_n and to obtain instruments with wide bandwidths without feedback. When this is done, however, that factor

$$\sqrt{\frac{\omega_c}{\omega_n}}$$

rarely exceeds $10^{\pm 1/2}$ so the errors in F that results from neglecting this factor are not too great.

$$(R\omega_c) \frac{1}{V} \frac{c}{\gamma\beta_g R}$$

$$\omega_c = \frac{c}{\gamma\beta_g R}$$

An attempt was made to write Eqs. 4.40 - 4.43 in terms of the most meaningful parameters, i. e., parameters that can easily be estimated in an actual instrument. For a rotary drive instrument α_{R_1} , α_{R_2} , β_1 , and β_g are easily estimated from the configuration of the drive member and gimbal. For a vibratory drive instrument γ can be estimated from the configuration and materials of the drive system and therefore β_g can be determined, and α_{V_1} or α_{V_2} can be determined from the configuration of the instrument with a small amount of effort.

Numerous conclusions can now be drawn from Eqs. 4.40 - 4.43 about materials and configurations for vibratory gyroscopes and the effect of size on Ω_T . The most important of these conclusions are listed below.

1. F_R and Ω_T^{-1} in all instruments are proportional to R^2 . Therefore, the bigger the instrument the lower the ultimate threshold which is certainly reasonable.
2. The material for the drive member of a rotary drive device should be selected to maximize the term $(\rho c \xi_{\max}^{1/2})^{1/2}$
3. The material for the drive member in a vibratory device should be selected to maximize the term $(\xi_{\max} \rho^{1/2} c^{1/2})$ for drive frequency output instruments
 $(\xi_{\max}^2 \rho^{1/2} c^{1/2})$ for double drive frequency output instruments
 if the drive member is made of a single material. If the energy storage material differs from that of the driven mass elements it must be kept in mind that $c = \frac{E}{\rho'}$ where E refers to the energy storage material and ρ' refers to both the energy storage and driven materials.
4. The material for the energy storage member in the sensing system should be selected for maximum Q. Furthermore, a high input impe-

dance transducer should be used to sense the output motion so that Q_s is large.

5. If there is a size constraint in a vibratory drive, drive frequency output instrument, the term $r_0(m\omega_c)^{1/2}$ should be maximized in order to minimize Ω_T . This usually occurs when m is made as large as possible and ω_c is made as small as possible; however, in some configurations, such as a tuning fork with uniform tines, the term $r_0\omega_c^{1/2}$ is constant. In this case, if Ω_T is the only quantity of importance, ω_c can be selected anywhere in a range of frequencies.
6. If there is a size and an input power constraint in a vibratory drive, drive frequency output instrument and the quality factor of the driven member is independent of frequency, Ω_T^{-1} is proportional to $(\frac{P_{in} Q_d}{\omega_c})^{1/2}$ where Q_d is the quality factor of the drive member and P_{in} is the input power. In this case, if Ω_T is the only quantity of importance, ω_c should be made as small* as possible.

Although numerous other conclusions could be drawn from the equations for Ω_T , it is probably more worthwhile at this time to apply the equations to the two configurations shown in Fig. 4.6. In this way an order of magnitude number can be obtained for the theoretical ultimate thresholds of rotary and vibratory drive devices of the same size. For the comparison a radius of 2 cm was selected since this would result in an instrument of approximately 2 - 2^{1/2} inches on a side which is a practical dimension. The ring dimensions, material properties (steel) and all the configuration constants mentioned in the chapter for the two devices shown in Fig. 4.6 are listed in Table 4.1. The calculated values of F and Ω_T are then listed in Table 4.2 for the two devices.

If there were no other sources of error in vibratory rate gyroscopes than thermal noise in the sensing system and output transducer and also noise in the signal processing electronics, it would not be unreasonable to expect that the

* The choice of high or low operating frequencies for vibratory drive instruments is a controversial topic. Other evidence supporting low operating frequencies is given in the following chapter.

Parameter	Rotary Drive Device Fig. 4.6a	Vibratory Drive Device Fig. 4.6b
R cm	2	2
S cm ²	1	1
$\rho \frac{\text{gm}}{\text{cm}^2}$	7.6	7.6
$c \frac{\text{cm}}{\text{sec}}$	5×10^5	5×10^5
ξ_{max}	10^{-3}	10^{-3}
Q	10^3	10^3
m gm	95	95
β_i	1	1
$k \frac{\text{erg}}{\text{cm}}$	1.38×10^{-16}	1.38×10^{-16}
T °K	300°	300°
$\sqrt{kT} \text{ erg}^{1/2}$	2×10^{-7}	2×10^{-7}
B.W. $\frac{\text{radians}}{\text{sec}}$	100	100
$\alpha_{R_1}, \alpha_{V_1}, \alpha_{V_2}$	1	1 and 1
β_ξ	2	1
γ	-	1
$V_{d\text{max}} \frac{\text{cm}}{\text{sec}}$	1.1×10^4	5×10^2
$\omega_d \frac{\text{radians}}{\text{sec}}$	5.5×10^3	2.5×10^5

Table 4.1 Parameters of Devices Shown in Fig. 4.6

Parameter	Rotary Drive Device-Fig. 4.6a	Vibratory Drive Device - Fig. 4.6b Output Frequency	
		Fundamental	Second Harmonic
$F_{R_1}, F_{V_1}, F_{V_2}$	4.5×10^4	6.1×10^2	1.5×10^{-1}
$\Omega_T \left(\frac{\text{radians}}{\text{sec}}\right)$	4.5×10^{-11}	3.3×10^{-9}	1.3×10^{-5}

Table 4.2 Ultimate Thresholds of Devices Shown in Fig. 4.6

thresholds of well engineered VRG's of this same size (2 - 2¹/₂ inches) would be approximately 100 times larger than the numbers for Ω_T in Table 4.2. These more practical numbers, changed into degrees per hour, are in Table 4.3 along with an estimate* of the threshold of a rotary drive, double drive frequency output device of the same size.

Type of Instrument	Size	Projected Practical Threshold
Rotary Drive Drive Frequency Output	2 ¹ / ₂ "	1 x 10 ⁻³ °/hr
Rotary Drive Double Drive Frequency Output	2 ¹ / ₂ "	2 x 10 ⁻³ °/hr
Vibratory Drive Drive Frequency Output	2 ¹ / ₂ "	7 x 10 ⁻² °/hr
Vibratory Drive Double Drive Frequency Output	2 ¹ / ₂ "	3 x 10 °/hr

Table 4.3 Projected Practical Thresholds of a 2¹/₂" Vibratory Rate Gyroscope

The numbers in Table 4.3 are simply crude estimates of the actual thresholds that one might expect for the four basic types of vibratory rate gyroscopes if there were no other sources of error than those mentioned in this chapter. These numbers have been presented because it is difficult to ascertain anything of this nature from Eqs. 4.40 - 4.43 alone. Furthermore, these members are probably conservative** because higher strength materials than the steel used in the examples are available and also higher Q's than 1000 can be obtained in sensing systems.

* The estimate is simply twice that of the rotary drive, drive frequency output device.

** These numbers are particularly conservative for the vibratory drive instruments.

It is apparent from Table 4.3 that the thresholds of both types of rotary drive instruments and of drive frequency output, vibratory drive instruments are small enough to more than satisfy the specifications of many applications. It is the unfortunate fact, however, that the unclassified thresholds of the best vibratory gyroscopes (assuming the numbers are scaled by a factor of R^2 and the bandwidths are the same) of these three types are at least three orders of magnitude larger. This indicates that factors other than thermal noise and amplifier noise are presently limiting the performance of these instruments. The most important of these "practical" performance limitations are examined in the following chapters.

CHAPTER V

PRACTICAL PERFORMANCE LIMITATIONS IN VIBRATORY RATE GYROSCOPES

5.1 INTRODUCTION

The precision of all vibratory rate gyroscopes that have been built to date has been more than three and often six orders of magnitude worse than the theoretical ultimate which is dictated by the thermal noise in the sensing system and the strain constraint in the drive member. The factors which are responsible for the low precision of these instruments are the instability and inhomogeneity of materials and the limited tolerances to which mechanical parts can be made. These "practical" performance limitations are examined in this chapter with the major emphasis being placed on vibratory drive instruments. This is accomplished by first examining in detail the sources of error in the tuning fork gyroscope and then generalizing the results to include all vibratory rate gyroscopes.

The tuning fork was selected as the configuration to examine in detail because it has been the most popular configuration for experimental and theoretical research on vibratory gyroscopes and it was the configuration used in the double modulation experiment which is reported in Chapter 6. This does not mean that the author considers the tuning fork to be superior to other configurations for laboratory type or practical instruments. On the contrary, the tuning fork is a poor* configuration for an actual instrument but it is excellent for demonstrating all the major sources of error in vibratory drive, vibratory rate gyroscopes.

* The tuning fork is a poor configuration for a vibratory gyroscope because it is extremely acceleration and temperature sensitive and it is difficult to accurately machine in a single piece.

5.2 TUNING FORK GYROSCOPE

5.2.1 History of Tuning Fork Gyroscopes

A conventional tuning fork gyroscope is a single axis, vibratory drive, vibratory rate gyroscope in which the drive member is a tuning fork and the carrier frequency of the output is the fundamental frequency of the tuning fork tine oscillation. This configuration has been the subject of much of the theoretical and experimental vibratory gyroscope research over the past 25 years because of its simplicity and potential high sensitivity. In the United States, Lyman¹ is credited with the basic invention of the tuning fork gyroscope and the Sperry Gyroscope Company in the development of the "Gyrotron"^{2, 3}. expended the major effort on it. This work started in the 1940's and resulted in an instrument with a long term stability of approximately 10^0 /hr and a short term stability more than an order of magnitude better; however, the instrument was not a low power, low cost device because the ambient temperature of the fork had to be closely controlled and a time consuming precision balancing operation was required for the tuning fork. Therefore the instrument was never put into production because it had few if any advantages over other types of gyroscopes for most applications.

Since the development of the "Gyrotron", numerous other experimental tuning fork gyroscopes have been built and at the present time research is being conducted on at least two separate tuning-fork type instruments. Although the reported drift rates of some of these instruments have been lower than that of the Gyrotron, long term drift data or even mention of it is noticeably absent in most of the reports on these instruments with the exception of those^{4, 5} on the A5 gyroscope currently being developed at the Royal Aircraft Establishment in Farnborough England. This instrument appears to be more than an order of magnitude better than the Gyrotron but it is an expensive instrument.

because the tuning fork requires hundreds of man hours to machine and balance. Also it is not a low power device because the temperature of the tuning fork must be maintained constant.

Reported data on the A5 gyroscope and the Gyrotron together with data obtained on the crude tuning fork gyroscope used in the double modulation experiment which is reported in Chapter VI are used throughout this chapter to give the reader an order of magnitude feeling for the various limitations that are discussed and to partially support some of the estimations and conclusions that are made. Data from the A5 and Gyrotron are used because these instruments are undoubtedly the two best tuning fork gyroscopes that have been built. The data from the instrument used in the double modulation experiment are used because this instrument is typical of many of the other tuning fork gyroscopes that have been built and the author is familiar with it and therefore able to make meaningful estimates of the characteristics of an improved version of this instrument. To simplify the repeated references to this instrument in this chapter it is called the UTFG-1 where the letters stand for unmodulated tuning fork gyroscope.

5.2.2 Cross Coupling

Many of the sources of error in the tuning fork gyroscope have been pointed out by Morrow⁶ and others.⁷ They have shown that the major barrier to high precision is associated with the forces required to drive the mass elements along their trajectories in the instrument reference frame. Minute fractions

of these forces couple into the sensing system of the instrument in an erratic manner which varies with the passage of time, linear accelerations and changes in the environmental parameters such as temperature. The torques produced by these cross-coupled forces which are at the same frequency as the Coriolis torques are distinguishable from the Coriolis torques only to the extent that they remain constant and can be bucked or calibrated out or else the signals produced by them can be discriminated against in the phase sensitive demodulation which occurs in the signal processing.

In order to understand more clearly how unwanted cross coupling can occur between the driving and sensing systems, an idealized model of a tuning fork gyroscope is now examined. Figure 5.1 shows an idealized double tuning fork in which the arcuate motion of the opposed tines has been idealized as the straightline motion of the masses represented by the dotted spheres. Ideally these two equal masses move periodically in colinear paths along the x_f axis and are always symmetric with respect to the stationary y_f axis. In practice, however, none of these conditions is true and the trajectories of the unequal masses are slowly varying functions of time, temperature, linear acceleration and other factors. To account for some of these effects, the masses are represented by the solid spheres in Fig. 5.1 and the trajectories are straight lines parallel to the x_f axis, but displaced a distance ΔR on each side of the x_f axis. The displacement ΔR causes a torsional cross coupling between the drive member and the sensing system. There are numerous other asymmetries which also produce unwanted cross coupling; however, they are not examined because this simple example clearly illustrates the difficulties that are encountered.

When the motion of the two masses in Fig. 5.1 is sinusoidal,

$$x_{m_1} = R_0 + r_0 \sin \omega_d t \quad (5.1)$$

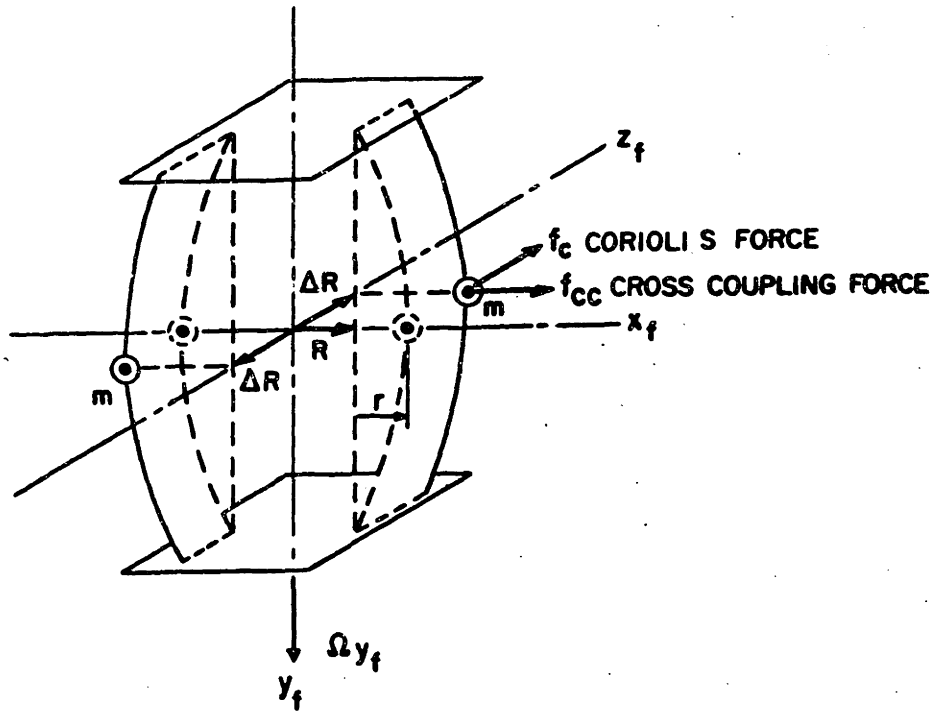


Fig. 5.1 Idealized Double Tuning Fork

BUSH-NEWTON

the fundamental component of the Coriolis torque which is exerted about the y_f axis of the tuning fork is

$$M_c = -\Omega_{y_f} 4mR_o r_o \omega_d \cos \omega_d t \quad (5.2)$$

and the cross-coupled torque about the same axis which results from the asymmetry ΔR is

$$M_{cc} = 2mr_o \omega_d^2 \Delta R \left(\sin \omega_d t + \frac{1}{Q_d} \cos \omega_d t \right) \quad (5.3)$$

where Q_d is the quality factor of the tuning fork in its driven mode. The first or sinusoidal term in Eq. 5.4 is caused by the reaction forces applied by the tines on the masses and center support to accelerate the masses. Therefore this term is 90° out of phase in time with the Coriolis torques which are proportional to the velocity of the masses. The second term in Eq. 5.3 is caused by the forces applied on the masses to supply the loss energy to the tuning fork. This term is in phase with the Coriolis torque and a factor of Q_d^{-1} smaller than the quadrature cross coupling. When the drive mechanism is carried by the tuning fork structure so that the drive forces act only on the tuning fork, the second term in Eq. 5.3 does not exist for the model of Fig. 5.1. In practice, however, because of asymmetries in the loss paths, a small inphase loss term appears even when the drive mechanism is carried by the tuning fork.

The total torque, M_s , acting on the tuning fork along the y_f axis is the sum of the Coriolis and cross-coupled torques and can be written as

$$M_s = -4mR_o r_o \omega_d \left[(\Omega_{y_f} + \Omega_d) \cos \omega_d t + \Omega_q \sin \omega_d t \right] \quad (5.4)$$

where Ω_d is an equivalent angular rate representing the component of the cross-coupled torque which is in phase with the Coriolis torque and Ω_q is an equivalent angular rate representing the component of the cross-coupled torque which is in quadrature with the Coriolis torque. For the model of Fig. 5.1, where only

torsional cross coupling exists and the drive mechanism is not carried by the tuning fork, Ω_q and Ω_d are defined as

$$\Omega_d \triangleq - \frac{\Delta R \omega_d}{2R_o \Omega_d} \quad (5.5)$$

$$\Omega_q \triangleq - \frac{\Delta R \omega_d}{2R_o} \quad (5.6)$$

In order to understand more clearly the magnitude of the cross-coupling problem, typical numbers for the RMS values of the quadrature and inphase cross coupling (denoted Ω_q and Ω_d respectively) in the A5, Gyrotron and UTFG-1 tuning fork gyroscopes are listed in Table 5.1. Also listed in this table are the authors estimates of cross coupling in an improved version of the UTFG-1 which is called the UTFG-H where H stands for the hypothetical. This hypothetical instrument is used to tabulate the authors estimates of the characteristics of a good cheap* tuning fork gyroscope where cheap means that in quantity the instruments would sell for approximately \$200. The reported or estimated values of Ω_q in these instruments are then used to determine $\frac{\Delta R}{R_o}$ by using Eq. 5.5.

Parameter	Instrument			
	A-5	Gyrotron	UTFG-1	UTFG-H
Ω_q ($^{\circ}$ /hr)	10	10^3 *	3×10^4	1.5×10^3
Ω_d ($^{\circ}$ /hr)	10	50 *	3×10^2	1.5×10^2
ω_d ($\frac{\text{radians}}{\text{sec}}$)	$2\pi(1000)$	$2\pi(1800)$	$2\pi(360)$	$2\pi(360)$
Q_d	10^4	5×10^3 *	2×10^3	8×10^3
R_o (inches)	0.8*	0.6*	0.5	0.5
$\frac{\Delta R}{R_o}$ (Eq. 5.6)	1.5×10^{-8}	8×10^{-7}	1.2×10^{-4}	6×10^{-6}
ΔR (inches)	1.2×10^{-8}	4.8×10^{-7}	6×10^{-5}	3×10^{-6}

Table 5.1 Torsional Unbalance Data

* Estimated numbers.

The tuning forks in all of these instruments were approximately the same size and the average length was approximately 2.5 inches (in the UTFG-1 it was actually approximately 5 inches but it was a double tuning fork). In the A5 the single piece tuning fork required hundreds of hours of machining and balancing time to reduce the cross coupling to the level indicated in Table 5.1 whereas the tines in the UTFG-1 were replaceable and clamped on and the balancing operation required approximately 3 man hours. In this operation the masses of the tines were adjusted to minimize the motion of the tuning fork center support in the x_f direction and then the centers of mass of the tines were adjusted in the z_f direction to minimize the torsional unbalance. Both of these adjustments were done with set screws in the tines. Smaller values of Ω_q were obtained on occasion for the UTFG-1 and it is conceivable that in an improved version of this instrument a patient balancer could reduce Ω_q to approximately 1.5×10^3 °/hr in a few hours time. This number is an engineering estimate based on the limited balancing experience of an impatient balancer--the author--using crude equipment.

Cross coupling magnitudes of 10^3 °/hr and 10^4 °/hr most likely appear large to readers who are unfamiliar with vibratory gyroscopes; however, they are not unreasonable for crude vibratory drive, vibratory rate gyroscopes of this size. Cross coupling magnitudes below 10^2 °/hr are extremely difficult to attain but they are possible in large low frequency instruments if the materials are free of defects and the tolerances on the tuning fork dimensions are very small and a precision balancing operation is performed on the tuning fork. The relationship between the size of an instrument and the magnitude of the cross coupling can be determined from Eqs. 5.5 and 5.6 for the model of Fig. 5.1. In a linearly scaled instrument as was considered in Chapter III, the drive frequency is inversely proportional to the size, the quality factor ideally is independent of the size, and normalized balances such as $\frac{\Delta R}{R_0}$ are much easier to obtain as the size increases. Therefore the magnitude of the cross coupling in a linearly scaled, vibratory drive instrument should be inversely proportional to the linear size of the instrument raised to a power between one and three.* If the instrument size is fixed, which is often the case, the cross coupling should be proportional to the drive frequency in which case it is desirable to design the drive member so that the drive frequency is low rather than high.

* The factor of three is obtained if ΔR is inversely proportional to R_0 .

If the cross-coupled torques were constant and the other components in a tuning fork gyroscope were ideal, cross coupling would not be troublesome because the torques or the signals resulting from them could be bucked out or else discriminated against in the demodulator. Unfortunately, in an actual instrument the cross coupling is not constant and the other components are not ideal and 'zero-rate errors' occur at the instrument output because of the cross-coupled torques. The most important causes of these zero rate errors in tuning fork gyroscopes are examined in the following three sections.

5.2.3 Time-Varying Cross Coupling

Since extremely small normalized asymmetries such as $\frac{\Delta R}{R_0}$ cause large cross-coupled equivalent rates, it is not difficult to see why the cross coupling in tuning fork gyroscopes is not constant. Non-uniform creep in the materials, which is aggravated by the large strains in the energy storage member in the drive system, causes slowly varying changes in the balance of the fork. Non-uniform thermal expansion and deflections under linear accelerations change both the balance of the fork and the alignment of the drive mechanism. These three causes of time-varying cross coupling are now examined separately.

Non-uniform creep in materials is a ~~severe~~ problem in vibratory drive, vibratory rate gyroscopes because it causes the balance of the drive member to change slowly with time. The standard deviation of the resulting time-varying cross coupling can be minimized by:

1. choosing the proper material for the drive member and then inspecting the material for defects,
2. machining the drive member to very close tolerances so that symmetrically located parts are as symmetric as possible, and
3. operating the drive member at low strain levels.

Unfortunately, the machining to close tolerances greatly increases the cost of the instrument and the operation at low strain levels reduces the signal power out of the sensing system transducer for a specific angular rate input. Also, the maximum allowable strain in stable materials such as quartz is orders of magnitude lower than it is in less stable steel alloys so if a stable material is selected, the noise in the sensing system, transducer and signal processing electronics becomes much more important than when less stable materials are used. This problem becomes more apparent by examining actual numbers for tuning fork gyroscopes.

The tuning forks in the instruments listed in Table 5.1 were made of steel alloys and the A5 and Gyrotron forks were operated at very low strain levels-- probably on the order of 10^{-4} . The changes in the cross coupling in these instruments with time corresponded to normalized unbalances $-\frac{\Delta R}{R_0}$ of approximately 10^{-9} and 10^{-8} respectively where the difference was due to better materials and machining in the A5. The forks in both of these instruments were very expensive because of the extensive machining required. However, even with all of this care, the long term drift of the A5 $-0.1^\circ/\text{hr}$ - was more than two* orders of magnitude larger than the thermal fluctuation threshold. This indicates that theoretically the amplitude of the drive oscillation and therefore the strain in the drive member could be reduced by almost a factor of 100 without seriously degrading the performance of the instrument. Therefore it appears advantageous to use materials such as quartz, which has a maximum allowable strain approximately two orders of magnitude smaller than that of steel, for the drive member because the creep and therefore the time-varying cross coupling should be much lower than in a steel drive member if both are made to the same tolerances.

* This number was estimated from data on the UTFG-1 in which $\alpha_{v_1} = 10^{-1}$ and $\xi = 10^{-3}$ and it assumes that in the A5 $\alpha_{v_1} > 10^{-2}$.

This theoretical conclusion has practical significance if the noise in the sensing system electromechanical transducer and signal processing electronics is not, or forseebly can be reduced to a point where it is not, several orders of magnitude larger than the thermal fluctuation noise which was analyzed in Section 4.4. At the present time this is known to be true only for capacitance type transducers. Hawley⁸ has reported that a Western Electric Model 640-AA condenser microphone can sense diaphragm displacements of 6×10^{-12} cm with a unity signal to noise ratio in a 10 cps bandwidth. This displacement is approximately equal to the standard deviation of the thermal fluctuations in the sensing system of a tuning fork gyroscope, such as the UTFG-1,* which has a moment of inertia about the output axis (I_o) of 10^3 gm cm², a drive frequency of 360 cps and is measured at 2 cm radius. (This number can be computed by using Eq. 4.10 and these members,) Although the reported sensitivities of other types of electromechanical transducers are lower than that of the condenser microphone, it does not seem unreasonable to predict that it should be possible in practice to reduce the noise in the sensing system, electromechanical transducer and signal processing electronics in a well designed vibratory rate gyroscope to a point where it is less than two orders of magnitude larger than the thermal fluctuation noise in the sensing system. If this is true, stable materials such as quartz should definitely be used for the drive members of low cost, vibratory drive instruments in which the balance is not as good as it could be. In expensive, high precision instruments the choice is not clear but it appears that the same conclusion is true; however, additional research on electromechanical transducers may be necessary.

* The UTFG-1 is used in this example because it is the only instrument for which the author knows the numerical value of I_o .

Non-uniform thermal expansion of materials is a severe problem in vibratory drive, vibratory rate gyroscopes because it causes the balance of the drive member and the alignment of the drive mechanism to change slowly with time unless the temperature is maintained constant. Furthermore, because energy is dissipated in the drive member and the drive mechanism when the instrument is operating, the warm up time is often large* even if the instrument temperature is maintained constant when the drive member is turned off. The magnitude of the time-varying cross coupling which is caused by these non-uniform thermal expansions can be minimized by:

1. inspecting the materials of the drive member and drive mechanism for defects,
2. machining the parts to very close tolerances so that symmetrically located parts are as symmetric as possible, and
3. thermally insulating the tuning fork environment and surrounding the tuning fork with a good heat conductor to minimize the temperature gradients seen by it.

Unfortunately, the machining to close tolerances greatly increases the cost of an instrument and the insulation and heat conductor increase the size and weight and even with these precautions it appears that a temperature controller is required if a high precision instrument must operate over a temperature range of more than a few centigrade degrees. This can be seen from the numbers listed in Table 5.2 for the first order temperature coefficients of the four instruments previously examined.

Stable materials such as quartz and configurations other than the tuning fork should reduce the temperature coefficients in vibratory drive instruments below the numbers listed in Table 5.2 but it is impossible with the data presently available to estimate the magnitude of these improvements. Other than this

* Many tens of minutes in high precision instruments like the Gyrotron and A. 5.

Instrument Parameter	A5	Gyrotron	UTFG-1	UTFG-H
First Order Temperature Sensitivity ($^{\circ}$ /hr/ $^{\circ}$ C)	?	4	300	20
First Order Acceleration Sensitivity ($^{\circ}$ /hr/g)	3	?	600	10

(g = acceleration of gravity)

Table 5.2 Approximate Temperature and Acceleration Sensitivities

vague possible solution, the temperature cross-coupling problem in high precision vibratory gyroscopes does not appear to have any solution other than temperature control. This solution is undesirable in some applications because it increases the power requirements of an instrument by orders of magnitude. The actual power required for temperature control depends on the temperature range and time characteristics of the environment in which the instrument must operate and the space available for insulation so it cannot be estimated.

Non-Uniform deflections under acceleration loading is a problem in vibratory drive, vibratory rate gyroscopes because it causes the balance of the drive member and the alignment of the drive mechanism to change in proportion to the acceleration. The magnitude of the time-varying cross coupling which is caused by these non-uniform deflections can be minimized by:

1. inspecting the materials of the drive member and the drive mechanism for defects,
2. machining all parts to close tolerances so that symmetrically located parts are as symmetric as possible, and
3. balancing the assembled instrument.

Unfortunately, the precision machining and balancing operations greatly increase the cost of an instrument and even after they are completed the acceleration coefficients of some configurations may be too large for some applications. This can be seen from the numbers listed in Table 5.2 for the acceleration coefficients

of the four tuning fork gyroscopes. The major cause of acceleration variations in the cross coupling in tuning fork gyroscopes is non-uniform deflections of the tines in the z_f direction (Fig. 5.1) which causes a torsional unbalance. This is one of the major disadvantages of the tuning fork configuration. In other configurations which are more symmetric, such as the cylinder shown in Fig. 4.6b, the acceleration sensitivity should be much lower than it is in tuning fork gyroscopes.

The acceleration cross-coupling problem in vibratory gyroscopes is not examined further because few data are available on configurations other than the tuning fork and in many applications the acceleration errors in vibratory gyroscopes should be small compared to the other errors. Also, in some applications, an acceleration signal is available so some of the acceleration errors can be compensated.

Regardless of the cause of the variations in the cross-coupled torques in tuning fork gyroscopes, the stochastic inphase torques cannot be distinguished from the Coriolis torques. Therefore zero-rate errors proportional to the stochastic inphase torques appear at the outputs of these instruments. Fortunately, the inphase stochastic cross-coupled torques in vibratory drive instruments are much smaller than the stochastic quadrature cross-coupled torques but the ratio is generally not as large as Eqs. 5.5 and 5.6 imply because it is much easier to reduce Ω_q by changing the mass distribution than it is to reduce Ω_d by changing the drive mechanism. Therefore in poorly balanced instruments $\Omega_q \gg \Omega_d$ (these are the RMS values of the cross-coupled components) while in high precision instruments, such as the A5, the two may be approximately equal. Also the standard deviation of the quadrature cross-coupled torques should be much larger than the standard deviation of the inphase cross-coupled torques in poorly balanced instruments while in well balanced instruments, such as the A5, the standard deviation of the quadrature cross coupling may only be an order

of magnitude larger than the standard deviation of the inphase cross coupling.

The spectrum of the zero-rate errors from the stochastic inphase cross coupling in a tuning fork gyroscope depends upon the spectrum of the acceleration applied to the instrument and the spectrum of the temperature of the tuning fork environment. When the accelerations are small and the temperature is approximately constant, the stochastic inphase cross coupling will depend primarily on the creep of the drive member and the spectrum will be extremely narrow with a bandwidth on the order of 10^{-3} cps.* In this case filtering is ineffective if the instrument is to remain a "rate" gyroscope with response to zero frequency. However, if the instrument response to very low frequency rates is not important, the slowly varying cross-coupled zero-rate errors that result from material creep and slowly varying temperature changes can be filtered electrically.

Ideally, the stochastic quadrature cross coupling and the steady components of both quadrature and inphase cross coupling are not troublesome in tuning fork gyroscopes because the quadrature components are discriminated against in the demodulator and the steady inphase components can be bucked out either before or after the demodulation. Unfortunately, in actual instruments, this is not true and all of these terms cause zero-rate errors. The most important ways in which this happens are examined in the following two sections.

5.2.4 Differential Frequency Drifts

The resonant frequency of any tuned system is a function of numerous factors and drifts in a stochastic manner with time. In tuning fork gyroscopes there are generally two resonant frequencies of interest--the tuning fork tine resonant frequency and the resonant frequency of the sensing system--and uncompensated differential frequency drifts between them cause significant

* This number was found in the Gyrotron.

zero-rate errors. Although some of this differential drift theoretically can be compensated if the temperature, linear acceleration, etc. of the instrument are known, this is usually not attempted in vibratory gyroscopes because it is difficult and usually both impractical and unnecessary. Therefore not only the truly random component of frequency drift (the drift of the resonant frequency when the temperature, linear accelerations, etc. are held constant) must be considered as **stochastic** when designing an instrument but also the uncompensated drifts due to temperature changes, linear accelerations etc. These stochastic differential frequency drifts are generally extremely small in well designed tuning fork gyroscopes (much smaller than the bandwidth of the resonant sensing system) but cause significant zero-rate errors because of the large unwanted quadrature cross coupling.

The zero rate errors which occur in tuning fork gyroscopes because of differential frequency drifts $-\Delta\omega$ -are primarily due to the differential phase shifts proportional to $\Delta\omega$ that occur in the sensing system. This differential phase angle is defined as

$$\theta(\Delta\omega) \triangleq \angle G_{s.c.}(j\omega_c) - \angle G_{s.c.}(j\omega_c + j\Delta\omega) \quad (5.7)$$

where $G_{s.c.}(s)$ is the transfer function of the sensing system and any electronics which precede the demodulator and ω_c is either equal to the natural frequency of the resonant sensing system or some slightly different frequency which is related to the natural frequency by a function of the time constant of the sensing system which is approximately equal to unity. If we use the same lumped parameter model for the sensing system that was used in Section 4.3, in which

$$G_{s.c.}(s) = \frac{B}{s^2 + \frac{\omega_c}{Q_s}s + \omega_c^2} \quad (5.8)$$

and we make $\omega_c = \omega_n$, then

$$\theta(\Delta\omega) = \frac{2Q_s \Delta\omega}{\omega_c} \quad (5.9)$$

if $\Delta\omega$ is small compared to $\omega_c Q_s^{-1}$. Now by using the small angle approximations for the sine and cosine of $\theta(\Delta\omega)$, the zero rate errors caused by a differential frequency drift can be written as

$$\Omega_{\Delta\omega} = \Omega_q \left(\frac{2Q_s \Delta\omega}{\omega_c} \right) - \frac{\Omega_d}{2} \left(\frac{2Q_s \Delta\omega}{\omega_c} \right)^2 \quad (5.10)$$

Since $\Omega_q > \Omega_d$ and $\left(\frac{2Q_s \Delta\omega}{\omega_c} \right) \ll 1$, the second term in Eq. 5.10 can be neglected.

It is now apparent from the first term, which can be rewritten as

$$\Omega_{\Delta\omega} = 2Q_s \left(\frac{\Delta\omega}{\omega_c} \right) \Omega_q, \quad (5.11)$$

that in order to minimize $\Omega_{\Delta\omega}$ it is necessary to

- (a) reduce both the steady and stochastic components of the quadrature cross-coupled torques applied to the sensing system. This can be accomplished by balancing the drive member and also bucking out the steady component of Ω_q by applying torques to the tuning fork. If bucking signals are applied afterward these errors will not be eliminated.
- (b) make the frequency stability $\left(\frac{\Delta\omega}{\omega} \right)$ of both the resonant frequencies as stable as possible. For this reason the resonant frequencies are generally determined by the mechanical properties of materials and any electrical feedback around a resonant system does not introduce any significant phase shifts at the resonant frequency.
- (c) make Q_s as small as possible. Unfortunately, this is inconsistent with our previous conclusion in Chapter IV that Q_s should be large to minimize the thermal noise in the sensing system and reduce the significance of noise which occurs in the signal processing electronics. Therefore, an optimum value of Q_s should exist for the sensing systems of vibratory rate gyroscopes.

In order to better understand the magnitude of the zero-rate errors caused by differential frequency drifts $\Delta\omega$, let us briefly examine typical numbers for the normalized frequency stability and the temperature and acceleration coefficients of tuning forks. Let us start with those of the UTFG-1 which are known but are excessively high because the tines of this tuning fork were clamped on. The numbers for this crude instrument are:

$$\begin{aligned} \left(\frac{\Delta\omega}{\omega}\right)^{T, A^*} &= 2 \times 10^{-4} \\ \left(\frac{\Delta\omega}{\omega}\right)/^{\circ}\text{C} &= 2 \times 10^{-4}/^{\circ}\text{C} \\ \left(\frac{\Delta\omega}{\omega}\right)/g &= 2 \times 10^{-5}/g \end{aligned} \tag{5.12}$$

In tuning fork frequency standards⁹ in which the tuning forks are made of composite materials, **temperature coefficients of $10^{-6}/^{\circ}\text{C}$ and long term stabilities of 10^{-6} are common but this long term stability often requires days to attain. Therefore, since no data are available on these three coefficients for the Gyrotron and A5, let us estimate the following numbers for a well designed tuning fork gyroscope the size of the A5 and Gyrotron

$$\begin{aligned} \left(\frac{\Delta\omega}{\omega}\right)^{T, A} &= 10^{-5} \\ \left(\frac{\Delta\omega}{\omega}\right)/^{\circ}\text{C} &= 10^{-6}/^{\circ}\text{C} \\ \left(\frac{\Delta\omega}{\omega}\right)/g &= 10^{-5}/g \end{aligned} \tag{5.13}$$

The temperature coefficient is assumed to be quite low because it is possible to make both resonant systems out of the same materials and therefore obtain good differential temperature compensation.

* $()^{T, A}$ means the long term stability when the temperature and acceleration are constant.

** Two materials combined so that the Young's Modulus of the combination has a zero slope with temperature at the operating point.

When these numbers are used for the Gyrotron and $\tau_g = 100^{-1}$ and $\Omega_q = 10^3$ /hr, then the long term zero-rate drift, temperature and acceleration sensitivities caused by differential frequency drifts are 1°/hr, 0.1°/hr/°C and 1°/hr/g respectively. These numbers are smaller than the experimental values of 10°/hr, 3°/hr/°C and ? for the Gyrotron but they certainly are not insignificant.

It is difficult to draw meaningful numerical conclusions from these few numbers but it does appear that differential frequency drifts in well designed tuning fork gyroscopes the size of the A5 and Gyrotron cause zero-rate errors which are not insignificant but most likely are smaller than the zero-rate errors caused by the stochastic inphase cross coupling and other factors. The most important of these "other factors" is examined in the following section. In small or high frequency instruments, however, in which Q_g is larger, this may not necessarily be true and differential frequency drifts may cause zero-rate errors which are important. In any case, an optimum value of Q_g generally will exist for tuning fork gyroscopes because the zero-rate errors due to thermal fluctuation noise, noise in the electromechanical transducer in the sensing system and noise in the signal processing electronics decrease as Q_g increases while zero-rate errors due to differential frequency drifts increase as Q_g increases.

5.2.5 Differential Phase Shift Drift

The phase shift at a fixed frequency of any electronic circuit or electromechanical transducer is a function of numerous factors and drifts in a stochastic manner with time. In tuning fork gyroscopes in which a large quadrature signal is present, a differential phase shift in either the signal or reference channels of the demodulator causes significant zero-rate errors. The equation for these errors was determined in Section 2.4 and is

$$\Omega_{\Delta\theta} = \Omega_q \Delta\theta - \frac{\Omega_q}{2} (\Delta\theta)^2 \quad (5.14)$$

where $\Delta\theta$ is by definition a differential phase shift. Since $\Omega_q = \Omega_d$ and $\Delta\theta \ll 1$, the second term in Eq. 5.14 can be neglected. Therefore the zero-rate errors caused by differential phase shifts in tuning fork gyroscopes are approximately equal to $\Omega_q \Delta\theta$.

Almost no data are available on the stability of differential phase shifts in tuning fork gyroscopes. This stability depends to a large extent on the quality of the components, the range of the environmental parameters and the sophistication of the circuitry while the zero-rate errors which result from differential phase shifts depend to some extent on the method used to null the instrument output. The output can be nulled either by adding an adjustable backing signal before or after the demodulator and then adjusting the phase of the demodulator reference to minimize the long term drift or else the backing signal can be omitted and the output nulled by adjusting the phase of the demodulator reference signal. The former has definite advantages in high precision instruments but the latter is much cheaper and simpler and works almost as well in instruments in which $\Omega_q \gg \Omega_d$. These fine points are not discussed further because in almost any scheme an adjustable phase shifter is required and it is the source of the major differential phase shift drifts.

In the Gyrotron bucking amplifier and the UTFG-1 demodulator reference it was necessary to shift the phase of the drive frequency signal approximately 90° . There are numerous ways of doing this but they all require reactive and resistive components. In order to estimate the magnitude of the phase shift drifts and temperature coefficients in these instruments, let us examine the extremely simple phase shifter shown in Fig. 5.2.

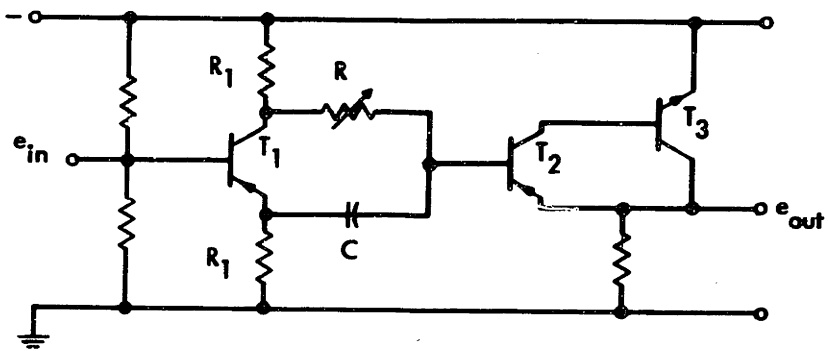


Fig. 5.2 Simple Phase Shifter

If $R \gg R_1$ and the input impedance of the $T_2 T_3$ combination is much larger than R_1 then the transfer function of the phase shifter is approximately,

$$\frac{e_2(s)}{e_1(s)} = \frac{1 - RCs}{1 + RCs} \quad (5.15)$$

Thus when $\omega_c RC = 1$, the phase shift is 90° and small changes can be achieved by changing R a small amount. The most important components in this circuit are R and C . If a change occurs in either of them, a differential phase shift results.*

$$\Delta\theta = \frac{1}{2} \left(\frac{\Delta C}{C} \right) = \frac{1}{2} \left(\frac{\Delta R}{R} \right) \quad (5.16)$$

Therefore if a long term stability of 0.2% and a temperature coefficient of 0.01%/°C are assumed** for R and C , the following two numbers are obtained for the long term drift and temperature coefficient of the phase shifter (assuming statistical independence).

$$\Delta\theta^T = 1.4 \times 10^{-3}$$

$$\Delta\theta / ^\circ C = 2.8 \times 10^{-4} / ^\circ C$$

It is interesting to note that this estimated temperature coefficient compares favorably with the observed temperature coefficient of the Gyrotron backing amplifier ($3 \times 10^{-4} / ^\circ C$) and the long term drift compares favorably with the ratio of the long term drift of the UTFG-1 in a temperature controlled environment to the RMS value of the quadrature cross coupling (3×10^{-3})

* A differential phase shift also occurs when ω_c changes an amount $\Delta\omega$. It is equal to $\frac{\Delta\omega}{2\omega}$ which is generally much smaller than $\theta(\Delta\omega)$ computed in the previous section so it can be neglected in most cases.

** These numbers are based on the following data on Sprague¹⁰ "highly stable" components of an appropriate size.

Styrachon Film Capacitors - Temperature Coeff. 120 ppm/°C,
drift 0.2 %

Molded Filmistor - Metal Film - Temperature Coeff
0.01 %/°C
drift 0.25 %

In practice, the long term stability of differential phase shifts in demodulators is generally higher than the number given in Eq. 5.17 and 10^{-2} is often quoted for the quadrature rejection capability. However, in vibratory gyroscopes the drive frequency reference signal should be quite clean so the number 10^{-2} often used is conservative for well designed demodulators. This was found to be true in the UTFG-1 and is documented in the following chapter. In that instrument the long term quadrature rejection was always better than 5×10^{-3} . Therefore, since no other data on this topic are known by the author, let us assume that an optimistic but not unreasonable number for the long term stability of differential phase shifts in well designed tuning fork gyroscopes is 3×10^{-3} or approximately twice the number given in Eq. 5.16.

Although it is difficult to draw meaningful numerical conclusions from these few numbers, it appears that differential phase shift drifts cause significant zero-rate errors in tuning fork gyroscopes. In well balanced instruments, such as the A5 and Gyrotron in which the standard deviation of the inphase cross coupling may not be more than two orders of magnitude smaller than the RMS value of the quadrature cross coupling, these errors may be small compared to the zero-rate errors caused by the stochastic inphase cross coupling. However, in smaller, higher frequency and poorer balanced instruments in which the RMS value of the quadrature cross coupling is more than two orders of magnitude larger than the standard deviation of the inphase cross coupling, the zero-rate errors caused by differential phase shifts will be a significant part of the total zero-rate error.

5.2.6 Tuning Fork Gyroscope Summary

The major factors which are presently limiting the performances of tuning fork gyroscopes are zero-rate errors caused by unwanted cross coupling

between the drive member and the sensing system and the noise in the sensing system electromechanical transducer and signal processing electronics. Both the inphase and quadrature components of cross coupling produce zero-rate errors. This is because the stochastic inphase cross-coupled torques cannot be distinguished from the Coriolis torques and the signals produced by the steady and stochastic quadrature cross-coupled torques cannot be perfectly discriminated against in the demodulation process because of differential phase shift drifts in the signal processing electronics and also differential frequency drifts between the tuning fork drive frequency and the resonant frequency of the sensing system. With the exception of the zero-rate errors caused by linear accelerations, the cross-coupled zero-rate errors in tuning fork gyroscopes generally vary very slowly with time. The higher frequency noise in these instruments is generally the noise from the sensing system electromechanical transducer and signal processing electronics. This noise appeared to be insignificant in the previous chapter but this is not the case because it is impossible to operate tuning forks at their maximum allowable strain levels as was assumed in Chapter IV. Instead it is necessary to operate them at very low strain levels and even use stable materials, such as quartz, which have much lower maximum allowable strain levels than steels in order to minimize the creep in the drive member and therefore the RMS value of the cross coupling.

The magnitude of the cross coupling in tuning fork gyroscopes can be reduced by using large, low frequency, accurately machined, well balanced tuning forks made out of stable materials and operated at low strain levels; but, it appears that it is also necessary to control the temperature in order to obtain high precision (low long term drift). This is necessary because the stochastic inphase cross coupling, which is the major cause of long term

drift in a well balanced instrument, and the differential phase shifts in the signal processing electronics, are both dependent upon temperature and temperature gradients. Therefore a high precision tuning fork gyroscope, such as the A5, will be both expensive because of the time consuming machining and balancing operations, and it will not necessarily be a low power device because of the temperature controller required. Also it will require long warm up times when the tuning fork is initially turned on. Therefore, in view of the precision, reliability, power requirements and cost of present day conventional gyroscopes, it is doubtful that there are many applications where a "high precision" tuning fork gyroscope would be better than a conventional gyroscope. One exception to this statement is an application in which extremely long life expectancies and continuous operation are required.

The above conclusion is not necessarily true for less expensive, tuning fork gyroscopes, however, because there are numerous applications for small, medium precision, vibratory rate gyroscopes which have short warm up times. Therefore let us estimate some of the characteristics of a low cost tuning fork gyroscope. On the basis of the conclusions presented in the previous parts of this section, it appears that in such an instrument the tuning fork should be made of a stable material such as quartz and the configuration should be one in which the drive frequency is low and also one in which the critical parts are easily machined to small tolerances and then balanced when the instrument is assembled. In this way the RMS cross coupling and the zero-rate errors associated with the cross coupling should be minimum for a limited amount of machining and balancing. The major cross-coupled zero-rate errors in such an instrument will be due to both the stochastic inphase cross coupling and the signals from the quadrature cross coupling which are not discriminated against in the demodulator. Because the

machining and balancing operations are not as precise as in expensive instruments, the zero-rate errors from the inphase and quadrature cross coupling will probably be approximately equal. Thus, if stable materials and good electronic components are used and the instrument cost is limited to a few hundred dollars, it is not unreasonable to expect that at constant temperature the standard deviation of the stochastic, inphase, cross coupling will be on the order of $3 \times 10^{-3} \Omega_q$ (the RMS value of the quadrature cross coupling) and the stability of the phase shift drifts will also be approximately 3×10^{-3} . Therefore an estimate of the long term drift of this hypothetical, low cost, tuning fork gyroscope is approximately

$$\overline{(\Omega^2)}^{1/2} = 4 \times 10^{-3} \Omega_q \quad (5.18)$$

Now if we assume that the balance of this instrument is characterized by a normalized unbalance $\frac{\Delta R}{R_0} = 10^{-5}$, which is approximately the number estimated for the UTFG-H in Table 5.1, the standard deviation of the long term drift can be rewritten as

$$\overline{(\Omega^2)}^{1/2} = 2 \times 10^{-8} \omega_d \text{ radians/sec} \quad (5.19)$$

by using Eq. 5.6. If the drive frequency is 1000 cps, which is reasonable for an instrument of this size, the long term drift in a temperature controlled environment will be approximately $30^\circ/\text{hr}$. On the basis of the characteristics of other tuning fork gyroscopes, it is not unreasonable to estimate that the temperature drift coefficient of the same instrument will be approximately $3^\circ/\text{hr}/^\circ\text{C}$ and the power required for this instrument without a temperature controller will be less than 10 milliwatts.

It must be kept in mind that these numbers are only the authors estimates of the characteristics of a low cost tuning fork gyroscope and are not data on an actual instrument. These numbers have been presented because they tie together the lengthy discussions and estimations that are contained in the

earlier sections and also they indicate that medium precision may be obtainable in low cost tuning fork gyroscopes. If this indication is accurate, it presently appears that in many applications these low-cost instruments have significant advantages over conventional rotary type gyroscopes of comparable precision. Since it was previously concluded that more expensive and higher precision tuning fork gyroscopes may not have many advantages over conventional rotary type gyroscopes of comparable precision it appears that future research on tuning fork gyroscopes should be directed toward the development of low cost, medium precision instruments.

5.2.7 Comments

The factors that are presently limiting the precision of tuning fork gyroscopes have now been thoroughly examined. This was done in far more detail than required because the conclusions of Section 5.2.6 are valid for vibratory drive instruments with other configurations and also because these same factors limit the precision of other types of vibratory rate gyroscopes. It should now be apparent that the most important of these factors is unwanted cross coupling because without it the errors caused by random phase shift drifts and frequency drifts would be insignificant. For this reason the cross coupling problem in other types of vibratory rate gyroscopes is briefly examined in the following section.

5.3 CROSS COUPLING IN VIBRATORY RATE GYROSCOPES

Unwanted cross coupling between the drive motion and the sensing system is one of the major sources of error in all types of vibratory rate gyroscopes. In vibratory drive instruments it is caused by mass unbalances in the drive member and asymmetries in the driving mechanism and misalignments or

nonlinearities of the electromechanical transducer in the sensing system. In rotary drive instruments it is caused by mass unbalances in the rotating parts and nonlinearities or asymmetries in the bearings or bearing supports. In both types of instruments the cross coupling magnitude and phase are a function of temperature, linear accelerations and numerous other factors so that the cross coupling has both steady and a stochastic components.

In vibratory drive, drive frequency output instruments the cross coupling caused by mass unbalances in the drive member and asymmetries in the drive mechanism is the same as in the tuning fork. Since this cross coupling was thoroughly examined in Section 5.2 it is not reviewed here. The other type of cross coupling in these instruments is caused by a misalignment or an asymmetry of the sensing system transducer such that it senses some of the drive motion. A simple example of this type of cross coupling occurs when a pressure sensor in the acoustic gyroscope shown in Fig. 3.5 is not located exactly at the pressure mode of the drive wave. Regardless of which of these two types of cross coupling occurs or predominates in a vibratory drive, drive frequency output instrument, the cross coupling that is in quadrature with the Coriolis forces (or the signals resulting from them) is generally much larger than the inphase cross coupling. Therefore it is possible to discriminate against much of the cross coupling by phase sensitive demodulation in both one and two axes instruments. Unfortunately, it is not possible to discriminate perfectly and the inphase component is not negligible so cross coupling produces large zero-rate errors in these instruments.

In vibratory drive, double drive frequency output instruments the major cross coupling still occurs at the fundamental of the drive frequency; but, smaller, second harmonic cross coupling also occurs for the reasons previously listed. The relationship between the magnitude of the fundamental and second harmonic cross coupling can be seen by examining the tuning fork model shown in Fig. 5.1.

When the trajectories of the point masses in this model are changed so that they are linear and planar but no longer parallel, a second harmonic cross coupling torque occurs because the torsional unbalance ΔR varies sinusoidally with time. If this torsional unbalance is written as

$$\Delta R = \Delta R_0 + \left(\frac{\Delta R'}{R}\right) r_0 \sin \omega_d t \quad (5.20)$$

where r_0 is the amplitude of the drive oscillation and $\left(\frac{\Delta R'}{R}\right)$ is the angle between the two trajectories, it can be shown that the second harmonic, cross-coupled, equivalent rates are a factor $\left(\frac{\Delta R'}{\Delta R}\right)$ smaller than the fundamental cross-coupled, equivalent rates given by Eqs. 5.5 and 5.6. Therefore the cross coupling in this type of instrument which is caused by mass unbalances and asymmetries in the drive mechanism will probably be as troublesome as it is in drive frequency output, vibratory drive instruments. Furthermore this same conclusion is true for the other types of cross coupling that occur in vibratory drive instruments. Therefore, since the signal strength in these instruments is much lower than it is in drive frequency output instruments of the same size and the noise from the sensing system transducer and signal processing electronics are approximately the same in the two types of vibratory drive instruments, it appears that double drive frequency output instruments will generally be poorer than drive frequency output instruments of the same size.

In rotary drive instruments, cross coupling is caused by static and dynamic unbalances in the rotating parts, asymmetries in the spin axis bearings, asymmetries or nonlinearities in the bearing supports and also angular rates about the other input axis. In a rotary drive instrument an unbalanced rotor causes the rotor and the rest of the instrument to translate and rotate in a periodic manner along and about the two instrument axes that

are normal to the spin axis. When the bearings and bearing supports in such an instrument are ideal, these motions occur at the fundamental of the drive frequency when the rotor is symmetric (drive frequency output instruments) and at the fundamental and odd harmonics of the drive frequency when the rotor is asymmetric (double drive frequency output instruments). Motion at these frequencies is not troublesome because in drive frequency output instruments it appears as an input angular rate at the drive frequency while in double drive frequency output instruments it appears as an output oscillation at these frequencies. Unfortunately, in practice bearings and bearing supports are not ideal and the motions of the rotor that result from unbalances do not occur at only the fundamental and odd harmonics of the drive frequency. Instead the motion has even harmonic terms that are small but nevertheless troublesome because:

1. When the rotor of a rotary drive, drive frequency output instrument rotates about an instrument axis normal to the axis of rotation at twice the rotation frequency, the instrument cannot distinguish between this motion and a constant input angular rate about the same axis. The relationship between the angle of this oscillation and the zero-rate error in this type of instrument is derived in Appendix C.
2. When the rotor and gimbal of a rotary drive, double drive frequency output instrument rotates about an instrument axis normal to the axis of rotation at twice the rotation frequency, the instrument cannot distinguish between this motion and the output oscillation that results from a constant input angular rate.

In both types of instruments the stochastic cross coupling of this type cannot be discriminated against in the demodulator because the phase angle of the stochastic cross coupling is uniformly distributed between θ and 2π radians.

The other type of cross coupling that occurs in rotary drive instruments is caused by input angular rates about the second input axis because these instruments are always two axis sensors. This cross coupling is in

quadrature with the angular rate signals about the input axis of interest but it can cause errors at the output because of differential phase shift drifts in the signal processing electronics or differential frequency drifts between the drive frequency and the resonant frequency of the sensing system. However, in many applications this ~~inter-axial~~ cross coupling is not nearly as troublesome as the cross coupling that causes zero-rate errors when the instrument is inertially stationary. Therefore let us examine the former type of cross coupling in more detail.

In both drive frequency output and double drive frequency output, rotary drive instruments the most troublesome cross coupling is that which causes the rotor to twist about an instrument axis normal to the axis of rotation at twice the rotation frequency. The major source of this cross coupling is bearing asymmetries that result from machining limitations, material creep and wear, dirt, temperature changes and linear accelerations. Therefore the problem of minimizing the cross coupling and the zero rate errors associated with it in rotary drive instruments is similar to that in vibratory drive instruments. However, there are important differences between the two. The most important of these are:

1. The critical machining problem in rotary drive instruments is the bearing surfaces and this is almost trivial compared to the machining problem in vibratory drive instruments in which the entire drive member must be made to very close tolerances.
2. The major cross coupling in vibratory drive instruments is in quadrature with the angular rate signals so it is possible to discriminate against much of it in the signal processing. In rotary drive instruments the quadrature and inphase cross coupling of the type being considered are statistically the same so no discrimination is possible.
3. The statistical properties of the cross coupling in rotary and vibratory drive instruments probably are greatly different. In vibratory drive instruments the cross coupling varies very slowly. The author is not familiar with any data on cross coupling in rotary drive instruments but it probably varies much faster with time so that its spectrum is much wider in the frequency range of interest.

The cross coupling problem in rotary drive instruments has now been examined but no examples have been given because the author is not aware of any published data on this subject. Therefore let us make some crude estimates of the magnitude of cross coupling that might exist in a well designed drive frequency output instrument similar to the one shown in Fig. 4.6a. Let us assume that the instrument is approximately 2 1/2 inches in size and the frequency of rotation is approximately 10^3 cps. Also let us assume that the bearings are round to within 10^{-6} inches which is well within the present state of the art. Then if the total displacement of the rotor at the second harmonic of the drive frequency is 0.1 % of 10^{-6} inches, a statistical estimate of the angle of the second harmonic torsional oscillation of the rotor is

$$\phi_0 = 10^{-9} \text{ radians.} \quad (5.21)$$

Since a torsional oscillation of the rotor at the second harmonic of the drive frequency causes a zero-rate error of $\frac{\phi_0 \omega_d}{2}$ radians*/sec., it follows that the long term drift of this hypothetical instrument is approximately 0.5°/hr. This is certainly a respectable number but it must be kept in mind that it is only a crude estimate presented here because of the lack of actual experimental data. Nevertheless, this low drift number does indicate that rotary drive, vibratory rate gyroscopes may be more sensitive than vibratory drive instruments of comparable cost. Rotary drive instruments should also be less sensitive to temperature variations than vibratory drive instruments but they will require more operating power than vibratory drive instruments and also will have shorter life expectancies because of the spin axis bearings.

The conclusions of the previous paragraph are all based on the single drift estimate made for a drive frequency output, rotary drive instrument. This was done because it is difficult to estimate the cross-coupled zero-rate errors in rotary drive, double drive frequency output instruments and also because almost no experimental data are available in the open literature on this problem.

* See Appendix C.

At the present time, it appears that additional theoretical and experimental research on the cross coupling problem in rotary drive vibratory rate gyroscopes is necessary so that the sensitivities and the relative merits of the two types of rotary drive instruments can be determined. It also appears that such research may be very fruitful because a low cost, rotary drive, vibratory rate gyroscope with a long term drift rate on the order of the estimated $0.5 \text{ }^\circ/\text{hr}$. would be a useful instrument in many applications.

5.4 CONCLUSIONS

The factors that are presently limiting the performance of both rotary and vibratory drive, vibratory rate gyroscopes are the instabilities and inhomogenities of materials and the limited tolerances which mechanical parts can be made. These factors cause unwanted cross coupling to take place between the driving motion of the mass elements and the output sensing system. The most troublesome cross coupling is that which occurs at the carrier frequency of the suppressed carrier angular rate information. This cross coupling can be distinguished from an input angular rate only to the extent that it remains constant and can be bucked or calibrated out or else can be discriminated against in the phase sensitive demodulation that occurs in the signal processing. It is also possible to electrically filter some of the signals resulting from unwanted cross coupling but generally the major cross-coupled zero-rate errors that occur at the output of vibratory rate gyroscopes vary very slowly with time so that filtering is ineffective if the instruments must remain true "rate" gyroscopes.

It appears that there are only two basic ways to improve the performances of vibratory rate gyroscopes. They are:

1. reduce the magnitude of the unwanted cross coupling (assuming the magnitude is normalized in terms of equivalent input angular rates), and
2. reduce the errors that occur at the output of vibratory rate gyroscopes because of the cross coupling.

The magnitude of the cross coupling in vibratory gyroscopes can be reduced using better materials, more accurate machining and balancing techniques and better configurations. These problems have been briefly examined in this chapter. Although a more detailed analysis of these very important problems is necessary, the present research is primarily concerned with the second method for improving the performances of vibratory rate gyroscopes.

The zero-rate errors that occur at the outputs of vibratory rate gyroscopes because of unwanted cross coupling between the drive motion and the sensing system can be reduced in at least two, greatly different ways. The first approach to be examined is called "double modulation" and it involves a radical change in the design of vibratory rate gyroscopes. The second approach to be examined is called "optimum processing of suppressed carrier signals" which is self explanatory. These two approaches are previewed in the following paragraphs and then thoroughly examined in the following three chapters.

Double modulation is a process by which the Coriolis forces in a vibratory rate gyroscope are modulated twice rather than once as in the more conventional types of vibratory rate gyroscopes. The additional modulation can be accomplished by either rotating or vibrating the basic drive member of a more conventional vibratory rate gyroscope about a rate insensitive axis. The purpose of the additional modulation is to make the carrier frequencies of the suppressed carrier angular rate information different than the frequencies of the cross coupling so that it is possible to discriminate against the cross coupling in the signal processing. The results of theoretical and experimental research on double modulation are reported in Chapter VI.

In conventional vibratory gyroscopes and vibratory gyroscopes that employ double modulation, the angular rate information is suppressed carrier modulated so that it is necessary to demodulate this information in order to obtain an unmodulated output signal. Furthermore the suppressed carrier

angular rate signals are corrupted by additive cross-coupled signals, and noise from the thermal fluctuations in the sensing system, the sensing system electromechanical transducer and the signal processing electronics so it is necessary to filter and demodulate the signals from the sensing system transducer. Since there are numerous ways of accomplishing this signal processing, it is desirable to know which way is best or optimum and how good it is. This information is useful because it allows one to determine if an instrument can satisfy the specifications of a particular problem and also it shows how the unspecified components in the instrument should be designed to achieve this optimum performance.

At first glance this might appear to be a simple filter problem but a closer examination shows that it is not because of the suppressed carrier modulation and fact that the carrier frequency, the resonant frequency of the sensing system, and the phase of the demodulator reference are not constant but drift in a stochastic manner with time. Since these "practical" problems cause significant errors in vibratory gyroscopes they must be taken into account when estimating the optimum performance of an instrument and designing the sensing system and signal processing electronics. For these reasons a filter theory for suppressed carrier systems is developed in Chapter VII that can handle this and other suppressed carrier problems. This filter theory is then applied to some aspects of the vibratory gyroscope signal processing problem in Chapter VIII.

CHAPTER VI

REDUCTION OF ERRORS IN VIBRATORY GYROSCOPES BY DOUBLE MODULATION

6.1 INTRODUCTION

Double modulation refers to the introduction of a second modulation of the Coriolis forces in a vibratory rate gyroscope through modulation of the angular rates to be sensed. The first modulation is associated with the drive motion imparted by the basic drive member and the second modulation is associated with a vibration or a rotation of the basic drive member about a rate insensitive axis. The purpose of second modulation is to separate the carrier frequency of the suppressed carrier Coriolis forces containing the input angular rate information from the fundamental and harmonics of the basic drive frequency and the double modulation frequency. When these frequencies are separated through double modulation, the "zero rate" errors associated with unwanted cross-coupling between the drive member and the sensing system may be reduced; however, other new sources of error occur because of the double modulation motion. Both of these aspects of double modulation are examined in this chapter and experimental evidence in support of the double modulation concept is presented.

Double modulation in vibratory gyroscopes was originally proposed by Newton¹ as a means of reducing the errors associated with unwanted cross coupling in a vibrating ring device. By rotating the ring about its rate insensitive axis (the ring axis of rotation) it was hoped that the long term drift rate of the ring device could be greatly improved. Although preliminary designs of a doubly modulated vibrating ring gyroscope were completed by Newton, experimental verification of the double modulation concept was not initiated until 1961 when a National Aeronautics and Space Administration Research Grant was given the MIT Electronic Systems Laboratory for research on doubly modulated and other types of vibratory gyroscopes. At that time it was decided that an experimental double-modulated tuning fork device would be built initially to obtain experimental data on a double-modulated device and also some practical experience with vibratory gyroscopes. After this preliminary research was completed, additional research on double modulation was to be carried out if it was necessary and justified and also other types of vibratory gyroscopes were to be investigated. Unfortunately, the

basic research on double modulation proceeded at a much slower rate than originally anticipated and it became necessary for the author to complete much of the basic experimental research as a part of this thesis rather than some of the originally proposed thesis research. Even more unfortunate is the fact that this research is not yet entirely completed to the author's satisfaction although enough has been completed to draw some meaningful conclusions on double modulation and to evaluate some of the theoretical work on suppressed carrier signal processing that is reported in the following two chapters.

The data that have been obtained by the author on the experimental double-modulated tuning fork gyroscope as a part of this thesis are reported in this chapter. Future data and a summary of all data obtained on the instrument will be reported in a final report to the National Aeronautics and Space Administration on Research Grant NsG-149-61 sometime in 1965.

6.2 DOUBLE-MODULATED TUNING FORK GYROSCOPE

The way in which double modulation can reduce the errors associated with unwanted cross coupling in vibratory gyroscopes is by making the frequencies of the suppressed angular rate information differ from the frequency of one of the major components of cross coupling. This concept is most easily understood by examining a simple vibratory drive configuration. * Figure 6.1 shows how a single axis double tuning fork instrument can be double modulated by rotating it about its x_f axis at a constant angular rate ω_m . From Fig. 6.1 it is seen that the angular rate around the sensitive axis of the tuning fork is

$$\Omega_{y_f} = \Omega_y \cos \omega_m t + \Omega_z \sin \omega_m t \quad (6.1)$$

where Ω_y is the angular rate around the y axis of the doubly modulated instrument and Ω_z is the corresponding rate around the z axis.

In order to understand how double modulation affects the total torque around the sensing axis it is necessary to rewrite Eq. 5.4 as

* This description of a rotary doubly modulated tuning fork gyroscope is taken from a paper² by Bush and Newton which is scheduled to appear in the IEEE Transactions on Automatic Control in October, 1964.

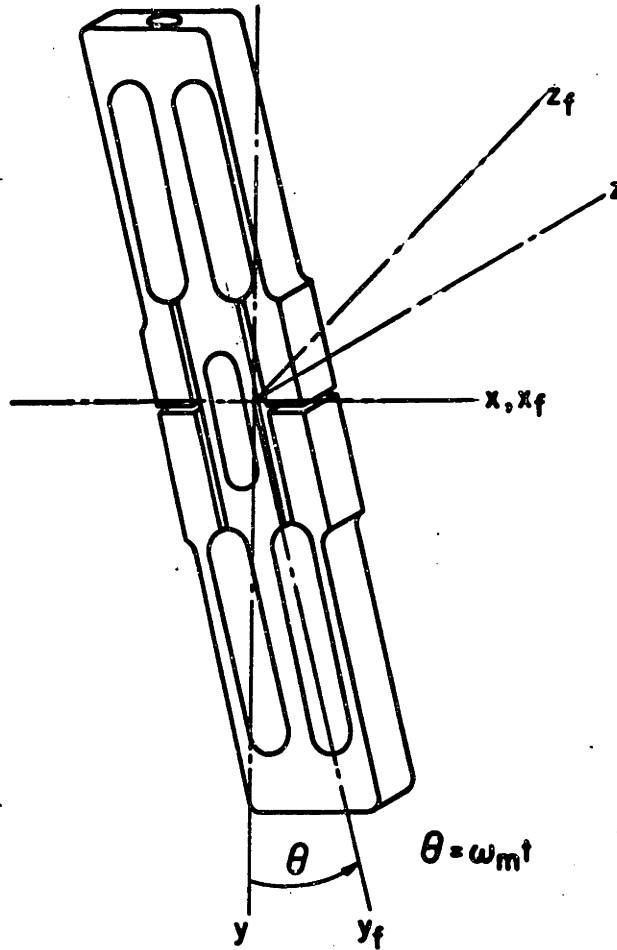


Fig. 6.1 Double Modulation of Tuning Fork

FIG. 3
PUSH TO

$$M_s = a_y \omega_d I_o \left[\Omega_{y_f} \cos \omega_d t + \Omega_n (\sin \omega_d t + \theta_n) + \Omega_a (\sin \omega_d t + \theta_a) \right] \quad (6.2)$$

In this equation the quadrature and direct components of the cross-coupled torque have been combined and then separated into a non-acceleration sensitive component represented by the equivalent rate Ω_n at a phase angle θ_n and an acceleration sensitive component Ω_a at a phase angle θ_a where both Ω_a and θ_a are functions of the acceleration vector. The non-acceleration dependent term is a slowly varying stochastic function of time and temperature (both the magnitude and the phase angle vary but θ_n is usually a very small angle) which was examined in Section 5.22. The acceleration dependent term arises primarily from unsymmetrical deflection of the tines under an acceleration load which was examined in Section 5.2.3. The major contribution to the acceleration component will be caused by unequal deflection of the tines in the z_f direction caused by an acceleration of the instrument in the z_f direction. Thus we can write approximately

$$\Omega_a \cong k_a a_{z_f} = k_a [-a_y \sin \omega_m t + a_z \cos \omega_m t] \quad (6.3)$$

where k_a is the first order acceleration sensitivity. Now by combining Eqs. 6.1 through 6.3 the following expression can be obtained for the total torque about the sensing axis of the tuning fork in Fig. 6.1

$$M_s = M_{s+} + M_{sd} + M_{s-} \quad (6.4)$$

The three components, at frequencies $\omega_+ = \omega_d + \omega_m$ (upper sideband), ω_d (drive frequency) and $\omega_- = \omega_c - \omega_m$ (lower sideband), respectively, are:

$$M_{s+} \triangleq \frac{a_y \omega_c I_o}{2} \left[(\Omega_y + k_a a_y') \cos \omega_+ t + (\Omega_z + k_a a_z') \sin \omega_+ t \right] \quad (6.5)$$

$$M_{sd} \triangleq -a_y \omega_c I_o \Omega_n \cos(\omega_d t + \theta_n) \quad (6.6)$$

$$M_{s-} \triangleq -\frac{a_y \omega_c I_o}{2} \left[(\Omega_y + k_a a_y') \cos \omega_- t - (\Omega_z + k_a a_z') \sin \omega_- t \right] \quad (6.7)$$

where

$$a_y' = a_y e^{j\Theta_a} \quad (6.8)$$

$$a_z' = a_z e^{j\Theta_a} \quad (6.9)$$

It can be seen from Eqs. 6.5 through 6.7 that the effect of rotary double modulation of the single axis tuning fork instrument about one of its rate insensitive axes is to produce torque components about the sensing axis at three frequencies instead of the single frequency that is present in the absence of double modulation. The upper and lower sideband torque components contain the desired rate information for a rotation of the instrument about an axis in the plane of the double modulation motion. The component at the tuning fork drive frequency is a non-doubly modulated torque which is proportional to the non-acceleration dependent cross-coupling- Ω_n . By suitable signal processing it is possible to have the overall instrument respond to either or both the sideband torques and exclude the other component(s). In this way double modulation, at least theoretically, is able to discriminate between the desired rate signal and the undesired non-acceleration dependent cross coupling in such a way as to exclude the latter. Since both of the sideband torque components contain acceleration dependent cross-coupling terms, however, double modulation does not discriminate between the rate signal and the acceleration dependent cross coupling. This occurs because both the acceleration vector in the yz plane and the angular rates seen by the single axis tuning fork device are modulated by the double modulation motion.

The above description of the double modulation method of discriminating against non-acceleration dependent cross coupling is a simplified discussion for the rotary method of double modulation that is valid for constant angular rates, small amplitudes of the tune vibration and for a negligible amplitude of torsional oscillation of the fork about the sensing axis. A more sophisticated analysis of the tuning fork device with rotary double modulation is given in Appendix A.

6.3 SOME ASPECTS OF DOUBLE MODULATION

Double modulation is the introduction of an additional modulation of the Coriolis forces in a vibratory rate gyroscope for the purpose of

discriminating against one major component of unwanted cross coupling between the basic drive member and the sensing system. This component is the non-acceleration dependent cross coupling which is one major component of cross coupling that in the past has severely limited the precision of vibratory drive, vibratory rate gyroscopes. Since the double modulation motion also modulates the acceleration vector seen by the basic drive member, acceleration dependent cross-coupling will not be discriminated against by double modulation and therefore must be handled by other means.

Double modulation can be accomplished by either rotating or vibrating the basic drive member of a vibratory rate gyroscope about a rate insensitive axis. When rotation is used the Coriolis forces on the driven mass elements are modulated the addition time at the double modulation frequency while if vibratory double modulation is used the modulation occurs at harmonics of the double modulation frequency. With both types of double modulation a single axis instrument is turned into a two axis instrument where both input axes are in the plane of the double modulation. The rates about the two orthogonal axes in this plane can be separated in the signal processing (see Section 3 5.4).

Although in principle, this concept of double modulation is applicable to both rotary and vibratory drive, vibratory rate gyroscopes, the discussion in this chapter is limited to vibratory drive devices. The reasons for this restriction are that cross-coupling problems are more severe in vibratory drive devices than in rotary drive devices and all of the theoretical and experimental work that has been done on doubly modulated vibratory gyroscopes has been on vibratory drive devices.

6.4 SOURCES OF ERROR IN VRG'S EMPLOYING DOUBLE MODULATION

Most of the sources of error and performance limitations in vibratory double-modulated rate gyroscopes are the same as those examined in Chapters 4 and 5 for the more conventional types of VRG's. These and the sources of error introduced by the double modulation motion are briefly examined in this section. The following three sections of this chapter then review the results of the experimental research on vibratory double-modulated rate gyroscopes and compare the experimental results with the theory.

The ultimate performance limitations in VRG's employing double modulation are the same as in the more conventional VRG's; namely

1. Thermal noise in the sensing system
2. Strain constraints in the drive member

3. Suppressed carrier signal processing

In doubly modulated instruments these limitations may be of greater significance than in other types of VRG's because:

1. The errors due to unwanted cross coupling may be significantly reduced by the double modulation.
2. The double modulation frequency, which is the important frequency from signal processing considerations, will generally be much lower than the basic drive frequency because it is necessary to move much larger masses. Therefore signal processing limitations must be taken into account if the spectrum of the input angular rates is broad. This is one of the reasons for the theoretical work on suppressed carrier signal processing which is done in the following chapter.

The theoretical ultimate threshold of rotary doubly modulated VRG's will generally be approximately a factor of 2 larger than that of the basic drive member which was examined in Chapter 4 because generally only one sideband is used to produce the output. With vibratory double modulation the ultimate threshold may be considerably larger than that of the basic drive member because the thermal noise is unchanged but the Coriolis torques are attenuated by a factor of $J_n(\Theta_0)$ where J_n is a Bessel function of order n and Θ_0 is the angle of the double modulation oscillation.

In addition to these ultimate performance limitations, cross coupling will also be troublesome in doubly modulated instruments. The cross coupling that occurs at the same frequencies as the Coriolis forces containing the rate information causes zero-rate errors because:

1. The cross coupling is not constant but varies with time.
2. The frequencies ω_d and ω_m are not constant but vary with time.
3. The components in the sensing system and signal processing electronics are not ideal but vary with time.
4. The phase of the demodulator(s)* reference signal is not constant but varies in a stochastic manner with time, temperature, etc.

These practical performance limitations were examined in Chapter 5 where it was shown that they are responsible for the major "zero rate" errors in conventional vibratory rate gyroscopes.

In addition to the cross coupling that occurs at the same frequencies as the rate information, cross coupling also occurs at harmonics of the drive frequency and double modulation frequency in VRG's employing double modulation. The signals caused by this cross coupling are troublesome, not because they occur at the same frequencies as the rate information, but

* Two demodulators may be used in a double-modulated instrument to process the signals in one channel and the overall two axis instrument may use 3 demodulators.

because they are often quite large and cause the sensing system transducer or signal processing electronics to saturate. Therefore a high Q resonant sensing system is often desirable in doubly modulated instruments.

In order to understand more clearly the types of cross coupling that can occur in a doubly modulated instrument let us now examine the tuning fork configuration of Fig. 6.1 and separate the cross coupling into three basic types.

Type 1 Cross Coupling - Type 1 cross coupling is between the basic drive member and the sensing system. This cross coupling was examined in Section 5.22. The non-acceleration dependent component occurs at the drive frequency while the acceleration dependent component is doubly modulated and occurs at the same frequencies as the suppressed carrier rate information.

Type 2 Cross Coupling - Type 2 cross coupling is between the sensing system and the double modulation motion and is independent of the drive motion of the basic drive member. The output signals from the sensing system for this type of cross-coupling occur at harmonics of the double modulation frequency. By way of example, an acceleration dependent component of this type of cross-coupling occurs in the tuning fork shown in Fig. 6.1 if the center of gravity of the tuning fork is displaced along the x_f axis from the y_f sensing axis when the instrument is accelerated in the yz plane. In other configurations such as the ring, this component of cross coupling should not be as large as in the tuning fork configuration.

When the tuning fork in Fig. 6.1 is double-modulated by vibration rather than rotation about the x_f axis, a non-acceleration dependent component of this type of cross coupling occurs if the center of gravity of the tuning fork is displaced from the point of intersection of the sensing axis (y_f) and the double modulation axis (x_f) along both the y_f and z_f axes. This component of cross coupling is similar to the torsional cross coupling that was examined in Section 5.22. Another cause of non-acceleration dependent cross coupling of this type occurs when the axis of the vibratory double modulation motion of the basic drive member is not orthogonal to the sensing axis. This misalignment causes type 2 cross coupling at the double modulation frequency.

When the tuning fork in Fig. 6.1 is double modulated by rotation rather than vibration about the x_f axis, a non-acceleration dependent component of this type of cross coupling occurs if the rotating member in the instrument is unbalanced and the structure supporting the rotating member is not

uniform in all directions or the bearings on the double modulation axis are not perfect. When a nonuniformity exists, either because of imperfections in the bearings or actual mechanical design, the tuning fork in Fig. 6.1 is twisted about the y_f or z_f axis in a periodic manner. The twisting about the z_f axis will cause cross coupling if the center of gravity of the tuning fork is misaligned in the $y_f z_f$ plane along both axes whereas the twisting about the y_f axis will cause type 2 cross coupling without any misalignments. This cross coupling occurs at harmonics of the double modulation frequency if bearings without balls or rollers are used. If the bearings contain balls or rollers, this type of cross coupling occurs at harmonics of the double modulation frequency which are not integers and depend on the dimensions of the races and balls or rollers and the number of balls or rollers.

Type 3 Cross Coupling - Type 3 cross coupling is between the sensing system and the double modulation motion and it is dependent on the drive motion of the basic drive member. The output signals from the sensing system for this type of cross coupling occur at many frequencies, including those at which the suppressed carrier rate information occurs.

In the vibratory double-modulated tuning fork that was examined in the previous paragraphs on type 2 cross coupling, a misalignment of the double modulation axis with respect to the sensing axis causes a component of the double modulation motion to be about the sensing axis of the tuning fork. This twisting appears as an angular rate at the double modulation frequency which is indistinguishable from a double modulated angular rate about the z axis. However, it was shown in Section 3.11 that the y axis rates are double modulated at twice the double modulation frequency so these rates are distinguishable from this type of cross-coupling.

In the rotary double-modulated tuning fork that was examined in the previous paragraphs on type 2 cross coupling, a misalignment of the double modulation axis with respect to the sensing axis causes a component of the double modulation motion to be about the sensing or input axis of the tuning fork. This misalignment causes type 3 cross coupling at the tuning fork frequency. Since there is already a large amount of quadrature cross coupling (Ω_n) at this frequency, this additional cross coupling may not be too troublesome even though ω_m is large.

The type 3 cross coupling that is troublesome in rotary double-modulated devices is that which is caused by an unbalance of the rotating member and/or non-uniform bearings and rotor supports which cause the

tuning fork to twist about the z_f and x_f axes at harmonics of the double modulation frequency. The twisting which occurs about the y_f axis which is at the double modulation frequency cannot be distinguished from a double modulated angular rate in the instrument yz plane. When the twisting motion of the rotor which causes this type 3 cross coupling in a rotary double-modulated instrument is viewed from the instrument (xyz) axes, it appears to be a twisting of the rotating member about an axis normal to the double modulation axis at twice the double modulation frequency. This twisting can be minimized by balancing the rotating member, using good bearings which are symmetric and making the bearing supports symmetric so that rotor unbalance causes twisting at only odd harmonics of the double modulation frequency.

Of the three types of cross coupling that have been examined, the second two will probably be the most troublesome in well designed instruments. By properly choosing the drive and double modulation frequencies and the bearings on the double modulation axis, the long term drift in a rotary double-modulated instrument should be caused primarily by type 3 cross coupling and should be much lower than the long term drift of the basic single axis instrument being double-modulated. The short term drift in the same instrument will depend on the quality of the bearings and the balance of the basic drive member and the rotating member.

In vibratory doubly modulated instruments the long term drift should also be better than that of the basic single axis instrument being double-modulated. However, the factor $J_2(\theta_0)$ in the sensitivity expression which results from using the second harmonic term rather than the first to avoid the misalignment problem may be quite small. In order to increase this factor it is necessary to decrease the double modulation frequency ω_m . Therefore, signal processing errors may be significant in this type of instrument.

Now that the sources of error in double-modulated vibratory rate gyroscopes have been examined, it is desirable to know the magnitude of the various cross-coupled errors and whether double modulation actually does improve the performance of a vibratory rate gyroscope. In order to answer this question, a rotary double-modulated tuning fork gyroscope was built. This instrument and the experimental data obtained on it are the subject of the following two sections.

6.5 EXPERIMENTAL APPARATUS

A vibratory gyroscope that can be operated either with or without double modulation was constructed so that the merits of double modulation could be evaluated experimentally. The goals of the experimental program were to compare the performances of the instrument in the two modes of operation and investigate the causes of zero-rate errors with double modulation. The instrument is solely a laboratory model of convenient size to permit easy modification. It can be separated into three parts for descriptive purposes: the single-axis vibratory gyroscope (tuning-fork unit), the double modulation carriage, and the signal processing electronics.

The tuning fork unit shown in Fig. 6.2 is a single-axis vibratory gyroscope designed so that it can be rotated about an axis perpendicular to its rate-sensitive input axis. A double tuning fork is used as the drive member and a mechanical tuned system is used in the sensing mode. Tuning fork drive and vibration detector electronic circuits are required for its operation as a single-axis vibratory gyroscope.

A double tuning fork, similar to that shown in Fig. 6.1 was selected because its symmetry simplified support problems when it is rotated and because it facilitates comparison with the considerable results of past research on tuning-fork gyroscopes. Replaceable tines are used in the actual instrument to simplify machining and to reduce fatigue problems. Set screws are provided for rough balancing by changing the mass of the tines and moving the center of mass normal to the plane of vibration. The double modulation rotation axis is perpendicular to the fork input axis y_f and coincides with the x_f axis when the fork is inertially stationary. The adjacent tines of the two tuning forks are coupled together by small springs and are magnetically driven. The amplitude and frequency of the fork vibration are maintained constant by the tuning fork drive circuit that has as its input a signal from reluctance type pickups sensing the motion of the tuning fork tines and as its output the coil currents of the magnetic drivers.

When angular rates are applied to the unit about the y_f axis, torques proportional to the product of these rates and the sinusoidal velocity of the tuning-fork tines are impressed on the tuning fork about the y_f axis. The resulting twisting motion of the tuning fork with respect to the case is usually sensed to provide the electrical output. This motion can be maximized by making the torsional resonant frequency of the fork and suspension about



Fig. 6.2 Tuning-Fork Unit

the y_f axis equal to the carrier frequency of the torques. Resonant sensing lowers the instrument threshold; however, the resulting twisting of the fork through large angles is undesirable because it increases the interaxis cross-coupling of rates from the x_f axis and decreases the useful dynamic range of the instrument. In order to alleviate these difficulties, while still maintaining the beneficial characteristics of resonant sensing, a counterpoise sensing scheme is used in this instrument. The counterpoise is an inertia much smaller than the tuning-fork inertia which is attached to the tuning fork by means of a torsion bar along the y_f axis. When the system is properly designed, the counterpoise and tuning fork vibrate in phase opposition and the amplitude of the counterpoise vibration is much larger than that of the fork when torques are applied to the tuning fork at the upper resonant frequency of the combination. Higher Q's (quality factor) are achievable with this scheme than with the signal tuned system and the undesirable motion of the drive member is greatly reduced.

The counterpoise element, shown in Fig. 6.3 is attached to the torsion rod which is along the y_f axis of the tuning fork. It is in the form of a plate which together with insulated plates attached to the tuning fork center piece forms a differential capacitor. This capacitor is in a capacitance bridge which is excited at a low radio frequency from the vibration detector circuit. The bridge output is then amplified, rectified and filtered to provide an electrical signal proportional to the angle between the tuning fork and counterpoise.

The torsion rod on which the counterpoise is mounted is attached to the tuning fork by collets at the ends of the fork and the ends, together with four spokes attached to each end of the fork, provide support for the tuning fork with respect to the case. The fork is therefore rigidly supported in all modes except the torsional mode about the y_f axis.

The upper resonant frequency of the sensing system is tuned to the lower sideband of the Coriolis torques which are applied to the tuning fork when the unit is double modulated by rotation. These torques occur at the tuning fork drive frequency minus the double modulation frequency. The numerous cross-coupled torques applied to the tuning fork at other frequencies justify discarding the energy in the upper sideband in order to obtain better filtering and not saturate the vibration detector. When this instrument is operated without double modulation, however, it is off resonance. Since the cross-coupled torques in this mode of operation are at the same frequencies as the Coriolis torques and we are well above the threshold of the detector,

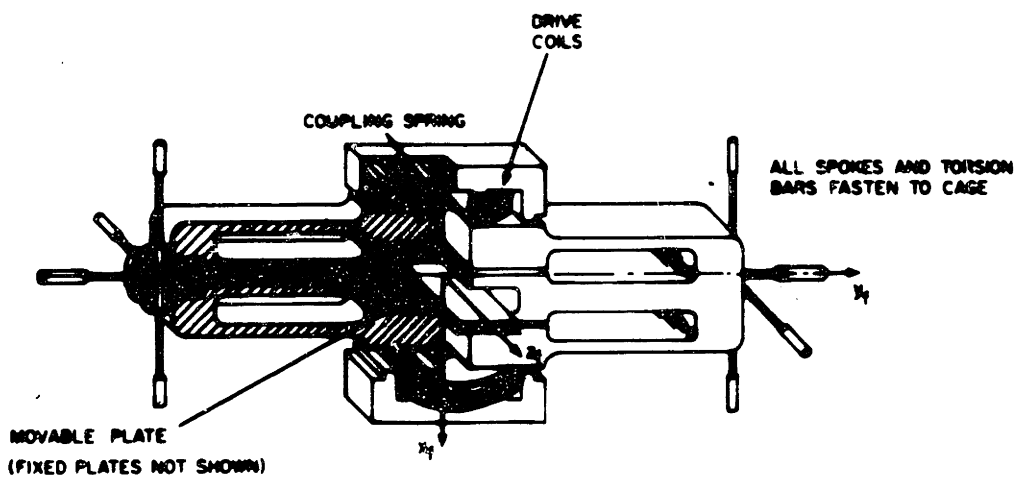


Fig. 6.3 Experimental Double Tuning Fork

this untuned condition is desirable because it reduces the zero rate errors due to shifts in the resonant frequencies of the tuning fork and sensing system. Therefore it is possible to use the same tuning fork unit for the comparison tests. This unit is shown in Fig. 6.2 without its outer steel shield and without the tuning fork drive and vibration detector electronics which are presently mounted on the unit inside the shield. The schematics for these circuits are in Appendix B.

Double Modulation Carriage

The double modulation carriage shown in Fig. 6.4 rotates the tuning-fork unit about an axis that can be oriented at any angle with respect to vertical. The carriage is turned by a printed circuit motor driven by silicon controlled rectifiers and is phase locked to a stable low-frequency oscillator. Tapered roller bearings were used in the initial tests and ball bearings are presently being investigated.

Accurate control of the carriage speed is necessitated by the high Q mechanical resonance in the sensing mode of the tuning-fork unit. It is accomplished by sampling the output of a stable, low-frequency oscillator twice each revolution of the carriage and using this error signal to control the firing angle of the SCR's driving the motor. The sampling occurs when two diametrically opposed slots in a disc rotating with the carriage pass a light source and allow the light to reach a photodiode. The schematics for the speed control system of the carriage are in Appendix B.

Signal Processing Electronics

Figure 6.5 shows a block diagram of all the electronic circuits used in the experiment. The signal processing electronics consist of two ring demodulators, one phase shifter and several linear amplifiers and filters. The vibration detector, tuning-fork drive and speed control have been discussed in the sections on the tuning-fork unit and the double modulation carriage.

The signal from the vibration detector, which is proportional to the angle between the counterpoise and tuning fork, is first demodulated using as the reference a phase-shifted signal from the tuning-fork drive. When the instrument is operated without double modulation the output of the demodulator is passed through a low-pass filter to obtain a d-c signal proportional to the applied rate. The transfer function for the instrument relating the input rate

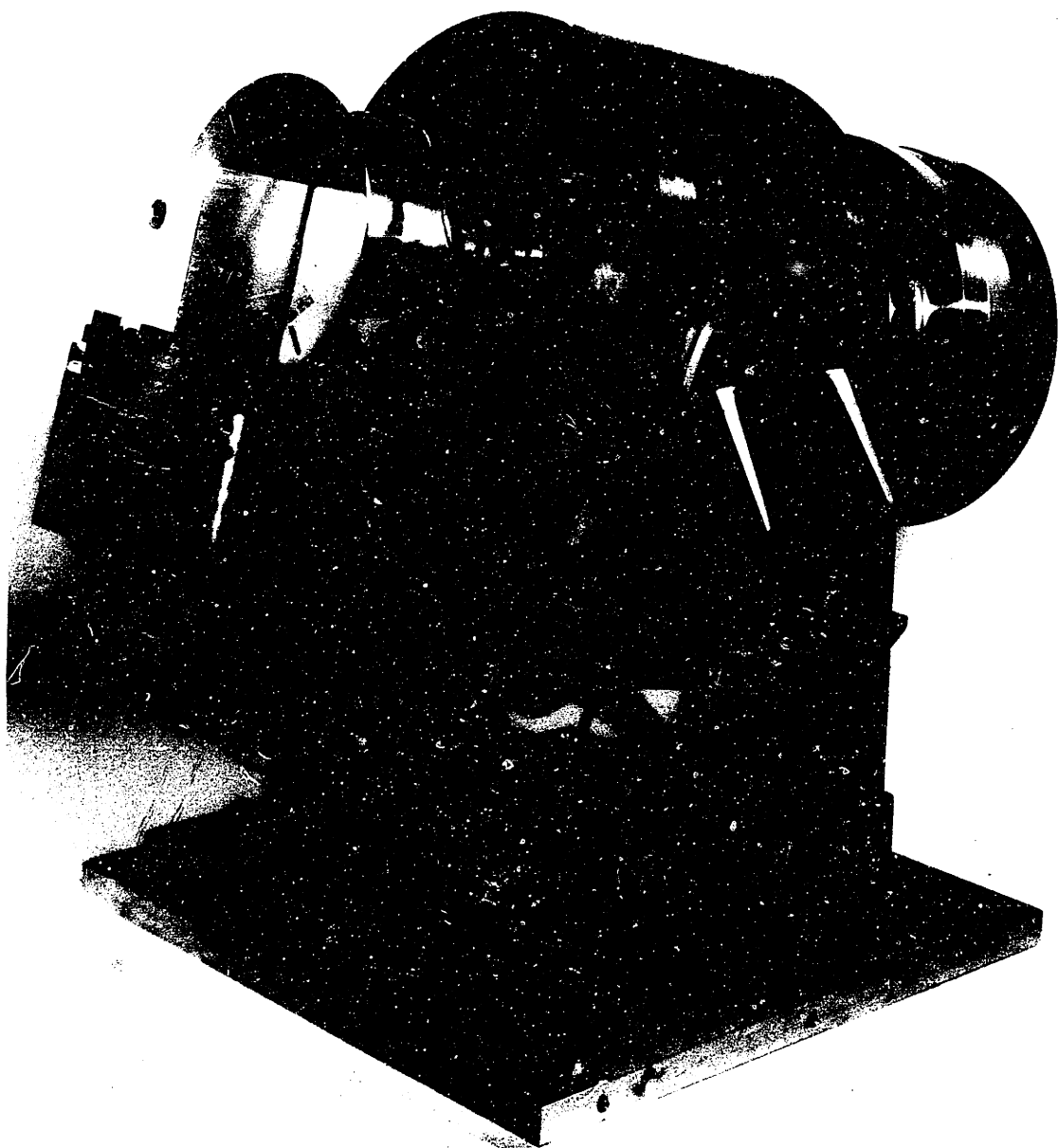


Fig. 6.4 Double Modulation Carriage

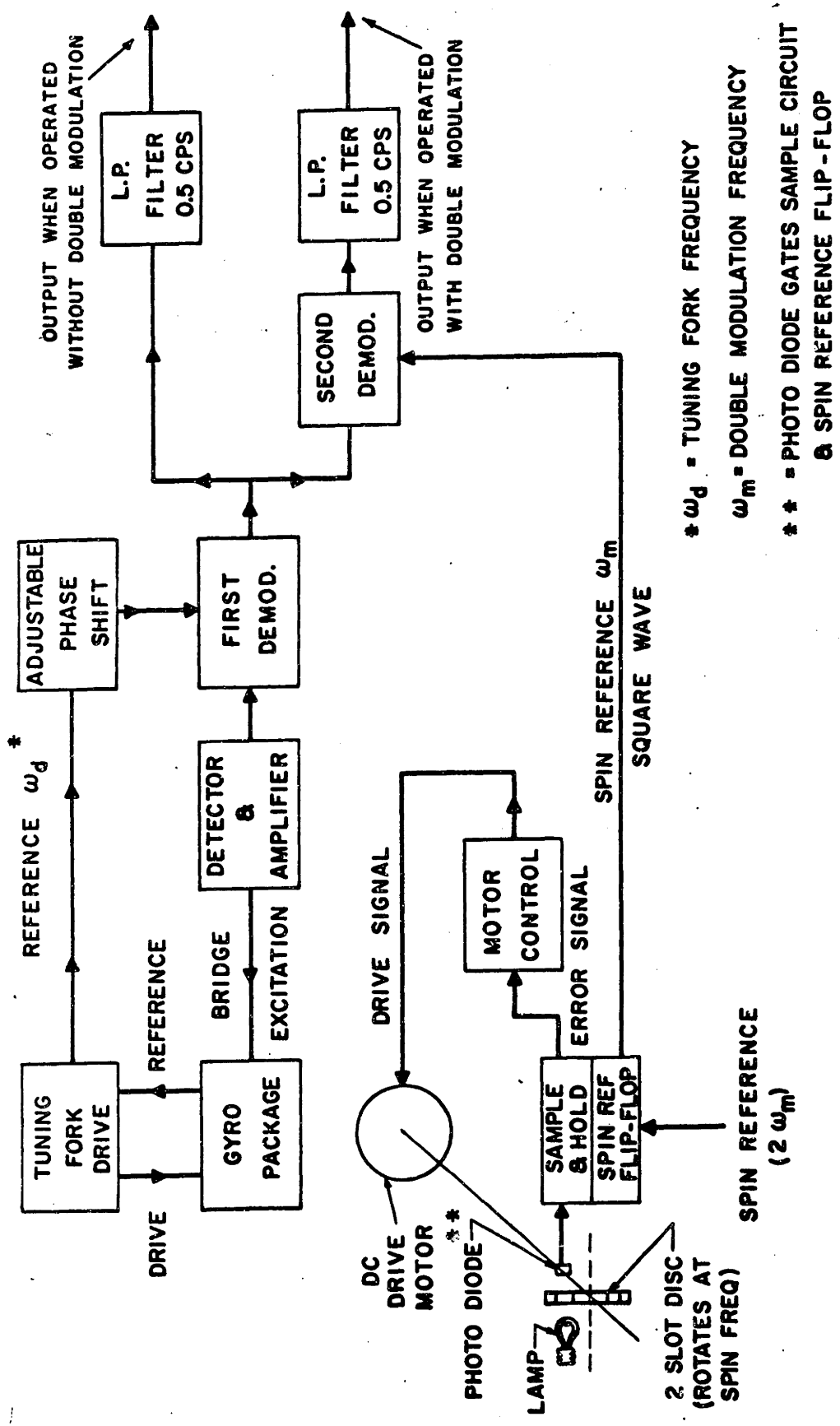


Fig. 6.5 Block Diagram of the Experimental Electronics

910.5

BUSH-NEWTON 215

65

and the output voltage has the frequency characteristics of the low-pass filter.

When the instrument is operated with double modulation the output of the first demodulator is transmitted through a band-pass filter centered at the double-modulation frequency. This signal is then demodulated using as the reference a square wave phase locked to the carriage. The output of the demodulator is passed through a low-pass filter to obtain a d-c signal proportional to the applied rate about either the y or z axes depending on the phase of the reference of the second demodulator. The transfer function relating the input rate and output voltage now depends on the quality factor Q_s of the mechanical sensing system, the band-pass filter following the first demodulator, and the low-pass filter following the second demodulator. The band-pass filter characteristics can be neglected since the filter is needed only to separate the signal from d-c and the second harmonic of the tuning fork frequency. For data transmission, the mechanical resonance contributes a first order pole with a time constant expressed by

$$\tau = \frac{2Q_s}{\omega_s} \tag{6.10}$$

where

Q_s = quality factor of the sensing system at frequency ω_s

ω_s = upper resonant frequency of the suspension

The schematics of the two demodulators presently used for signal processing are in Appendix B.

6.6 EXPERIMENTAL DATA

The experimental tuning fork gyroscope which is described in Section 6.4 has been successfully operated with and without double modulation. Although double modulation greatly reduced the long term drift of the tuning fork unit when it was operated in free air, the present data are not sufficient to allow final quantitative conclusions to be drawn. Research is continuing on the instrument so that a better understanding of the performance limitations with double modulation can be obtained and then final conclusions will be drawn. This future research will be reported in a final report on NASA Research Grant No. NsG-149-61 which will be written in 1965.

The data which presently are available on the instrument are sufficient to draw meaningful conclusions however. Figures 6.6 and 6.7 show typical output recordings from the instrument when it was operated in free air with some 5000 eru (earth rate units of approximately $15^{\circ}/\text{hr}$) of quadrature cross coupling. The instrument parameters for these tests -- dated July 1963 -- are listed in Table 6.1. The drift characteristics of this instrument for both modes of operation are listed in Table 6.2. These numbers were taken from the traces shown in Figs. 6.6 and 6.7.

The short term noise visible on the recordings shown in Figs. 6.6 and 6.7 is different in the two modes of operation because the signal processing bandwidths and the noise sources are different in the two modes of operation. The bandwidth of the instrument (sensing system plus signal processing electronics) was approximately 0.5 cps without double modulation and approximately 0.15 cps with double modulation because of the resonant sensing system employed with double modulation.

The standard deviation of the short term noise when the instrument was operated without double modulation was approximately 6 eru. This noise was due to random cross coupling in phase with the Coriolis forces and a combination of the large quadrature cross coupling and random phase shifts of the demodulator reference signal. Better tuning-fork balance would reduce the latter component of drift but tests have shown that the major source of short term drift or noise in the instrument used for these tests was random inphase cross coupling. This drift could be reduced (see Chapter 5 for estimates of the performance of a greatly improved version of this same instrument) by using a better tuning fork* but the present fork is thought to be adequate for the present double-modulation experiment.

The standard deviation of ~~the short term noise~~ when the instrument was operated with double modulation was approximately 6 eru. This noise was due to the spin axis bearings which were tapered roller bearings. It can be seen in Fig. 6.7 that much of the short term noise was a 0.8 cps component which is thought to come from the bearings. If the source of this noise was known and removed (such as by using a different type of bearing or similar bearings with different race dimensions or number of rollers), the standard

* This fork was quite crude. It was made out of 4340 steel and the replaceable tines were clamped in place. Furthermore, the balancing operation was also quite crude and the drivers were not mounted on the tuning fork.

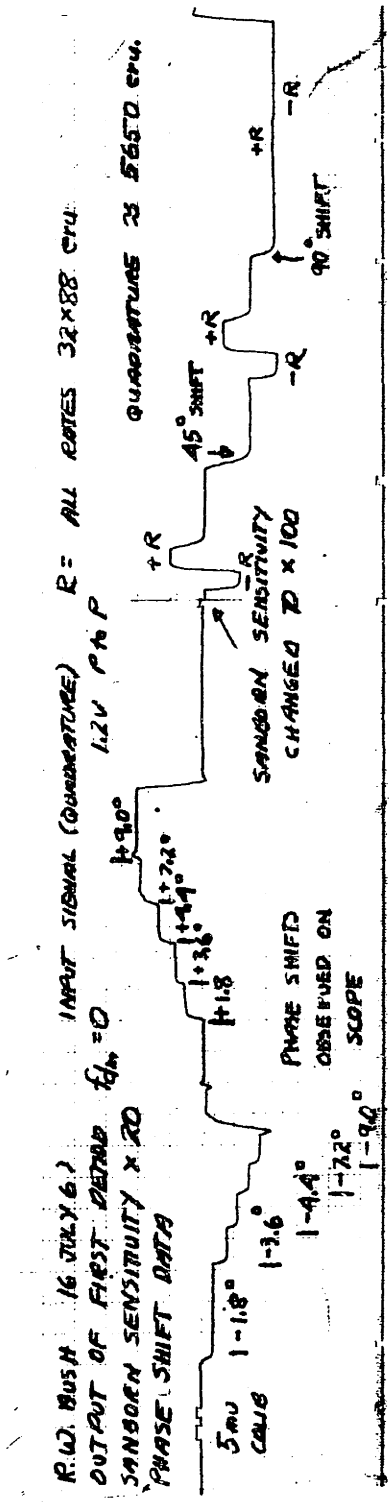
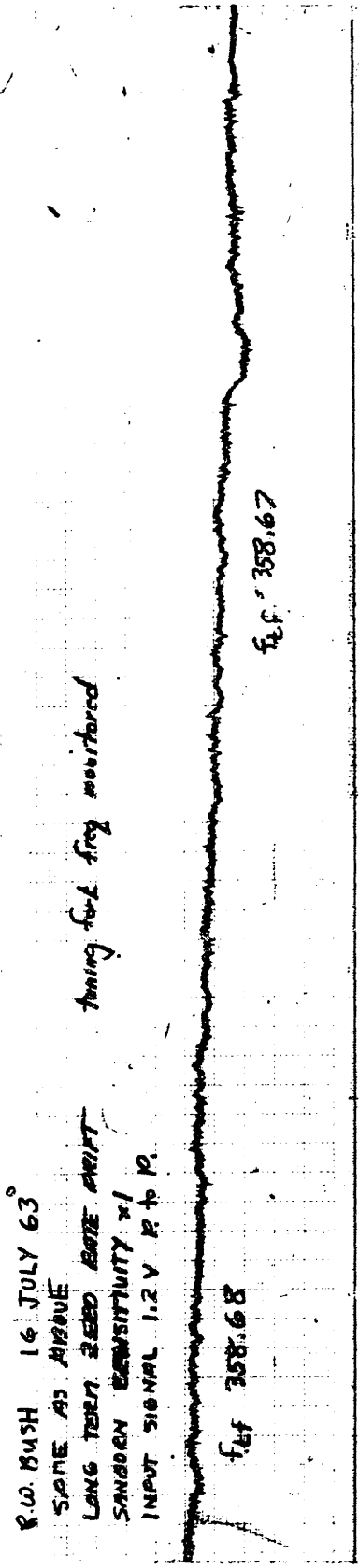
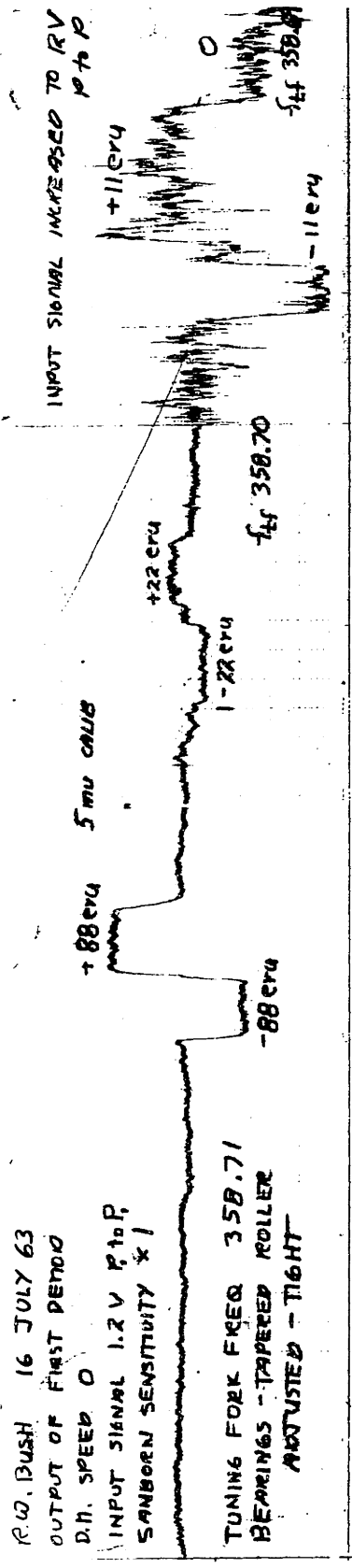


Fig. 6.6 Experimental Data Without Double Modulation

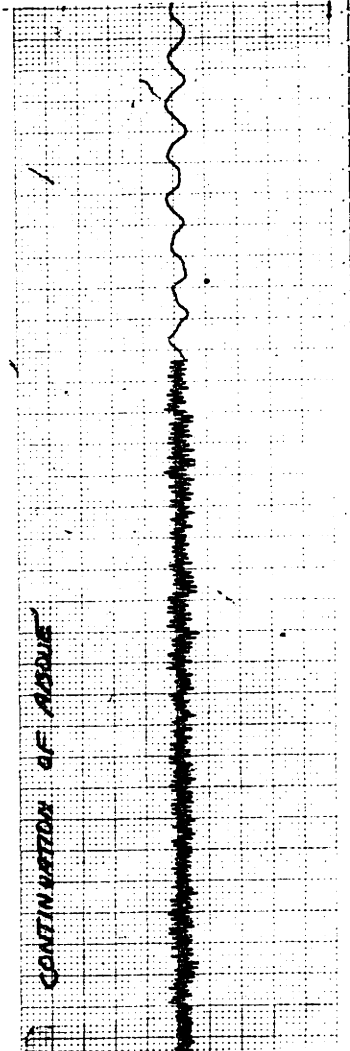
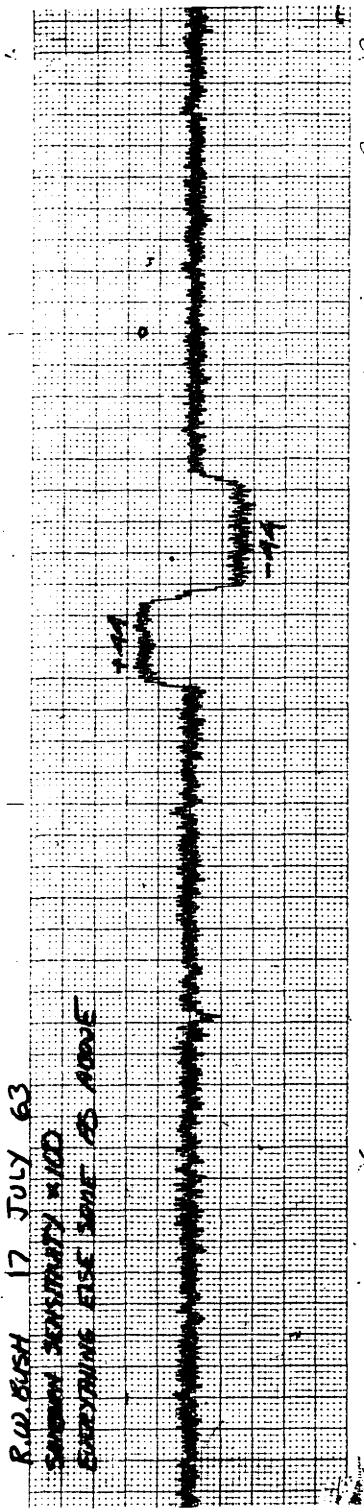
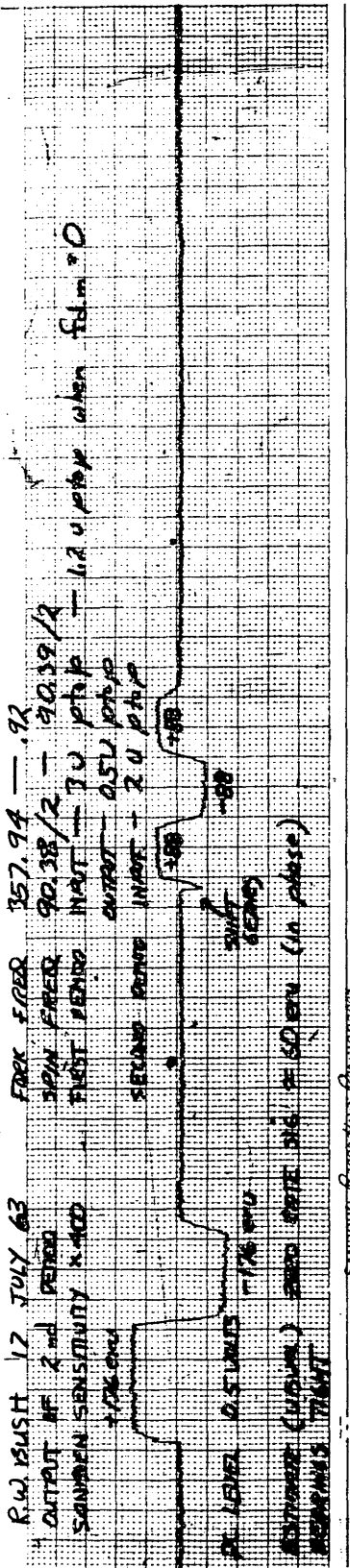


Table 6.1 Instrument Parameters

Parameter	Symbol	Value
Tuning fork drive frequency	$\frac{\omega_d}{2\pi}$	358.6 cps when $\omega_m = 0$
Upper resonant frequency of the sensing system	$\frac{\omega_s}{2\pi}$	312 cps when $\omega_m = 0$
Lower resonant frequency of the sensing system	$\frac{\omega_l}{2\pi}$	31 cps when $\omega_m = 0$
Double modulation frequency	$\frac{\omega_m}{2\pi}$	45.19 cps
Tuning fork average inertia about the y_f axis	I_y	1000 gm cm ² (approx.)
Ratio of the modulated inertia of the tuning fork about the y_f axis to the unmodulated inertia	α_y	0.03 (approx.)
Quality factor of the sensing system at the upper resonant frequency	Q_s	580
Bandwidth of the electrical filtering following the demodulators	BW	0.5 cps

Table 6.2 Drift Characteristics

Parameter	Without Double Modulation	With Double Modulation
Standard deviation of the short term drift	6 eru	6 eru
Standard deviation of the long term drift	44 eru	Not observable on trace
Magnitude of the cross-coupling	5650 eru	60 eru
Phase shift sensitivity	100 $\frac{\text{eru}}{\text{degree}}$	1 $\frac{\text{eru}}{\text{degree}}$

deviation of the short term drift with double modulation would probably be approximately 3 eru which compares favorably with the 6 eru without double modulation since the square root of the ratio of the bandwidths in the two modes of operation is 1.8.

On the basis of the data shown in Figs. 6.6 and 6.7, it appears that the limiting factors determining the short term drift of the tuning fork gyroscope used in these tests are the relatively crude tuning fork and bearings used in the experiment. The short term noise in both modes of operation could probably be reduced by two orders of magnitude in a more sophisticated instrument; however, it is felt that the instrument used in these tests and the data obtained on it are partial to neither mode of operation and the relative short term performances are characteristic of rotary double-modulated vibratory rate gyroscopes. Additional data supporting this conclusion are presented after the long term drift data from Figs. 6.6 and 6.7 are examined.

The long term drift visible on the recordings shown in Figs. 6.6 and 6.7 is different in the two modes of operation because double modulation successfully reduced the cross coupling from 5650 eru to 60 eru. Of the remaining 60 eru, 40 eru of it was due to the acceleration of gravity which was exactly equal to the predicted^{*} value and was inphase with the rate signal as the theory predicts.

The standard deviation of the long term drift when the instrument was operated without double modulation was approximately 44 eru. Much of this drift was caused by temperature-balance effects in the tuning fork which was operated in free air; however, this does not imply that controlling the temperature of the tuning fork would eliminate or reduce the large long term drift in this instrument because the drift caused by material creep and drift of electrical components becomes important over longer periods of time. This has been found to be true in other vibratory drive instruments and was examined in Chapter 5. On the basis of data extrapolated from other tuning fork gyroscopes it is not unreasonable to expect that the long term drift of the present instrument in a temperature controlled environment would be approximately 7×10^{-3} times the quadrature cross coupling. Therefore the number of 44 eru listed in Table 6.2 is not unreasonable for the crude tuning fork used in the test. Data supporting this conclusion are presented later in this section.

* 40 eru of drift were predicted on the basis of the measured acceleration sensitivity of the tuning fork unit.

The standard deviation of long term drift of this same tuning fork unit when it was operated with double modulation in free air was much lower than 44 eru. It is difficult to estimate a number for this drift from the trace of Fig. 6.7 because of all the 0.8 cps noise. If this noise is filtered out by using an eyeball low pass filter with a time constant of 30 seconds or so, the standard deviation of the long term drift would be approximately 3 eru. This drift was caused by the 40 eru of acceleration dependent type 1 cross coupling and the 20 eru of types 2 and 3 cross coupling. Most of this drift or noise appeared to be associated with types 2 and 3 cross coupling because the signal out of the vibration detector did not change noticeably when the tuning fork was turned off. Over a longer period of time, however, the acceleration dependent type 1 cross coupling would undoubtedly be a more important source of drift because it is in phase with the rate signals.

Unfortunately, insufficient data are available to fully support all of these conclusions or allow us to analyze the various components of the drift in the double-modulated instrument. Furthermore, lack of temperature control may appear to be a means of deliberately biasing the experimental data in favor of the double modulation. For these reasons and also the author's dissatisfaction with the repeatability of the data and overall reliability of the equipment, a major modification of the electronics was undertaken and the bearings were changed to ABEC no. 9 ball bearings. Also a temperature controlled environment was assembled so that the temperature of the overall instrument could be controlled. The need for this environment should be apparent from the trace shown in Fig. 6.8 in which the output of the tuning fork gyroscope without double modulation made an approximate periodic oscillation over a 24 hour period with a peak to peak variation of approximately 160 eru when it was operated in free air.

When the tuning fork gyroscope was operated in an environment in which the temperature was controlled to within $\pm 1^{\circ}\text{F}$ and the signal processing demodulators were operated in an environment in which the temperature was controlled to within $\pm 7^{\circ}\text{F}$, the long term drift without double modulation was greatly reduced. This can be seen from the trace in Fig. 6.9. During this test the quadrature cross coupling was approximately 1600 eru and the time constant of the low pass filter following the first demodulator was 2 seconds. The standard deviation of the long term drift over a period of 16 hours was approximately 5 eru. This was the best data obtained on the instrument but

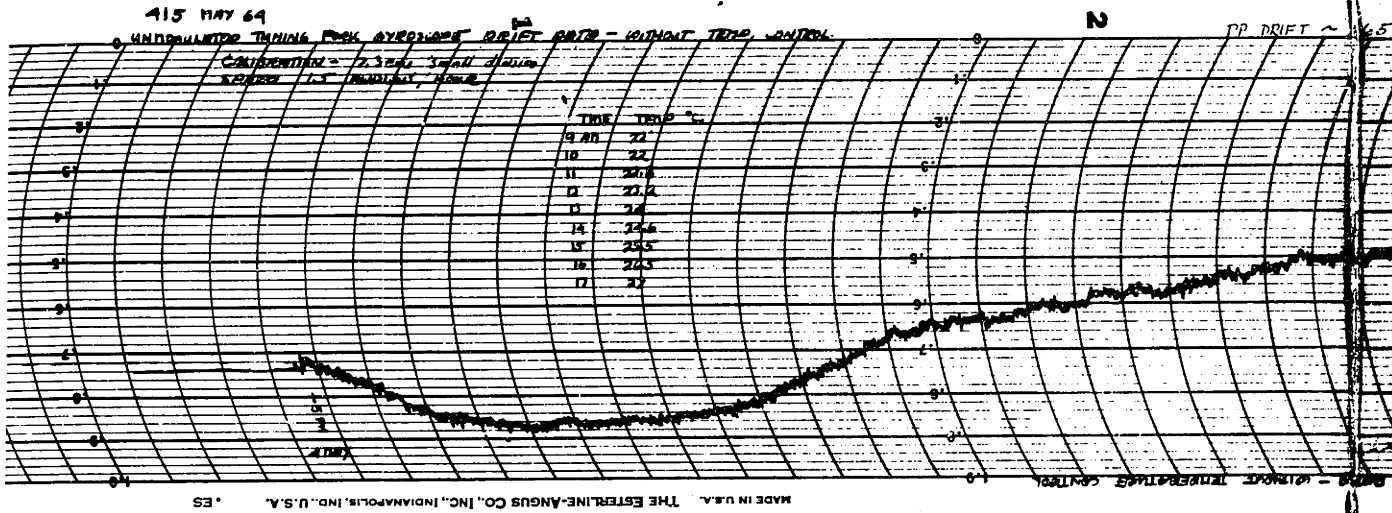


Fig. 6.8 Drift Data Without Double Modulation and Without Temperature Control

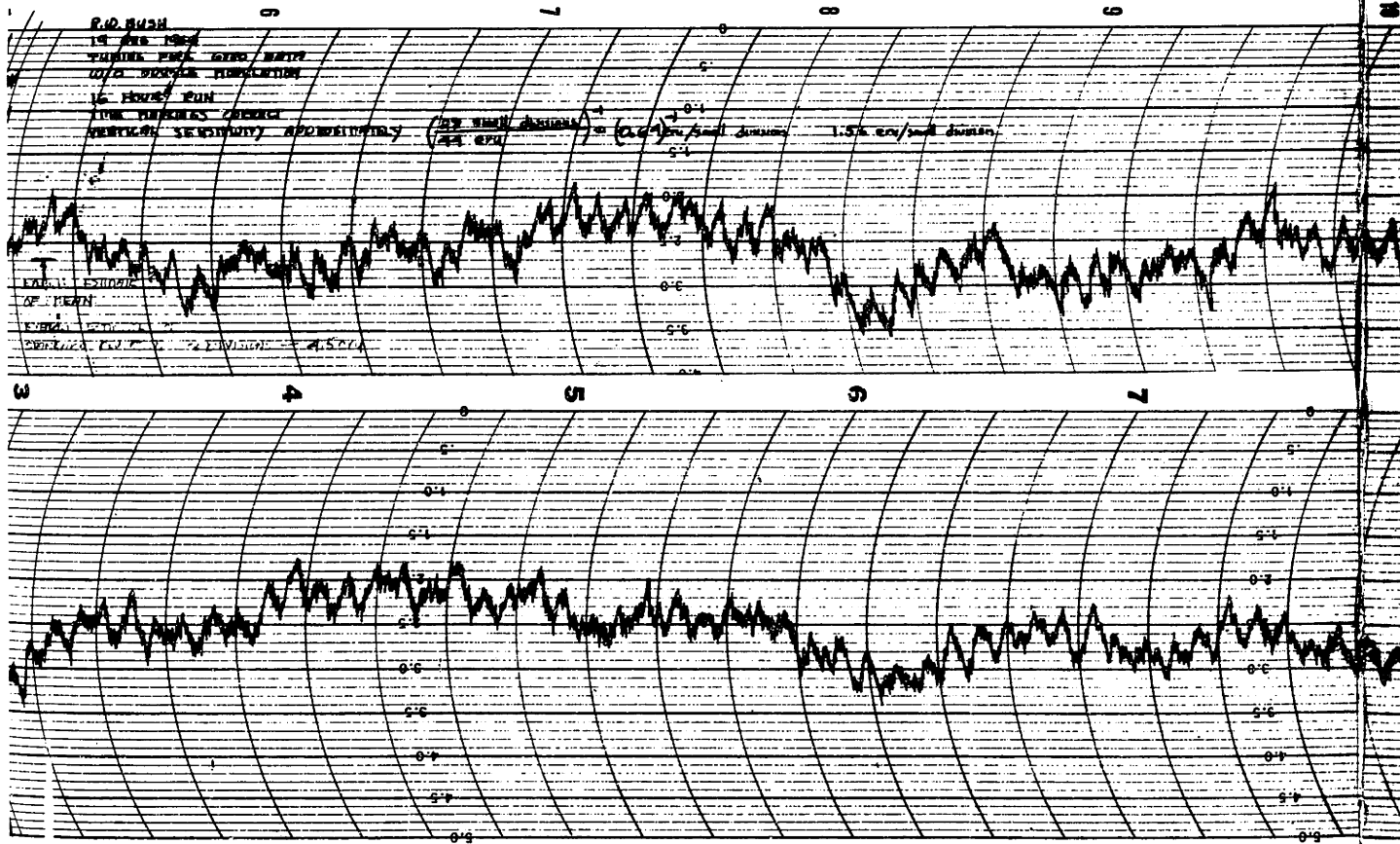
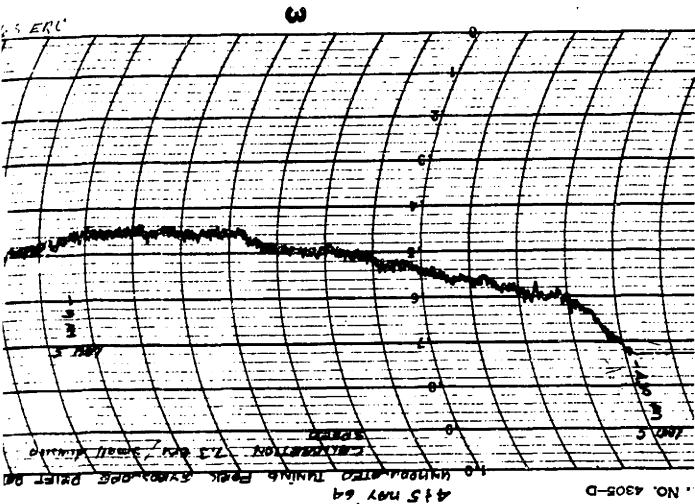
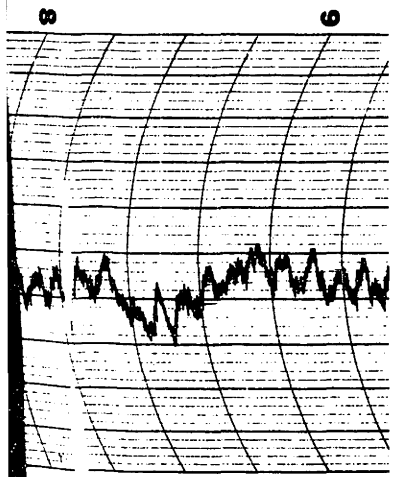
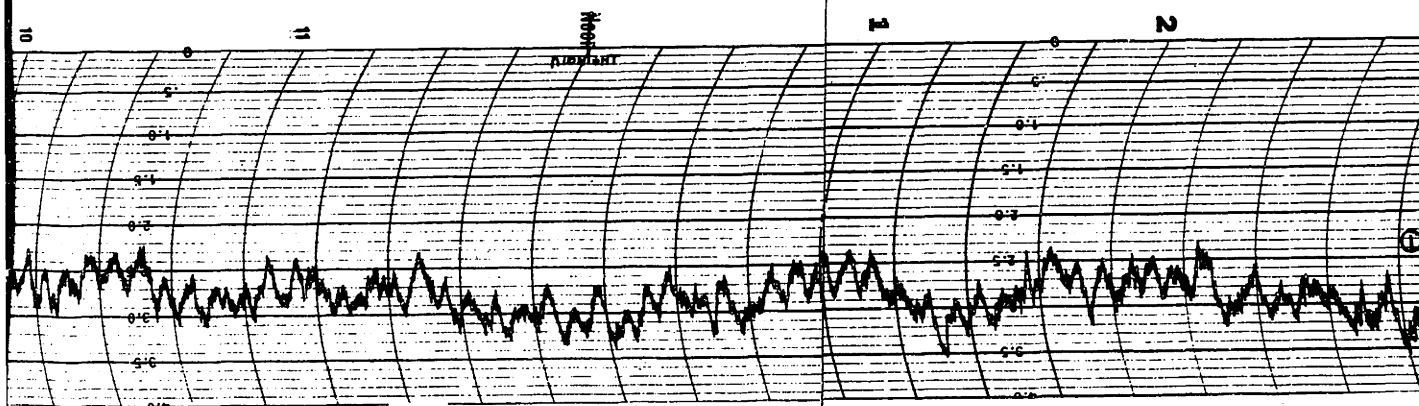


Fig. 6.9 Drift Data Without Double Modulation and With Temperature Control



QUADRATURE CROSS COUPLING APPROXIMATELY 2000 eru
 VERTICAL CALIBRATION 7.3 eru/SMALL DIVISION
 TIME SCALE: 1.5 DIVISIONS/HOUR
 TEMPERATURE NOT CONTROLLED, INSTRUMENT IN
 FREE AIR AND TEMPERATURE CHANGES APPROXIMATELY 8°C
 OVER A 24 HOUR PERIOD



QUADRATURE CROSS COUPLING APPROXIMATELY 2000 eru
 VERTICAL CALIBRATION 1.56 eru/SMALL DIVISION
 TIME SCALE: 4 DIVISIONS/HOUR
 TEMPERATURE CONTROLLED TO WITHIN $\pm 1^\circ\text{C}$

typical drift rates over a 24 hour period were always less than 10 eru. and over a period of several days the drift rates were not much larger. Therefore the previous estimate that the long term drift would be approximately 7×10^{-3} times the quadrature cross coupling is valid for the present instrument.

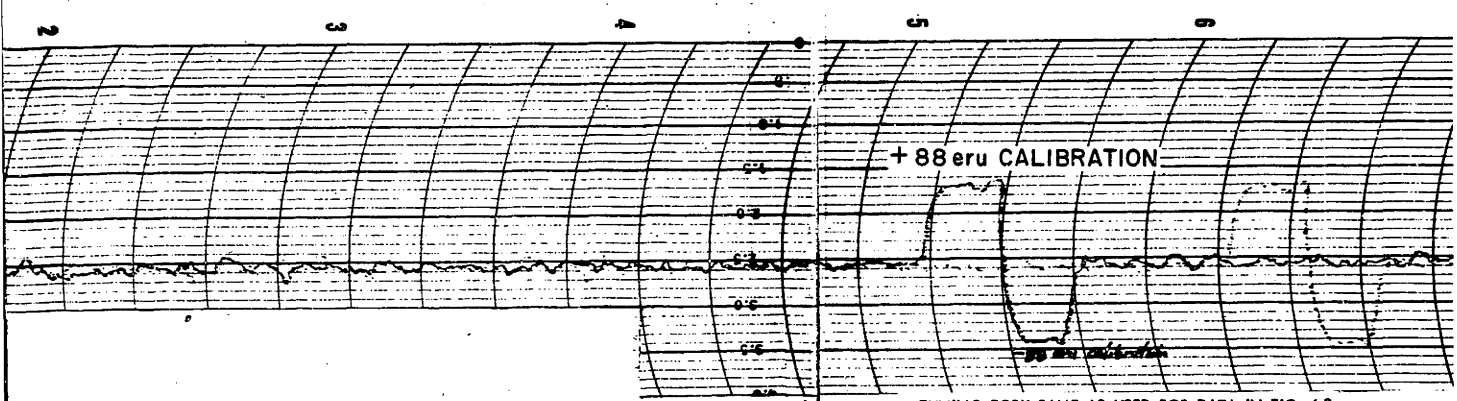
When this same tuning fork gyroscope was operated with double modulation, the equipment was found to be extremely reliable but the types 2 and 3 cross coupling were excessively large. This problem hopefully will be solved in the coming months. Some double modulation data are available on the instrument in its present form, however, which are of interest and are described below.

When the tuning fork gyroscope was operated with double modulation and ABEC no. 9 ball bearings and the resonant frequency of the sensing system was changed so that it did not coincide with any of the frequencies of the three major* types of ball bearing noise, the output noise was very similar to that shown in Fig. 6.7. This noise is shown in Fig. 6.10 where the quality factor of the sensing system was approximately 200 and the time constant of the low pass filter following the second demodulator was 2 seconds. The standard deviation of the short term drift in Fig. 6.10 is approximately 3 eru and does not change when tuning fork drive amplitude and therefore the factor a_y in Eq. 6.2 is increased by a factor of 1.4. Therefore this short term noise was caused by type 2 cross coupling or noise associated with the bearings and unbalance of the carriage. This is encouraging because no attempt was made to accurately balance the carriage. Because this test was made without temperature control and the recording time was relatively short it is impossible to determine the long term drift of the double modulated instrument. This will be done in the future and it is expected that it will be approximately 2 - 3 eru.

6.7 CONCLUSIONS

Double modulation is a process by which all non-acceleration dependent cross coupling between the drive member and the sensing system (type 1 cross coupling) can theoretically be discriminated against in vibratory gyroscopes. Experimental tests with a crude tuning fork gyroscope operating with and without rotary double modulation have shown that this is true. Unfortunately, double modulation also introduces other types of cross coupling so the

* The three types of bearing noise that are referred to here are caused by lumps or holes on the inner race, outer race, or any of the balls.



+ 88 eru CALIBRATION

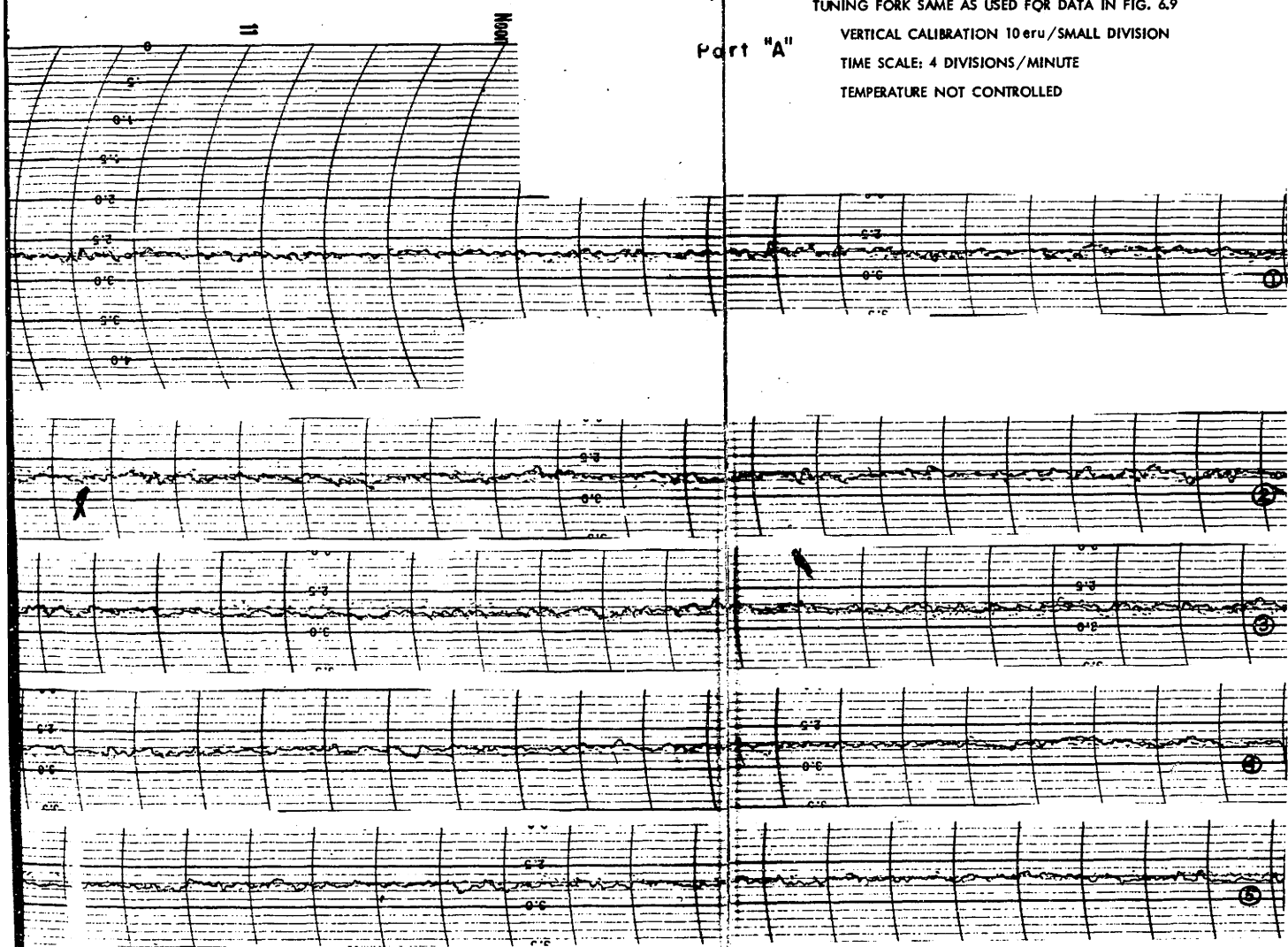
TUNING FORK SAME AS USED FOR DATA IN FIG. 6.9

Part "A"

VERTICAL CALIBRATION 10 eru / SMALL DIVISION

TIME SCALE: 4 DIVISIONS / MINUTE

TEMPERATURE NOT CONTROLLED



Moon

TUNING FORK SAME AS USED FOR DATA IN FIG. 6.9

Part 3"

VERTICAL CALIBRATION 7 eru / SMALL DIVISION

TIME SCALE: 4 DIVISIONS / MINUTE

TEMPERATURE NOT CONTROLLED

overall improvement of the long term drift by double modulation is not necessarily large and depends to a large extent on whether or not the temperature of the instrument is controlled. This is because the non-acceleration dependent cross coupling in vibratory drive instruments is very dependent on temperature and temperature gradients.

Experimental tests with the same crude tuning fork gyroscope showed double modulation improved the long term drift by more than two orders of magnitude when the temperature was not controlled and the instrument was operated in the ordinary laboratory environment; however, when the temperature was controlled to within $\pm 1^{\circ}\text{F}$ the improvement was less than an order of magnitude. Unfortunately, even with the present data and experience with rotary double modulation it is difficult to predict what the long term drift of a well designed rotary doubly modulated instrument is likely to be because:

1. The improvement obtained on the tuning fork configuration by double modulation is not necessarily representative of the improvement that could be obtained with a ring configuration because the Type 2 cross coupling, which was very large and troublesome in the tuning fork instrument, should be small in a ring configuration.
2. The bearings used in the tuning-fork double-modulation experiment were ball bearings which are known to be noisy.

However, on the basis of these experiments and the reported low zero-rate drift of the A5 tuning-fork gyroscope (better than $1^{\circ}/\text{hr}$), it appears doubtful that the long term drift of a rotary double-modulated instrument could be significantly better than that of a super precision, temperature controlled, vibratory drive instrument. If this proves to be so, then the advantages of the double modulated instrument over the more conventional vibratory drive instrument are likely to be:

1. A much lower cost if a ring configuration is used because the super precision drive members are extremely expensive to fabricate and balance whereas rather crude basic drive members can be used in double-modulated instruments.

2. Shorter warm up times because the type 1 cross coupling in vibratory drive devices requires considerable time (at least an hour) to stabilize whereas the rotary double modulation rotation can be started up quite quickly (a few seconds at most).

3. A lower acceleration sensitivity because the Type 1 cross coupling in the low frequency vibratory drive configurations that are necessary for super precision, vibratory drive instruments is more acceleration sensitive than the Types 1 and 3 cross coupling in rotary double modulated instruments.

On the other hand the double-modulated instrument would undoubtedly be less reliable because the rotation requires bearings, some type of a motor and more electronics. Furthermore, the power required for the double modulation motion would be several orders of magnitude larger than that of the basic drive member and certainly the overall instrument would require more than a watt whereas the power required to control the temperature of the more conventional vibratory drive device would also be large but would depend on the environment and the space available for insulation.

The major sources of error in rotary double-modulated instruments are acceleration dependent type 1 cross coupling between the basic drive member and the sensing system and types 2 and 3 cross coupling between the double modulation motion and the sensing system. The types 1 and 2 cross coupling can be reduced to acceptable levels for most applications by the proper choice of the drive member configuration and by balancing the drive member. Therefore the type 3 cross coupling should be the major source of error in rotary double-modulated instruments. This cross coupling can be minimized by balancing the basic drive member and the rotating member and by choosing symmetric bearings and bearing supports but it presently appears that additional research on bearings, or other means of suspending a rotating member, would have to be conducted before a high precision, rotary double-modulated, vibratory gyroscope could be perfected.

At this point it is interesting to note that the type 3 cross coupling that is troublesome in rotary double-modulated VRG's is very similar to the cross coupling that is troublesome in simple, rotary drive, vibratory rate gyroscopes (see Section 5.3). In both types of instruments a torsional oscillation of the rotating member about a rate sensitive axis at twice the frequency of rotation causes zero-rate errors at the instrument output which cannot be distinguished from angular rate inputs. Therefore any research on bearings should be directed toward this problem in such a way that the results are applicable to both rotary drive, vibratory gyroscopes and rotary double-modulated, vibratory gyroscopes. One of the results of this research should be an answer to the following question: Does rotary double modulation, which includes disadvantages of both rotary and vibratory drive instruments, have any significant advantages over the more conventional rotary drive, vibratory gyroscopes? A crude analysis of the non-acceleration dependent cross coupling in these two types of instruments is presented in Appendix C and it indicates that the answer may be "no" in some applications. This is because a torsional oscillation of amplitude ϕ_o and frequency $2\omega_m$ of the rotating member in a rotary double-modulated, vibratory gyroscope causes a zero rate error of $\frac{\phi_o \omega_m}{2}$ while the same oscillation of the rotating member in a rotary drive, drive frequency output, vibratory gyroscope causes a zero-rate error of $\frac{\omega_d \phi_o}{2a_{R_1}}$ where $\omega_d = \omega_m$ and a_{R_1} was defined in Eq. 4.16. Since a_{R_1} has a maximum value of 2 and generally is approximately 1, this indicates that the magnitude of the cross coupling in the double-modulated instrument will be approximately the same as that in the rotary drive instrument. Admittedly this analysis is very crude and does not consider the acceleration dependent cross coupling, which is a severe problem in rotary drive drive frequency output instruments, and does not consider rotary drive, double drive frequency output instruments in which the output torques must be transmitted through the bearings. However, the results are presented here because this is an important question which should be investigated before any additional research is conducted on rotary double-modulated instruments

The conclusions of the previous paragraphs indicate that rotary double modulation is a questionable practical value because:

1. The types 2 and 3 cross coupling associated with rotary double modulation cause significant zero-rate errors. Thus the precision obtainable in a double-modulated instrument may not be any better than that of a high precision vibratory gyroscope without double modulation.
2. The double modulation rotation decreases the reliability of a vibratory drive instrument because it requires bearings, a motor and additional electronics. Also the double modulation rotation greatly increases the power requirements of a vibratory drive device so that the overall instrument is not a low power device. As a result, a rotary double-modulated vibratory gyroscope has few advantages over conventional rotating type gyroscopes and perhaps also over rotary drive vibratory gyroscopes.

However, the latter two reasons are not applicable to instruments which employ the second or vibratory type of double modulation. This is because a second vibration should be as reliable as the first and require approximately the same amount of power. Therefore, the future of double modulation in vibratory gyroscopes may well depend on what precision can be obtained in vibratory double-modulated, vibratory drive instruments. No experimental research has been conducted on these instruments; however, vibratory double modulation does not appear too promising even on paper because:

1. A misalignment of the double modulation axis with respect to the input axis of the basic vibratory drive configuration being double modulated causes type 3 cross coupling. This cross coupling is in quadrature but at the same frequency as the angular rates which are double modulated at the double modulation frequency (See Section 3.5.4 for a description of the harmonic modulations that are associated with vibratory double modulation). Therefore it is necessary to use the second harmonic double-modulation rather than the fundamental if this cross coupling is large.
2. The second harmonic double modulation attenuates the angular rate signals in the basic vibratory drive configuration being double modulated by a factor $-2J_2(\theta_0)$ while the noise in the sensing system and signal processing electronics is almost unchanged. Since it is a difficult mechanical problem to make the double-modulation angle θ_0 large, high precision is even theoretically difficult to achieve in this type of instrument.
3. Second harmonic type 3 cross coupling can occur in these instruments. It results from a misalignment of the double modulation axis and a non-linearity of some type that causes even harmonic rather than just odd harmonic signals from the sensing system transducer. This nonlinearity would not have to be very large to be as troublesome as the fundamental cross coupling which made it necessary to use the second harmonic double modulation because the ratio

of the Coriolis torques at the fundamental and second harmonic of the double modulation frequency --

$\frac{J_2(\theta_o)^*}{J_1(\theta_o)}$ - is a small number for most practical angles θ_o .

Therefore it appears that high precision may not be obtainable in vibratory double-modulated instruments because of the cross-coupling problem and the signal attenuation caused by the factors $J_1(\theta_o)$ or $J_2(\theta_o)$. This is only a preliminary conclusion and there are no experimental data to support it but the numerical values of the Bessel functions given in Table 3.3 are certainly not encouraging.

θ_o	$J_1(\theta_o)$	$J_2(\theta_o)$	$\frac{J_1(\theta_o)}{J_2(\theta_o)}$
0.1	.0499	.0012	41.5
0.2	.0995	.0050	19.8
0.3	.1483	.0112	13.2
0.4	.1960	.0197	9.95
0.5	.2423	.0306	7.90

Table 6.3 Bessel Functions of Small Angles

CHAPTER VII

OPTIMUM PROCESSING OF SUPPRESSED CARRIER SIGNALS

7.1 INTRODUCTION

Suppressed carrier signals occur in communication systems, AC servo-mechanisms, vibratory gyroscopes and numerous other instruments. In all of these applications a suppressed-carrier, amplitude modulated signal which is corrupted by noise is demodulated and filtered. The choice of the signal processor to use for this purpose in a particular problem is always based on the signal processing ability, cost and reliability of the different applicable types, the designers experience and other factors. Usually, this choice is quickly narrowed to a few basic types of signal processors and at that time it is useful to know approximately the signal processing ability of each of them. This is often determined by choosing a meaningful 'goodness' criterion as a measure of the performance of the entire system or instrument and then determining either analytically or experimentally the signal processor of each type which maximizes the "goodness" criterion. These signal processors are by definition the optimum signal processors and the numerical values of their "goodness" are also optimum or maximum.

A 'goodness' criterion which is often used for determining an "optimum signal processor" is a weighted function of the error between the actual output of the system or instrument being investigated and the desired output for typical input signals and noise. Since there are numerous, greatly different

error criteria and also numerous greatly different types of signal processors, it is impossible to develop one analytical design procedure which is meaningful for all suppressed carrier problems. Therefore any theoretical work on optimum signal processing is only applicable to a small class of problems. This is particularly true of the theory developed in the present chapter because only one error criterion and only one basic type of signal processor are considered. This error criterion is the ensemble mean square error and the signal processors are restricted to linear time-varying systems or filters. This choice was made because: the minimum mean square error is a useful figure of merit for comparing the performances of vibratory rate gyroscopes and some other suppressed carrier systems; linear time-varying filters are usually used to process the signals in synchronous, suppressed-carrier systems and the equations which must be solved to determine the optimum signal processor and minimum mean square error are linear and can be solved analytically or else numerically on a digital computer.

The motivation for this research on optimum signal processing of suppressed carrier signals was the vibratory gyroscope signal processing problem in which an accurate model of the instruments is available and the statistical properties of the input angular rates, environmental parameters, and corrupting noise are either known or can easily be determined for a given application. What are not always known are the optimum performance that can be obtained from a given instrument for a given problem; and, how to design the parameters that are relatively easy to change in that instrument, such as the sensing system resonant frequency and time constant and signal processing electronics, to obtain the optimum performance. Much of the previous work done on suppressed carrier filter problems is not particularly

applicable to the vibratory gyroscope problem. This is because the systems analyzed were either not synchronous or quadrature signals were absent or an infinite delay which is associated with good bandpass filtering was allowable¹ or else an elegant theory was developed² which was not applicable to practical suppressed carrier systems in which the carrier frequency, the resonant frequencies of bandpass filters and the phase of the demodulator reference signal drift with time, temperature and numerous other factors. Therefore it was necessary to develop an analytical design technique for the vibratory gyroscope, signal processing problem which takes into account noise, quadrature signals and drifts of the carrier frequency, the resonant frequency of the sensing system and also the phase of the demodulator reference signal. However, since this problem is very similar to the AC servomechanism problem and some other suppressed carrier problems, a more general problem than required for the optimization of vibratory gyroscopes was selected. This problem is examined in the following section. The following chapter then applies the theory developed in the present chapter to the vibratory gyroscope signal processing problem.

7.2 STATEMENT OF THE PROBLEM

A block diagram of the suppressed carrier problem which is analyzed in this chapter is shown in Fig. 7.1 where

$m(t)$ is the message signal which has an autocorrelation function $\phi_{mm}(\tau)$.

$q(t)$ is the quadrature signal which has an autocorrelation function $\phi_{qq}(\tau)$.

$n(t)$ is the noise which has an autocorrelation function $\phi_{nn}(\tau)$.

ω_c is the carrier frequency.

$g_{f1}(\xi_1)$ is the impulse response of a linear time-invariant system located before the optimum filter.

$g_o(t, \tau)$ is the impulse response of the optimum linear time-varying signal processor or filter which has a carrier frequency synchronizing signal as an input in addition to the normal input.

$g_{f2}(\xi_2)$ is the impulse response of a linear time-invariant system located after the optimum filter.

$d(t)$ is the desired output which is usually the input message signal convolved with the desired transfer function-- $g_d(\tau)$ --of the overall system.

$e(t)$ is the error between the actual output and the desired output of the system being analyzed.

The problem is to develop an analytical design technique which will allow us to determine either exactly or approximately the optimum linear time-varying filter and the minimum mean square in the suppressed carrier problem of Fig. 7.1 when:

1. the form of the optimum filter is either completely unspecified, which results in the absolute minimum mean square error, or else restricted so that the minimum mean square error is larger than the absolute minimum but the optimum processor is simpler and therefore cheaper and more reliable.
2. the carrier frequency is not a constant but a slowly varying random variable with a mean of ω_c .
3. the impulse responses of the fixed elements are not constant but vary slowly with time in a random manner.
4. the phase of the synchronizing input to the optimum signal processor is not constant but varies slowly with time in a random manner.
5. the linear time-invariant filters are not completely linear but saturate for large input signals.

The reasons for developing this design technique are to improve the design of suppressed carrier systems and to learn more about the performance limitations of such systems.

The words exactly or approximately are used in the above problem statement because the time and effort required to obtain exact solutions to practical suppressed carrier filter problems, like many other problems, are often much larger than the time and effort required to obtain good approximate solutions. For this reason, ~~analytical~~ techniques for determining upper and lower bounds on the minimum mean square error in the suppressed carrier problem of Fig. 7.1 are also developed in this chapter. These techniques are found to be very worthwhile because:

1. in most suppressed carrier problems in which the exact solution is difficult to obtain, the upper and lower bound on the minimum mean square error are very close together.
2. insight as to the most important performance limitations in suppressed carrier problems is often lost in the lengthy calculations required to obtain the exact answers.
3. in some practical suppressed carrier problems, such as the vibratory gyroscope problem, it is necessary to modify the error criterion slightly to fit the actual problem and this is often difficult to do in the exact formulation but relatively easy to handle approximately.

The model of Fig. 7.1 is a good representation of a vibratory rate gyroscope and also an AC servomechanism. In vibratory gyroscopes the noise $n(t)$ represents both the noise which enters in the sensing system and the noise which enters after the sensing system from the sensing system transducer and the signal processing electronics. The power density spectrum of this "all purpose" noise is

$$\Phi_{nn}(s) = \Phi_{n_1 n_1}(s) + \frac{\Phi_{n_2 n_2}(s)}{G_{f_1}(s) G_{f_1}(-s)} \quad (7.1)$$

where

$\phi_{n_1 n_1}(s)$ is the spectrum of the noise entering at the input or in the sensing system of a vibratory rate gyroscope.

$\phi_{n_2 n_2}(s)$ is the spectrum of the noise entering after the fixed element $g_{f_1}(\xi_1)$.

$G_{f_1}(s)$ is the Laplace transform of $g_{f_1}(\xi_1)$ which is the sensing system transducer combination in a vibratory rate gyroscope.

The analytical techniques required to solve the suppressed carrier filter problem which has just been described are developed in this chapter. This is accomplished by first thoroughly analyzing the simplest possible suppressed carrier filter problem of the type shown in Fig. 7.1. In this analysis all the tools required for the analysis of more complex problems are developed and demonstrated. After this is done the analytical techniques required for solving the more difficult problems involving fixed elements, saturation constraints, carrier frequency drift, component drift and drift of the phase of the synchronizing signal into the optimum filter are developed one at a time. No examples of these more complicated problems are given in this chapter because the following chapter uses many of the analytical techniques to study the vibratory gyroscope problem. Admittedly, some of the theory developed in this chapter is not demonstrated because it appears to be of questionable practical value except in problems where a digital computer is actually going to be used for signal processing. The reason for this apparent lack of usefulness is that the computations required to obtain a solution in a practical problem are lengthy and the optimum performance is difficult to achieve with signal processors other than digital processors. Naturally there is a limit to the complexity of calculations which are practical and when such a limit is approached in the following work it is pointed out and when it is obviously surpassed, alternative, simpler and less exact approaches to the same problem are investigated.

7.3 FREE CONFIGURATION PROBLEM WITH IDEAL COMPONENTS

7.3.1 Review of Linear Time-Varying Filter Theory

When there are no fixed elements or saturation constraints in the suppressed carrier system of Fig. 7.1 and the components are ideal and the form of the optimum LTV (linear time-varying) filter is unrestricted, the problem of minimizing the mean square error is called a free configuration problem with ideal components. Ideal in this sense means that the components in the optimum filter and those that determine the carrier frequency and the phase of the optimum filter synchronizing signal do not drift with time and environmental changes but are constant. In the following work the word ideal is omitted and unless otherwise stated it should be assumed that the components are ideal. When all of these conditions are true, the resulting suppressed carrier filter problem can be analyzed by using the conventional, linear, time-varying, statistical filter theory. This theory, which was originally developed by Booten, is an extension of Wiener's linear time-invariant statistical filter theory. Because the conventional linear time-invariant theory is probably familiar to most readers and the time-varying theory is straightforward extension of it, the time-varying theory is simply reviewed at this time rather than examined in detail. Readers who desire more background on the time-varying statistical filter theory should refer to References 4 by R.C. Booten and 5 by L.A. Zadeh and some of the references* listed in the extensive bibliography which accompanies the latter reference.

* Because this lengthy list of references on the subject exists in the literature, only the references directly pertaining to the present work are listed in this chapter.

The output of a linear time-varying filter is obtained by using a convolution equation which is similar to the one used for linear time-invariant systems. Two different forms of this equation are

$$f_o(t) = \int_0^{\infty} f_i(t-\tau_1)g(t, \tau_1) d\tau_1 = \int_{-\infty}^t f_i(\xi)g(t, \xi) d\xi \quad (7.2)$$

where $f_i(t)$ is the input time function

$f_o(t)$ is the output time function

$g(t, \tau_1)$ is the impulse response of the linear time-varying filter at time t to an impulse occurring τ_1 seconds before at time $t-\tau_1$. $g(t, \xi)$ is another way of writing this expression. The only difference is the change of variables $\xi = t-\tau_1$.

In the following work the notation $g(t, \tau_1)$ is used as the definition of the impulse response of a linear time-varying filter.

When the error out of the system at time t is defined as the difference between actual output and the desired output $d(t)$,

$$e(t) = d(t) - f_o(t) \quad (7.3)$$

the ensemble mean square error at time t is minimized when $g(t, \tau_1)$ is a solution of the following integral equation.

$$\overline{f_i(t-\tau_2) \left[d(t) - \int_{-\infty}^{\infty} d\tau_1 f_i(t-\tau_1)g(t, \tau_1) \right]} = 0, \quad \tau_2 \geq 0 \quad (7.4)$$

This equation simply says that the mean square error is minimum when the error at time t is uncorrelated with the input for all past time. It is necessary to use ensemble averaging in this equation because the input and output autocorrelation functions and the ensemble mean square error are all functions of time so that time averaging is not interchangeable with ensemble averaging.

The expression for the minimum mean square error at time t is

$$\overline{e_{\min}^2(t)} = \overline{d(t) d(t)} - \int_0^{\infty} d\tau_1 \overline{f_i(t-\tau_1) d(t) g_o(t, \tau_1)} \quad (7.5)$$

and is always a function of time in the present problem. This expression is immediately apparent from the equation for the mean square error and the equation for $g_o(t, \tau_1)$

* The straight line is used to represent ensemble averaging.

The integral equation which must be solved for $g_o(t, \tau_1)$ can be obtained from the equation for the mean square error at time t by defining

$$g(t, \tau_1) = g_o(t, \tau_1) + \epsilon y(t, \tau_1) \quad (7.6)$$

where $g_o(t, \tau_1)$ is the optimum filter

$y(t, \tau_1)$ is a physically realizable LTV filter

and then setting the first partial derivative of this equation with respect to ϵ equal to zero. This is the conventional calculus of variations approach to minimizing a functional.

When the correlation functions that are obtained as the result of ensemble averaging in Eq. 7.4 are stationary (independent of time)

$$\overline{d(t) d(t)} = \phi_{dd}(t, 0) = \phi_{dd}(0)$$

$$\overline{f_i(t-\tau_1) f_i(t-\tau_2)} = \phi_{ii}(t, \tau_1, \tau_2) = \phi_{ii}(\tau_2 - \tau_1)$$

$$\overline{f_i(t-\tau_2) d(t)} = \phi_{id}(t, \tau_2) = \phi_{id}(\tau_2)$$

Equation 7.4 is the standard Wiener-Hopf equation and the solution can be written as

$$G_o(s) = \frac{1}{\Delta^+(s)} \left[\frac{\Gamma(s)}{\Delta^-(s)} \right]_+ \quad (7.7)$$

where $\Delta^+(s) \Delta^-(s) = \phi_{ii}(s) =$ Exponential Transform of $\phi_{ii}(\tau_2 - \tau_1)$

$$\Delta^+(-s) = \Delta^-(s)$$

$$\Gamma(s) = \phi_{id}(s) =$$
 Exponential Transform of $\phi_{id}(\tau_2)$

$[]_+ =$ Laplace transform or the transform of the positive time portion of the inverse exponential or two sided transforms of $[]$ when the integration required for the inverse transform is carried out along the $s = j\omega$ axis.

$G_o(s) =$ Laplace transform of the optimum filter which in this case is independent of time.

The minimum mean square error in this time-invariant problem is

$$\overline{e^2(t)}_{\min} = \phi_{dd}(0) - \mathcal{L}_{t=0}^{-1} [\Phi_{dd}(s) G_o(s)] \quad (7.8)$$

where $\mathcal{L}_{t=0}^{-1}$ is the inverse Laplace or one sided transform of the bracketed term evaluated at $t=0$. Readers who are not familiar with the Wiener-Hopf equation, exponential transforms or the spectrum factorization method of solution (Eq. 7.7) should refer to reference 6 by Newton, Gould and Kaiser.

When the correlation functions in Eq. 7.5 are not stationary, the optimum filtering problem becomes much more difficult to solve and explicit solutions can usually only be obtained in special cases. In these special cases the input autocorrelation function can be factored as

$$\phi_{ii}(t, \tau) = \int_{-\infty}^{\infty} d\xi h(t, \xi) h(t+\tau, \xi+\tau) \quad (7.9)$$

This is equivalent to requiring that $\phi_{ii}(t, \tau)$ be the autocorrelation function of the output of a linear time-varying filter with an impulse response $h(t, \xi)$ when the input is white noise. Unfortunately, this factoring technique is usually extremely difficult or even impossible to carry out in most signal processing problems with non-stationary inputs. This is particularly true if the input is the sum of a signal and noise. In fact, it has not yet been shown that it is always possible to factor any real valued, symmetric autocorrelation function. Therefore, in the following work on suppressed carrier signal processing, no attempt* is made to factor the autocorrelation function except in the special case where the autocorrelation function of the message signal is equal to that of the quadrature signal.

* This does not mean the author did not try to factor $\phi_{ii}(t, \tau)$.

7.3.2 Free Configuration Problem Integral Equation

In this section and the following ones in this chapter, the integral equation which must be solved in order to determine the optimum LTV filter is simply stated when it is in the form of Eq. 7.4. Furthermore it is assumed that the autocorrelation functions of m, q, n , and d (see Fig. 7.1) are stationary and all cross-correlation functions except $\phi_{md}(\tau)$ are zero. These conditions exist in most suppressed carrier problems and when they are true the resulting equations are much easier to write down and solve.

For the free configuration suppressed carrier problem, the integral equation which must be solved for the optimum filter is

$$\phi_{md}(\tau_2) \cos \omega_c(t-\tau_2) = \frac{1}{2} \int_0^{\infty} d\tau_1 g_o(t, \tau_1) \left\{ \begin{array}{l} [\phi_{mm}(\tau_2-\tau_1) + \phi_{qq}(\tau_2-\tau_1)] \cos \omega_c(\tau_2-\tau_1) \\ [\phi_{mm}(\tau_2-\tau_1) - \phi_{qq}(\tau_2+\tau_1)] \cos \omega_c(2t-\tau_1-\tau_2) \\ 2 \phi_{nn}(\tau_2-\tau_1) \end{array} \right\} \quad (7.10)$$

for $\tau_2 \geq 0$

and the minimum mean square error associated with the optimum filter is

$$\overline{e^2(t)}_{\min} = \phi_{dd}(0) - \int_{-\infty}^{\infty} d\tau_1 g_o(t, \tau_1) \phi_{md}(\tau_1) \cos \omega_c(t-\tau_1) \quad (7.11)$$

Because of the time varying term $-\cos \omega_c(2t-\tau_2-\tau_1)$ in the input autocorrelation function in Eq. 7.10, it appears impossible to obtain an explicit solution for $g_o(t, \tau_1)$ unless

$$\phi_{mm}(\tau) = \phi_{qq}(\tau) \quad (7.12)$$

Although this appears to be a severe restriction, this condition is often approximately true for rotary drive vibratory gyroscopes, gyroscopes with rotary double modulation and some synchronous communication systems. It is fortunate that it is true for these two types of vibratory gyroscopes because in these instruments the carrier frequency is generally much lower than in other

types of instruments and therefore signal processing difficulties due to overlapping spectra may be encountered. In other types of vibratory g gyroscopes and other suppressed carrier problems in which $\phi_{mm}(\tau) \neq \phi_{qq}(\tau)$, an approximation technique must be used to solve Eq. 7.10. These two cases are examined in the following sections and then approximate techniques for bounding the mean square error are developed.

7.3.3 Special Case When $\phi_{mm}(\tau) = \phi_{qq}(\tau)$

When $\phi_{mm}(\tau) = \phi_{qq}(\tau)$ the input autocorrelation function in Eq. 7.10 is stationary. When a solution of the following form is assumed, the integral equation

$$g_o(\tau_1, t) = g_{o_s}(\tau_1) \sin \omega_c t + g_{o_c}(\tau_1) \cos \omega_c t \quad (7.13)$$

for $g_o(t, \tau_1)$ separates into two time-invariant Wiener-Hopf type equations.

$$\phi_{md}(\tau_2) \cos \omega_c \tau_2 = 2 \int_0^{\infty} d\tau_1 g_{o_c}(\tau_1) [\phi_{mm}(\tau_2 - \tau_1) \cos \omega_c(\tau_2 - \tau_1) + \phi_{nn}(\tau_2 - \tau_1)] \quad (7.14)$$

($\cos \omega_c t$ equation which must be satisfied for $\tau_2 \geq 0$)

$$\phi_{md}(\tau_2) \sin \omega_c \tau_2 = 2 \int_0^{\infty} d\tau_1 g_{o_s}(\tau_1) [\phi_{mm}(\tau_2 - \tau_1) \cos \omega_c(\tau_2 - \tau_1) + \phi_{nn}(\tau_2 - \tau_1)] \quad (7.15)$$

($\sin \omega_c t$ equation which must be satisfied for $\tau_2 \geq 0$)

The reason we were able to assume the correct form of the optimum filter is developed in the following section. The solutions of these Wiener-Hopf type equations for the two linear time-invariant filters are

$$\begin{aligned} G_{o_s}(s) &= \frac{1}{\Delta^+(s)} \left[\frac{\Gamma_s(s)}{\Delta^-(s)} \right]_+ = I_m^* \left\{ \frac{1}{\Delta^+(s)} \left[\frac{\phi_{md}(s-j\omega_c)}{\Delta^+(s)} \right]_+ \right\} = \text{Im} \{ G_{-j\omega_c}(s) \} \\ &= -I_m \left\{ \frac{1}{\Delta^+(s)} \left[\frac{\phi_{md}(s+j\omega_c)}{\Delta^-(s)} \right]_+ \right\} = \text{Im} \{ G_{+j\omega_c}(s) \} \end{aligned}$$

* This notation means the imaginary part of the bracketed term when s is considered as a real number. Re on the following page means the real part under the same condition.

$$G_{O_c}(s) = \frac{1}{\Delta^+(s)} \left[\frac{\Gamma_c(s)}{\Delta^-(s)} \right] = \operatorname{Re} \left\{ \frac{1}{\Delta^+(s)} \left[\frac{\phi_{md}(s-j\omega_c)}{\Delta^-(s)} \right] \right\} = \operatorname{Re} \{ G_{-j\omega_c}(s) \} \quad (7.16)$$

$$= \operatorname{Re} \left\{ \frac{1}{\Delta^+(s)} \left[\frac{\phi_{md}(s+j\omega_c)}{\Delta^-(s)} \right] \right\} = \operatorname{Re} \{ G_{+j\omega_c}(s) \} \quad (7.17)$$

where

$$\Delta^+(s) \Delta^-(s) = \Delta(s) = \phi_{mm}(s-j\omega_c) + \phi_{mm}(s+j\omega_c) + 2\phi_{nn}(s)$$

$$\Gamma_s(s) = \frac{1}{2j} [\phi_{md}(s-j\omega_c) - \phi_{md}(s+j\omega_c)] = \operatorname{Im} \{ \phi_{md}(s-j\omega_c) \}$$

$$\Gamma_c(s) = \frac{1}{2} [\phi_{md}(s-j\omega_c) + \phi_{md}(s+j\omega_c)] = \operatorname{Re} \{ \phi_{md}(s-j\omega_c) \} \quad (7.18)$$

$$G_{-j\omega_c}(s) = \frac{1}{\Delta^+(s)} \left[\frac{\phi_{md}(s-j\omega_c)}{\Delta^-(s)} \right] +$$

$$G_{+j\omega_c}(s) = \frac{1}{\Delta^+(s)} \left[\frac{\phi_{md}(s+j\omega_c)}{\Delta^-(s)} \right] +$$

The optimum filter which results is shown in Fig. 7.2. An equivalent filter, shown in Fig. 7.3, is the more conventional representation. The two linear time-invariant filters in Fig. 7.3 are related to those in Fig. 7.2 as

$$G'_{O_c}(s) = 2 [\operatorname{Re} G_{O_c}(s-j\omega) + \operatorname{Im} G_{O_s}(s-j\omega_c)] = 2 \operatorname{Re} G_{+j\omega_c}(s-j\omega_c)$$

$$G'_{O_s}(s) = 2 [\operatorname{Re} G_{O_s}(s-j\omega) - \operatorname{Im} G_{O_c}(s-j\omega_c)] = -2 \operatorname{Im} G_{+j\omega_c}(s-j\omega_c) \quad (7.19)$$

An example of the optimum filters that occurs in this special case is determined in Section 7.38.

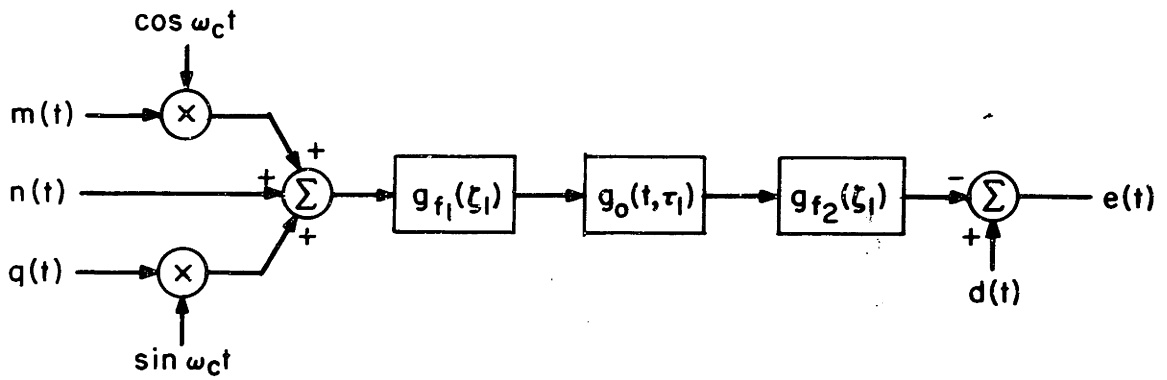


Fig. 7.1 Suppressed Carrier Problem

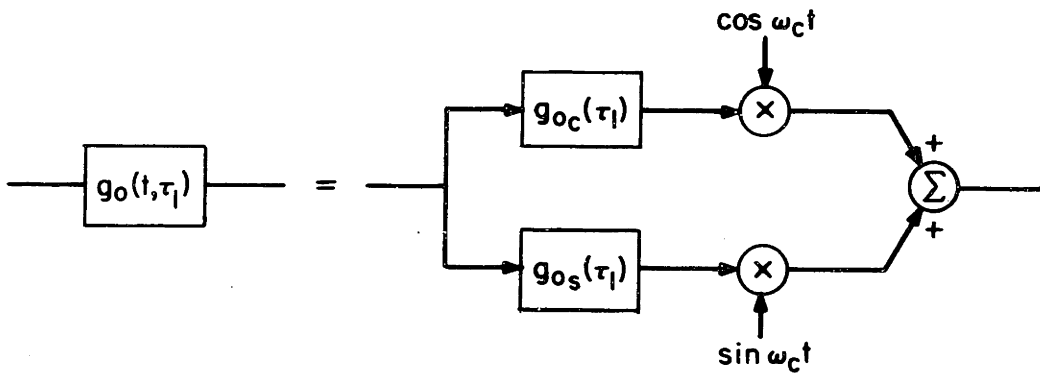


Fig. 7.2 Optimum Filter in the Special Case

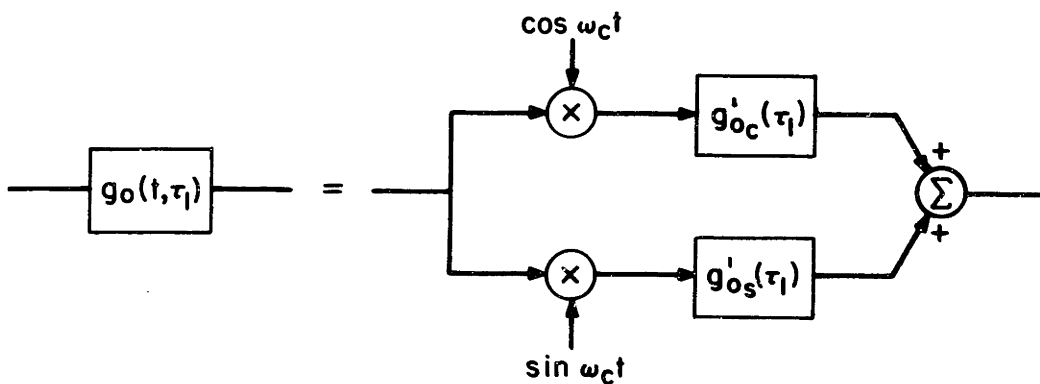


Fig. 7.3 Alternate Form for the Optimum Filter in the Special Case

7.3.4 General Case When $\phi_{mm}(\tau) \neq \phi_{qq}(\tau)$

When the input autocorrelation function is not stationary, the integral equations (Eq. 7.9) for $g_0(t, \tau_1)$ cannot be solved explicitly. Therefore an approximation technique must be employed. Two approximate methods which are applicable to the present problem are:

1. Approximate the optimum LTV filter by a discrete filter (discrete along both the τ_1 and t axes) with a finite memory and solve the resulting set of linear algebraic equations on a digital computer.
2. Represent the optimum LTV filter by means of a complete set of linear time-varying filters of an appropriate form. Then truncate the series and solve the resulting set of linear integral equations for an approximately optimum LTV filter.

When the optimum LTV filter is constrained to be a discrete filter with a finite memory, the minimum mean square error is naturally larger than when a continuous unrestricted LTV filter with an infinite memory is used. As the memory time of the discrete filter becomes longer and the distance between the discrete points becomes smaller, the minimum mean square error associated with the discrete filter converges to the absolute minimum. Unfortunately, it may take a large number of discrete points to adequately approximate the continuous filter. This practical consideration is examined after the equations for the optimum discrete filter are derived.

Let us first examine the equation which must be solved in order to find the optimum discrete filter at time $t = t_k$. This filter is a series of N impulses along the τ_1 axis which are separated by time T_1 and have an area $T_1 g_k(nT_1)$ where $nT_1 = \tau_1$. Therefore the output of this filter can be written as

$$f_o(t_k) = \sum_{n=0}^{N-1} T_1 g_k(nT_1) f_i(t_k - nT_1) \quad (7.20)$$

and the ensemble mean square error at time t_k can be written in matrix notation as

$$\overline{e(t_k)e(t_k)} = \phi_{dd}(0) - 2T_1 \overline{g_k(nT_1)} \phi_{id}(t_k, nT_1) + T_1^2 \overline{g_k(nT_1)} [\phi_{ii}(t_k, nT_1, mT_1)] g_k(mT_1) \quad (7.21)$$

where the autocorrelation function matrix

$$\begin{aligned} [\phi_{ii}(t_k, nT_1, mT_1)] &= \overline{f_i(t_k - nT_1) f_i(t_k - mT_1)} \\ &= \left[\left\{ \frac{\phi_{mm}(n-m)T_1}{2} + \frac{\phi_{qq}(n-m)T_1}{2} \right\} \cos \omega_c(n-m)T_1 + \phi_{nn}(n-m)T_1 \right] + \\ &\quad \left[\left\{ \frac{\phi_{mm}(n-m)T_1}{2} - \frac{\phi_{qq}(n-m)T_1}{2} \right\} \cos \omega_c 2t - (n+m)T_1 \right] \end{aligned} \quad (7.22)$$

is an $N \times N$ square symmetric matrix and the cross-correlation function matrix

$$\phi_{id}(t_k, nT_1) = \overline{f_i(t_k - nT_1) d(t_k)} = \phi_{md}(nT_1) \cos \omega_c(t_k - nT_1) \quad (7.23)$$

is an N element vector. Now in order to find the optimum discrete filter, the expression for the mean square error must be partially differentiated with respect to each of the unknowns and this expression equated to zero.

$$0 = \frac{\overline{\partial e^2(t_k)}}{\partial g_k(0)} \Delta g_k(0) + \frac{\overline{\partial e^2(t_k)}}{\partial g_k(1)} \Delta g_k(1) + \dots + \frac{\overline{\partial e^2(t_k)}}{\partial g_k(N-1)} \Delta g_k(N-1) \quad (7.24)$$

Since this equation must be satisfied for all $\Delta g_k(nT_1)$, the partial derivatives must all equal zero to minimize* the mean square error. It can be shown that this requires that the optimum filter satisfy the following matrix equation.

$$\phi_{id}(t_k - nT_1) = T_1 [\phi_{ii}(t_k, nT_1, mT_1)] g_k(mT_1) \quad (7.25)$$

Therefore the discrete optimum filter and associated minimum mean square error at time t_k are

$$T_1 g_k(mT_1) = [\phi_{ii}(t_k, nT_1, mT_1)]^{-1} \phi_{id}(t_k - nT_1) \quad (7.26)$$

$$\overline{e^2(t_k)}_{\min} = \phi_{dd}(0) - T_1 \overline{\phi_{id}(t_k, mT_1)} [\phi_{ii}(t_k, nT_1, mT_1)]^{-1} \phi_{id}(t_k - nT_1) \quad (7.27)$$

* This is a minimum and not a maximum because the second partial derivative is always greater than zero.

** This same equation could have been obtained by substituting the discrete form of the optimum filter into Eq. 7.4.

It is now apparent that an NxN square symmetric matrix must be inverted at each point t_k in order to determine the optimum discrete filter $g_k(nT_1)$. Fortunately this task can be performed by a digital computer by using standard matrix inversion programs. All that must be determined are T_1 and N and the interval T' along the t axis between the discrete points. This interval is defined as

$$t_{k+1} - t_k \triangleq T', \quad k = \text{an integer} \quad (7.28)$$

Let us now determine some of the characteristics of the optimum filter before analyzing these practical considerations.

Two important characteristics can be determined from the following two equations

$$\phi_{id}(t_k, nT_1) = \phi_{id}(t_k + \frac{2\pi}{\omega_c}, nT_1) = -\phi_{id}(t_k + \frac{\pi}{\omega_c}, nT_1) \quad (7.29)$$

$$\phi_{ii}(t_k, mT_1, nT_1) = \phi_{ii}(t_k + \frac{2\pi}{\omega_c}, mT_1, nT_1) = \phi_{ii}(t_k + \frac{\pi}{\omega_c}, nT_1, mT_1) \quad (7.30)$$

The first characteristic is that $g_o(t, \tau_1)$ is periodic in time with a frequency ω_c .

The second characteristic is that

$$g_k(mT_1) = -g_{k + \frac{\pi}{\omega_c T_1}}(mT_1) \quad (7.31)$$

Third, it can be shown that when N is chosen so that NT_1 is an integral number of periods of the carrier frequency and T' is made equal to T_1 , the following relationship is true.

$$[\phi_{ii}(\ell T, nT, mT)] = [\phi_{ii}(\frac{\pi}{\omega_c} - \ell T, nT_1, mT_1)] [I_s], \quad \ell=0, 1, \dots, \frac{\pi}{T'\omega_c} \quad (7.32)$$

$$\text{where } [I_s] = \begin{bmatrix} 0 & & \\ & 1 & \\ & & 1 & \\ & & & 0 \end{bmatrix} = \delta(i+j = N)$$

Because of these three characteristics, when $T' = T_1 = T$ and N is chosen so that NT is an integral number of periods of the carrier frequency

and there are K discrete points per half cycle of the carrier frequency, it is only necessary to invert $\frac{K}{2}$ (K even) or $\frac{K+1}{2}$ (K odd) $N \times N$ square symmetric real matrices in order to determine the optimum, discrete, linear time-varying filter with a finite memory.

Let us now return to the practical considerations of selecting the values of N and T to use in an actual problem. Fortunately, it is possible to estimate these parameters without resorting to the trial and error approach on a computer. Since the input message is modulated at ω_c , it follows that the impulse response of $g_o(t, \tau_1)$ will also oscillate with a frequency ω_c back along the τ_1 axis in synchronism with $\cos \omega_c(t - \tau_1)$. Therefore T must be less than $\frac{\pi}{\omega_c}$ and as a first approximation it should probably be on the order of $\frac{\pi}{10\omega_c}$

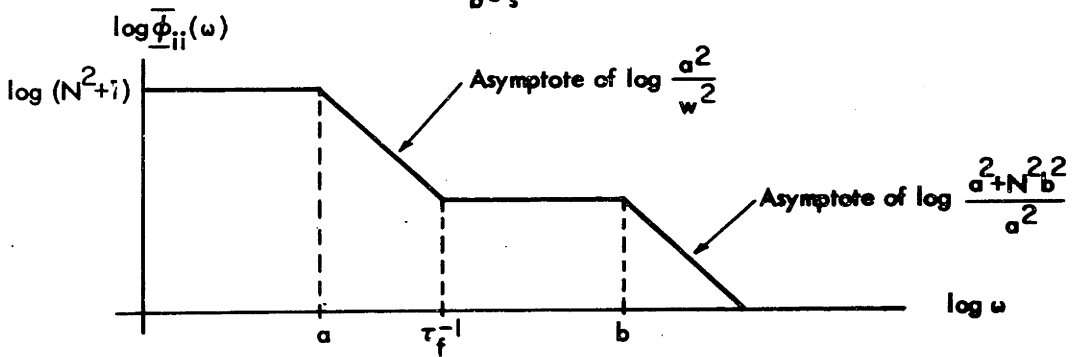
$$\frac{\pi}{10\omega_c} \geq T \quad (7.33)$$

Intuitively the envelope of $g_o(t, \tau_1)$ should be approximately equal to the impulse response of the optimum filter that one obtains by minimizing the mean square error between the desired output $d(t)$ and the output of a linear time-invariant filter in which the input is $m(t) + n'(t)$ where $n'(t)$ is the noise $n(t)$ in the frequency range around the carrier frequency. Actually we are only interested in the time constant of this filter and it can be estimated in numerous ways. One of the easiest is the graphical spectrum factorization method that is demonstrated in Fig. 7.4. It is shown in this figure that by plotting $\log \hat{q}_{11}(j\omega)$ vs. $\log \omega$ using asymptotic approximation techniques and then observing where the break frequencies differ from those of $\hat{q}_{mm}(j\omega)$, it is possible to estimate the bandwidth of the optimum filter. Therefore, in this manner it is possible to estimate the time constant (τ_f) of the optimum suppressed carrier filter. Once this has been estimated the memory time of optimum discrete filter (NT) can be made three times this time constant as a good first approximation.

1. Plot $\bar{\phi}_{ii}(s) = \bar{\phi}_{mm}(s) + \bar{\phi}_{nn}^i(s)$ on log log paper using asymptotic approximation techniques.

Example $\bar{\phi}_{mm}(s) = \frac{a^2}{a^2 - s^2}$

$\bar{\phi}_{nn}^i(s) = \frac{N^2 b^2}{b^2 - s^2}$



2. If $\bar{\phi}_{dd}(s) = \bar{\phi}_{mm}(s)$ the filter bandwidth is τ_f^{-1} and can be determined approximately by this asymptotic approximation technique.

Fig. 7.4 Graphical Spectrum Factorization

It then follows that as a good first approximation one should use

$$N = 10 \tau_f \omega_c$$

It is immediately apparent that in problems in which the time constant of the optimum filter is more than an order of magnitude larger than period of the carrier frequency, N is large and the time required for matrix inversion on a digital computer, which increases as N^3 for large N , may be excessively large and expensive. In these cases it is necessary to use some other method to obtain an exact answer or else settle for bounds on the mean square error. Fortunately, in cases where N is large the upper and lower bounds that are derived later in this section are quite close together.

A second method for obtaining an approximate solution to the integral equation for $g_o(t, \tau_1)$ involves representing $g_o(t, \tau_1)$ in terms of a sinusoidal expansion along the time axis. This representation is shown in Fig. 7.5 and is

$$g_o(t, \tau_1) = \sum_{n=1}^{\infty} g_{2n-1}^c(\tau_1) \cos (2n-1)\omega_c t + g_{2n-1}^s(\tau_1) \sin (2n-1)\omega_c t \quad (7.34)$$

This expansion is a valid representation of $g_o(t, \tau)$ because $g_o(t, \tau_1) = -g_o(t + \frac{\pi}{\omega_c}, \tau_1) = g_o(t + \frac{2\pi}{\omega_c}, \tau_1)$ and the sine-cosine series is complete. Now instead of minimizing the ensemble mean square error at all times, it is necessary to minimize the time average of the ensemble mean square error. This can be done by using the conventional calculus of variations technique on each of the linear time-invariant filters in the equation for $e^2(t)$ where the second line implies a time average. When the series is complete the resulting minimum mean square error as a function of time is a minimum at all points. When the series is not complete but is truncated after a finite number of terms, the same approach is valid but the resulting minimum mean square error at any time t

* $N = 100$ is a reasonable dividing line between large and small although some people talk seriously of inverting matrices in which $N = 2 \times 10^4$.

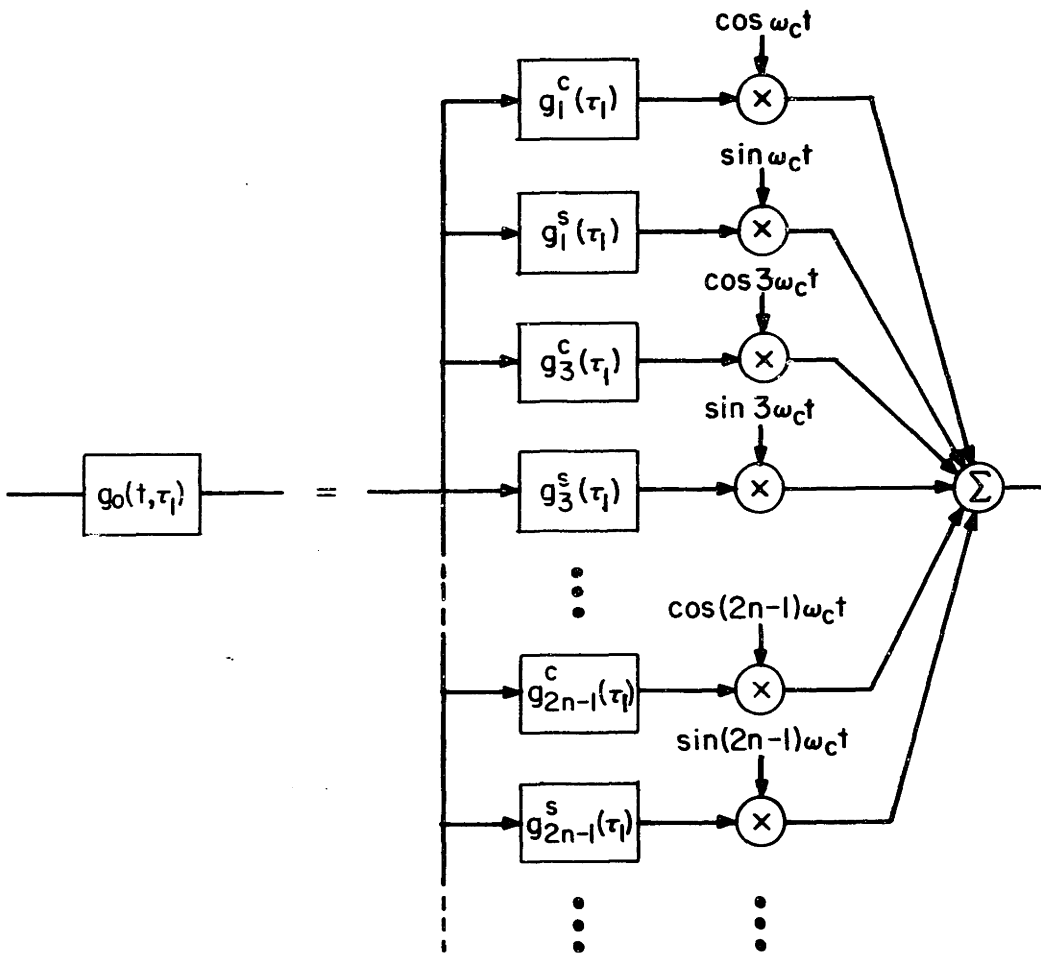


Fig. 7.5 Series Form of the Optimum Filter

is equal to or greater than the absolute minimum and therefore the time average of the mean square error is greater than the absolute minimum.

When this expansion is used the solution is quite easy to obtain in the special case where $\phi_{mm}(\tau) = \phi_{qq}(\tau)$ because the integral equations for $g_{2n-1}^C(\tau_1)$ and $g_{2n-1}^S(\tau_1)$ are uncoupled. In the more general problem when $\phi_{mm}(\tau) \neq \phi_{qq}(\tau)$ the equations are coupled together and the solutions are difficult to obtain. This can be seen from the following equations for g_1^C , g_1^S , g_m^C and g_m^S where $m = 2n-1$.

$$\begin{aligned} \phi_{md}(\tau_2) \cos \omega_c \tau_2 &= \int_0^{\infty} d\tau_1 \left\{ \begin{aligned} &\phi_{11}(\tau_2 - \tau_1) g_1^C(\tau_1) + \\ &\frac{1}{2} \phi_{22}(\tau_2 - \tau_1) \left\{ \begin{aligned} &\cos \omega_c(\tau_2 + \tau_1) [g_3^C(\tau_1) + g_1^C(\tau_1)] \\ &\sin \omega_c(\tau_2 + \tau_1) [g_3^S(\tau_1) + g_1^S(\tau_1)] \end{aligned} \right\} \end{aligned} \right\} \quad (7.35) \\ \tau_2 &\geq 0 \\ (g_1^C \text{ equation}) \end{aligned}$$

$$\begin{aligned} \phi_{md}(\tau_2) \sin \omega_c \tau_2 &= \int_0^{\infty} d\tau_1 \left\{ \begin{aligned} &\phi_{11}(\tau_2 - \tau_1) g_1^S(\tau_1) + \\ &\frac{1}{2} \phi_{22}(\tau_2 - \tau_1) \left\{ \begin{aligned} &\cos \omega_c(\tau_2 + \tau_1) [g_3^S(\tau_1) - g_1^S(\tau_1)] \\ &\sin \omega_c(\tau_2 + \tau_1) [-g_3^C(\tau_1) + g_1^C(\tau_1)] \end{aligned} \right\} \end{aligned} \right\} \quad (7.36) \\ \tau_2 &\geq 0 \\ (g_1^S \text{ equation}) \end{aligned}$$

$$\begin{aligned} 0 &= \int_0^{\infty} d\tau_1 \phi_{11}(\tau_2 - \tau_1) g_m^C(\tau_1) + \\ \tau_2 &\geq 0 \\ (g_m^C \text{ equation}) \frac{1}{2} \phi_{22}(\tau_2 - \tau_1) &\left\{ \begin{aligned} &\cos \omega_c(\tau_2 - \tau_1) [g_{m-2}^C(\tau_1) + g_{m+2}^C(\tau_1)] \\ &\sin \omega_c(\tau_2 - \tau_1) [-g_{m-2}^S(\tau_1) + g_{m+2}^S(\tau_1)] \end{aligned} \right\} \quad (7.37) \end{aligned}$$

$$\begin{aligned} 0 &= \int_0^{\infty} d\tau_1 \phi_{11}(\tau_2 - \tau_1) g_m^S(\tau_1) + \\ \tau_2 &\geq 0 \\ (g_m^S \text{ equation}) \frac{1}{2} \phi_{22}(\tau_2 - \tau_1) &\left\{ \begin{aligned} &\cos \omega_c(\tau_2 + \tau_1) [g_{m-2}^S(\tau_1) + g_{m+2}^S(\tau_1)] \\ &\sin \omega_c(\tau_2 + \tau_1) [g_{m-2}^C(\tau_1) + g_{m+2}^C(\tau_1)] \end{aligned} \right\} \quad (7.38) \end{aligned}$$

where $\phi_{11}(\tau_2 - \tau_1) = [\phi_{mm}(\tau_2 - \tau_1) + \phi_{qq}(\tau_2 - \tau_1)] \cos \omega_c(\tau_2 - \tau_1) + 2\phi_{mn}(\tau_2 - \tau_1)$

$$\phi_{22}(\tau_2 - \tau_1) = [\phi_{mm}(\tau_2 - \tau_1) - \phi_{qq}(\tau_2 - \tau_1)]$$

$$m = 1, 3, 5 \dots \text{odd}$$

It is immediately evident that when $\phi_{22}(\tau) \neq 0$, the integral equations are coupled together and even when the series is truncated the integral equations are difficult to solve in a practical problem. To make matters worse, all the filters change when the number of terms in the expansion increases or decreases. Therefore, this method will not be examined further because it appears to involve more algebraic manipulations than are practical for even an academic exercise: and, if numerical techniques must be employed, the previous method is undoubtedly more practical.

7.3.5 Upper Bound on the Minimum Mean Square Error

In the suppressed carrier problem as well as many others, an exact answer is often difficult to obtain but bounds can be put on the answer with much less difficulty. If the upper and lower bounds are not too far apart, it is often unnecessary to solve the exact problem. We have just seen that in the suppressed carrier optimization problem, the minimum mean square error and optimum linear time-varying filter are difficult to determine except when $\phi_{mm}(\tau) = \phi_{qq}(\tau)$ or when the bandwidth of $g(t, \tau_1)$ and the carrier frequency are within an order of magnitude or so of each other such that N is less than 100 when the discrete approximation approach is employed. In other problems the number of samples required to approximate the optimum filter with sufficient

accuracy may become so large that matrix inversion on a digital computer is excessively expensive. Fortunately, in most of these problems the upper and lower bounds that can be determined easily and explicitly without the aid of a digital computer do not differ greatly. This is true because harmonic errors from the message and quadrature components can be filtered out when the carrier frequency is much larger than the bandwidth of the modulating signals and the delays of $\frac{\pi}{2\omega_c}$ imposed by sampling demodulators do not cause significant errors.

An upper bound on the minimum mean square error in the free configuration suppressed carrier problem can be determined by using sampled data theory. When the input signal $f_1(t)$ in Fig. 7.1 is sampled twice per cycle when $\sin \omega_c t = 0$ and the sampled pulses alternate in sign and then these pulses are passed through an optimum discrete time-invariant linear filter and the output held between samples, the resulting mean square error establishes an upper bound on the minimum that can be obtained using an unrestricted linear time-varying filter. This system is shown in Fig. 7.6. The equation for the optimum discrete filter in Fig. 7.6 is

$$H_o(z_o) = \frac{1}{\Delta^+(z)} \left[\frac{F(z)}{\Delta^-(z)} \right]_+ \quad (7.39)$$

where

$$\Delta(z) = \phi_{mm}(z) + \phi_{nn}(-z)^*$$

$$\Gamma(z) = \phi_{md}(z)$$

$$\Delta^+(z) = \Delta^-(z^{-1})$$

$$z = e^{-s} \frac{\pi}{\omega_c}$$

* $\phi_{nn}(-z)$ rather than $\phi_{mm}(z)$ is correct because the samples alternate in sign.

This equation is simply the discrete version of Eq. 7.7 for a Wiener type optimum filter. Readers who are not familiar with this z transform notation or sampled data theory should refer to any sampled data text or Reference 7 by DeRusso.

Once $H_o(z)$ is known, the minimum mean square error at the sampling time, $t = nT$, is also known. It is

$$\overline{e^2(nT)} = \phi_{dd}(0) - \mathcal{L}_{t=0}^{-1} \left\{ \phi_{dm}(z) H_o(z) \right\} \quad (7.40)$$

where $\mathcal{L}_{t=0}^{-1}$ means the inverse Laplace or z transform of the positive time portion of the bracketed expression evaluated at $t = 0$.

When the output of the discrete filter is held between sample times, the mean square error increases from a minimum at the sample time to a maximum immediately before the next sample. At a time $t = nT + \Delta$ the equation for the mean square error is

$$\overline{e^2(nT + \Delta)} = \phi_{dd}(0) - 2 \int_{m=0}^{\infty} \phi_{md}(mT + \Delta) h_o(mT) + \int_{m=0}^{\infty} \int_{n=0}^{\infty} h_o(mT) h_o(nT) [\phi_{mm}(m-n)T + (-1)^{m-n} \phi_{nn}(m-n)T] \quad (7.41)$$

where $T = \frac{\pi}{\omega_c}$. Therefore if the following two conditions are true

$$\begin{aligned} \phi_{md}(z) &= \phi_{mm}(z) G_d(z) \\ \phi_{mm}(\tau) &= M^2 e^{-b|\tau|} \end{aligned}$$

the mean square error at time $t = nT + \Delta$ is

$$\overline{e^2(nT+\Delta)} = \phi_{dd}(0) - (2e^{-b\Delta} - 1) \mathcal{L}_{t=0}^{-1} \{ \phi_{md}(z) H_o(z) \}. \quad (7.42)$$

Therefore the time average minimum mean square error can be obtained by averaging Eq. 7.42 over the interval $0 \leq \Delta \leq \frac{\pi}{\omega_c}$ and this average is approximately equal to

$$\overline{e^2(t)} = \phi_{dd}(0) - \left(1 - \frac{b\pi}{2\omega_c}\right) \mathcal{L}_{t=0}^{-1} \{ (\phi_{dm}(z) H(z)) \}. \quad (7.43)$$

In this manner an upper bound on the time average of the minimum mean square error can be obtained in a free configuration suppressed carrier problem. However, this bound will be unnecessarily high when the spectrum of $n(t)$ is flat over a frequency band much larger than the carrier frequency because the sampler cannot tell the difference between noise occurring at a frequency ω_c and noise occurring at odd harmonics of ω_c . Therefore in order to improve the bound it is necessary to add a low pass filter in the noise channel (see Fig. 7.6) with a cutoff frequency larger than the carrier frequency. This filter, if located in front of $g_o(t, \tau_1)$, would greatly increase the calculations required to obtain a solution but would have a negligible effect on the spectra of the modulated message and quadrature if the cutoff frequency is larger than the carrier frequency plus the bandwidth of the modulated signals. Therefore it can be placed only in the noise channel. since we are only interested in bounding the answer.

7.3.6 Lower Bound on the Minimum Mean Square Error

A lower bound on the minimum mean square error in the free configuration suppressed carrier problem can be obtained by using an ideal demodulator like the one shown in Fig. 7.7. This demodulator consists of a sinusoidal sensitivity function and an ideal bandpass filter which eliminates all harmonics resulting from the modulated $m(t)$ and $q(t)$ signals. Following the demodulator is a linear time-invariant optimum filter which is designed in the conventional manner to minimize the time average of the ensemble mean square error. The equation for this optimum filter is

$$G_o(s) = \frac{1}{\Delta_+^+(s)} \left[\frac{\Gamma(s)}{\Delta^-(s)} \right]_+ \quad (7.44)$$

where

$$\Delta(s) = \phi_{mm}(s) + \phi_{nn}(s+j\omega_c) + \phi_{nn}(s-j\omega_c)$$

$$\Gamma(s) = 2\phi_{md}(s)$$

The resulting time average mean square error is

$$\overline{e^2(t)} = \phi_{dd}(0) - \int_{t=0}^{-1} \left\{ \frac{G_o(s) \phi_{md}(s)}{2} \right\} \quad (7.45)$$

and it establishes a lower bound on the time average of the minimum mean square error in the free configuration suppressed carrier problem. An example of the lower and upper bounds that one obtains in suppressed carrier problems is worked out in the following section.

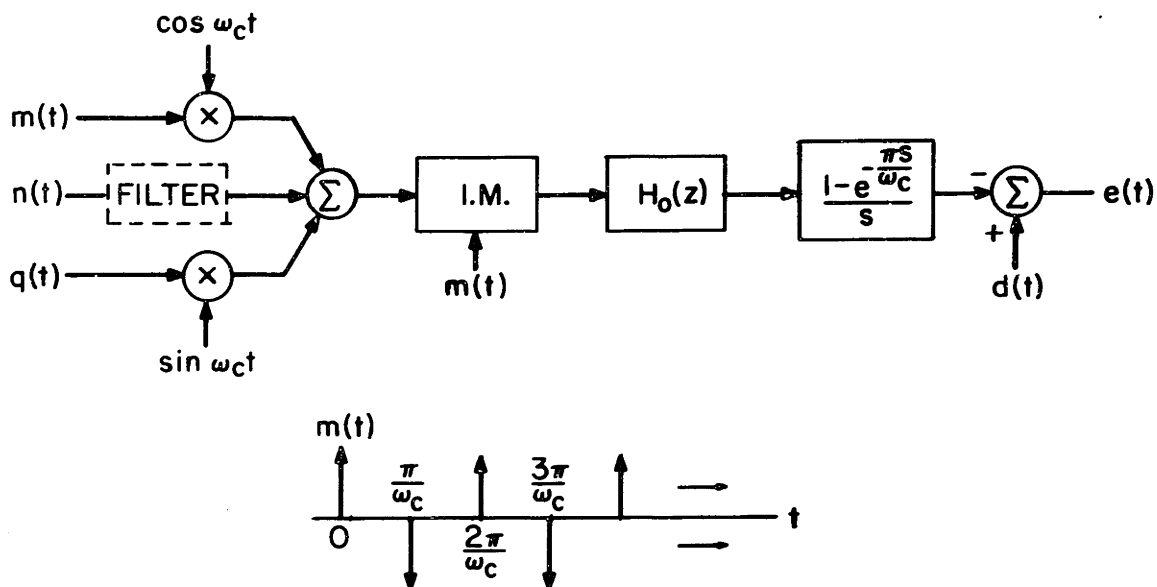


Fig. 7.6 Sampling Demodulator and Optimum Discrete Filter

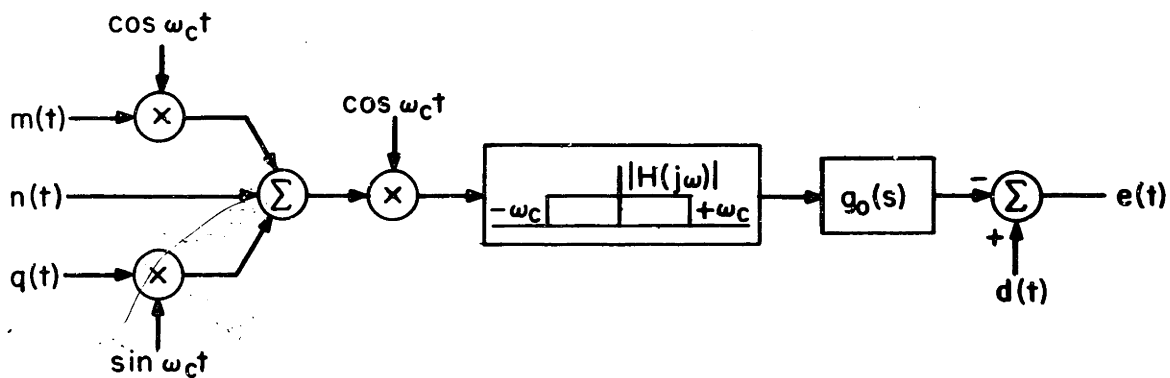


Fig. 7.7 Ideal Demodulator and Optimum Linear Filter

7.3.7 Examples of the Developed Methods

Example 1. The example which is examined first in the section is the suppressed carrier filtering problem when there is no unmodulated noise at the input and the autocorrelation functions of the message and quadrature signals are exponential. In the special case of this problem, when message and quadrature autocorrelation functions are equal, the optimum filter can be determined by applying Eqs. 7.16 - 7.18. The steps of this procedure are outlined below.

$$\phi_{mm}(\tau) = \phi_{qq}(\tau) = e^{-|\tau|}$$

$$\Phi_{mm}(s) = \frac{-2}{(s-1)(s+1)}$$

$$\Delta(s) = 2 \operatorname{Re} \left\{ \frac{-2}{(s-1-\omega_c)(s+1-j\omega_c)} \right\} = \frac{-4(s+a)(s-a)}{(s+1+j\omega_c)(s+1-j\omega_c)(s-1+j\omega_c)(s-1-j\omega_c)}$$

$$\text{where } a^2 \triangleq \omega_c^2 + 1$$

$$\Phi_{md}(s) = \Phi_{mm}(s)$$

$$G_{O_c}(s) = \operatorname{Re} \left\{ \frac{(s+1+j\omega_c)(s+1-j\omega_c)}{-2(s+a)} \left[\frac{2(s-1+j\omega_c)(s-1-j\omega_c)}{(s-1-j\omega_c)(s+1-j\omega_c)(s-a)} \right]_+ \right\} = 1$$

$$G_{O_s}(s) = I_m \{ \text{Same} \} = + \frac{\omega_c(s-a)}{(1+a)(s+a)}$$

$$\overline{e^2(t)} = \frac{2}{1+a} \sin^2 \omega_c t$$

$$\overline{e^2(t)} = \frac{1}{1+\sqrt{1+\omega_c^2}} = \frac{1}{\omega_c}, \quad \omega_c \gg 1$$

Although this is a relatively simple example, several important conclusions can be drawn from the answers. The most important of these conclusions are listed on the following page.

1. The optimum LTV filter is not impractical to build because it consists of two multipliers and a phase shifter similar to the one shown in Fig. 5.2.
2. The way in which the optimum LTV filter minimizes the ensemble mean square error is by cancelling out the harmonic errors from the message and quadrature signals by using two demodulation channels. Although this demodulation scheme will also work when $\phi_{mm}(\tau) \neq \phi_{qq}(\tau)$, the mean square error will not be minimum because it is possible to do better by using higher harmonic terms in the expansion of Eq. 7.34. This will become apparent when the bounds on the minimum mean square error in the present problem are determined.
3. The time average of the minimum mean square error is not small and the major errors are due to harmonics from the message and quadrature signals which are not perfectly cancelled. The errors due to the quadrature signal is approximately $\frac{\phi_{qq}(0)}{2}$ for $\omega_c > 1$ because

$$\phi_{qq}(0) \int_0^{\infty} d\tau_1 \int_0^{\infty} d\tau_2 e^{-|\tau_2 - \tau_1|} g_0(t, \tau_1) g_0(t, \tau_2) = \frac{\phi_{qq}(0)}{2}$$

Therefore if this same filter is used in a problem in which the message and quadrature are related as

$$\phi_{qq}(\tau) = k\phi_{mm}(\tau)$$

then the time average mean square error will be approximately

$$\overline{e^2(t)} \approx \frac{k+1}{\omega_c}$$

4. The optimum filter allows some quadrature cross coupling in order to minimize the mean square error. Although this cross coupling is small in the present problem because

$$G_o(j\omega_c) = - \frac{\omega_c}{1+a} \angle \left(\frac{\pi}{2} + \frac{1}{2\omega_c} \right)$$

it demonstrates one of the major limitations of the ensemble mean square error criterion in suppressed carrier problems. This limitation is that harmonic errors are weighted the same as noise errors, quadrature cross-coupled errors and message distortion errors. Therefore in some problems the mean square error criterion may not be a good error criterion to use because harmonic errors which occur far outside the message spectrum are generally not nearly as troublesome as the errors that appear at the output at the same frequencies as the message signals.

Let us now examine the bounds that can be put on the time average minimum mean square error in the present problem. Obviously the lower bound is 0 and when no quadrature signals are present this performance can be realized by using an optimum filter with an impulse response

$$g_o(t, \tau_1) = -2 \delta(\tau_1) \sum_{n=1}^{\infty} (-1)^{2n-1} \cos(2n-1)\omega_c t$$

This performance is considerably better than that of the optimum filter determined for the special case where $\phi_{mm}(\tau) = \phi_{qq}(\tau)$ because with that filter the mean square error is approximately $\frac{1}{2\omega_c}$ when no quadrature signals are present.

The upper bound on the time average mean square error can be determined by using sampled data theory. Since the input is assumed to be free of noise

* The symbol \angle is used to represent the phase angle of $G_o(s)$ when $s = j\omega_c$.

the optimum filter in this case is a simple sample and hold circuit and

$$H_o(z) = 1$$

$$g_o(t, \tau) = \mathcal{L}^{-1} \left\{ \frac{1-z}{(1+z)^s} \right\}, \quad z = e^{-\frac{\pi}{\omega_c} s}$$

At the discrete sampling times the mean square error is zero and at the time Δ away from the sampling time the mean square error is

$$\overline{e^2(\Delta)} = 2 \left[1 - e^{-\Delta} \right], \quad 0 \leq \Delta \leq \frac{\pi}{\omega_c}$$

Therefore by averaging over Δ , the time average mean square error is found to be

$$\overline{e^2(t)} = 2 \left[1 + \frac{1 - e^{-\frac{\pi}{\omega_c}}}{\frac{\pi}{\omega_c}} \right] = \frac{\pi}{\omega_c}, \quad \omega_c \gg 1$$

This indicates that the bound is off by a factor of approximately π in the special case when the autocorrelation function of the message and quadrature signals are equal. However, when the quadrature signals are very much larger than the message signals the bound becomes much better and the optimum filter approaches a sample and hold demodulator.

In conclusion, it has been shown that the time average of the ensemble mean square error in a suppressed carrier problem in which

$$\phi_{nn}(\tau) = 0$$

$$\phi_{md}(\tau) = \phi_{mm}(t) = e^{-|\tau|}$$

is between the limits

$$0 \leq \overline{e^2(t)} \leq \frac{\pi}{\omega_c}, \quad \omega_c \geq 1$$

where the lower bound represents the case where the quadrature signals are much smaller than the message signals and the upper bound represents the

case where the quadrature signals are much larger than the message signals. In between these two extremes when the spectra of the message and quadrature signals are equal, the time average of the minimum mean square error is approximately ω_c^{-1} for $\omega_c > 1$. In situations other than these three, it is necessary to use the discrete approximation method of Section 7.3.4 to determine the exact answer.

Example 2. The second example which is examined in this section is the suppressed carrier filtering problem when there is gaussian white noise at the input and the autocorrelation functions of the message and quadrature signals are exponential. The purpose of this relatively simple problem is to demonstrate the techniques for bounding the error and also to remove the impression which might have been obtained from the pervious example that the algebraic manipulations required to obtain the exact answer and the upper bound are relatively simple.

Let us first consider the problem in which the spectra of the message and quadrature signals are equal and the mean square value of the noise is N^2 .

$$\phi_{mm}(s) = \phi_{qq}(s) = \phi_{md}(s) = \frac{-2}{(s-1)(s+1)}$$

$$\phi_{nn}(s) = N^2$$

$$\Delta(s) = 2\text{Re} \left\{ \phi_{mm}(s-j\omega_c) + N^2 \right\} = \frac{2N^2(s+c+jd)(s+c-jd)(s-c+jd)(s-c-jd)}{(s+1+j\omega_c)(s+1-j\omega_c)(s-1+j\omega_c)(s-1-j\omega_c)}$$

where $d^2 - c^2 \triangleq \omega_c^2 - \left(1 + \frac{1}{N^2}\right)$

$$d^2 + c^2 \triangleq (1 + \omega_c^2) \left(1 + \omega_c^2 + \frac{2}{N^2}\right)$$

$$G_{O_c}(s) + j G_{O_s}(s) = \frac{2(1-j\omega_c)(s+1+j\omega_c)}{N^2 \left[(1+c)^2 + (d^2 - \omega_c^2) - 2j\omega_c(1+d) \right] (s+c+jd)(s+c-jd)}$$

When one attempts to find each of the optimum filters and then determine the minimum mean square error it soon becomes apparent that even in the special

case in which $\phi_{mm}(\tau) = \phi_{qq}(\tau)$, the time required to obtain exact solutions to actual problems is long. Therefore let us examine the bounds that can be put on the time average of the minimum mean square error.

The upper bound on the time average of the ensemble mean square error in the present problem can be determined by using sampled data theory.

$$\phi_{mm}(\tau) = e^{-|\tau|}$$

$$\phi_{mm}(z) = \sum_{n=-\infty}^{\infty} \phi_{mm}(n\tau) z^n = \frac{1-e^{-2\tau}}{(1-e^{-\tau}z^{-1})(1-e^{-\tau}z)}$$

$$\phi_{nn}(\tau) = \frac{N^2 f}{2} e^{-f|\tau|}$$

where f is the bandwidth of the noise

$$\phi_{nn}(-z) = \sum_{n=-\infty}^{\infty} \phi_{nn}(n\tau)(-1)^n z^n = \frac{N^2 f(1-e^{-2f\tau})}{2(1+e^{-f\tau}z^{-1})(1+e^{-f\tau}z)}$$

$$\Delta(z) = \frac{N^2 \gamma(1-e^{-g\tau}z^{-1})(1-e^{-g\tau}z)}{(1-e^{-\tau}z^{-1})(1-e^{-\tau}z)(1+e^{-f\tau}z^{-1})(1+e^{-f\tau}z)}$$

where $\gamma \triangleq \frac{N^2 f e^{-\tau}(1-e^{-2f\tau}) - e^{-f\tau}(1-e^{-2\tau})}{2N^2 e^{-f\tau}}$

$$(e^{-g\tau} + e^{+g\tau}) \triangleq \frac{N^2 f(1-e^{-2f\tau})(1+e^{-2\tau}) + 2(1-e^{-2\tau})(1+e^{-2f\tau})}{N^2 f e^{-\tau}(1-e^{-2f\tau}) - e^{-f\tau}(1-e^{-2\tau})}$$

$$H_o(z) = \frac{(1-e^{-2\tau})(1-e^{-(f+1)\tau})(1-e^{-f\tau}z)}{N^2 (1-e^{-(g+1)\tau})(1-e^{-g\tau}z)}$$

When one attempts to determine $H_o(z)$ and $\overline{e^2(t)}$ for various values of N^2 , f and ω_c it soon becomes apparent that the time required to obtain an upper bound on the time average minimum mean square error in these problems is also long. Therefore let us examine the lower bound.

The lower bound on the mean square error in this same problem can be determined by using the method of Section 7.3.7. The steps of the procedure are outlined below

$$\phi_{mm}(s) = \frac{-2}{(s+1)(s+1)} = \phi_{md}(s)$$

$$\phi_{nn}(s) = N^2$$

$$\Delta(s) = \frac{2N^2(s-b)(s+b)}{(s+1)(s-1)}$$

where $b^2 \triangleq 1 + \frac{1}{N^2}$

$$G_o(s) = \frac{(s+1)}{2N^2(s+b)} \frac{-4}{(s+1)(s-b)} + \frac{2}{N^2(b+1)(s+b)}$$

$$\overline{e^2(t)} = 1 - \int_{t=0}^{-1} \frac{-2}{N^2(b+1)(s+b)(s+1)(s-1)} = \frac{2}{b+1}$$

It should now be apparent that this method is by far the easiest of the methods examined in this chapter to apply to actual problems. Also the filter time constant is approximately equal to the time constant of the optimum filter ($b = c$). However, it must be kept in mind that the answer is only a lower bound on the minimum mean square error and not an exact answer. Furthermore this method completely neglects harmonic errors which is undesirable in some problems but desirable in many others.

7.3.8 Summary

All of the mathematical "tools" required to solve either exactly or approximately the problem of minimizing the ensemble mean square error in suppressed carrier problems have now been presented and all except the finite memory discrete approximation method have been demonstrated.

It was shown in the second example of Section 7.3.7 that determining the absolute minimum mean square error or an upper bound on it is a time consuming task even in the special case when $\phi_{mm}(\tau) = \phi_{qq}(\tau)$. It was also shown that the ensemble mean square error criterion may not necessarily be the best criterion for some of the suppressed carrier problems of interest because the harmonic errors in suppressed carrier problems are large and the error criterion weights them equally with the cross-coupled errors, the errors from the unmodulated noise and distortion errors. Unfortunately it is impossible to change the weighting of the various errors in the exact method of analyzing the problem. However, in the approximate methods presented for bounding the time average of the ensemble mean square error, the harmonic errors are neglected or weighted with 0 which is the other extreme. In the vibratory gyroscope problem and many other suppressed carrier problems this latter weighting is preferable to the former because the harmonic errors which occur at even harmonics of the carrier frequency are relatively unimportant compared to the errors which occur at the output at the same frequencies as the message signal. Therefore in the following sections the method developed for establishing the lower bound on the time average minimum mean square error is improved so that it can be used to determine a "practical", approximately optimum, LTV filter and the time average mean square error associated with it when the harmonic errors from the message and quadrature signals are neglected.

7.4 PROBLEM WITH FIXED ELEMENTS LOCATED BEFORE THE SIGNAL PROCESSOR

7.4.1 Unrestricted Optimum Filter Problem

When a linear time-invariant fixed element is located in front of the optimum LTV filter as shown in Fig. 7. 8 , the integral equation which must be solved for the optimum filter is

$$\int_0^{\infty} d\xi_2 g_{f_1}(\xi_2) \phi_{md}(\tau_2 - \xi_2) \cos \omega_c(t - \tau_2 - \xi_2) = \quad (7.46)$$

$$\int_0^{\infty} d\xi_2 \int_0^{\infty} d\xi_1 \int_0^{\infty} d\tau_1 g_{f_1}(\xi_1) g_{f_1}(\xi_2) g_o(t, \tau_1) [\phi_{11}(x) + \phi_{22}(x) \cos \omega_c(2t - \tau_1 - \tau_2 - \xi_1 - \xi_2)]$$

where $\phi_{11}(x) = [\phi_{mm}(x) + \phi_{qq}(x)] \cos \omega_c x + 2\phi_{nn}(x)$

$$\phi_{22}(x) = [\phi_{mm}(x) - \phi_{qq}(x)] \cos \omega_c x$$

$$x = \tau_2 + \xi_2 - \tau_1 - \xi_1$$

where $g_{f_1}(\xi_1)$ is the impulse response of the fixed element. The associated minimum mean square error is

$$\overline{e^2(t)}_{\min} = \phi_{dd}(0) - \int_0^{\infty} d\tau_1 \int_0^{\infty} d\xi_1 g_{f_1}(\xi_1) g_o(t, \tau_1) \phi_{md}(\tau_1 - \xi_1) \cos \omega_c(t - \tau_1 - \xi_1) \quad (7.47)$$

As long as the fixed element is ideal, realizable and minimum phase (all poles and zeros in the right half of the s plane) and the optimum LTV filter is unrestricted, a fixed element does not impose any limitations on the system because the optimum filter simply cancels out the poles and zeros with equal zeros and poles so that the problem returns to the free configuration problem which was examined in Section 7.3. When the fixed element is non-minimum phase or a time delay this cancellation requires that the optimum filter have poles in the left half of the s plane or prediction capability--both of which are not allowable. Therefore when these fixed elements are in a system they cause the minimum mean square error to be larger than it would be in a similar but free configuration system.

When the fixed element is a time delay, all the methods presented in Section 7.3 for determining $g_o(t, \tau_1)$ either exactly or approximately and the methods for bounding the time average of the minimum mean square error are valid if $\phi_{md}(\tau)$ is replaced by $\phi_{md}(\tau + t_d)$ where t_d is the delay time. When the fixed element is non-minimum phase and $\phi_{mm}(\tau) = \phi_{qq}(\tau)$ the optimum filters can be determined by using Eqs. 7.17 and the following expressions which have a subscript ₁ instead of Δ, Γ_c and Γ_s .

$$\begin{aligned} \Delta_1^+(s) \Delta_1^-(s) &= \Delta_1(s) = G_{f_1}(s) G_{f_1}(-s) \Delta(s) \\ \Gamma_{s_1}(s) &= G_{f_1}(-s) \Gamma_s(s) \\ \Gamma_{c_1}(s) &= G_{f_1}(-s) \Gamma_c(s) \end{aligned} \quad (7.48)$$

where $G_{f_1}(s)$ is the Laplace transform of $g_{f_1}(\xi_1)$. When $\phi_{mm}(\tau) \neq \phi_{qq}(\tau)$ and there is a non-minimum phase fixed element $g_{f_1}(\xi_1)$, it is necessary to approximate the optimum filter by a discrete filter with a finite memory and solve a set of algebraic equations in order to determine $g_o(t, \tau_1)$ and the minimum mean square error. The equations are not presented here because they are easily derived from Eqs. 7.26 and 7.46 and they have questionable practical value because of the complexity of the solution and the lack of actual problems which fall into this classification. Usually the non-minimum phase fixed elements in suppressed carrier problems are caused by some kind of time delay and are best treated as such.

7.4.2 Restricted Optimum Filter Problem

A problem which arises quite often in suppressed carrier systems which involves fixed elements located before the signal processor is the following. A

suppressed carrier signal, $f_1(t)$, is passed through a bandpass filter in which there are a fixed number of poles and zeros with adjustable numerical values. The output of this filter is then demodulated by a demodulator that consists of a multiplier with a sensitivity function $p(t)$ followed by a linear time-invariant filter, as shown in Fig. 7.9. In some problems the final filter and sensitivity function may assume any form while in others--such as the AC servomechanism problem in which a two phase servo motor is the demodulator--either or both may be specified. The problem is to optimize the system by properly choosing whatever is not fixed so that the time average of the ensemble mean square error or some other similar error criterion is minimized.

If the sensitivity function of the demodulator is defined as

$$p(t) \triangleq \sum_{n=1}^{\infty} a_{2n-1} \cos [(2n-1) \omega_c t - \psi] \quad (7.49)$$

where ψ is the phase angle between the $\cos \omega_c t$ modulating signal and the sensitivity function, the following equation can be written for the time average of the ensemble mean square error..

$$\overline{e^2(t)} = \phi_{dd}(0) - \int_0^{\infty} d\tau_2 \int_0^{\infty} d\xi_2 \phi_{md}(\tau_2 - \xi_2) g_{f_1}(\xi_2) g_o(\tau_2) a_1 \cos(\omega_c(\xi_2 - \psi) +$$

$$\frac{1}{2} \int_0^{\infty} d\tau_1 d\tau_2 d\xi_1 d\xi_2 g_o(\tau_1) g_o(\tau_2) g_{f_1}(\xi_1) g_{f_1}(\xi_2) \left\{ \begin{array}{l} \phi_{11}(x) \sum_{n=1}^{\infty} a_{2n-1}^2 \cos(2n-1)\omega_c(\tau_2 - \tau_1) - \\ \phi_{22}(x) \left\{ \begin{array}{l} \frac{1}{2} a_1^2 \cos\{\omega_c(\xi_2 + \xi_1) - 2\psi\} \\ \sum_{n=1}^{\infty} a_{2n-1} a_{2n+1} \cos 2n\omega_c(\tau_2 - \tau_1) \cdot \\ \cos \omega_c(\xi_2 + \xi_1) \end{array} \right\} \end{array} \right\}$$

(7.50)

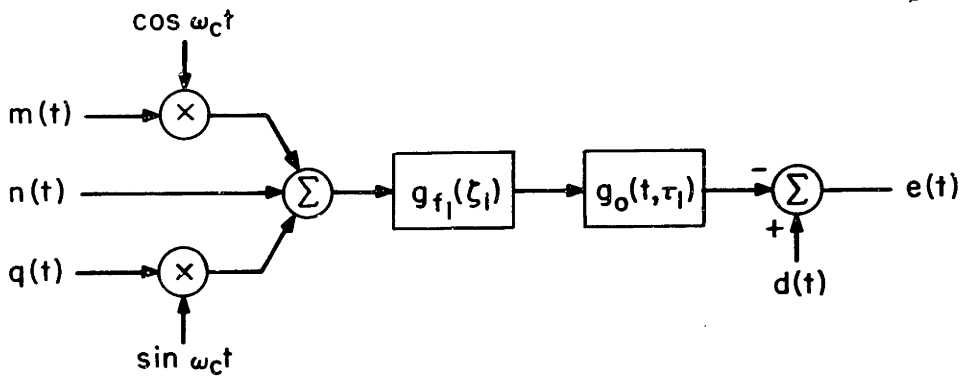


Fig. 7.8 Fixed Element Before the Signal Processor

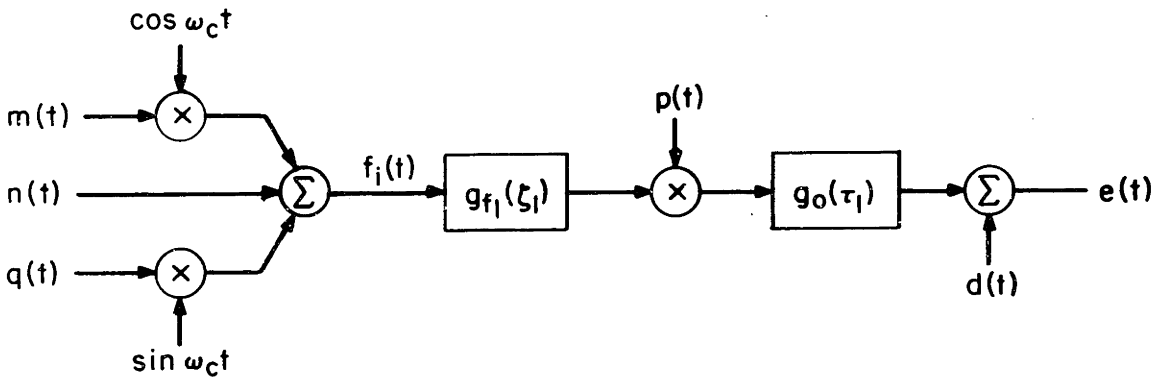
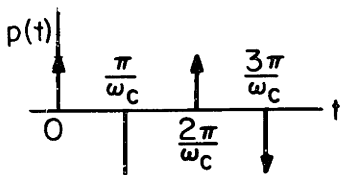
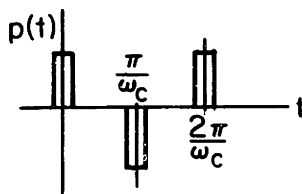


Fig. 7.9 Fixed Element Before a Restricted Signal Processor

Sensitivity function of an ideal full wave sampling demodulator



Sensitivity function of a practical full wave sampling demodulator



Square wave sensitivity function

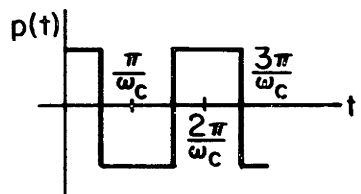


Fig. 7.10 Practical Sensitivity Functions

where

$$\begin{aligned} \phi_{11}(x) &= \frac{1}{2} [\phi_{mm}(x) + \phi_{qq}(x)] \cos \omega_c x + \phi_{nn}(x) \\ \phi_{22}(x) &= \frac{1}{2} [\phi_{mm}(x) - \phi_{qq}(x)] \\ x &= (\tau_2 + \xi_2 - \tau_1 - \xi_1) \end{aligned}$$

By using the conventional calculus of variations technique on this expression, the following integral equation can be obtained for $g_o(\tau_1)$ when it is not specified

$$\begin{aligned} &\int_0^\infty d\xi_2 g_{f_1}(\xi_2) \phi_{md}(\tau_2 + \xi_2) a_1 \cos(\omega_c(\xi_2 - \psi)) = \\ &\int_0^\infty d\xi_1 \int_0^\infty d\xi_2 \int_0^\infty d\tau_1 g_{f_1}(\xi_1) g_{f_1}(\xi_2) g_o(\tau_1) \left[\begin{array}{l} \text{same as the bracketed} \\ \text{term in Eq. 7.50} \end{array} \right] \quad (7.51) \end{aligned}$$

This equation can be solved for $g_o(\tau_1)$ in terms of the coefficients a_{2n-1} and the unknown parameters of $g_{f_1}(\xi_1)$ by using the conventional spectrum factorization approach.* Then the time average of the minimum mean square error can be determined in terms of the same unknowns and this expression minimized by choosing the optimum values for the unknowns. Unfortunately, this requires a great deal of time and patience even if n is a small number. Furthermore in most practical demodulators the coefficients a_{2n-1} are usually constrained because practical sensitivity functions are rectangular as shown in Fig. 7.10 and the fastest they fall off is in the square wave where $\frac{a_{2n+1}}{a_{2n-1}} = \frac{2n+1}{2n-1}$. Therefore let us take a step backward and away from this elegant mathematical solution and intuitively examine the sensitivity function problem. First, if $g_{f_1}(\xi_1)$ is a bandpass filter so that the spectrum of the signal into the multiplier is band limited or approximately band limited, this signal can be written as

$$f'_1(t) = [m'(t) + n'(t)] \cos(\omega_c t - \psi) + [q'(t) + n'(t)] \sin(\omega_c t - \psi) \quad (7.52)$$

* This technique cannot be used to determine the optimum $g_{f_1}(\xi_1)$ if it is completely unspecified because of the terms $\cos \omega_c(\tau_2 + \xi_1)$ in Eq. 7.50.

where $\psi \triangleq -\angle G_{f_1}(j\omega_c)$ (7.53)

and $m'(t)$, $q'(t)$, $n'(t)$ are functions of $m(t)$, $q(t)$, $n(t)$ and the bandpass filter.

Now when $f_i'(t)$ is multiplied by $p(t)$ in Eq. 7.49, the following terms result.

$$f_i'(t)p(t) = a_1[m'(t)+n'(t)] + \sum_{n=1}^{\infty} \left\{ \begin{aligned} & a_{2n-1}[m'(t) + n'(t)] + a_{2n+1}[m'(t)+n'(t)] \cos 2(n\omega_c t - \psi) \\ & + a_{2n-1}[q'(t)+n'(t)] - a_{2n+1}[q'(t)+n'(t)] \sin 2(n\omega_c t - \psi) \end{aligned} \right\}$$

(7.54)

It is now apparent that the power density spectrum of the input to the linear time-invariant filter, $g_o(\tau_1)$, has large components centered at 0 frequency and the even harmonics of the carrier frequency. Therefore the output of $g_o(\tau_1)$ also has power predominality at these frequencies. The problem is to reduce the harmonic power by the proper or optimum selection of the coefficients a_{2n-1} and the filter transfer function. When the filter transfer function contains band rejection characteristics at the harmonics this is a difficult optimization problem because it is necessary to factor the bracketed term in Eq. 7.51 and follow the previously outlined procedure. In practice, however, it is usually impractical and unnecessary to use band rejection filters at the even harmonics of the carrier frequency because the harmonic signals are usually filtered sufficiently by a simple low pass filter and the other components in the system. When this is the case, the coefficients a_{2n-1} should be chosen to minimize the time average mean square harmonic error at each of the harmonics. This is accomplished when the coefficients are related in the following manner.

$$\frac{a_{2n+1}}{a_{2n-1}} = - \frac{\phi'_{mm}(0) - \phi'_{qq}(0)}{\phi'_{mm}(0) + \phi'_{qq}(0) + 2\phi'_{nn}(0)} = A = \frac{\alpha - 1}{\alpha + 1} \quad (7.55)$$

where

$$\alpha \triangleq \frac{\phi'_{qq}(0) + \phi'_{nn}(0)}{\phi'_{mm}(0) + \phi'_{nn}(0)}$$

When the Fourier coefficients of the sensitivity functions are related in this manner, the time average of the mean square harmonic error from the bracketed term in Eq. 7.54 is minimized and the sensitivity function can be written in a closed form as

$$p(t) = \frac{2(1-A) \cos \omega_c t}{(1+A^2) - 2A \cos 2\omega_c t} = \frac{2\alpha(1+\alpha) \cos \omega_c t}{(1+\alpha^2) + (1-\alpha^2) \cos 2\omega_c t} \quad (7.56)$$

This is an approximately optimum sensitivity function for problems of this type in which the input signal to the signal processor is bandlimited and the filter following the sensitivity function does not have band rejection characteristics at even harmonics of the carrier frequency.

In a special class of problems of this type in which $\phi'_{qq}(T) = \gamma \phi'_{mm}(T)$ and $\phi'_{nn}(T) = 0$, $p(t)$ is the optimum sensitivity function. This class of problems has been analyzed by Booten and Goldstein⁸ in some detail; however, they did not obtain the explicit solution for $p(t)$ which is given by Eq. 7.56 when $\alpha = \gamma$. In this special case the sensitivity function approaches an impulse train in which the impulses alternate in sign when $\gamma \gg 1$ and when $\gamma \ll 1$ it approaches $(\cos \omega_c t)^{-1}$. In between these extremes, when $\gamma = 1$, it is equal to $\cos \omega_c t$ as Booten and Goldstein pointed out.

Now that the sensitivity function problem is understood, it is necessary to return to Eq. 7.52 and examine the effect of the bandpass filter on $\phi'_{mm}(0)$, $\phi'_{qq}(0)$ and $\phi'_{nn}(0)$. If the noise $n(t)$ is flat around the carrier frequency and the bandpass filter has an adjustable bandwidth but is constrained such* that the gain bandwidth product is approximately constant, $\phi'_{nn}(0)$ is proportional to product of the filter time constant and $\Phi_{nn}(j\omega_c)$ while $\phi'_{mm}(0)$ and $\phi'_{qq}(0)$ are proportional to the product of the square of the filter time constant and $\phi_{mm}(0)$ and $\phi_{qq}(0)$ respectively as long as the filter bandwidth is greater than the bandwidths of the power density spectra of the message and quadrature signals. Therefore, A , a and $p(t)$ are functions of the bandwidth of $g_{f_1}(\xi_1)$ so the sensitivity function cannot be determined independently of $g_{f_1}(\xi_1)$. However, if the filter following the sensitivity function does not have band rejection characteristics, the harmonic errors generally are much less important than the other errors in the problem. When this is true the optimum parameters of $g_{f_1}(\xi_1)$ and the optimum filter $g_o(\tau_1)$ can be approximated by using the method of section 7.4.4 and then an approximately optimum sensitivity function can be determined. It must be kept in mind that these two methods are only applicable in problems in which $g_o(\tau_1)$ does not have band rejection characteristics and the input $f_1'(t)$ is band limited.

* See Eq. 4.9 and associated discussion.

When the filter $g_o(\tau_1)$ contains band rejection elements, $p(t)$ is not the optimum sensitivity function but only an approximation to it. However, the determination of the optimum sensitivity function and the optimum filter is a ~~horrendous~~ problem in this case. Therefore let us examine the bounds that can be put on the minimum mean square error by using the techniques developed in Section 7.3.5 and 7.3.6.

7.4.3 Upper Bound on the Minimum MSE in the Restricted Problem

An upper bound on the minimum MSE in the restricted problem of Fig. 7.9 can be determined by using sampled data theory and assuming the output of $g_{f_1}(\xi_1)$ is sampled twice each cycle of the carrier and then passed through an optimum discrete filter and held between sampling pulses. The equation which must be solved for the optimum discrete filter is

$$H_o(z) = \frac{1}{\Delta^+(z)} \left[\frac{\Gamma(z)}{\Delta^-(z)} \right] + \quad (7.57)$$

where $\Delta(z) = \phi_{mm}(z) G_{f_1}^r(z) G_{f_1}^r(z^{-1}) + \phi_{qq}(z) G_{f_1}^i(z) G_{f_1}^i(z^{-1}) + \phi_{nn}(-z) G_{f_1}(z) G_{f_1}(z^{-1})$

$$\Gamma(z) = \phi_{md}(z) G_{f_1}^r(z)$$

$$G_{f_1}^r(z) = [\text{Re } e^{-j\psi} G_{f_1}(s - j\omega_c)]^*$$

$$G_{f_1}^i(z) = [\text{Im } e^{-j\psi} G_{f_1}(s - j\omega_c)]^*$$

* Means the z Transform of the Bracketed Term

$$z = e^{-sT}, \quad T = \frac{\pi}{\omega_c}$$

These relations can be determined by representing an impulse train with pulses that alternate in sign as

$$i(t) = \sum_{n=1}^{\infty} \cos(2n-1) \omega_c t \quad (7.58)$$

and an impulse train with pulses that are all positive as

$$i(t) = \frac{1}{2} + \sum_{n=1}^{\infty} \cos 2n\omega_c t \quad (7.59)$$

and then applying trigonometric identities to the sampled output of $g_{f_1}(\xi_1)$.

The minimum mean square error associated with the optimum discrete filter at the sample times is

$$\overline{e^2(nT)} = \phi_{dd}(0) - \sum_{t=0}^{-1} \phi_{dm}(z) G_{f_1}^r(z) H_d(z) \quad (7.60)$$

and the time average MSE can be determined in the same manner as in Section 7.3.5 by averaging over the period between samples. This number will be in terms of the adjustable coefficients of the bandpass filter. These coefficients can then be chosen to minimize this upper bound on the time average mean square error. A meaningful minimum will usually exist because as the bandwidth of the bandpass filter increases the errors due to overlapping spectra increase while the errors due to the cross coupling between $m(t)$ and $q(t)$ decrease. When a sampling demodulator is used, which is quite often in practice, the

optimum parameters for the bandpass filter and the minimum time average mean square error are known. When other types of demodulators are used, these parameters for the bandpass filter will not be exactly optimum but not too far from it as long as the input signal to the demodulator is band limited. Since Eqs. 7.57 and 7.60 are generally orders of magnitude easier to solve than Eq. 7.51, these small errors may be overlooked if one is only seeking an approximate answer.

7.4.4 Lower Bound on the Minimum MSE in the Restricted Problem

The lower bound on the minimum MSE in the restricted problem of Fig. 7.9 can be determined by using the ideal demodulator approach examined in Section 7.3.6 and shown in Fig. 7.6. The equation which must be solved for the optimum linear time-invariant filter which follows the demodulator is

$$G_o(s) = \frac{1}{\Delta^+(s)} \left[\frac{F(s)}{\Delta^-(s)} \right]_+ \quad (7.61)$$

where
$$\Delta(s) = \phi_{mm}(s) G_{f_1}^r(s) G_{f_1}^r(-s) + \phi_{dd}(s) G_{f_1}^i(s) G_{f_1}^i(-s) +$$

$$2\text{Re} \{ \phi_{mn}(s+j\omega_c) [G_{f_1}(s-j\omega_c) G_{f_1}(-s+j\omega_c)] \}$$

$$F(s) = 2\phi_{md}(s) G_{f_1}^r(s)$$

$$G_{f_1}^r(s) = \text{Re} [e^{-j\psi} G_{f_1}(s+j\omega_c)]$$

$$G_{f_1}^i(s) = \text{Im} [e^{-j\psi} G_{f_1}(s+j\omega_c)]$$

and the time average minimum mean square error associated with the filter is

$$\overline{e^2(t)} = \phi_{dd}(0) - \frac{1}{2} \int_{t=0}^{-1} \phi_{dm}(s) G_{f_1}^r(s) G_o(s) \quad (7.62)$$

The two transfer functions $G_{f_1}^r(s)$ and $G_{f_1}^i(s)$ are the envelope transfer functions for the message and quadrature signals respectively and are functions of $G_{f_1}(s)$. The minimum mean square error and $G_0(s)$ are also functions of $G_{f_1}(s)$ and it is necessary to choose the variable parameters of $G_{f_1}(s)$ so that the value of $\overline{e^2(t)}$ determined from Eq. 7.62 is minimum. Because of the nature of the approximation, however, this minimum will often occur when the bandwidth of $G_{f_1}(s)$ is as large as possible because the cross-coupled errors decrease as the filter bandwidth increases while the errors from the noise $n(t)$ can usually be filtered equally well before or after the sinusoidal sensitivity functions. This is not true in all problems though. In problems such as the vibratory gyroscope problem which is examined in the following chapter, wide band noise enters after the bandpass filter so that the preamplification of the message signals with respect to this noise is proportional to the bandwidth of the bandpass filter. In problems of this type the optimum parameters for $G_{f_1}(s)$ are not those associated with the largest bandwidth.

The difficulty which has just been encountered arises because overlapping noise spectra are not troublesome when a sinusoidal sensitivity function is used while they are troublesome when practical sensitivity functions are used. For this reason the lower bound on the minimum mean square error which is obtained by using Eqs. 7.61 and 7.62 is low and the bandwidth of the bandpass filter may be larger than optimum. Fortunately it is possible to improve this bound in a relatively easy manner by adding more noise to the expression for $\Delta(s)$ in Eq. 7.61. The exact expression which should be used for the noise in Eq. 7.61 is the noise term in the brackets of Eq. 7.51. The time domain expression for all the noise in $\Delta(s)$ is

$$\phi_{nn}(x) g_o(\tau_1) g_{f_1}(\xi_1) g_{f_1}(\xi_2) \sum_{n=1}^{\infty} a_{2n-1}^2 \cos(2n-1)\omega_c(\tau_2 - \tau_1)$$

where

$$x = \tau_2 - \tau_1 + \xi_2 - \xi_1$$

The transform of the new components of this term is

$$G_o(s) \sum_{\substack{m=\text{odd} \\ 3, 5, 7, \dots}} a_m^2 \left\{ \frac{1}{2} \phi_{nn}(s-jm\omega_c) G_{f_1}(-s-jm\omega_c) G_{f_1}(s+jm\omega_c) + \frac{1}{2} \phi_{nn}(s+jm\omega_c) G_{f_1}(-s+jm\omega_c) G_{f_1}(s-jm\omega_c) \right\}$$

where the bracketed term can be simplified as either of the following expression

$$R_e \{ \phi_{nn}(s-jm\omega_c) G_{f_1}(-s-jm\omega_c) G_{f_1}(s+jm\omega_c) \}$$

or

$$R_e \{ \phi_{nn}(s+jm\omega_c) G_{f_1}(s-jm\omega_c) G_{f_1}(-s+jm\omega_c) \}$$

as was done in Eq. 7.61. This term could now be added to $\Delta(s)$ in Eq.

7.61 but this would only result in an unnecessarily complicated solution.

Therefore, let us again step back from this elegant mathematical solution and intuitively examine the present problem.

In most practical problems $G_{f_1}(s)$ is a bandpass filter such that $G_{f_1}(s+jm\omega_c)$ is proportional to ω_c^{-1} or ω_c^{-2} for m larger than unity. Therefore, if the bandwidths of the message and quadrature signal are much smaller than ω_c , the higher harmonic noise terms can be approximated as

$\phi_{nn}(jm\omega_c) G_{f_1}(-jm\omega_c) G_{f_1}(jm\omega_c)$ and all but the first few harmonic terms dropped. Since these noise terms are just numbers it means that the only change in Eq. 7.61 is that a constant defined as N is added to $\Delta(s)$.

* The a_1 term is already in Eq. 7.61 so it is not included in the following equations.

When this is done, $G_o(s)$ and $e^2(t)$ are functions of the variable parameters of $G_{f_1}(s)$ and it is necessary to determine the values of the variable parameters which minimize the time average of the mean square error. If the coefficients a_m are specified this is a relatively easy problem. However; if the coefficients are not specified but are chosen to satisfy Eq. 7.4, it is necessary to use an iterative type of solution to determine the approximately optimum parameters of $G_{f_1}(s)$, $G_o(s)$ and a_m . Although this is a more difficult problem to solve than the one that was presented earlier in this section, it is orders of magnitude easier to solve and the solution is usually much more practical than that which is obtained by solving Eq. 7.51. However, because of the approximations made with respect to the harmonic noise, the answer can no longer be considered a bound. Instead it is an approximate solution to the problem of minimizing the time average mean square error in the restricted problem of Fig. 7.9 when harmonic errors from the message and quadrature signals are neglected. Since this error criterion is better for analyzing many problems than the conventional mean square error criterion which weights the harmonic errors equally with the other errors, it is used in the vibratory gyroscope example of Chapter VIII.

7.4.5 Summary

Most suppressed carrier problems with fixed elements located before the signal processor can be analyzed by using the methods developed in this section. The exact solution to the problem of minimizing the ensemble mean square error is difficult to obtain but an approximate solution which minimizes the time average of the ensemble mean square error but neglects harmonic errors from the message and quadrature signals

is relatively easy to obtain. No illustrative examples are included in this section because they would only duplicate the vibratory gyroscope example of Chapter VIII. Furthermore the methods used to obtain the solutions are the same as those used in the simpler examples worked out in Section 7.3 and therefore need not be demonstrated.

7.5 PROBLEM WITH FIXED ELEMENTS LOCATED AFTER THE SIGNAL PROCESSOR

When linear time-invariant fixed elements are located after the optimum filter as shown in Fig. 7.11, the conventional linear time-varying filter theory is no longer directly applicable because the output of the fixed element at time t_k is a function of the optimum LTV filter for all past time. In the problems previously studied the output was only a function of the optimum filter at time t_k so the calculus of variations could be used on the expression for the ensemble mean square error to determine an integral equation for $g_o(t_k, \tau_1)$. However, because of the periodic nature of the optimum filter in the present problem, an approximate solution can always be obtained by minimizing the time average mean square error and an exact solution can be obtained in the special case where $\phi_{mm}(\tau) = \phi_{qq}(\tau)$.

It should be obvious that when the fixed element, $g_{f_2}(\xi_1)$, is minimum phase it does not have any effect on the minimum mean square error because the optimum filter can cancel out the poles and zeros of the fixed element with equal zeros and poles and the problem returns to the free configuration problem of Section 7.3. When the fixed element is a time delay it can be handled in the same way as a time delay before the optimum filter which was discussed in Section 7.41. Therefore the only problem of interest is the one in which a non-minimum phase fixed element, such as

$$G_{f_2}(s) = \frac{s-a}{s+a}, \quad (7.63)$$

is located after the optimum LTV filter.

In the special case where $\phi_{mm}(x) = \phi_{qq}(x)$ and a sinusoidal expansion form of $g_o(t, \tau_1)$ is used, the following two integral equations must be solved for $g_{oc}(\tau_1)$ and $g_{os}(\tau_1)$ in order to minimize the time average of the ensemble mean square error.

$$\int_0^{\infty} d\xi_2 g_{f_2}(\xi_2) \phi_{md}(\tau_2 + \xi_2) \cos \omega_c \tau_2 = \frac{1}{2} \int_0^{\infty} d\xi_1 d\xi_2 d\tau_1 g_{f_2}(\xi_1) g_{f_2}(\xi_2) \phi_{11}(x) [g_{oc}(\tau_1) \cos \omega_c (\xi_2 - \xi_1) + g_{os}(\tau_1) \sin \omega_c (\xi_2 - \xi_1)] \quad (7.64)$$

$$\int_0^{\infty} d\xi_2 g_{f_2}(\xi_2) \phi_{md}(\tau_2 + \xi_2) \sin \omega_c \tau_2 = \frac{1}{2} \int_0^{\infty} d\xi_1 d\xi_2 d\tau_1 g_{f_2}(\xi_1) g_{f_2}(\xi_2) \phi_{11}(x) [-g_{oc}(\tau_1) \sin \omega_c (\xi_2 - \xi_1) + g_{os}(\tau_1) \cos \omega_c (\xi_2 - \xi_1)]$$

where $\phi_{11}(x) = \frac{1}{2} [\phi_{mm}(x) + \phi_{nn}(x)] \cos \omega_c x + \phi_{nn}(x)$

$$x = (\tau_2 + \xi_2 - \tau_1 - \xi_1) \quad (7.65)$$

These equations can be put into a form which can easily be solved by adding $\pm j$ times the second equation to the first one. The solutions of these two modified equations are

$$G_c^+(s) = \text{Re} \left\{ \frac{1}{G^+(s-j\omega_c) \Delta^+(s)} \left[\frac{\phi_{md}(s-j\omega_c) G_{f_2}(-s+j\omega_c)}{G^-(s-j\omega_c) \Delta^-(s)} \right]_+ \right\} \quad (7.66)$$

$$G_s(s) = \text{Im} \quad \{ \text{Same Term} \} \quad (7.67)$$

where $G_{f_2}(s+j\omega_c) G_{f_2}(-s-j\omega_c) = G^+(s+j\omega_c) G^-(s+j\omega_c)$
 $G_{f_2}(s-j\omega_c) G_{f_2}(-s+j\omega_c) = G^+(s-j\omega_c) G^-(s-j\omega_c)$

It is much easier to solve these equations in the form of optimum filter shown in Fig. 7.3, in which case

$$G'_c(s) = \text{Re} \left\{ \frac{1}{G^+(s) \Delta^+(s+j\omega_c)} \left[\frac{\phi_{md}(s) G_{f_2}(-s)}{G^-(s) \Delta^-(s+j\omega_c)} \right]_+ \right\} \quad (7.68)$$

$$G'_s(s) = \text{Im} \quad \{\text{Same Term}\} \quad (7.69)$$

When $g_{f_2}(\xi_2)$ is minimum phase, $G_{f_2}(-s) = G^-(s)$, and the term $[G^+(s)]^{-1}$ factors out of both equations as expected. When the fixed element is non-minimum phase this does not occur and the minimum mean square error is larger than in the free configuration problem.

In the general case when $\phi_{mm}(x) \neq \phi_{qq}(x)$ and non-minimum phase fixed elements are located after $g_o(t, \tau_1)$, it is necessary to approximate the optimum LTV filter by h discrete linear time-invariant filters and h periodic multipliers. This can be done because

$$g_o(t, \tau_1) = g_o\left(t + \frac{2\pi}{\omega_c}, \tau_1\right) = -g_o\left(t + \frac{\pi}{\omega_c}, \tau_1\right)$$

The approximate LTV filter is represented as

$$g_o(t, nT) = \sum_{k=0}^{h-1} T g_k(nT) p_k(t) \quad (7.70)$$

where $p_k(t)$ is the periodic waveform shown in Fig. 7.12,

The error at time pT can now be written as

$$e(pT) = d(pT) - \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} T [f_1(pT - nT - mT)] \sum_{k=0}^{h-1} g_f^k(mT) g_{p-k}(nT) \quad (7.71)$$

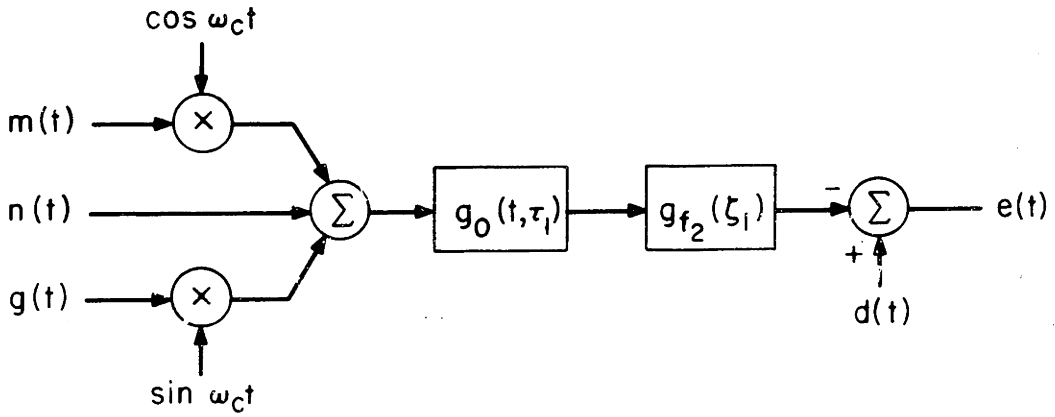


Fig. 7.11 Fixed Element Following the Signal Processor

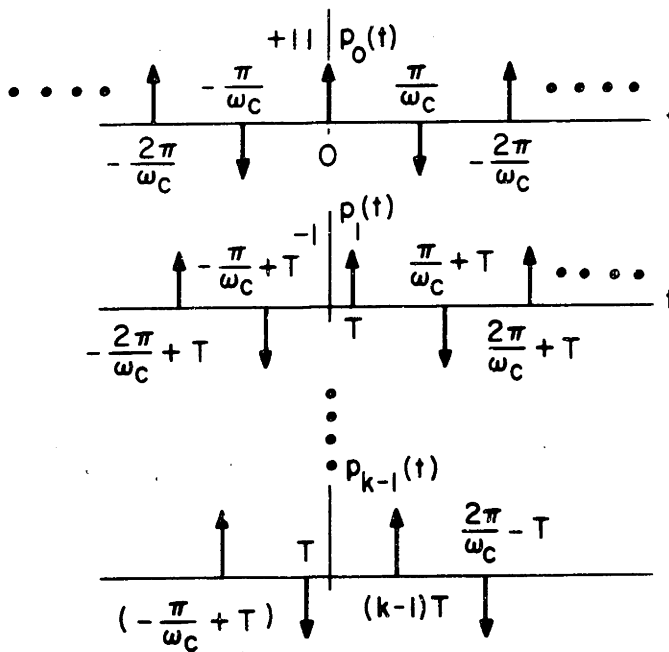


Fig. 7.12 Periodic Multipliers

where
$$g_f^k(mT) = g_f(\tau_1) p_k(\tau_1) \quad (7.72)$$

and
$$g_{-q}(nT) = -g_{n-q}(nT) \quad (7.73)$$

The time average of the ensemble mean square error can now be minimized by using the calculus of variations.

$$\frac{\partial}{\partial g_i} \left(\sum_{p=0}^{h-1} e^{2(pT)} \right) \quad i = 0, 1, \dots, h-1 \quad (7.74)$$

This differentiation produces a set of h linear matrix equations in terms of the h optimum linear time-invariant filters. These equations are derived in Appendix D. In order to determine the h optimum filters it is necessary to solve the h equations simultaneously. This can be done on a digital computer but the computation time is proportional to $N^3 h^3$ for large N and h where N is the number of discrete points in each of the h discrete filters. Therefore in most practical problems the computation time required to obtain a solution will be very large.

At this point it is interesting to note that this same technique can be used to solve some problems in which the fixed element following the optimum filter is linear time-varying. If the fixed element can be represented as a periodic multiplier followed by a linear time-invariant filter and the multiplier has the characteristics

$$m(t) = m\left(t + \frac{2\pi}{\omega_c}\right) = -m\left(t + \frac{\pi}{\omega_c}\right)$$

then the multiplier can be lumped with $p_k(t)$ and a solution to the problem obtained. Although this is a severe restriction on the form of

of the fixed element there is at least one important problem that satisfies it. It is the AC servomechanism problem in which a two phase motor is the fixed element.

In conclusion, a method for determining the optimum LTV filter and minimum ensemble mean square error in suppressed carrier problems with fixed elements following the optimum filter has been derived. Unfortunately, it appears to be of questionable practical value unless a digital computer is going to be used as the signal processor because the computation time required to obtain a solution is long and it may be difficult to realize the optimum filter without a digital computer. The time average of the minimum mean square error in this problem can be bounded by using the techniques of Sections 7.3.5 and 7.3.6 but when these methods are used the problem is simply that of a linear time-invariant or sampled data system with fixed elements. Since this problem has been thoroughly analyzed by Newton, Gould and Kaiser⁶ in the continuous case and by DeRusso⁷ in the discrete case it is not reviewed in this section.

7.6 SATURATION CONSTRAINTS ON SIGNALS FOLLOWING THE OPTIMUM LTV FILTER

In practice, components such as amplifiers and motors saturate when the input signal or its derivative is large and when this occurs the output is no longer

linearly related to the input and the performance of the system is usually degraded. It is often desirable to know approximately how well a system will perform with a given saturating element so that the most reliable and economical system can be built. Intuitively, the optimum system should adapt to the magnitude of the input signal but such a system is not linear and is beyond the scope of the present work. If we confine the optimum filter in a system with a saturating element to be a linear filter it is impossible to solve the problem exactly by the techniques which have been developed in the previous sections of this chapter because when an element saturates the output is no longer linearly related to the input. Therefore it is necessary to use some other approach or reformulate the problem so that the present techniques are applicable.

Newton⁹ has shown that it is possible to analyze the performance of linear time-invariant systems with saturating elements by constraining the mean square value of the saturating signal to be equal to or below a certain value. This approximate technique is useful because constraining the mean square value of a signal limits the percentage of the time the element is saturated. Furthermore the equations that must be solved when this technique is employed are similar to those of the free configuration problem and the same techniques can be used to solve them. These equations arise because the problem of minimizing the mean square error and constraining the mean square value of a saturating signal is a minimization problem which can be solved by using the calculus of variations and Lagrange multipliers. The way in which this problem differs from the free configuration problem is that the input and error are no longer uncorrelated and the optimum filter which is obtained by solving the Wiener Hopf type equation contains Lagrange multipliers in it. These multipliers--one for each constraint--are independent of time and must be evaluated such that the constraint is satisfied and the mean square error minimized. This can be done by use of a trading curve in which the mean square

value of the saturating signal and the mean square error are plotted as functions of the Lagrange multipliers.

This same technique can also be applied to the suppressed carrier filter problem; however, the calculations involved quickly get out of hand. This can be reasoned by analyzing the system shown in Fig. 7.13 where $g_{s\epsilon}(\xi_1)$ is the saturating element. If we attempt to minimize the mean square error at all time and also constrain the mean square value of the saturating signal $f_{s\epsilon}(t)$ to be less than or equal to a constant, the following functional must be minimized

$$F[g_o(t, \tau_1), \rho(t)] = \overline{e^2(t) + \rho(t) f_{s\epsilon}^2(t)} \quad (7.75)$$

where $\rho(t)$ is a Lagrange multiplier which is a function of time. By using a discrete approximation for $g_o(t, \tau_1)$, $g_{f_2}(\xi_1)$, $g_s(\eta)$ and $\rho(t)$ it is possible to determine an optimum discrete filter by using the methods developed in the previous section; however, this filter and therefore the mean square error and mean square value of the saturating signal are all functions of the discrete values of the Lagrange multiplier. Therefore it is necessary to determine the discrete values of the Lagrange multiplier which satisfy the saturation constraint and minimize the mean square error which is an extremely difficult non-linear problem.

Since minimizing the mean square value of a saturating signal is actually a substitute problem for the actual problem of constraining the magnitude of the saturating signal, it is not too unreasonable to choose another substitute problem in the hope of reducing the complexity of the solution without seriously degrading the significance of the results. One possible alternative is to constrain the time average of the saturating signal. When this technique is employed it is necessary to minimize the following functional

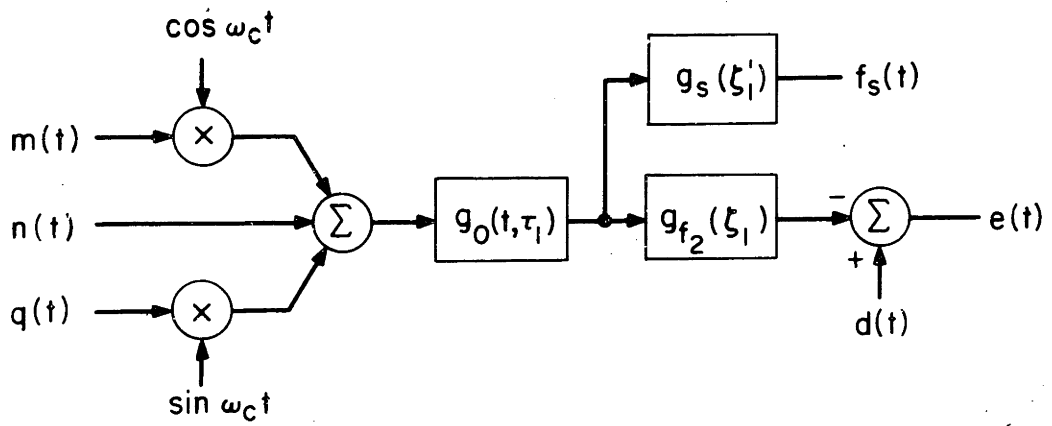


Fig. 7.13 Saturating Signal Following the Signal Processor

$$F[g_o(t, \tau_1),] = e^{\frac{\rho}{2}t} + \rho f_s^{\frac{\rho}{2}}(t) \tag{7.76}$$

where ρ is a constant rather than a function of time. An optimum discrete filter can now be determined by using the method of Section 7.5 and Appendix D by replacing the matrix

$$\int_0^\infty d\xi_1 g_{f_2}(\xi_2) \int_0^\infty d\xi_2 g_{f_2}(\xi_2) [\phi_{ii}(i, n, j)]$$

in Eq. D.14 by the following matrix.

$$\left\{ \int_0^\infty d\xi_1 g_{f_2}(\xi_1) \int_0^\infty d\xi_2 g_{f_2}(\xi_2) + \int_0^\infty d\xi_1 g_s(\xi_1) \int_0^\infty d\xi_2 g_s(\xi_2) \right\} [\phi_{ii}(i, n, j)]$$

In the special case where $\phi_{nm}(\tau) = \phi_{qq}(\tau)$ an explicit solution can be obtained but in other problems it is necessary to solve a set of h matrix equations simultaneously. Therefore the required computation time will be large because it is necessary to solve the problem for several values of ρ in order to determine the value which minimizes the time average of the mean square error and also satisfies the saturation constraint. Therefore it is doubtful that this analytical technique has much practical value.

The saturation problem of Fig. 7.13 can be analyzed by the approximate methods of Sections 7.3.5 and 7.3.6 in order to obtain bounds on the answer. However, when these methods are used the problem is simply that of a linear time-invariant or sampled data system with a saturating element. Since this problem has been thoroughly analyzed by Newton, Gould and Kaiser⁶ in the continuous case and by DeRusso⁷ in the discrete case it is not reviewed in this section.

7.7 RANDOM DRIFT OF THE CARRIER FREQUENCY

7.7.1 Optimization without Fixed Elements

The carrier frequency in most suppressed carrier problems drifts in a random manner at a rate that is usually quite slow with respect to the bandwidth of the signal being modulated. The errors caused by this drift are often large and must be allowed for when designing a system.

(e.g., lag compensating in AC servomechanisms by using second order resonant filters is almost impossible in practice although it is theoretically fine as long as the carrier frequency and components of the filter are constant.)

The effect of random frequency drift on the optimum LTV filter and minimum mean square error in the suppressed carrier problem of Fig. 7.1 is investigated in this section by making the carrier frequency a random variable, ω , which is constant in each ensemble member. The only difficulty with this approach is that it considers only the amplitude statistics associated with the frequency drift and neglects the time statistics. In most practical problems, however, this does significantly affect the results when the mean square error criterion is used because it means that terms such as $\phi_{mm}(\tau) \phi_{ww}(\tau)$ are approximated as $\phi_{mm}(\tau) \phi_{ww}(0)$, where $\phi_{ww}(\tau)$ is the autocorrelation function of the carrier frequency which is generally much wider than the autocorrelation function of the modulated signals.

Since a carrier frequency reference signal is available to the optimum LTV filter, it is unrealistic to require a fixed optimum filter which does not adapt in some manner to a change in the carrier frequency. Therefore two methods of adapting are analyzed. The choice of the procedure to use in studying an actual problem will depend on the type of signal processor being used.

The best possible adaptive filter for a suppressed carrier system is one that responds to changes in the carrier frequency by scaling the optimum filter along both the τ_1 and t axes. When this occurs the optimum filter is a function of the carrier frequency and it is necessary to average over ω in order to determine the ensemble mean square error. However, in order to obtain a meaningful average it is necessary to average the errors out of the ensemble members when the phase of the carrier frequency reference signals into the optimum signal processors are the same. Therefore it is necessary to write the error at time t from an ensemble member as

$$e(t) = d(t) - \int_{-o}^{\infty} d\tau_1 g_o(t, \tau_1) [m(t-\mu\tau_1) \cos \omega_c(t-\tau_1) + q(t-\mu\tau_1) \sin \omega_c(t-\tau_1) + n(t-\mu\tau_1)] \quad (7.77)$$

where $\mu \triangleq \frac{\omega_c}{\omega}$ (7.78)

is a stochastic scaling factor along the τ_1 axis which is used so that the optimum filters in each ensemble member are identical. This equation can be used to obtain an expression for the ensemble mean square error and then the calculus of variations can be used to obtain an integral equation for $g_o(t, \tau_1)$. When this is done, the only difference between the integral equation and Eq. 7.10 is that the correlation functions must be averaged over the random variable μ in the following manner.

$$\int_{-\infty}^{\infty} d\mu p(\mu) \phi(\mu x) = \phi'(x) \quad (7.79)$$

where $p(\mu)$ is the probability density function of the random variable μ . This averaging does not change the value of the correlation function at $x = 0$ but it does at the other values of x . These changes are most easily understood by considering the power density spectra of $\phi(x)$ and $\phi'(x)$. If $\phi(x)$ is exponential and the mean frequency is ω_c so that $\overline{\mu^{-1}} = 1$, then the spectra are equal at $s = 0$, ~~they are~~ equal and the spectrum of $\phi'(x)$ falls off faster at low frequencies and slower at high frequencies than the spectrum of $\phi(x)$. In most practical problems this is an insignificant change. Therefore the minimum mean square error should not differ significantly from its constant frequency value when the optimum filter adapts perfectly along both the τ_1 and t axes.

Unfortunately, it is often impossible or unpractical to use filters that respond in this perfectly adaptive manner to changes in the carrier frequency. Besides digital computers which can be programmed to be perfectly adaptive in this sense, the next most adaptive LTV filters are synchronous demodulators which are not followed by band rejection filters or preceded by band pass filters or else LTV filters which employ synchronously switched elements. As long as there are no fixed elements preceding a LTV filter of these types, the minimum mean square error in these systems should not change significantly when the carrier frequency drifts. However, when there are fixed elements in problems of this type the mean square error does change significantly when the carrier frequency drifts. This problem is analyzed in Section 2.7.2.

The least adaptive type of filter to be considered is one that scales only along the t axis in synchronism with the carrier frequency. Although this type of scaling is usually unrealistically poor, it is briefly analyzed because it establishes an upper bound on the minimum mean square error in synchronous suppressed carrier systems. With this type of scaling the optimum LTV filters are the same in all ensemble members and the conventional calculus of variations techniques can be used to obtain an integral equation for $g_o(t, \tau_1)$. The only difference between the integral equation and Eq. 7.10 is that the cosine terms must be averaged over the carrier frequency ω with the product ωt considered a constant. When $p(\omega)$ is gaussian with a mean of ω_c and a variance σ^2 , the averaging results in the following expressions

$$d\omega p(\omega) \cos \omega(t-\tau_2) = e^{-\frac{\sigma^2 \tau^2}{2}} \cos \omega_c(t-\tau_2) \quad (7.80)$$

$$d\omega p(\omega) \cos \omega(\tau_2-\tau_1) = e^{\frac{-\sigma^2(\tau_2-\tau_1)^2}{2}} \cos \omega_c(\tau_2-\tau_1) \quad (7.81)$$

$$d\omega p(\omega) \cos \omega(2t-\tau_2-\tau_1) = e^{\frac{-\sigma^2(\tau_2+\tau_1)^2}{2}} \cos \omega_c(2t-\tau_2-\tau_1) \quad (7.82)$$

Since the exponential terms in Eqs 7.81 and 7.82 are not equal, the optimum filter is unable to get a sample of $m(t)$ in the past without getting some quadrature cross coupling. Therefore the minimum mean square error will increase as expected. When numerical methods are used to solve the integral equation for $g_o(t, \tau_1)$, the exact expressions for the averaged cosines should be used. However, when an explicit solution is sought in the special case when $\phi_{mm}(\tau) = \phi_{qq}(\tau)$, the autocorrelation function in Eq. 7.81 should be approximated as an exponential autocorrelation function which falls off faster than the original autocorrelation function.

The least adaptive type of filter to be considered is one that scales only along the t axis in synchronism with the carrier frequency. Although this type of scaling is usually unreallistically poor, it is briefly analyzed because it establishes an upper bound on the minimum mean square error in synchronous suppressed carrier systems. With this type of scaling the optimum LTV filters are the same in all ensemble members and the conventional calculus of variations techniques can be used to obtain an integral equation for $g_o(t, \tau_1)$. The only difference between the integral equation and Eq. 7.10 is that the cosine terms must be averaged over the carrier frequency ω with the product ωt considered a constant. When $p(\omega)$ is gaussian with a mean of ω_c and a variance σ^2 , the averaging results in the following expressions

$$d\omega p(\omega) \cos \omega(t-\tau_2) = e^{-\frac{\sigma^2 \tau^2}{2}} \cos \omega_c(t-\tau_2) \quad (7.80)$$

$$d\omega p(\omega) \cos \omega(\tau_2-\tau_1) = e^{\frac{-\sigma^2(\tau_2-\tau_1)^2}{2}} \cos \omega_c(\tau_2-\tau_1) \quad (7.81)$$

$$d\omega p(\omega) \cos \omega(2t-\tau_2-\tau_1) = e^{\frac{-\sigma^2(\tau_2+\tau_1)^2}{2}} \cos \omega_c(2t-\tau_2-\tau_1) \quad (7.82)$$

Since the exponential terms in Eqs 7.81 and 7.82 are not equal, the optimum filter is unable to get a sample of $m(t)$ in the past without getting some quadrature cross coupling. Therefore the minimum mean square error will increase as expected. When numerical methods are used to solve the integral equation for $g_o(t, \tau_1)$, the exact expressions for the averaged cosines should be used. However, when an explicit solution is sought in the special case when $\phi_{mm}(\tau) = \phi_{qq}(\tau)$, the autocorrelation function in Eq. 7.81 should be approximated as an exponential autocorrelation function which falls off faster than the original autocorrelation function.

7.7.2 Optimization with Fixed Elements

In most practical suppressed carrier systems there are fixed elements which precede the signal processor and also the signal processor is some type of synchronous demodulator. Therefore the problem is the same as shown in Fig. 7.9. When the fixed element is a linear time-invariant bandpass filter, the minimum mean square error may increase considerably when the carrier frequency drifts. This is particularly true when significant quadrature signals are present (as in vibratory drive, vibratory gyroscopes) or when the standard deviation of the carrier frequency is a significant fraction of the filter bandwidth (as in AC servomechanisms with lag compensation). The effect of frequency drift in these problems can be analyzed by allowing the carrier frequency to be a random variable which is constant in each ensemble member. Then when the method of Eq. 7.44* is used to analyze the problem, the introduction of a stochastic carrier frequency makes it necessary to average the following terms over ω .

$$\phi_{md}(\tau_2 - \xi_2) \cos(\omega \xi_2 - \psi)$$

$$\phi_{mm}(\tau_2 + \xi_2 - \tau_1 - \xi_1) \cos(\omega \xi_1 - \psi) g_{f_1}(\xi_1) \cos(\omega \xi_2 - \psi) g_{f_1}(\xi_2)$$

$$\phi_{qq}(\tau_2 + \xi_2 - \tau_1 - \xi_1) \sin(\omega \xi_1 - \psi) g_{f_1}(\xi_1) \sin(\omega \xi_2 - \psi) g_{f_1}(\xi_2)$$

$$\phi_{nn}(\tau_2 + \xi_2 - \tau_1 - \xi_1) \cos \omega(\tau_2 - \tau_1) g_{f_1}(\xi_1) g_{f_2}(\xi_2)$$

If the probability density function of ω is known, these quantities can be averaged; however, with most practical probability distributions the resulting expressions are in forms that make it impossible to obtain explicit solutions for the optimum or approximately optimum filter and associated

* When the discrete approach to analyzing this problem is used to obtain a bound or else the exact answer when a sampling demodulator is used, the equations that result are the discrete versions of the equations presented in this section. Therefore they are not presented separately.

time average of the ensemble mean square error. Therefore it is necessary to use some approximate method to solve the equations. A convenient approximation is to represent the probability density function of ω as a finite set of discrete impulses because when this is done the optimum or approximately optimum filters can be determined by the spectrum factorization technique.

When the probability density function of the carrier frequency is approximated as a finite set of discrete impulses and the method of Section 7.4.4 is used to find an approximately optimum filter, Eq. 7.61 must be solved and $\Delta(s)$ and $\Gamma(s)$ must be averaged in the following manner.

$$\Delta(s) = \Phi_{mm}(s) \sum_i P_i G_{f_1}^r(s, \Delta\omega_i) G_{f_1}^r(-s, \Delta\omega_i) \quad (7.83)$$

$$+ \Phi_{qq}(s) \sum_i P_i G_{f_1}^i(s, \Delta\omega_i) G_{f_1}^i(-s, \Delta\omega_i)$$

$$+ 2 \operatorname{Re} \left\{ \sum_i P_i \Phi_{nn}(s+j\omega_i) G_{f_1}(s-j\omega_i) G_{f_1}(-s+j\omega_i) \right\} + 2N$$

$$(s) = \sum_i P_i G_{f_1}^r(s, \Delta\omega_i) \quad (7.84)$$

where $\omega_i \triangleq \omega_c + \Delta\omega_i$

$P_i =$ Probability that $\omega = \omega_i$

$$N \triangleq \sum_{n=1}^{\infty} \sum_i 2P_i \Phi_{nn} [j(2n+1)\omega_i] G_{f_1} [-j(2n+1)\omega_i] G_{f_1} [j(2n+1)\omega_i]$$

In most problems N changes very little with frequency and can be considered a constant. The most important changes are in the averaged terms involving $G_{f_1}^r(s)$ and $G_{f_1}^i(s)$. These changes are most easily understood by examining the fixed element used in Chapter IV to describe the resonant sensing system of a vibratory rate gyroscope.

$$G_{f_1}(s) = \frac{s}{s^2 + 2as + \omega_c^2} \quad (7.85)$$

If $\psi = 0$ the approximate envelope transfer function for the message and quadrature signals are given by Eqs. 7.86 and 7.87 respectively.

$$G_{f_1}^r(s, \Delta\omega_1) = \text{Re} \{ e^{-j\psi} G_{f_1}(s - j\omega_1) \} = \frac{s + a}{2[(s+a)^2 + \Delta\omega_1^2]} \quad (7.86)$$

$$G_{f_1}^i(s, \Delta\omega_1) = \text{Im} \{ e^{-j\psi} G_{f_1}(s - j\omega_1) \} = \frac{s^2 - 2\omega_c \Delta\omega_1}{4\omega_c [(s+a)^2 + \Delta\omega_1^2]} \quad (7.87)$$

where $\omega_1 \triangleq \omega_c + \Delta\omega_1$ and it is assumed that $\omega_c > s, a,$ and $\Delta\omega_1$. The following two important conclusions can be drawn from these equations.

- (1) $G_{f_1}^r(s, \Delta\omega_1)$ is a good low pass filter if and only if (a) in Eq. 7.85 is larger than $\Delta\omega_1$. For this reason lag compensation is extremely difficult to achieve in AC servomechanisms with resonant linear time-invariant filters.
- (2) The quadrature cross-coupling which appears at the output is proportional to $\Delta\omega_1$. For extremely small frequency drifts $\Delta\omega_1 < a$, the ratio of the two transfer functions

$$\frac{G_{f_1}^i(s, \Delta\omega_1)}{G_{f_1}^r(s, \Delta\omega_1)} = \frac{s^2 - 2\omega_c \Delta\omega_1}{2\omega_c (s+a)} \quad (7.88)$$

is equal to $(\frac{\Delta\omega_1}{a\omega_c})$ at $s = 0$. Therefore the low frequency

quadrature cross-coupled errors are proportional to

$q(t) (\frac{\Delta\omega}{a\omega_c})$ which is the same relationship that was

determined in Section 5.2.4.

If the probability density function of ω_1 is approximated as two impulses of area $\frac{1}{2}$ at frequencies $\omega_c \pm \Delta\omega$, where $\Delta\omega$ is the RMS value of ω_1 , then the averaging which is required in Eq. 7.83 is relatively easy.

$$\sum_i P_i G_{f_1}^r(s, \Delta\omega_i) G_{f_1}^r(-s, \Delta\omega_i) = \frac{-(s+a)(s-a)}{4[(s+a)^2 + \Delta\omega^2][(-s+a)^2 + \Delta\omega^2]} \quad (7.89)$$

$$\sum_i P_i G_{f_1}^i(s, \Delta\omega_i) G_{f_1}^i(-s, \Delta\omega_i) = \frac{s^4 + 4\omega_c^2 \Delta\omega^2}{16[(s+a)^2 + \Delta\omega^2][(-s+a)^2 + \Delta\omega^2] \omega_c^2} \quad (7.90)$$

$$\sum_i P_i G_{f_1}^r(s, \Delta\omega_i) = \frac{s+a}{2[(s+a)^2 + \Delta\omega^2]} \quad (7.91)$$

The noise term cannot be averaged unless the noise spectrum is known.

This problem is analyzed further in the following chapter and was presented here only to show that it is relatively easy to perform the averaging if a discrete probability density function is selected for the carrier frequency.

7.7.3 Conclusions

By using the analytical techniques developed in this section, the effects of carrier frequency drift on the optimum filter and minimum mean square error in a suppressed carrier problem can be determined. When harmonic errors are relatively unimportant compared to cross-coupled errors and noise errors, as in vibratory gyroscopes, the method of Section 7.7.2 should be used. This method is relatively easy to apply if a simple probability density function is used for the carrier frequency. In the following chapter the effect of carrier frequency drift in vibratory gyroscopes is analyzed by using this analytical procedure.

7.8 RANDOM DRIFT OF THE FIXED ELEMENTS

The characteristics of any fixed element vary with time, environmental parameters and other factors. When the fixed elements are in linear systems, these changes or drifts cause the poles and ~~zeros~~ of the transfer function of the linear system to move in a random manner in the s plane. Usually, these changes are relatively small and are not troublesome if good components and designs are used. However, if the poles and zeros of the transfer function are located near the $j\omega$ axis in the s plane in order to obtain band pass or band rejection characteristics, a random drift of the poles or zeros parallel to the $j\omega$ axis can cause important errors in carrier systems. Fortunately, the time constants associated with these drifts are generally very long compared to the widths of the autocorrelation functions of the signals so it is possible to introduce this problem into the model of Fig. 7.1 by assuming that the fixed elements are constant in each ensemble member. Then the only changes that occur in the integral equations that have been developed in previous sections are that terms containing the impulse response of the random fixed elements must be averaged. If a discrete probability density function is selected for the fixed elements, this is a relatively easy problem to solve. However, in practice the most troublesome fixed elements are band pass filters and their resonant frequencies generally drift but the damping remains approximately constant. When this is the case, the drift can be statistically added to the carrier frequency drift since the differential drift between the two is really the important drift in carrier systems. Then the analytical techniques of section 7.72 can be used, which is an even easier method of analysis.

7.9 RANDOM PHASE SHIFTS ON THE REFERENCE SIGNAL

In the suppressed carrier problem of Fig. 7.1, the optimum filter is synchronized with the carrier frequency modulating signal. In practice, the phase of the reference or synchronizing signal is usually adjustable so that the best quadrature rejection and signal reception can be achieved in a given system. Unfortunately, the phase of the synchronizing signal with respect to modulating signal is not always constant and the reference signal may be noisy. Both of these imperfections can cause significant errors in suppressed carrier systems. The statistics associated with the random component of the phase shift are usually very slow because they are due to material aging and temperature variations while the spectrum of the noise on the synchronizing signal varies from problem to problem.

In vibratory gyroscopes, the noise on the synchronizing signal is usually quite small¹⁰ and can be neglected in most well designed instruments. The random component of the phase is also small but cannot be neglected because of the large quadrature signals that usually are present. The effect of this randomness on the optimum filter and minimum mean square error can be analyzed by using the model of Fig. 7.1 and adding a phase angle $\theta(t)$ to the modulating sinusoids as

$$\sin \omega_c t \rightarrow \sin [\omega_c t + \theta(t)]$$

$$\cos \omega_c t \rightarrow \cos [\omega_c t + \theta(t)]$$

When this is done, the sinusoidal terms in the integral equation (7.10) for $g_0(t, \tau_1)$ become

- a) $\phi_{md}(\tau_2) \cos [\omega_c(t-\tau_2) - \theta(t-\tau_2)]$
- b) $\phi_{11}(\tau_2) \cos [\omega_c(\tau_2-\tau_1) + \theta(t-\tau_1) - \theta(t-\tau_2)]$
- c) $\phi_{22}(\tau_2) \cos [\omega_c(2t-\tau_2-\tau_1) + \theta(t-\tau_1) + \theta(t-\tau_2)]$

These expressions must be averaged over θ since θ is a random variable in the problem. This averaging can be done by assuming that $\theta(t)$ is constant in each ensemble member and then averaging over the ensemble as was done in the previous two sections or the time statistics of $\theta(t)$ can be introduced into the problem by using the small angle approximations.

$$\sin \theta = \theta$$

$$\cos \theta = 1 - \frac{\theta^2}{2}$$

When this latter approach is used and terms involving products of angles to powers larger than 2 are dropped, the following relationships are found for the expressions in Eq. 7.10.

$$a) \quad \phi_{md}(\tau_2) \left[1 - \frac{\overline{\theta^2}}{2} \right] \cos \omega_c(t - \tau_2)$$

$$b) \quad \phi_{11}(\tau_2 - \tau_1) \left[1 - \overline{\theta^2} + \phi_{\theta\theta}(\tau_2 - \tau_1) \right] \cos \omega_c(\tau_2 - \tau_1)$$

$$c) \quad \phi_{22}(\tau_2 - \tau_1) \left[1 - \overline{\theta^2} - \phi_{\theta\theta}(\tau_2 - \tau_1) \right] \cos \omega_c(2t - \tau_2 - \tau_1)$$

These expressions can be used in Eq. 7.10 to obtain the optimum LTV filter and minimum mean square error when the harmonic errors are weighted equally with the other errors. When the harmonic errors are not weighted in this manner but instead are neglected, the effect of phase shift drift can be analyzed by using the approximate method of Sections 7.3.6 and 7.4.4. When there are no fixed elements in the system the answer is obvious and need not be discussed. When there are fixed elements and a synchronous demodulator is used so that the problem is similar to that shown in Fig. 7.9, then the effect of phase shift drift

can be introduced into the problem by representing the angle ψ between the cosine modulating signal and the fundamental of the sensitivity function as

$$\psi = \psi_0 + \Delta\theta \tag{7.92}$$

where ψ_0 steady or average component and $\Delta\theta$ is the random component which has a zero mean. When this is done and $\Delta\theta$ is assumed to be constant in each ensemble member, the only terms in Eq. 7.61 that change are $G_{f_1}^r(s)$ and $G_{f_1}^i(s)$. When small angle approximations are made

$$\sin \Delta\theta = \Delta\theta \tag{7.93}$$

$$\cos \Delta\theta = 1 - \frac{\Delta\theta^2}{2} = 1$$

the following expressions are obtained.

$$G_{f_1}^r(s, \Delta\theta) = G_{f_1}^r(s) + \Delta\theta G_{f_1}^i(s) \tag{7.94}$$

$$G_{f_1}^i(s, \Delta\theta) = G_{f_1}^i(s) - \Delta\theta G_{f_1}^r(s) \tag{7.95}$$

When the method of Section 7.4.4* is used to find an approximately optimum filter, Eq. 7.61 must be solved and $\Delta(s)$ and $\Gamma(s)$ must be averaged over $\Delta\theta$. This can easily be done but actually it is unnecessary because the present method of analyzing the problem is more involved than necessary. What actually happens when a differential phase shift occurs is that the inphase signal is no longer $m(t)$ but is $m'(t)$ given by Eq. 7.96 and the quadrature signal is no longer $q(t)$ but is $q'(t)$ given by Eq. 7.97.

$$\begin{aligned} m'(t) &= m(t) + \Delta\theta(t) q(t) \\ \phi_{mm}^i(\tau) &= \phi_{mm}(\tau) + \phi_{\theta\theta}(\tau) \phi_{qq}(\tau)** \end{aligned} \tag{7.96}$$

* When the discrete approach to analyzing this problem is used to obtain a bound or else the exact answer when a sampling demodulator is used, the equations that result are the discrete versions of the equations presented in this section. Therefore they are not presented separately.

** If $\Delta\theta$ is assumed to be a constant in each ensemble member, then $\phi_{\theta\theta}$ must be replaced by $\Delta\theta^2$.

$$q'(t) = q(t) - \Delta\theta(t) m(t) \quad (7.97)$$

$$\phi_{qq}''(\tau) = \phi_{qq}(\tau) + \phi_{\theta\theta}(\tau) \phi_{mm}(\tau)$$

where

$$\phi_{m'q'}(\tau) = \phi_{mq}(\tau) = 0 \quad (7.98)$$

When these relationships are used for the inphase and quadrature autocorrelation functions and

$$\overline{\Delta\theta^2} < 1$$

$$\overline{\Delta\theta} = 0$$

one obtains the same equation for the optimum filter.

7.10 SUMMARY OF OPTIMUM PROCESSING OF SUPPRESSED CARRIER SIGNALS

An analytical design technique has now been developed for the suppressed carrier problem shown in Fig. 7.1. By using it one is able to determine

1. the optimum LTV filter to minimize the ensemble mean square error at all times when harmonic errors from the message and quadrature signals are weighted equally with the other errors.
2. the approximately optimum restricted LTV filter to minimize the time average of the ensemble mean square error when harmonic errors from the message and quadrature signals are neglected.
3. an upper bound on the time average of the minimum mean square error.
4. a lower bound on the time average of the minimum mean square error.

In problems in which harmonic errors are as important as the other errors, one should use the techniques developed for obtaining the exact answer. However, in other problems the harmonic errors may not be important on the form of the signal processor may be restricted. In either of these cases the techniques developed for obtaining the approximately optimum LTV filter should be used.

CHAPTER VIII
EXAMPLE OF AN OPTIMUM FILTER
FOR A VIBRATORY RATE GYROSCOPE

8.1 INTRODUCTION

Most vibratory rate gyroscopes can be represented by a block diagram similar to the one shown in Fig. 8.1. However, in some instruments

1. a position or an acceleration sensitive transducer may be used rather than the velocity sensitive transducer that is indicated,
2. the sensing system may be designed so that the resonant frequency does not coincide with the carrier frequency but is an order of magnitude or so away so that the transfer function of the fixed elements is flat around the carrier frequency, or
3. feedback may not be used around the sensing system to change the bandwidth.

Because these numerous different schemes exist, it is impossible to analyze one model and draw conclusions about all of the different types of sensing schemes. With this in mind, let us now thoroughly examine one example. Admittedly, one is a small number of examples, but recall that the primary purpose of this work is to develop an analytical design technique for suppressed carrier problems, such as vibratory gyroscopes. The theory was developed in the preceding chapter and this chapter simply demonstrates how to apply the theory to one type of problem that occurs in vibratory gyroscopes: This example was selected because it cannot be solved by just plugging numbers into the derived equations but instead it is necessary to modify one of the analytical techniques slightly to fit the problem.

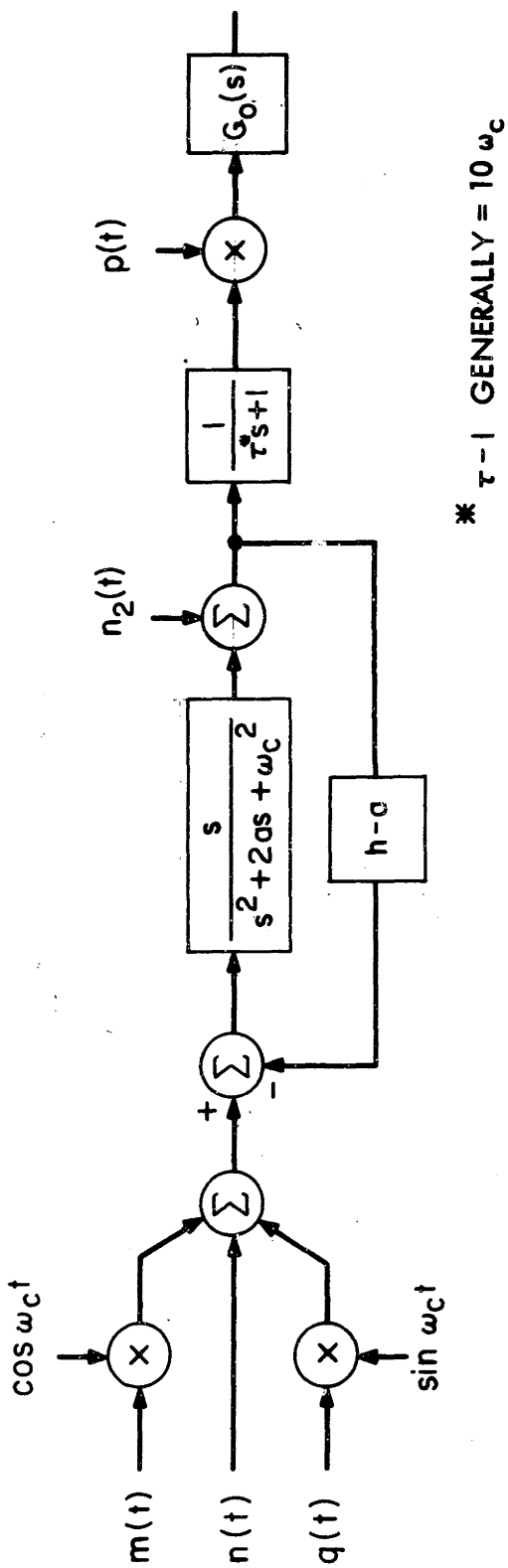


Fig. 8.1 Vibratory Gyroscope Model

The following example is analyzed by using the approximate method of Section 7.4.4 to minimize the time average of the ensemble mean square error using a restricted form of optimum LTV filter. This error criterion and form of the optimum LTV filter are used rather than the straight mean square error criterion because:

1. the harmonic errors that result from the message and quadrature signals in vibratory gyroscopes are generally much less important than the errors that occur at the output at the same frequencies as the rate information.
2. In most practical vibratory gyroscopes, the carrier frequency is more than 10 times the bandwidth of the applied angular rates so overlapping spectra are not troublesome although they still cause significant mean square errors.
3. The signal processors generally used in vibratory gyroscopes are similar to the restricted form of signal processor shown in Fig. 8.1. If the instrument is a two axis instrument, then two demodulators are used but generally they are both similar to the one shown in Fig. 8.1.

8.2 VIBRATORY DRIVE, VIBRATORY RATE GYROSCOPE EXAMPLE

In vibratory drive, vibratory rate gyroscopes, the major source of long term drift or very low frequency noise at the output is unwanted cross coupling between the drive member and the sensing system. As was explained in Chapter V, this noise is primarily caused by the signals produced by

1. Inphase cross-coupled torques.
2. Quadrature cross-coupled torques which are not perfectly discriminated against in the demodulation process.

The spectra of the output signals caused by the cross coupling is generally extremely narrow and centered around DC and cannot be filtered electrically if the instrument is to be a "rate" gyroscope with response to zero frequency. Therefore these cross-coupled, zero-rate errors, which can be written approximately as

$$\Omega_{cc} = \Omega_d + \Omega_q (\Delta\theta + 2\Omega_s \frac{\Delta\omega}{\omega})$$

from Chapter V, must not be introduced directly into the equation for the mean square error because when they are, they force the transfer characteristics of the optimum filter-- $G_o(s)$ in Fig. 8.1--to be very small at low frequencies. When the cross-coupled, zero-rate errors are omitted and the method of Section 7.4.4 is used to analyze the problem, the major sources of error are additive noise and message distortions. These errors are functions of the bandwidth of the resonant sensing system which is determined by the gain in the feedback loop. Since the cross-coupled, zero-rate errors are also functions of the bandwidth of the resonant sensing system, an optimum bandwidth can be determined by adding the two types of errors and determining the minimum of the sum. This method is demonstrated in the following example in which

$m(t) = \Omega_{in}(t) + \Omega_d(t)$ where $\Omega_{in}(t)$ is the input angular rate which has an autocorrelation function of $e^{-|t|}$

- a = one half the bandwidth of the resonant sensing system without feedback
- h = one half of the bandwidth of the resonant sensing system with feedback. This is the parameter we are attempting to optimize.
- $n_1(t)$ = noise in the sensing system which is assumed to be negligible compared to the noise in the signal processing electronics.
- $n_2(t)$ = noise in the signal processing electronics which is assumed to be white and 10 times larger than the noise that would occur at the point it enters because of thermal fluctuations of the sensing system components
- $p(t)$ = square wave sensitivity function because the Fourier coefficients decrease faster with this waveform than with other practical sensitivity functions.

Now we can write down the following expressions from Chapters V and VII

$$\Phi_{\Omega_{in} \Omega_{in}}(s) = \frac{-2}{(s-1)(s+1)}$$

$$\Phi_{n_2 n_2}(s) = \frac{10N_1}{a^2} = N_2 \frac{(0/hr)^2}{\text{radian/sec}^2}$$

N_1 = Amplitude of the thermal fluctuation noise spectrum

$$\Phi_{md}(s) = \Phi_{\Omega_{in} \Omega_{in}}(s) G_d(s)$$

$$G_d(s) = 1$$

When the method of Section 7.4.4 is used, it can be shown that it is necessary to solve Eq. 7.61

$$G_o(s) = \frac{1}{\Delta^+(s)} \left[\frac{\Gamma(s)}{\Delta^-(s)} \right] +$$

where

$$\Delta(s) = \frac{1}{2(s-1)(s+1)(s-h)(s+h)} + 2N_2 \frac{(s-a)(s+a)}{(s-h)(s+h)} + 2N$$

$$N = \sum_{n=1}^{\infty} N_2 \left(\frac{1}{2n+1} \right)^2 \left[\frac{10^2}{10^2 + (2n+1)^2} \right] = 0.2 * N_2$$

$$\Gamma(s) = \frac{2}{(s-1)(s+1)(s-h)}$$

Let us now define

$$\Delta(s) \triangleq \frac{2.4N_2^2 (s-b)(s-c)(s+b)(s+c)}{(s-1)(s-h)(s+1)(s+h)}$$

$$\text{where } b^2 + c^2 \triangleq \frac{4(a^2+1) + 0.8(h^2+1)}{4.8}$$

$$b^2 c^2 \triangleq \frac{N_2^{-1} + 4a^2 + 0.8 h^2}{4.8}$$

* This number is simply an estimate based on the first four coefficients $\frac{1}{9} + \frac{1}{25} + \frac{4}{5} + \frac{1}{49} \left(\frac{2}{3} \right) + \frac{1}{63} \left(\frac{2}{3} \right)$. It shows that the noise from the harmonics of $p(t)$ is small

Now we can determine the optimum filter and the time average of the minimum mean square error.

$$G_o(s) = \frac{1}{1.2 N_2^1 (b+1)(c+1)} \frac{(s+b)}{(s+b)(s+c)}$$

$$\overline{e^2(t)} = 1 - \frac{1}{4.8 N_2^1 (b+1)^2 (c+1)^2}$$

The answer checks when $N_2^1 = 0$ because as N_2^1 becomes small the product $N_2^1 b^2 c^2$ approaches 4.8^{-1} and $b > 1$ and $c > 1$.

Now we have an expression for the time average of the minimum mean square error as a function of the gain of the feedback loop ($h^1 + a = h$). By putting in values for a and N_2^1 we can plot $\overline{e^2(t)}$ as a function of h . However, since we have used an input spectrum with the bandwidth of 1 radian/sec, it is necessary to perform some frequency scaling in order to obtain realistic numbers. If we assume we have an instrument in which

$$\omega_c = 2\pi \times 10^3$$

$$Q_s = 300$$

$$a = \frac{\omega_c}{2Q_s} = 10$$

it is convenient to scale (a) by a factor of 100 so that it is 10^{-1} which is equivalent to increasing the bandwidth of the input angular rates to 100 radians second. Now let us make a $N_1^1 = \frac{(10^{-4} \text{ o/hr})^2}{\text{radians/sec}}$ and $\overline{\Omega_q^2} = 10^5 (\text{o/hr})^2$. Therefore $N_2^1 = \frac{10N_1^1}{a^2} = \frac{10}{a^2} \frac{(10^{-4} \text{ o/hr})^2}{\text{radians/sec}}$ but since we have scaled the frequency it is necessary to use scale N_2^1 by a factor of 10^6 so that

$$N_2^1 = 10^{-3}$$

Now we can plot $\overline{e^2(t)}$ as a function of h , (which is also scaled by a factor of 100 from the actual instrument). First, however let us define some new variables

because for the numbers just chosen, b and c are imaginary for small values of h.

$$b \triangleq d + je$$

$$c \triangleq d - je$$

When this is done it can be shown that

$$d^2 - e^2 = \frac{4(a^2 + 1) + 0.8(h^2 + 1)}{9.6}$$

$$d^2 + e^2 = \left(\frac{N_2^{-1} + 4a^2 + 0.8h^2}{4.8} \right)^{1/2}$$

$$\frac{\overline{e^2(t)}}{e^2} = 1 - \frac{1}{4.8 N_2^1 [(d+1)^2 + e^2]^2}$$

In Fig. 8.2, $\frac{\overline{e^2(t)}}{e^2}$ and $\Omega_q^2 (2Q \frac{\Delta\omega}{\omega})^2$ are plotted as functions of h where it is assumed that $\frac{\Delta\omega}{\omega} = 10^{-5}$. The parameters of the optimum filter for several values of h are also recorded. These two curves are then added together to produce a third curve which is a plot of the mean square error from the noise, message distortion and quadrature cross coupling which is not perfectly discriminated against because of differential phase shifts. It is apparent from the third curve that a minimum occurs approximately when

$$h = 3.3$$

$$d = 2.81$$

$$e = 2.57$$

Therefore the optimum filter for the model of Fig. 8.1 is

$$G_o(s) = 40 \left[\frac{s + 3.3}{(s + 2.81 + j2.57)(s + 2.81 - j2.57)} \right]$$

which shows that the optimum filter weights the low frequency signals more than the high frequency ones. This is because in the model we have selected, the low frequency signals are preamplified before the H_2 noise enters by the

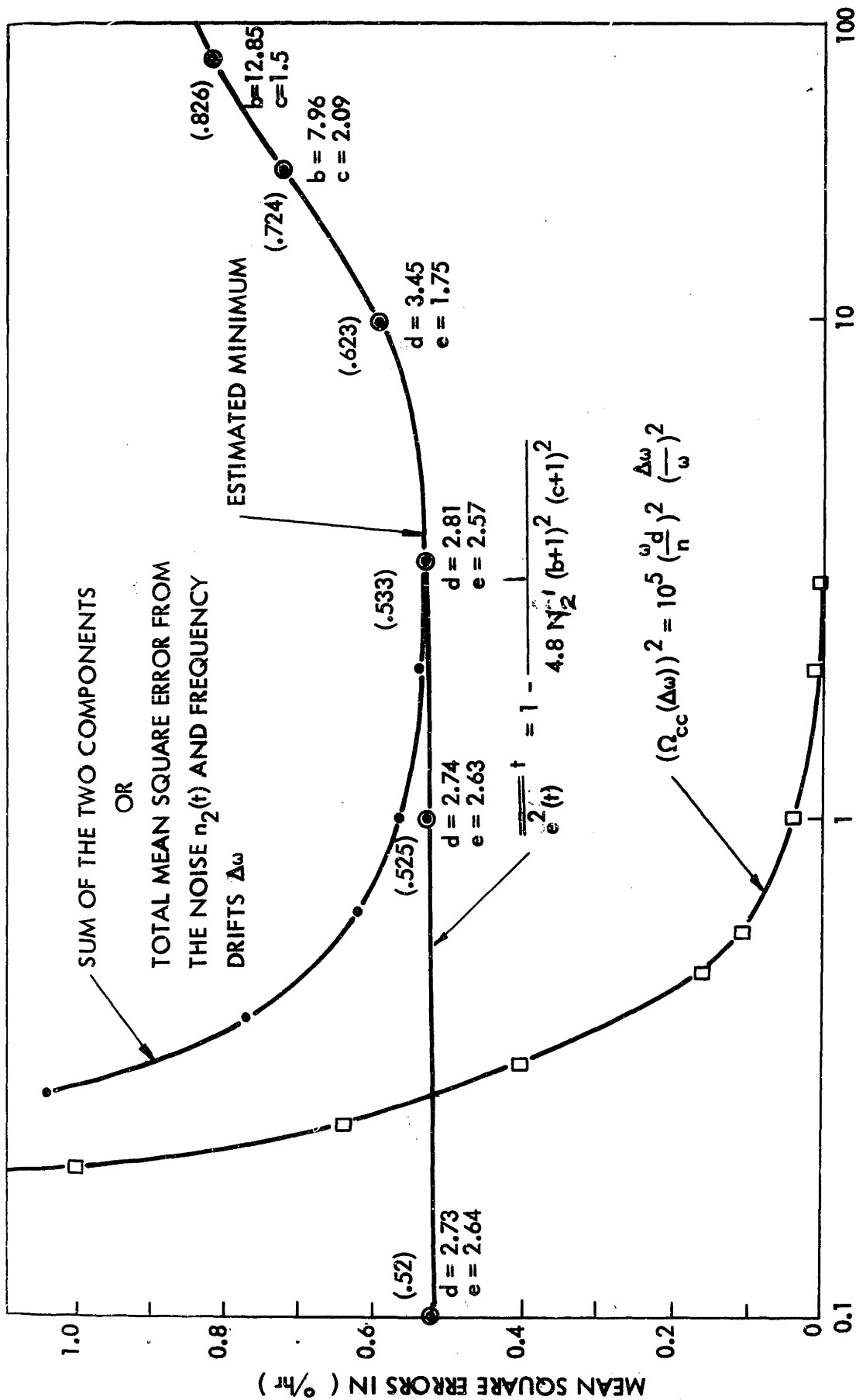


Fig. 8.2 Two Components of the Mean Square Error in the Vibratory Gyroscope Example

resonant sensing system while the higher frequency components in the input spectrum are not.

Let us now estimate what the mean square error is from the harmonics of the message and quadrature. The largest harmonic error results from the quadrature signal and it is approximately

$$\overline{e_{qq}^2}(t) = \frac{105}{4} \left[\frac{40}{(2\pi \times 20)^2} \right]^2 = 1.5(0/\text{hour})^2$$

which is a significant number but certainly not larger than the mean square error that results from cross coupling and noise in the electronics which is tabulated below and is approximately $11.4(0/\text{hr})^2$

$$\overline{e^2}(t) = (0.53)(0/\text{hr})^2 \quad \text{From noise } N_2$$

$$\overline{(\Omega_q^2)} \left(2\theta \frac{\Delta\omega}{\omega} \right)^2 = < 0.01(0/\text{hr})^2$$

$$\overline{(\Omega_q^2)} (\Delta\theta^2) = 10^5 \times 9 \times 10^{-6} = 0.9(0/\text{hr})^2$$

$$\overline{(\Omega_d^2)} = 10^{-4} \overline{(\Omega_q^2)} = 10(0/\text{hr})^2$$

$$= 11.4(0/\text{hr})^2$$

It is apparent from this example that the feedback gain does have an optimum setting but the total mean square error in the instrument changes only a few percentage points if the gain is off by a factor of 10 because the major source of error in the example is the stochastic inphase cross coupling. Furthermore the time average of the mean square harmonic errors that result from the message and quadrature signals are smaller than the other errors and therefore the decision to neglect them when formulating the problem was sound. In order to further reduce the mean square in the instrument it is necessary to filter out the low frequency cross-coupled errors.

The optimum filter for this job can easily be determined by the standard techniques. Conceivably such a filter could reduce the mean square error (neglecting harmonic errors) to approximately the value of $\frac{\sigma^2}{2\pi f^2}$ plotted in Fig. 8.2 but then the instrument would not have response to zero frequency.

8.3 SUMMARY

The performance of a vibratory gyroscope can be improved by processing the signals in an optimum manner but this improvement most likely will be small if the cross-coupled zero-rate errors are large and the instrument must have a flat response to 0 frequency. This conclusion is consistent with the example of section 8.2 and also the conclusions in Chapter V. However, this does not mean that the theory developed in the previous chapter leads to poor signal processors. The reason for the relative ineffectiveness of optimum filtering techniques in vibratory gyroscopes is the fact that the cross-coupled noise occurs at very low frequencies and filtering is often not allowable at these frequencies. In other suppressed carrier filter problems in which the spectra of the message and quadrature signals are similar and the same order of magnitude, optimum filtering in the manner of Chapter VII will lead to significant improvements.

APPENDIX A

THEORY OF OPERATION OF TUNING-FORK GYROSCOPE EMPLOYING DOUBLE MODULATION

The following two differential equations describe the ideal double modulated tuning-fork gyroscope when inertial rates are applied about the y and x axes shown in Fig. A-1. Terms involving rates about the z axis are omitted because of their similarity to the y axis terms.

$$\begin{aligned}
 & I_y(1+a_y \sin \omega_d t) \ddot{\theta}_1 + (f_1 + f_2 + a_y I_y \omega_d \cos \omega_d t) \dot{\theta}_1 + \{K_1 K_2 \\
 & + [I_x - I_z(1+a_z \sin \omega_d t)] \omega_m^2\} \theta_1 = f_2 \dot{\theta}_2 + K_2 \theta_2 \\
 & + I_y(1+a_y \sin \omega_d t) [\Omega_y \omega_m \sin \omega_m t - \dot{\Omega}_y \cos \omega_m t] \\
 & - I_y a_y \omega_d \cos \omega_d t [\Omega_y \cos \omega_m t] + [I_z(1+a_z \sin \omega_d t) - I_x] \\
 & [\Omega_x^2 \theta_1 + \Omega_y^2 \theta_1 \sin^2 \omega_m t - \Omega_y(\omega_m + \Omega_x) \sin \omega_m t] \quad (A-1)
 \end{aligned}$$

$$\begin{aligned}
 & I_{y_c} \ddot{\theta}_2 + f_2 \dot{\theta}_2 + [K_2 + (I_{x_c} - I_{z_c}) \omega_m^2] \theta_2 = f_2 \dot{\theta}_1 + K_2 \theta_1 \\
 & + I_{y_c} [\Omega_y \omega_m \sin \omega_m t - \dot{\Omega}_y \cos \omega_m t] + [I_{z_c} - I_{x_c}] \\
 & [\Omega_x^2 \theta_2 + \Omega_y^2 \theta_2 \sin^2 \omega_m t - \Omega_y(\omega_m + \Omega_x) \sin \omega_m t] \quad (A-2)
 \end{aligned}$$

where

- I_x = Moment of inertia of the tuning fork about the x_f axis
- $I_y(1+a_y \sin \omega_d t)$ = Moment of inertia of the tuning fork about the y_f axis
- $I_z(1+a_z \sin \omega_d t)$ = Moment of inertia of the tuning fork about the z_f axis

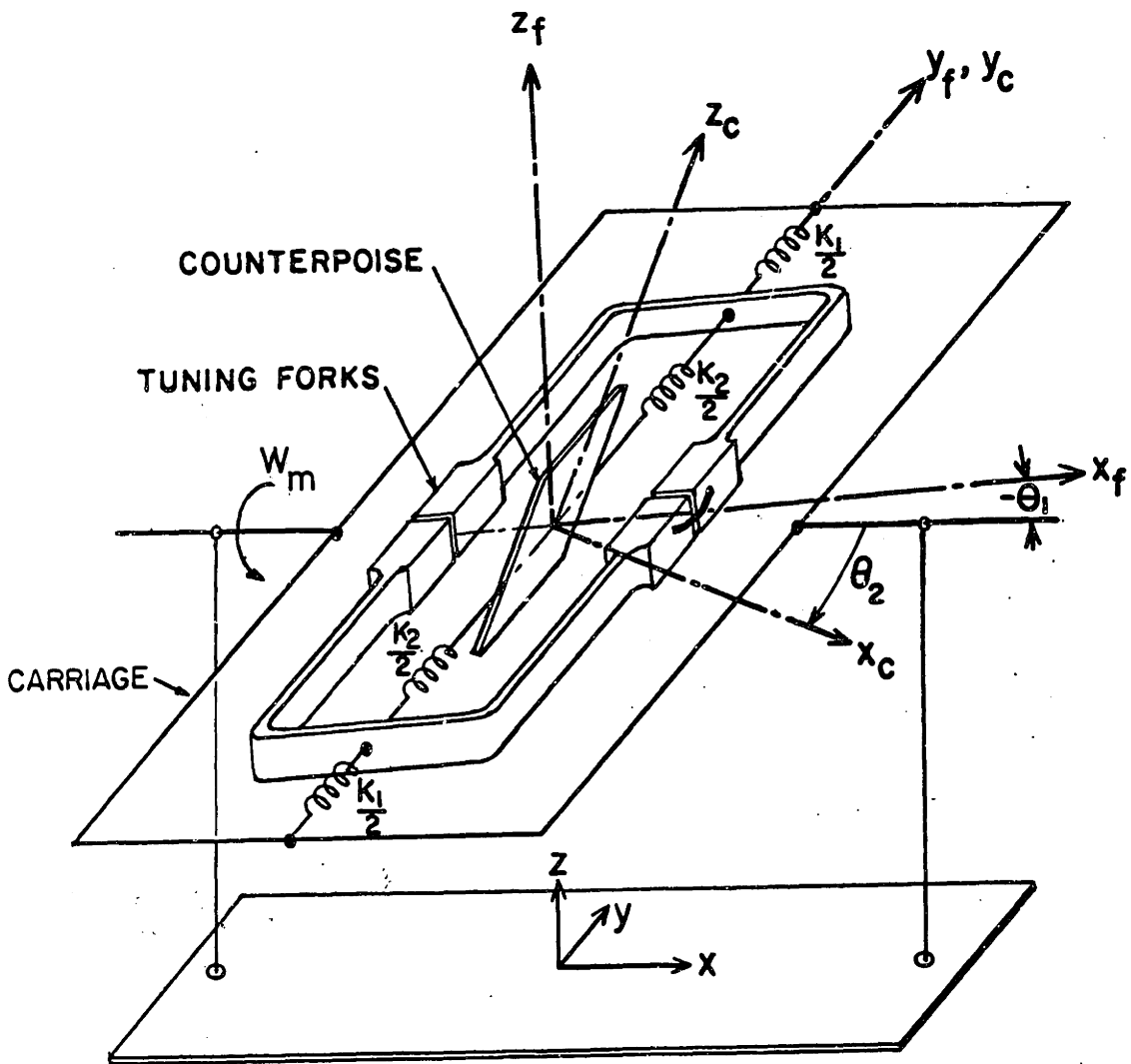


Fig. A-1 Instrument Coordinate Systems

APP.

I_{x_c} = Moment of inertia of the counterpoise about the x_c axis

I_{y_c} = Moment of inertia of the counterpoise about the y_c axis

I_{z_c} = Moment of inertia of the counterpoise about the z_c axis

K_1 = Spring constant for torsion rod and spokes supporting the tuning fork

K_2 = Spring constant for torsion rod between tuning fork and counterpoise

f_1 = Loss term associated with K_1 spring

f_2 = Loss term associated with K_2 spring

Ω_y = Inertial rate applied about the y axis

Ω_x = Inertial rate applied about the x axis

θ_1 = Angle between the tuning fork and the case

θ_2 = Angle between the counterpoise and the case

ω_m = Double modulation frequency

ω_d = Tuning fork or drive frequency

The differential equations were derived using Lagrange's method and making the following assumptions.

1. Higher harmonic terms of the tuning fork inertias are negligible.
2. Double modulation is planar.
3. The tuning fork and counterpoise are rigidly constrained except in the torsional mode about the y_f axis.

The last term of Eq. 2 can be eliminated by making $I_{x_c} = I_{z_c}$. The remaining time-varying term represents a torque applied to the counterpoise member at the double modulation frequency. Since we are using resonant sensing and synchronous demodulation this term can be neglected and the resulting equation shows that the counterpoise element acts in the intended manner as a vibration damper.

The last term of Eq. 1 contains the interaxis cross-coupling terms and second order terms in the applied rates. Even when these terms are dropped, the time-varying coefficients make it difficult to solve the equations exactly. An approximate solution can be obtained by using the average values of the coefficients. A more exact solution can be obtained by using properties of Mathieus equation.

In the following analysis the simplified theory is used to determine the approximate behavior of the instrument. Forcing functions at the double modulation frequency, interaxis cross-coupling terms and second order terms are dropped. $F(t)$ is the forcing function at the two sidebands $(\omega_d \pm \omega_m)$ and is expressed by

$$F(t) = \frac{1}{2} I_y a_y \omega_d \left[\left(1 + \beta \frac{\omega_m}{\omega_d}\right) \cos(\omega_d - \omega_m)t + \left(1 - \beta \frac{\omega_m}{\omega_d}\right) \cos(\omega_d + \omega_m)t \right] \quad (A-3)$$

where β is a function of the inertia of the tuning fork which is approximately zero since $I_y a_y \approx I_z a_z$. When these approximations are made the following two equations result.

$$I_y \ddot{\theta}_1 + (f_1 + f_2) \dot{\theta}_1 + [K_1 + K_2 + (I_x - I_z) \omega_m^2] \theta_1 = f_2 \dot{\theta}_2 + K_2 \theta_2 - F(t) \quad (A-4)$$

$$I_{y_c} \ddot{\theta}_2 + f_2 \dot{\theta}_2 + K_2 \theta_2 = f_2 \dot{\theta}_1 + K_2 \theta_1 \quad (A-5)$$

Since the equations are linear, the following transfer function relating the output angle $(\theta_2 - \theta_1)$ to the forcing function $F(t)$ can be obtained.

$$\frac{\theta_2(s) - \theta_1(s)}{F(s)} = \frac{s^2}{I_y (s^2 + \frac{\omega_l}{Q_l} s + \omega_l^2) (s^2 + \frac{\omega_s}{Q_s} s + \omega_s^2)} \quad (A-6)$$

where

ω_l = Lower resonant frequency

ω_s = Upper resonant frequency

Q_l = Quality factor associated with the lower resonant frequency

Q_s = Quality factor associated with the upper resonant frequency

If ω_m is made equal to the drive frequency minus the upper resonant frequency of the mechanical system and $\omega_l \ll \omega_s$, the envelope transfer function for the lower sideband of $F(t)$ is expressed by Eq. 7

$$\frac{[\theta_2(s) - \theta_1(s)]_E}{\Omega_y(s)} = \frac{a_y \omega_d \tau}{4\omega_s (\tau s + 1)}, \quad \tau = \frac{2Q_s}{\omega_s} \quad (A-7)$$

The upper sideband is separated from the resonant frequency by $2\omega_m$ and the cross-coupled tuning fork drive signal by ω_m so they are attenuated with respect to the lower sideband from which we are extracting the rate information.

APPENDIX B

This appendix contains the schematics of the electronic circuits used in the double modulation experiment as of August 1964. No attempt was made to minimize the power requirements of these circuits.

The schematics found on the following pages are tabulated below.

Figure B.1 Tuning Fork Drive

Figure B.2 Vibration Detector

Figure B.3 Error Angle Detector

Figure B.4 SCR Motor Control

Figure B.5 First Demodulator

Figure B.6 Second Demodulator

A block diagram of the instrument is in Fig. 6.5.

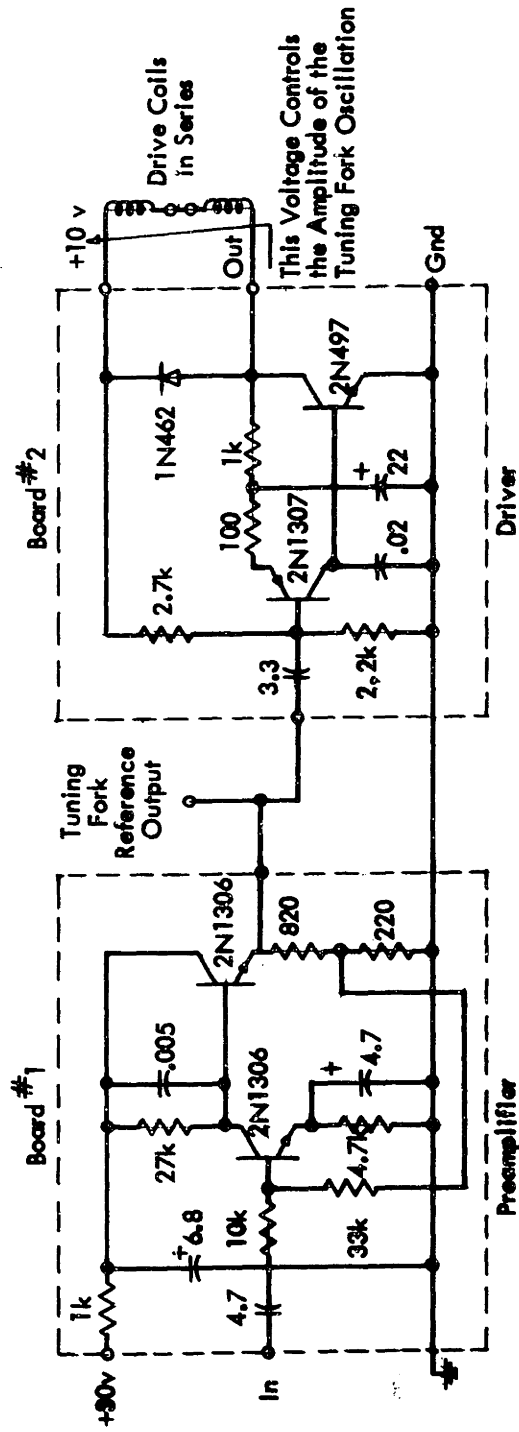


Fig. B-1 Tuning Fork Drive

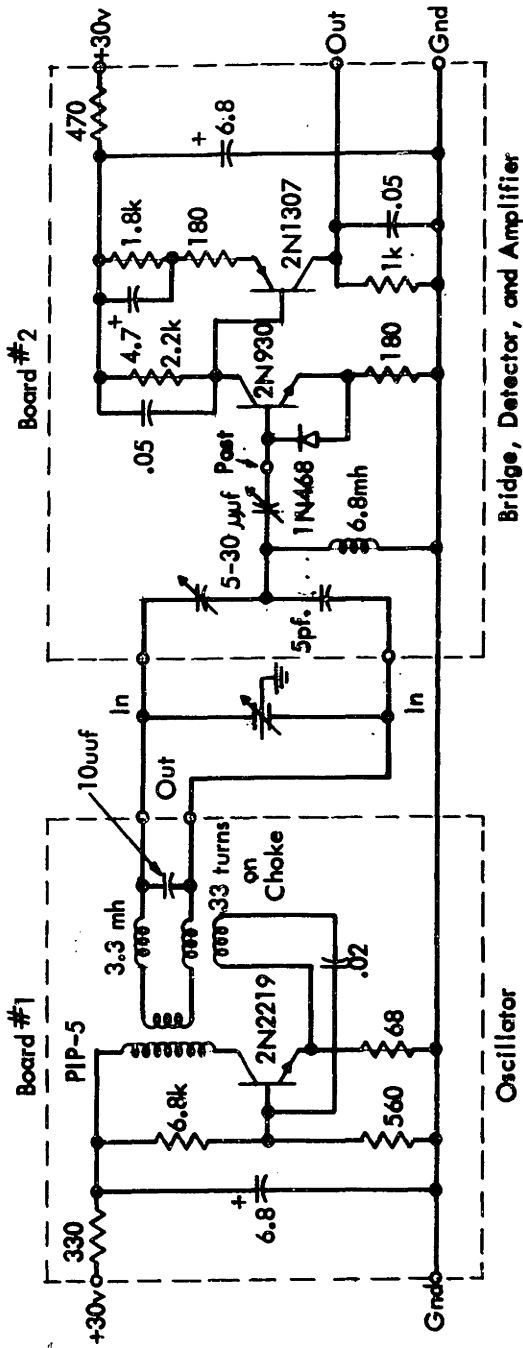


Fig. B-2 Vibration Detector

B

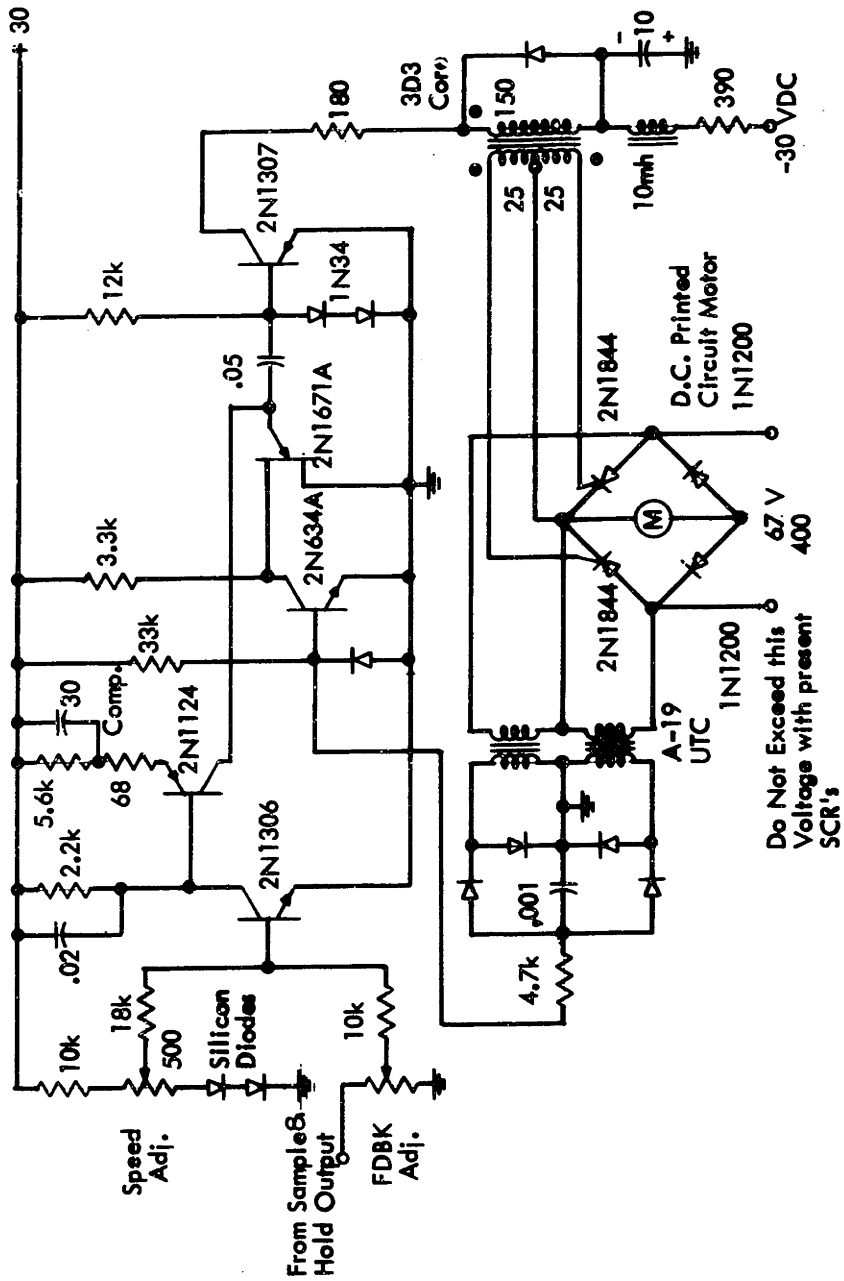


Fig. B-4 SCR Motor Control

APPENDIX C

CROSS COUPLING COMPARISON

In a rotary double-modulated, vibratory rate gyroscope a torsional oscillation of the rotating member about a rate sensitive axis at twice the double modulation frequency cannot be distinguished from an input angular rate. If the amplitude of this oscillation is ϕ_o , the angular rate seen by the input axis of the basic double-modulated instrument is

$$\frac{\partial}{\partial t} [(\phi_o \cos 2\omega_m t) \cos \omega_m t] = \frac{\theta_o \omega_m}{2} [\sin \omega_m t + 3 \sin 3\omega_m t] \quad (C.1)$$

Therefore, a zero rate error of $\frac{\theta_o \omega_m}{2}$ appears at the instrument output.

In a rotary drive, drive frequency output, vibratory rate gyroscope a torsional oscillation of the rotating member about a rate sensitive axis at twice the double modulation frequency couples energy into the sensing system at the drive frequency and its third harmonic. If a lumped parameter sensing system is assumed and the instrument configuration is similar to that of Figure 4.6a, the approximate transfer function relating the amplitudes of the fundamental oscillation of the sensing system and oscillation of the rotating member is

$$\frac{\theta(s)}{\phi_o(s)} = \frac{s^2}{2(s^2 + \frac{\omega_n}{Q_s} s + \omega_n^2)}, \quad (s = j\omega_d) \quad (C.2)$$

where ϕ_o is the amplitude of the torsional oscillation of the rotating member and θ is the amplitude of the instrument output oscillation. Since the transfer function for constant input angular rates in the same instrument is

$$\frac{\theta(s)}{\Omega_i(s)} = \frac{a_{R1} \omega_d}{(s^2 + \frac{\omega_n}{Q_s} s + \omega_n^2)}, \quad (s = j\omega_d)$$

the torsional oscillation ϕ_o causes a zero rate error of $\frac{\omega_d \phi_o}{2a_{R1}}$.

APPENDIX D

OPTIMIZATION PROBLEM WITH FIXED ELEMENTS FOLLOWING THE OPTIMUM FILTER

When the optimum LTV filter is approximated by h discrete linear time-invariant filters $g_k(nT)$ and h periodic multipliers $p_k(t)$ as

$$g_o(t, nT) = \sum_{k=0}^{h-1} T g_k(nT) p_k(t) \quad (D-1)$$

where $p_k(t)$ is the periodic waveform shown in Fig. 7.14, then the error from an ensemble member at time pT can be written as

$$e(pT) = d(pT) - \sum_{h=0}^{\infty} \sum_{m=0}^{\infty} T [f_i(pT - nT - mT)] \sum_{k=0}^{h-1} g_f^k(mT) g_{p-k}(nT) \quad (D-2)$$

where

$$g_f^k(mT) = g_f(\tau_1) p_k(\tau_1) \quad (D-3)$$

$$g_{-q}(nT) = -g_{n-q}(nT) \quad (D-4)$$

The time average mean square error can now be minimized by using the calculus of variations technique on the expression for the time average mean square error.

$$\frac{\partial}{\partial g_i(nT)} \left[\sum_{p=0}^{h-1} \overline{e^2(pT)} \right] = 0, \quad i = 0, 1, \dots, h-1 \quad (D-5)$$

If we define $p - k \triangleq -i$ and carry out this partial differentiation, the following equation results and must be satisfied for $n'T \geq 0$

$$\sum_{p=0}^{h-1} \sum_{m'} g_f^{p+i}(m'T) \overline{d(pT) f_i(pT - n'T - m'T)} = \sum_{m'} g_f^{p+i}(m'T) \overline{f_i(pT - n'T - m'T)} \sum_n \sum_m \overline{f_i(pT - n'T - m'T)} \sum_{k'=0}^{h-1} g_f^{k'}(mT) g_{p-k'}(nT) \quad (D-6)$$

Since m' can only take on the following discrete values

$$m' = (p+i), (p+i) + h, (p+i) + 2h, \dots \quad (D-7)$$

it can be shown that the summation of the first term in Eq. D-6 over p results in the following expression,

$$\sum_{m'} g_f(m'T) \phi_{md}(m'T + n'T) \cos \omega_c(iT + n'T + m'T) \quad (D-8)$$

if

$$f_1(t) = m(t) \cos \omega_c t + q(t) \sin \omega_c t + n(t) \quad (D-9)$$

If we define $p - k' \triangleq -j$, it follows from Eq. D-3 that m can only take on the following discrete values.

$$m = (p + j), (p + j) + h, (p + j) + 2h \dots \quad (D-10)$$

Therefore the summation of the second term in Eq. D-6 over p results in the following expression.

$$T \sum_{m'} \sum_{m} \sum_n g_f(m'T) g_f(mT) \{\phi\} \sum_{j=0}^{h-1} g_{-j}(nT) \quad (D-11)$$

where

$$\phi \triangleq \left\{ \begin{array}{l} \left[\frac{1}{2} \phi_{mm}(x) + \frac{1}{2} \phi_{qq}(x) + \phi_{nn}(x) \right] \cos \omega_c [(j-i)T + mT - m'T + nT - n'T] \\ \left[\frac{1}{2} \phi_{mm}(x) - \frac{1}{2} \phi_{qq}(x) \right] \cos \omega_c [(i+j)T + mT + m'T + nT + n'T] \end{array} \right\}$$

If the summations over m and m' are replaced by integrations, we obtain the following discrete equation which must be satisfied for $n'T \geq 0$.

$$\int_0^{\infty} d\xi' g_f(\xi') \phi_{md}(\xi' + n'T) \cos \omega_c(iT + n'T + \xi') =$$

$$T \sum_{j=1}^{h-1} g_{-j}(nT) \int_0^{\infty} d\xi' g_f(\xi') \times \int_0^{\infty} d\xi g_f(\xi).$$

$$\left\{ \begin{array}{l} \left[\frac{1}{2} \phi_{mm}(x) + \frac{1}{2} \phi_{qq}(x) + \phi_{nn}(x) \right] \cos \omega_c [(i-j)T + \xi - \xi' + nT - n'T] \\ \left[\frac{1}{2} \phi_{mm}(x) - \frac{1}{2} \phi_{qq}(x) \right] \cos \omega_c [(i+j)T + \xi + \xi' + nT + n'T] \end{array} \right\}$$

where $x = \xi - \xi' + nT - n'T$

(D-13)

Recall that this is only the i th equation and all h of these i equations must be solved simultaneously in order to determine the h optimum filters. Therefore it is necessary to solve the following matrix equation.

$$\phi_{id}(i, n) = [\phi_{ii}(i, n, j)] g_{-j}(n) \quad (D-14)$$

where

$$\phi_{id}(i, n) \triangleq \int_0^{\infty} d\xi' g_f(\xi') \phi_{md}(\xi' + nT) \cos \omega_c(\xi' + iT + nT) \quad (D-15)$$

$$[\phi_{ii}(i, n, j)] \triangleq \int_0^{\infty} d\xi' g_f(\xi') \int_0^{\infty} d\xi g_f(\xi) \cdot \left\{ \begin{array}{l} \left[\frac{\phi_{mm}(x)}{2} + \frac{\phi_{qq}(x)}{2} + \phi_{nn}(x) \right] \cos \omega_c[(i-j)T + \xi - \xi' + (n-n')T] \\ \left[\frac{\phi_{mm}(x)}{2} - \frac{\phi_{qq}(x)}{2} \right] \cos \omega_c[(i+j)T + \xi + \xi' + nT + n'T] \end{array} \right\} \quad (D-16)$$

$$g_{-j}(n) \triangleq T g_{-j}(nT) \quad (D-17)$$

These h matrix equations can be written as

$$\begin{array}{lll} \phi_{id}(0, n') & [\phi_{ii}(0, n', n, 0)] & [\phi_{ii}(0, n', n, h-1)] \\ \phi_{id}(1, n') & [\phi_{ii}(1, n', n, 0)] g_0(nT) + \dots + & [\phi_{ii}(1, n', n, h-1)] g_{-h+1}(nT) \\ \vdots & \vdots & \\ \phi_{id}(h-1, n') & [\phi_{ii}(h-1, n', n, 0)] & [\phi_{ii}(h-1, n', n, h-1)] \end{array} \quad (D-18)$$

When the h optimum filter are constrained such that they have only a finite memory

$$0 \leq n \leq (N-1) \quad (D-19)$$

it is necessary to invert an $h \times h$ square symmetric matrix in which each element is an $N \times N$ square symmetric matrix in order to determine the hN

elements of the optimum filter. This can be done on a digital computer by using the Gauss Jordan reduction procedure or a modification thereof to reduce roundoff errors.

Recall that the Gauss Jordan procedure for solving a set of h simultaneous equations is just a systematic procedure for combining the equations to eliminate variables and finally end up with a matrix equation of the same form but in which the square $h \times h$ matrix is no longer symmetric but instead has zeros for all the elements below the diagonal. Therefore one unknown is immediately known from the last equation and it can be used to determine another from the next to last equation, etc., etc., until all the unknowns have been determined.

When this procedure is used in the present problem, the computation time is roughly proportional to $N^3 h^3$ for large N and h because the multiplication of two $N \times N$ matrices requires that the computer perform N^3 multiplications and N^3 additions and the number of times it is necessary to do this is proportional to h^3 in order to reduce the equations to a form in which the elements below the diagonal are zero. Once the equations are in this form the computation time to determine the Nh elements is proportional to $h^2 N^3$. Therefore the computation time required to solve the filtering problem with fixed elements following the optimum demodulator will always be large since the number $h^3 N^3$ is large. Therefore it is doubtful that this method is of much practical use except in problems where a digital computer is actually going to be used for the signal processor.

REFERENCES

- 1.1 Pringle, J.W.S., Phil. Trans. Roy. Soc., London, Ser. B. Vol. 23, 347, 1948.
- 1.2 Goldstein, H., Classical Mechanics, 139, 142, Addison Wesley, 1953.
- 1.3 Ising, G., "A Natural Limit for the Sensitivity of Galvanometers," Philosophical Magazine, Vol. 1, 827, 1926.
- 1.4 Newton, G.C., Jr., "Comparison of Vibratory and Rotating-Wheel Gyroscope Rate Indicators," AIEE Trans., Part II, 79, 143-150, 1960.
- 3.1 Goldstein, H., Classical Mechanics, 139, 142, Addison Wesley, 1953.
- 3.2 Quick, W.H. and Barnes, G.A., "Inertial Gyro Research," Technical Documentary Report ASD-TDR-62-898, 1963 (Secret).
- 3.3 Slater, J., "Gyro and Free Rotor Gyro," Control Engineering, November, 1962.
- 3.4 Diamantides, N.D., "The Gyrovibratory," IRE Trans. ANE-6, 16-25, March, 1959
- 3.5 Birdsall, F.H., "Means and Apparatus for Utilizing Gyrodynamic Energy", U.S. Patent No. 2, 716, 893, 1955.
- 3.6 Bowden, B.V., "Gyroscopes," U.S. Patent No. 2, 991, 659, 1961.
- 3.7 Newton, G.C., Jr., "A Rotary-Drive, Vibratory-Output Gyroscopic Instrument," Second International Congress on Automatic Control, Basel, Switzerland, Paper 402, 1963.
- 3.8 Grangvist, C.E., "Gyroscope Device with Vibratory Gas Particles or Particles of Another Sound Transferring Medium," U.S. Patent No. 2, 999, 389, 1961.
- 3.9 Swerdlow, R. and Whitman J.C., Jr., "A Vibratory Acoustic Gyroscope," Electronic Systems Laboratory, M.I.T., Technical Memo. ESL-TM-162, 1962.
- 3.10 Bush, R.W., "Memorandum to File--Acoustic Gyroscope--Mod. No. 3," Electronic Systems Laboratory, M.I.T., February 4, 1964.
- 3.11 Newton, G.C., Jr., "The Gyroflex Principle for Angular Motion Measurement," Report 1503, Lake States Oil Company, Iron Mountain, Michigan, 1958.

REFERENCES (Continued)

- 3.12 Bush, R.W., "Vibrating Ring Analysis for Inertial Rate Sensor," S.M. Thesis, Department of Electrical Engineering, M.I.T., 1961.
- 3.13 Bush, R.W., "Preliminary Analysis of Five Possible Vibratory-Output Angular Motion Sensors," Electronic Systems Laboratory, M.I.T., Report ESL-SR-130, 1961.
- 3.14 Lyman, J., "New Space Rate Sensing Instrument," Aero. Eng. Rev., Vol. 12, 24-30, 1953.
- 3.15 Gulton Industries Inc., "Crystal Rate Gyro," ASD Technical Report 61-253, 1961.
- 3.16 Clevite Research Center, "Final Report on Piezoelectric Rate-of-Turn Indicator," Contract Nonr. 2210(00), Office of Naval Research, 1957.
- 3.17 Westinghouse Electric Corporation, "Solid State Gyroscope Technique Study," Technical Documentary Report ASD-TDR-67-81, 1962.
- 3.18
- 3.19 Op. Cit. (3.16).
- 3.20 Op. Cit. (3.13).
- 3.21 Bush, R.W. and Swerdlow, R., "Research on Vibratory-Output Angular Motion Sensors," Electronic Systems Laboratory, M.I.T., Report ESL-SR-130, 1961.
- 4.1 Truxal, J.C., Control System Synthesis, 399-403, McGraw-Hill, 1955.
- 4.2 Danthine, A.A.S., "Modulation Transfer Characteristics of Lightly Damped Distributed Parameter Systems," Electronic Systems Laboratory, M.I.T., Report ESL-R-121, 1961.
- 4.3 Eisberg, R.M., Fundamentals of Modern Physics, John Wiley and Sons, 1961.
- 4.4 Lawson, J.L., and Uhlenbick, G.E., Threshold Signals, Chapter 4, McGraw Hill, Radiation Lab. Series, 1950.
- 5.1 Op. Cit. (3.14).

REFERENCES (Continued)

- 5.2 Sperry Gyroscope Company, "Final Engineering Report on the Second Portion of Gyrotron II Angular Rate Tachometer Development Program," Document No. AD-2296 V (Report dated 1953), ASTIA.
- 5.3 Sperry Gyroscope Company, "Final Engineering Report on Gyrotron IA Vibratory Rate Gyro Development Program," Document No. AD-40494 (Report dated 1954), ASTIA.
- 5.4 Hunt, G.H., "Some Sources of Error in the Tuning Fork Gyroscope," Technical Note I. A. P. 1139, Royal Aircraft Establishment, Farnborough, Hants, April, 1962.
- 5.5 Hobbs, A.E.W., "The A-5 Tuning Fork Gyro as a Single Axis Rate and Angle Sensor," Technical Note I.E.E.34, Royal Aircraft Establishment, Farnborough, Hants, September, 1963.
- 5.6 Morrow, C.T., "Zero Signals in Sperry Tuning Fork Gyrotron," J. Acoustical Society of America, 27, 581-585, 1955.
- 5.7 Newton, G.C., Jr., "Vibratory Rate Gyroscopes," M.I.T., Electronic Systems Laboratory, Report ESL-R-161, 1962.
- 5.8 Hawley, M.S., "The Condenser Microphone as an Acoustic Standard," Bell Laboratories Record, 33, 6-10, 1955.
- 5.9 Time and Frequency Tuning Forks, Batavia, Illinois, Specification Sheet 10M-1/62.
- 5.10 Sprague Engineering Bulletin, North Adams, Massachusetts, No. 2053A, February, 1958 and No. 7025A, July, 1962.
- 5.11 Marlin Rockwell Corporation, Jamestown, N. Y., "MRC Super-Precision Ball Bearings for Machine Tools," Form 1531-1.
- 6.1 Newton, G.C., Jr., "A Program for Research on Vibratory Output Angular Motion Sensors," Electronic Systems Laboratory, M.I.T., March 28, 1960.
- 6.2 Bush, R.W. and Newton, G.C., Jr., "Reduction of Errors in Vibratory Gyroscopes by Double Modulation," Preprint-1964, Joint Automatic Control Conference, Stanford California, Paper 1 Session XVIII, 1964.
- 7.1 Viterbi, A.J., "Optimum Coherent Demodulation for Continuous Modulation System," Proc. Nat. Elect. Conf., Vol. 18, 498-508, 1962.
- 7.2 Weiss, G., "A Network Theory for Carrier-Suppressed Modulated Systems," IRE Trans. on Auto. Control, 54-66, February, 1961.

REFERENCES (Continued)

- 7.3 Thomas, J.B. and Wong, E., "On the Statistical Theory of Optimum Demodulation," IRE Trans. on Info. Theory, Vol. IT-6, No. 4, 420, 1960.
- 7.4 Booton, R.C., Jr., "An Optimization Theory for Time-Varying Linear Systems with Nonstationary Statistical Inputs," Proceedings of the I.R.E., 977-81, 1952.
- 7.5 Zadeh, L.A., "Time Varying Networks," Proc. IRE, Vol. 49, No. 10, 148-1503, 1961.
- 7.6 Newton, G.C., Jr., Gould, L.A., and Kaiser, J.F., Analytical Design of Linear Feedback Controls, Wiley and Sons, Inc., New York, 1957.
- 7.7 DeRusso, P.M., "Ultimate Performance Limitations of Linear Sampled-Data Systems," M.I.T., Electronic Systems Laboratory, Report 7849-R-8, 1959.
- 7.8 Booton, R.C., Jr., and Goldstein, M.H., Jr., "Design and Optimization of Synchronous Demodulators," IRE Wescon Conv. Record, Part 2, 154-170, 1957.
- 7.9 Newton, G.C., Jr., "A Compensating Network Theory for Feedback Control Systems Subject to Saturation," M.I.T. Doctorate Thesis, May, 1950, Section 7.2.
- 7.10 D'Appolito, J.A., "Zero-Rate Error in a Vibratory Double-Modulation Gyroscope," Electronic Systems Laboratory, M.I.T., ESL-R-207, 1964.

BIOGRAPHICAL SKETCH

Richard Wright Bush was born on October 25, 1937 in Abington, Pennsylvania. At the age of 2 he moved to La Grange, Illinois. He attended LaGrange and La Grange Park public schools and was graduated from Lyons Township High School in 1955. From the fall of 1955 to June 1959, he attended Purdue University. He graduated from Purdue in 1959 with a B.S. in Electrical Engineering.

After a summer with Minneapolis-Honeywell in Minneapolis, Minnesota, he entered the M.I.T. Graduate School of Electrical Engineering with a Honeywell Fellowship.

In 1961, he married Suzanne Smith from La Grange, Illinois. Later in 1961 he received his S.M. degree in Electrical Engineering with a thesis entitled "Vibratory Ring Analysis for an Inertial Angular Rate Sensor". In 1962 he received an Electrical Engineer degree from M.I.T. Since 1962 he has been a research assistant in the M.I.T. Electronic Systems Laboratory.

Mr. Bush is a member of Sigma Xi, Tau Beta Pi, Eta Kappa Nu and the Institute of Electrical and Electronic Engineers.