Technical Report No. 32-819

Theory and Practical Design of Phase-Locked Receivers

Volume I

Robert C. Tausworthe

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

GPO PRICE $  
CFSTI PRICE(S) $  
Hard copy (HC) $  
Microfiche (MF) $  

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA
February 15, 1966
Theory and Practical Design of Phase-Locked Receivers
Volume I

Robert C. Tausworthe

Richard M. Goldstein, Manager
Communications Systems Research Section

February 15, 1966
BLANK PAGE
# CONTENTS

## PART I

**THE THEORY OF PHASE-LOCKED LOOPS**

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction and History of the Phase-Locked Loop</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>References for Chapter 1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>Fundamental Concepts</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>2-A. Statistics</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>2-B. Linear Filtering</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>2-C. Noise Bandwidth</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>2-D. Sinusoidal Filter Inputs</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2-E. Fiducial Bandwidth</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2-F. Band-Pass Mixers</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2-G. Amplitude and Phase Detectors</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>2-H. Noise</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>1. Thermal Noise</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>2. Shot Noise and Other Problems in Electron Tubes</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>3. Noise With a 1/f Spectrum</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>References for Chapter 2</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>Behavior of the Loop in the Absence of Noise</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>3-A. The Basic Integro-Differential Equation</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>3-B. Tracking When the Loop Error Is Small</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>3-C. Acquiring Lock in the First-Order Loop</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>3-D. Acquiring Lock in the Second-Order Loop With Passive Loop Filter</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>3-E. Tuning the VCO</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Reference for Chapter 3</td>
<td>24</td>
</tr>
<tr>
<td>4</td>
<td>Behavior of Phase-Locked Loops With Stochastic Inputs</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>4-A. Development of a Mathematical Model and a Basic Loop Equation</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>4-B. Discussion of Mathematical Models</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>4-C. Spectrum of the Loop Noise</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>References for Chapter 4</td>
<td>27</td>
</tr>
<tr>
<td>5</td>
<td>The Linearized Analysis of Phase-Locked Systems</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>5-A. Behavior of a Linear Loop</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>5-B. Calculation of Loop Bandwidth</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>1. First-Order Loop</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>2. Second-Order Loop, Passive Integrator</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>3. Second-Order Loop, Perfect Integrator</td>
<td>31</td>
</tr>
</tbody>
</table>
CONTENTS (Cont'd)

5-C. Optimization of Loop Parameters .......................... 31
5-D. The Effects of VCO Noise ................................. 31
   1. Optimization of \( w_L \) and \( r \) ...................... 32
   2. An Example ........................................... 33
   3. Conclusion ............................................ 33

References for Chapter 5 ........................................... 33

Chapter 6. Optimized Design of Tracking Filters
(Linear Analysis) .............................................. 34

6-A. Tracking Loop Design .................................... 34
6-B. Optimum Filter for Random Phase Offset ................. 34
6-C. Optimum Filter for Frequency and Random
     Phase Offset ........................................... 34
     1. Choice of Parameters ................................ 35
     2. Evaluation of Transient Error ....................... 37
     3. The Effects of Doppler Rates in
        Second-Order Loops .................................. 37
     4. Comments on the Choice of \( r \) in
        Second-Order Tracking Loops ......................... 37
     5. The Case With VCO Noise Included ................. 37

Reference for Chapter 6 ........................................... 38

Chapter 7. The Design of Synchronous-Detector AGC
Systems .......................................................... 39

7-A. The Synchronous-Amplitude-Detector AGC Loop .......... 39
7-B. AGC Stability ............................................ 41
7-C. Calibration Equation ..................................... 41
7-D. Dynamic Behavior of AGC Loops ......................... 42

Reference for Chapter 7 ........................................... 43

Chapter 8. The Double-Heterodyne Phase-Locked Receiver .... 44

8-A. Basic Configuration of the Receiver .................... 44
8-B. Effects of Band-Pass Limiting .......................... 47

References for Chapter 8 ........................................... 48

Chapter 9. Behavior of Phase-Locked Loops
(Nonlinear Analysis) ........................................ 49

9-A. The Spectral Equation ................................... 49
9-B. Calculation of \( \gamma \) ................................ 51
9-C. Calculation of \( \gamma^2 \) ................................ 51
9-D. Behavior of the First-Order Loop ..................... 52
     1. Linear Loop ....................................... 52
     2. Quasi-linear Loop .................................. 52
     3. Spectral Approximation ............................. 52
     4. Reduction Modulo \( 2\pi \) ............................ 52
## CONTENTS (Cont'd)

9-E. Calculation of the Behavior of the Second-Order Loop ........................................ 53
1. Linear Analysis ............................................................................................................. 54
2. Quasi-linear Analysis ................................................................................................... 54
3. Linear Spectral Analysis ............................................................................................... 54
4. Conclusions About the Second-Order Loop ............................................................... 55

References for Chapter 9 .................................................................................................. 56

Chapter 10. Designing a Double-Heterodyne Tracking Loop .............................................. 57

10-A. Definition of Receiver Threshold ............................................................................. 57
10-B. Tracking Loop Performance of the Double-Heterodyne Receiver ......................... 57
10-C. Nonlinear Behavior of the Double-Heterodyne Receiver ....................................... 58
10-D. The Signal Level Producing $\sigma^2 = 1$ .................................................................. 59
10-E. Choice of Receiver Parameters ............................................................................... 60

## PART II

**SUMMARY OF PHASE-LOCKED LOOP DESIGN**

Chapter 1. Introduction and History of the Phase-Locked Loop ........................................ 62

Chapter 2. Fundamental Concepts ..................................................................................... 63

2-A. Statistics .................................................................................................................. 63
2-B. Linear Filtering ......................................................................................................... 63
2-C. Noise Bandwidth ....................................................................................................... 64
2-D. Fiducial Bandwidth ................................................................................................... 64
2-E. Sinusoidal Filter Inputs ............................................................................................ 64
2-F. Band-Pass Mixers ..................................................................................................... 64
2-H. Noise ....................................................................................................................... 65

Chapter 3. Behavior of the Loop in the Absence of Noise .................................................. 66

3-A. The Basic Integro-Differential Equation ................................................................... 66
3-B. Tracking When the Loop Error Is Small .................................................................. 66
3-C. Acquiring Lock in the First-Order Loop ................................................................... 66
3-D. Acquiring Lock in the Second-Order Loop With Passive Loop Filter .................... 66

Chapter 4. Behavior of Phase-Locked Loops With Stochastic Inputs .................................. 72
CONTENTS (Cont’d)

Chapter 5. The Linearized Analysis of Phase-Locked Systems  . 73
  5-A. Behavior of the Linear Loop  . 73
  5-B. Calculation of Loop Bandwidth  . 74
    1. First-Order Loop  . 74
    2. Second-Order Loop, Passive Integrator  . 74
    3. Second-Order Loop, Perfect Integrator  . 75
  5-C. Optimization of Loop Parameters  . 75
  5-D. Effects of VCO Noise  . 75

Chapter 6. Optimized Design of Tracking Filters
  (Linear Analysis)  . 76
  6-A. Tracking Loop Design  . 76
  6-B. Optimum Filter for Random Phase Offset  . 76
  6-C. Optimum Filter for Frequency and
      Random Phase Offset  . 76
    1. Choice of Parameters  . 76
    3. The Effects of Doppler Rates in
      Second-Order Loops  . 77
    4. Comments on the Choice of $r$ in
      Second-Order Tracking Loops  . 77

Chapter 7. Design of Synchronous Detector AGC Systems  . 79
  7-A. The Synchronous-Amplitude-Detector AGC Loop  . 79
  7-C. Calibration Equation  . 79

Chapter 8. The Double-Heterodyne Phase-Locked Receiver  . 81
  8-A. Basic Configuration of the Receiver  . 81
  8-B. Effects of Band-Pass Limiting  . 81

Chapter 9. Behavior of Phase-Locked Loops
  (Nonlinear Analysis)  . 83
  9-A. The Spectral Equation  . 83
  9-D. Behavior of the First-Order Loop  . 83
  9-E. Calculation of Behavior of the
      Second-Order Loop  . 86

Chapter 10. Designing a Double-Heterodyne Tracking Loop  . 87
  10-A. Definition of Receiver Threshold  . 87
  10-B. Tracking Loop Performance of the
      Double-Heterodyne Receiver  . 87
  10-D. The Signal Level Producing $\sigma^2 = 1$  . 88
  10-E. Choice of Receiver Parameters  . 88

Appendix. Nomenclature  . 90
FIGURES

PART I

1-1. Basic configuration of a simple phase-locked loop 2

1-2. A mathematically equivalent model of the simple phase-locked loop 2

2-1. Filtering device 5

2-2. Response of filter to a unit-impulse function 6

2-3. Equivalent noise bandwidth 7

2-4. Double-sided frequency response 7

2-5. Double- and single-sided spectra 7

2-6. The simple product-mixer 9

2-7. Choosing filter bandwidths to avoid the image noise problem 9

2-8. Thevenin and Norton equivalent circuits of noisy resistors 11

2-9. Total Johnson noise current- and voltage-squared relations in simple circuits 12

2-10. Tuned circuit for investigating the spectral density of $v_n(t)$ 13

3-1. The basic phase-locked loop 16

3-2. First-order loop pull-in behavior 18

3-3. Lock-in behavior of a second-order loop with imperfect integrator, $F(s) = (1 + \tau_r s)/(1 + \tau_is)$, for $\Omega_\omega/AK = 0.4$ and $AK\tau_\omega/\tau_1 = 2$ 20

3-4. Lock-in behavior of a second-order loop with imperfect integrator, $F(s) = (1 + \tau_r s)/(1 + \tau_is)$, for $\Omega_\omega/AK = 0.6$ and $AK\tau_\omega/\tau_1 = 2$ 21

3-5. Phase-plane trajectory of a second-order loop with perfect integrator to a doppler-rate input $\Lambda_\omega$, for $AK\tau_\omega/\tau_1 = 2$, and $\Lambda_\omega = AK/2\tau_1$ 23

3-6. Normalized maximum doppler rate, $k/r = \Lambda_\omega \tau_1/AK$, for which lock is guaranteed in the absence of noise, as a function of the loop parameter $r = AK\tau_\omega/\tau_1$ 24

4-1. Exact mathematical equivalent of the phase-locked loop 26

4-2. Equivalent exact mathematical model of phase-locked loop, with explicit reduction of $\phi \mod 2\pi$ 27

5-1. The linearized model of the phase-locked loop 29

5-2. Passive integrator loop filter 30

5-3. Variation of maximum loop response $L^*$, noise bandwidth $W_\omega$, fiducial bandwidth $W_\omega$, and frequency at maximum loop response, for second-order phase-locked loop, as a function of the parameter $r = AK\tau_\omega/\tau_1$ 30
FIGURES (Cont'd)

5-4. Factors governing relative contributions of VCO noise to output phase noise . 32

6-1. Response of first- and second-order loops to various inputs . 36

7-1. A synchronous-detector AGC loop, using a phase-locked loop to provide a coherent reference. . 39

7-2. Equivalent diagram of an AGC loop . 41

7-3. Measured AGC curve showing departure from linear behavior . 42

8-1. The double-heterodyne receiver . 45

8-2. Davenport's band-pass limiter zonal SNR curve . 47

8-3. The ratio of band-pass limiter output signal-to-noise spectral density to that at input . 48

9-1. Relationship between the variance of \( \phi \) (mod 2\( \pi \)) and that of a Gaussian stationary variate \( \phi \) . 51

9-2. Variation of the parameter \( \gamma \) as a function of the Gaussian variance \( \sigma^2 \), for various forms of \( R_{\phi}(\tau) \) . 52

9-3. Comparison of linear, quasi-linear, and linear-spectral approximate methods with the actual behavior of the first-order loop . 53

9-4. Comparison of linear and nonlinear theories for second-order, constant linear bandwidth loop . 54

9-5. Comparison of phase-noise variances by linear and linear-spectral approximations . 55

9-6. Variation in bandwidth and damping parameters, as a function of signal strength . 55

10-1. Variation in margin producing \( n^2 = 1 \) as a function of threshold design parameter, \( r_n = a_n K_a K_{\psi_{1/2}} M_{F_1}/\tau \) . 59

10-2. Comparison of linear and nonlinear approximations to loop rms phase error, as a function of loop margin . 60

10-3. Variation of loop bandwidths and damping factors, as a function of loop margin . 60

PART II

II-1-1. Basic configuration of a simple phase-locked loop . 62

II-2-1. Filtering device . 63
FIGURES (Cont’d)

II-2-2. Response of filter to a unit-impulse function. 63
II-2-3. Equivalent noise bandwidth. 64
II-2-4. Double-sided frequency response. 64
II-2-5. Double- and single-sided spectra. 64
II-2-6. The simple product-mixer. 65
II-2-7. Choosing filter bandwidths to avoid the image noise problem. 65
II-2-8. Thevenin and Norton equivalent circuits of noisy resistors. 65
II-3-1. The basic phase-locked loop. 67
II-3-2. First-order loop pull-in behavior. 67
II-3-3. Lock-in behavior of a second-order loop with imperfect integrator,
\[ F(s) = \frac{(1 + \tau_1 s)}{(1 + \tau_1 s)}, \text{ for } \Omega_n / AK = 0.4 \text{ and } AK \tau_2^2 / \tau_1 = 2 \] 68
II-3-4. Lock-in behavior of a second-order loop with imperfect integrator,
\[ F(s) = \frac{(1 + \tau_1 s)}{(1 + \tau_1 s)}, \text{ for } \Omega_n / AK = 0.6 \text{ and } AK \tau_2^2 / \tau_1 = 2 \] 69
II-3-5. Phase-plane trajectory of a second-order loop with perfect integrator to a doppler-rate input \( \Lambda D \), for \( AK \tau_2^2 / \tau_1 = 2 \), and \( \Lambda_n = AK / 2 \tau_1 \). 70
II-3-6. Normalized doppler rates \( \Lambda_n / \beta^2 \) and \( \Lambda_n / \omega_n^2 \) for which lock is guaranteed in the absence of noise, as a function of the loop parameter \( r = AK \tau_2^2 / \tau_1 \). 71
II-4-1. Exact mathematical equivalent of the phase-locked loop. 72
II-4-2. Equivalent exact mathematical model of phase-locked loop,
with explicit reduction of \( \phi \) (mod 2\pi). 72
II-5-1. The linearized model of the phase-locked loop. 73
II-5-2. Passive integrator loop filter. 74
II-5-3. Variation of maximum loop response \( L^2 \), noise bandwidth \( W_L \), fiducial bandwidth \( w_L \), and frequency at maximum loop response, for second-order phase-locked loop, as a function of the parameter \( r = AK \tau_2^2 / \tau_1 \). 74
II-5-4. Factors governing relative contributions of VCO noise to output phase noise. 75
II-6-1. Response of first- and second-order loops to various inputs. 78
II-7-1. A synchronous-detector AGC loop, using a phase-locked loop to provide a coherent reference. 79
II-7-2. Equivalent diagram of an AGC loop. 80
II-7-3. Measured AGC curve showing departure from linear behavior. 80
FIGURES (Cont'd)

II-8-1. The double-heterodyne receiver ........................................ 81
II-8-2. Davenport's band-pass limiter zonal SNR curve ...................... 82
II-8-3. The ratio of band-pass limiter output signal-to-noise spectral
density to that at input ............................................................... 82
II-9-2. Variation of the parameter $\xi$ as a function of the Gaussian
variance $\alpha^2$, for various forms of $R_\phi(t)$ .................................. 83
II-9-3. Comparison of linear, quasi-linear, and linear-spectral
approximate methods with the actual behavior of the
first-order loop ............................................................................... 83
II-9-4. Comparison of linear and nonlinear theories for second-order,
constant linear bandwidth loop ......................................................... 84
II-9-5. Comparison of phase-noise variances by linear and linear-spectral
approximations .............................................................................. 85
II-9-6. Variation in bandwidth and damping parameters, as a function
of signal strength ............................................................................. 86
II-10-1. Variation in margin producing $\sigma^2 = 1$ as a function of
threshold design parameter, $r_n = a.K/K_\psi MF_1$ $\frac{T_1}{\tau}$ ................ 87
II-10-2. Comparison of linear and nonlinear approximations to loop rms
phase error, as a function of loop margin ........................................ 88
II-10-3. Variation of loop bandwidths and damping factors, as a
function of loop margin ..................................................................... 89
PREFACE

To say that there have been many articles written over the past few years dealing with phase-locked devices seems almost an understatement. Besides this work, I am aware of two books being written on the subject, those of Viterbi and Van Trees, and there may be still others. This monograph was prompted by a need at the Jet Propulsion Laboratory for a practical handbook for the design, testing, evaluation, and philosophy of phase-locked systems. Such a book should possess several qualities: it should contain all the formulas needed for the accurate design of a complicated receiver; it should provide a unified notation with unambiguous definitions of receiver parameters; it should tell how to pick these parameters, based on any specific job required of a communications system; it should give special attention to the general system design problems and procedures, with all the details included; it should serve as a reference against which a receiver in the field can be checked to see whether or not it is working properly, and thereby meet its proper specification; and, finally, it should contain that philosophy which has evolved to ensure the successful fabrication of the most sensitive, flexible, and stable receiver in the world today. It has been my intention to fulfill these goals in the pages you see here.

The approach I have followed in trying to assemble words and formulas enough to succeed in my aim is that the presentation should expose the reader to the reasoning by which the theory evolved and to a few of the steps necessary for him to follow this evolution mathematically, and should set forth its results in as concise a treatment as precision could permit.

Thus, not only in intent, but also in approach, is the treatment in the following pages different from that to be found elsewhere. The two books mentioned above are theoretical treatments, rigorously argued, and excellent reading for one desiring specific insight to the purer aspects of communications. But the content of these works necessarily ends before most of the detailed systems engineering techniques are evoked. For example, the band-pass-limiter loop is nowhere treated with the detail it is here.
PREFACE (Cont’d)

While this is intended as a working book, by no means have I presumed to consider the practical aspects of building a better VCO, RF mixer, IF amplifier, etc. These are components that develop naturally with the state of the communications art. What is given is meant to provide the systems engineer with the tools he needs to effect a successful analysis of his particular job.

Any exposition of a broad topic mathematically is apt to frustrate the conscientious author almost beyond comprehension in the matter of notation. He soon exhausts both Roman and Greek alphabets, and, as the frustration becomes extreme, he begins to eye script, Russian, Hebrew, and Germanic characters in the desperate hope that he can use separate symbols for all the quantities to be defined, before all of these, too, are used up. Suffice it to say that I have retained only the Roman and Greek at some awkwardness, being forced away, in some cases, from a common notation because of multiple demands on the same letter. An attempt was made to resolve these conflicts by giving the more important quantity precedence. I have also tried to keep subscripts as simple as possible by keeping subscripted subscripts and other such complexities to a minimum.

The subject is treated in two volumes, the first of which consists of ten chapters dealing primarily with the detailed tracking aspects of phase-locked devices. Topics involving modulation in various modes are relegated to the second volume. Each of the volumes is further divided into two portions: the first, a section setting forth theory and design optimization; and the latter, a summary of formulas and methods. It is hoped that the reader will find in these pages as much insight into the mysterious behavior of the phase-locked loop as the author supposes he developed in writing them.
ABSTRACT

This is Volume I of a two-volume work on the theory of phase-locked receivers, with pertinent reference material on practical receiver design. Volume I is primarily devoted to the performance of carrier-tracking loops, including a rigorous treatment of narrow-band systems having IF limiters. The bulk of the work is based on a theoretical linear model, but a nonlinear method is also presented to predict behavior near the threshold. The emphasis throughout the work is toward completeness, simplicity, and internal consistency of the material assembled.

Part I of this Volume is an exposition of the theory, the resulting equations, and the design philosophy that have enabled the phase-lock concept to evolve into the basic principle underlying the most sensitive receivers in the world today. Part II is a condensed version of Part I, intended as a quick reference to formulas, definitions, and salient design considerations.

Author
PART I

THE THEORY OF PHASE-LOCKED LOOPS
CHAPTER 1

INTRODUCTION AND HISTORY OF THE PHASE-LOCKED LOOP

While the origins of automatic phase control date back to the 1920's and 30's, the first serious application of the concept began as a horizontal-line synchronizing device for television in the late 40's. Shortly thereafter, Jaffee and Rechtin showed how a phase-locked loop could be used as a tracking filter for a missile beacon, and how the loop parameters could best be specified. The first analysis including the effects of noise appeared in a paper published by Jaffee and Rechtin in 1955.

Basically, a phase-locked loop is an electronic servomechanism that operates as a coherent detector by continuously correcting the frequency of its local oscillator according to a measurement of the error between the phase of the incoming signal and that of its local oscillator. The simplest form of loop is shown in Fig. 1-1. The precise relationship between the input and response functions is a nonlinear integro-differential equation from which very little information concerning loop behavior is analytically available, in the general case at least. With a very few nonrestrictive assumptions mathematically the configuration given in Fig. 1-1 can be replaced by that in Fig. 1-2 (we shall indicate precisely why this is so a little later). This model first appeared, without proof, in a paper by Develet in 1956, but in the absence of noise it had been used by several authors, notably Gruen.

Viterbi found solutions for a number of loop filters and various input frequency functions in the no-noise case by analog simulation.

The case in which additive noise is present has been treated by a variety of approximate methods. The first approach, by Jaffe and Rechtin, essentially replaced the sinusoidal nonlinearity of the model of Fig. 1-2 by a linear amplifier of gain A, a case applicable when the phase error is very small. Margolis analyzed the nonlinear operation in the presence of noise by perturbation methods, obtaining a series solution for the loop differential equation, and, using only the first few terms of the series, he determined approximate moments of the phase error. Develet applied Booton's quasi-linearization technique to replace the sinusoidal nonlinearity by a linear amplifier whose gain is the expected gain of the device. More recently, Van Trees obtained a Volterra series representation of the closed-loop response by a perturbation method similar to the method employed by Margolis, but with the advantage of the simplified model he obtained more extensive results. Fokker-Planck, or continuous random-walk, techniques yield exact expressions for the statistics of the random phase error process. Unfortunately, expressions in closed form are available only for the first-order loop (i.e., when the filter is omitted). Such techniques were first applied to this problem by Tikhonov, who was able to determine the steady-state probability distribution of the first-order loop phase error and an approximate expression for the distribution when the loop contains a one-stage RC filter. Viterbi extended this work on the first-order
loop also to obtain the mean time to loss-of-lock (giving the frequency of skipping cycles).

In what follows, we shall review many of these analyses and add another, due to the author, wherein the spectral density of the phase process is approximated. This method, incidentally, is conceptually as simple as the linear approximation methods and yields results almost as exact as the results gotten by Fokker-Planck techniques over the entire useful range of the device.

These analyses have not been merely of academic interest. They have paved the way for building the most narrow-band, sensitive, flexible receivers in the world. Phase-locked loops are used as filters to "clean up" the output of frequency multipliers. The phase-lock principle has been used in ranging devices and in radar systems capable of tracking a planet (Venus) with range-jitter less than 500 meters! It has been used to synchronize telemetry data, to derive bit and word synchronization... and the list could go on at some length.

Because there is such a long list of jobs it can do so well, it is only natural that the phase-locked loop has received attention. Significant portions of the fruits of this attention are revealed in the ensuing chapters.

REFERENCES FOR CHAPTER 1


CHAPTER 2
FUNDAMENTAL CONCEPTS

In this Chapter we shall define such concepts as auto- and cross-correlation functions, spectral density, bandwidths, etc. This defining must be done carefully, as it can lead to one of the prime sources of reader confusion. We shall assume that the reader is familiar with certain statistical properties of signals and noise, so that the exposition here need not be a tutorial one, but merely descriptive.

2-A. Statistics

First, consider a single time function \( x(t) \). Ask yourself: Can \( x(t) \) be specified exactly at every value of \( t \)? If it is, \( x(t) \) is said to be deterministic. Otherwise, there is uncertainty about the character of \( x(t) \). In such cases, any particular observed waveform \( x(t) \) is said to be a sample function of a random process (others call \( x(t) \) a stochastic process). There are varying degrees of randomness one can think of here. For example, \( x(t) = \cos t \) is a well defined, deterministic process. However, \( x(t) = A \cos t \), in which \( A \) can assume any arbitrary value (according to some probability law) is a sample function from a random process, even though we can measure \( x(0) = A \), and from then on, \( x(t) \) is known exactly. The randomness here is evident, for if we were presented with another sample function from the same process, we would probably have a different \( A \).

Another example that occurs frequently is the sample function \( x(t) = \cos(t + \theta) \), where \( \theta \) can assume any value over \([0, 2\pi]\) with equal likelihood. Again, \( \theta \) can be found for any particular observed \( x(t) \) very easily, and forever thereafter \( x(t) \) is known exactly. However, there is a great difference in the types of processes represented by \( A \cos t \) and \( \cos(t + \theta) \). If one were to ask "what is the mean value of \( x(t) \) at time \( t = t_1 \)?" in the first case one would answer "\( E(A) \cos t_1 \)," where \( E(A) \) represents the expected value of \( A \) with respect to its random elements:

\[
E(A) = \int A \, p(A) \, dA \quad (2-1)
\]

whereas, in the second, one would answer "zero," because for every value of \( t_1 \),

\[
E[x(t_1)] = \frac{1}{2\pi} \int_0^{2\pi} \cos(t_1 + \theta) \, d\theta = 0. \quad (2-2)
\]

The averaging operator \( E \) here, we note, is not the same as the operator that averages over time. In fact, in the first case above, the mean value depends on \( t \), and in the second, it does not.

Generally, a process whose statistical behavior is independent of the time origin— as is \( \cos(t + \theta) \)—is said to be a stationary process, while others not exhibiting this property are nonstationary.

Another point that can be made is that, for \( A \cos t \), averages over the entire set, or ensemble, of sample functions are not the same as averaging with respect to time. A process in which the time averages involving all functions of \( x(t) \) are the same as averages taken over the ensemble of sample functions is said to be ergodic. This is a behavior different from stationarity, as witnessed by the fact that the function \( A \cos(t + \theta) \) has its time-mean-square value equal to \( A^2/2 \), whereas its ensemble mean-square value is \( E(A^2)/2 \). Ergodic processes are always stationary, but, as our example shows, the converse need not be true.

If \( x(t) \) is a well-behaved sample function, the expression

\[
R_{xx}(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x(t) x(t + \tau) \, dt \quad (2-3)
\]

defines the time-autocorrelation function of \( x(t) \). The entire process has a statistical analog

\[
R_{xx}(t_1, t_2) = E[x(t_1) x(t_2)] \quad (2-4)
\]

in which averaging is performed with respect to the statistical variables in \( x(t) \). When \( x(t) \) is from a stationary process, the latter is a function of \( t_1 - t_2 \), rather than \( t_1 \) and \( t_2 \) separately, in which case we write

\[
R_{xx}(\tau) = E[x(t) x(t + \tau)]. \quad (2-5)
\]

Any process \( x(t) \), stationary or not, satisfying this last particular equation is called wide-sense stationary. Under the further restriction that \( x(t) \) be ergodic, we have

\[
R_{xx}(\tau) = R_{xx}(\tau). \quad (2-6)
\]

*This is not strictly true under a more formalized mathematical definition of ergodicity, but this formalism is of no concern here.
The spectral density of a function \( x(t) \) is defined as the Fourier transform of its time autocorrelation function,

\[
S_{xx}(\omega) = \int_{-\infty}^{\infty} R_{xx}(\tau) e^{-j\omega \tau} \, d\tau
\]  

(2-7)

and the spectral density of a wide-sense stationary process is defined as

\[
S_{xx}(\omega) = \int_{-\infty}^{\infty} R_{xx}(\tau) e^{-j\omega \tau} \, d\tau.
\]  

(2-8)

When \( x(t) \) is a function from a stationary ergodic process, these spectral densities are equal with probability one,

\[
S_{xx}(\omega) = S_{xx}(\omega).
\]  

(2-9)

We shall usually not make such distinctions in notation in the future. For the most part, we deal with stationary ergodic processes whenever analysis requires a statistical treatment.

Having found \( S_{xx}(\omega) \), we may replace \( \omega \) by the complex frequency variable \( s \); the resulting function \( S_{xx}(s) \) is the analytic continuation of \( S_{xx}(\omega) \) to the entire complex plane. When \( S_{xx} \) is known, \( R_{xx} \) can be found by inverse transformation. It is usual to insert \( \omega = 2\pi f \) in \( S_{xx} \) and treat spectra as functions of \( f \). (Sometimes care must be taken in this step; e.g., \( |\omega| \to (-s)\omega \).)

The mean of \( x(t) \) (i.e., its expected behavior) is denoted

\[
\mu(t) = E[x(t)]
\]  

(2-10)

and its variance (i.e., the mean-square variation about \( \mu(t) \)), is denoted

\[
\sigma^2_x(t) = E[(x(t) - \mu(t))^2] = E[x^2(t)] - \mu^2(t)
\]  

(2-11)

For stationary processes, both \( \sigma^2_x \) and \( \mu \) are independent of \( t \), and, if \( R_{xx}(\infty) \) exists,

\[
\mu_x = R_{xx}(\infty) \quad \sigma_x^2 = R_{xx}(0) - R_{xx}(\infty)
\]  

(2-12)

Equations (2-3) through (2-12) can be altered to yield functions of average products of two different processes, say \( x(t) \) and \( y(t) \). Then, for example,

\[
R_{xy}(t_1, t_2) = E[x(t_1)y(t_2)]
\]

is the cross correlation function of the two processes; when stationarity prevails, we set \( t_2 = t_1 - \tau \):

\[
R_{xy}(\tau) = E[x(t)y(t-\tau)] = R_{yx}(-\tau).
\]

The Fourier transform \( S_{xy}(\omega) \) of \( R_{xy}(\tau) \) is then called a cross-spectral density.

### 2-B. Linear Filtering

Consider a device, which we shall label \( H \) as depicted in Fig. 2-1, having an input \( x(t) \), giving rise to an output \( y(t) \). Suppose first that the input is a very sharp pulse having unit weight; that is, an impulse function \( \delta(t) \):

\[
\delta(t) = \begin{cases} 
1 & \text{for } 0 \leq t \leq T \\
0 & \text{for all other values of } t.
\end{cases}
\]  

(2-13)

As \( T \) approaches zero, \( \delta(t) \) becomes infinite at the origin, and the result is a Dirac delta function \( \delta(t) \). One must be very careful in treating functions of this sort mathematically, because of the lack of uniform convergence. However, for our applications it has been shown that in all practical calculations we may treat \( \delta(t) \) as the limit of \( \delta(t) \) as \( T \to 0 \).

The response of the device \( H \) to \( \delta(t) \) is called the unit-impulse response of the filter, denoted \( h(t) \). That is, \( h(t) \) is the output (Fig. 2-2) of \( H \) at time \( t \) when an impulse was applied at time \( t = 0 \). The filter is called realizable if \( h(t) = 0 \) when \( t < 0 \) (i.e., no response before an input). The filter is said to be linear if superposition holds:

\[
y(t) = \int_{-\infty}^{\infty} x(t_1) h(t-t_1) \, dt_1
\]  

(2-14)
The law states that the output \( y \) at time \( t \) is the superposition of the input \( t \) seconds ago, weighted by the decay response of the filter after \( t \) seconds.

When \( x(t) \) and \( h(t) \) have Fourier or Laplace transforms (call them \( X(s) \) and \( H(s) \)), then for linear filters, \( y(t) \) has a transform \( Y(s) \) related to the others by

\[
Y(s) = X(s)H(s). \tag{2-15}
\]

However, many random processes (notably noise) do not possess transforms, so one cannot perform a system analysis based on \( X(s) \). Instead, one can compute statistics such as the correlation function of the output,

\[
R_y(\tau) = \int_{-\infty}^{\infty} x(t) h(t+\tau) dt. \tag{2-16}
\]

and this expression can be transformed to yield the output power spectral density of the process:

\[
S_{yy}(\omega) = |H(\omega)|^2 S_{xx}(\omega). \tag{2-17}
\]

This equation reveals one of the most important facts about linear filters: the spectral density of the output process is merely the product of the input density times the filter's response. By analytic continuation,

\[
S_{yy}(s) = H(s)H(-s)S_{xx}(s). \tag{2-18}
\]

The power appearing at the filter output can be found by inverse transformation

\[
P_y = R_y(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(\omega)|^2 S_{xx}(\omega) d\omega. \tag{2-19}
\]

### 2-C. Noise Bandwidth

Suppose \( n(t) \) is a zero-mean, ergodic random process of such a nature that \( n(t) \) and \( n(t+\tau) \) are completely uncorrelated for every \( \tau \neq 0 \); that is, \( R_n(\tau) = 0 \) for every \( \tau \neq 0 \). Such a process is often called a white noise process. It is often the character of such noises that

\[
R_n(\tau) = N_0 \delta(\tau). \tag{2-20}
\]

The spectral density of such a noise is thus uniform in the frequency domain

\[
S_n(\omega) = N_0 \quad \text{for all } -\infty < \omega < \infty. \tag{2-21}
\]

It should be pointed out that \( R_n(0) \) indicates \( n(t) \) has infinite power, while \( S_n(\omega) \) shows that over any finite frequency range, the power is finite. The \( N_0 \) above can be thought of either as the weight of the correlation function impulse, or else as the uniform height of the spectral density of the noise.

Now consider what happens when a white noise \( n(t) \) is put into a linear filter \( H(s) \): the output spectral density is

\[
S_{nn}(\omega) = N_0 |H(\omega)|^2 \tag{2-22}
\]

and the output noise power \( N \) from the filter is

\[
N = N_0 \left[ \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(\omega)|^2 d\omega \right]. \tag{2-23}
\]

Whenever the integral in brackets converges, the output noise power is brought down from an infinite value to a finite one by the action of the filter. The filter thus has the effect of limiting the noise, in some sense, to a band of frequencies. Thus we define the effective (or equivalent) noise bandwidth of \( H \) by

\[
W_H = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| H(\omega) \right|^2 d\omega. \tag{2-24}
\]

*Actually \( R_n(0) \) is the mean-square voltage of the process \( y(t) \), so the unit of power we are speaking of here is volts; if a 1-ohm resistance is assumed, the power is in watts. However, for the most part, we shall be dealing with power ratios, such as sign: \( \frac{y}{x} \)-noise, so the units are unimportant, as long as they are consistent.*
In these terms, the output noise power is

$$N = N_0 W_u |H(j\omega)|_{\text{max}}^2.$$  \hspace{1cm} (2-25)

Viewed graphically (Fig. 2-3), $W_u$ is the spectral width of an ideal band-pass filter whose maximum response is the same as that of $H(s)$ and whose output power (in the presence of white noise) is the same as that of $H(s)$.

Some authors prefer to consider a single-sided behavior of filters. They argue that negative frequencies are not "really" observable, and hence that all the concepts involving frequency should be defined accordingly, to involve only positive frequency. This is not efficient mathematically, for one immediately rules out such powerful analysis tools as the ordinary Fourier transform. Hence, these authors are led to a double-sided mathematical analysis and a single-sided interpretation.

For example, the filter response in Fig. 2-4 has noise bandwidth $W = 2B$, whereas most engineers would agree that the band of frequencies passed is only $B$ wide. Working with a single-sided spectral system, as in Fig. 2-5, one fold the negative frequency response onto the positive. The result is the single-sided spectrum, which we give a different notation:

$$G_{\text{ax}}(j\omega) = \begin{cases} 2S_{\text{ax}}(j\omega) & \omega \geq 0 \\ 0 & \omega < 0 \end{cases}$$  \hspace{1cm} (2-26)

![Fig. 2-3. Equivalent noise bandwidth](image)

![Fig. 2-4. Double-sided frequency response](image)

The single-sided equivalent bandwidth is also given a new symbol

$$B_u = \frac{1}{2\pi} \int_{\omega}^{\infty} \frac{|H(j\omega)|^2 d\omega}{|H(j\omega)|_{\text{max}}^2} = \frac{1}{2} W_u$$  \hspace{1cm} (2-27)

and in this new notation, the noise power spectral density is

$$G_{\text{ax}}(j\omega) = \begin{cases} N_0 & \text{for } \omega > 0 \\ 0 & \omega < 0 \end{cases}$$  \hspace{1cm} (2-28)

The resulting noise output is the same as (2-25) but appears written in single-sided parameters as

$$N = N_0 B_u |H(j\omega)|_{\text{max}}^2.$$  \hspace{1cm} (2-25)

There is really no practical advantage to be gained from a single-sided treatment once the engineer realizes that, since physical filters have both positive and negative frequency responses, it is not possible to measure only the positive side or only the negative side of a spectrum by using linear filters—to the contrary, it is only possible to measure the composite effects of both positive and negative frequencies with a filter. On the other hand, one can certainly build a device that finds the autocorrelation function of $x(t)$ and then takes its Fourier transform. With such a device, negative frequency spectra are separately observable; however, the result is a mirror image of...
of the positive frequency spectrum and hence yields no new information. We shall use both concepts more or less interchangeably in this work, and notation will make it clear which is meant in any particular case: \( S(j\omega) \) is a double-sided spectral density, and \( W \) is a double-sided bandwidth, while \( G(i\omega) \) refers to a single-sided spectral density, and \( B \) is a single-sided bandwidth.

2-D. Sinusoidal Filter Inputs

Let a sinusoidal carrier having power \( P \) be inserted into the filter \( H \). We represent this input as the stationary process

\[ x(t) = (2P)^{\text{th}} \cos(\omega_0 t + \theta) \]  

(2-29)

where \( \theta \) is a uniformly distributed random variable, and \( P \) is a constant (namely, the carrier power). Then we have

\[ R_x(\tau) = P \cos \omega_0 \tau \]

\[ S_{xx}(j\omega) = \pi P \left[ \delta(\omega - \omega_0) + \delta(\omega + \omega_0) \right] \]  

(2-30)

The factor of \( \pi \) is present in \( S_{xx} \) because we must satisfy

\[ R_x(0) = P = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{xx}(j\omega) d\omega. \]  

(2-31)

By our formula (2-17), the output spectrum is

\[ S_y(j\omega) = \pi P |H(j\omega)|^2 \left[ \delta(\omega - \omega_0) + \delta(\omega + \omega_0) \right] \]  

(2-32)

and the output signal power (call it \( S \)) is

\[ S = R_y(0) = P |H(j\omega_0)|^2. \]  

(2-33)

If white noise is also present at the input, so that \( x(t) = (2P)^{\text{th}} \cos(\omega_0 t + \theta) + n(t) \), then the output consists of a signal component with power \( S \) and noise with power \( N \):

\[ R_y(0) = S + N. \]  

(2-34)

The resulting output signal-to-noise-power ratio (SNR) is

\[ \rho_r = \frac{S}{N} = \frac{P}{N S W_N} \frac{|H(j\omega_0)|^2}{|H(j\omega)|^2_{\text{max}}}. \]  

(2-35)

If the carrier frequency is placed at the filter's maximum gain point, the SNR is maximized, and

\[ \rho_r(\text{max}) = \frac{P}{N S W_N}. \]  

(2-36)

Throughout this work, we shall consistently use the symbol \( \rho \) to denote signal-to-noise ratios; \( \rho_r \) then denotes the signal-to-noise ratio of the waveform \( y(t) \).

2-E. Fiducial Bandwidth

The ratio \( W_{\text{f}} |H(j\omega)|^2_{\text{max}} / |H(j\omega_0)|^2 \) appearing in (2-35) occurs very frequently in the theory of phase-locked loops, as we shall see in later chapters. It is therefore very convenient to give this quantity its own special notation

\[ \omega_{\text{f}} = 2b_{\text{f}} = \frac{W_{\text{f}} |H(j\omega)|^2_{\text{max}} / |H(j\omega_0)|^2}{\frac{1}{2\pi} \int_{-\infty}^{\infty} |H(j\omega)|^2 d\omega}. \]  

(2-37)

This quantity is much like the noise bandwidth of \( H \), and, in fact, reduces to \( W_{\text{f}} \) when the maximum filter response occurs at \( \omega_0 \). It is the same form of definition as that given to \( W_{\text{f}} \), except that it is referenced to an arbitrary frequency \( \omega_0 \) rather than to the filter's maximum response. For this reason, we shall refer to it as the fiducial bandwidth of \( H(s) \). This fiducial bandwidth is then the spectral width of an ideal band-pass filter whose response at \( \omega = \omega_0 \) is the same as that of \( H(s) \) and whose output power (with a white noise input) is the same as \( H(s) \).

2-F. Band-Pass Mixers

Now consider what happens when a sinusoidal signal in white noise is heterodyned (a nonlinear operation) into a different passband by a device such as that shown in Fig. 2-6. Let the premixing filter \( H(s) \) have bandwidth \( W_{\text{f}} = 2B_{\text{f}} \), as shown in the Figure. The filter input, denoted by \( x(t) \), is

\[ x(t) = (2P)^{\text{th}} \sin (\omega_0 t + \theta_s) + n(t), \]

where \( n(t) \) is white, with spectral density \( N_0 \), and \( P \) is the power in the sinusoid. The premixer filter output \( y(t) \) is then of the form

\[ y(t) = (2P)^{\text{th}} |H(j\omega_0)| \sin (\omega_0 t + \theta_s) + n_{\text{f}}(t). \]  

(2-38)

Here \( \theta_s = \theta_s + \arg H(j\omega_0) \), the new noise \( n_{\text{f}}(t) \) is band-limited by \( H(s) \) to the filter width \( W_{\text{f}} \), and \( S_{n_{\text{f}n_{\text{f}}}} (j\omega) = N_0 |H(j\omega)|^2 \).

We shall assume now that \( y(t) \) is multiplied by a unit-power\(^4\) sinusoid, \( 2^{\text{th}} \sin (\omega_0 t + \theta_2) \). Thus, the mixer output\(^4\) is

\[ v(t) = P^{\text{th}} |H(j\omega_0)| \cos \left[ (\omega_0 - \omega_2) t + (\theta_2 - \theta_s) \right] 
   - P^{\text{th}} |H(j\omega_0)| \cos \left[ (\omega_0 + \omega_2) t + (\theta_2 + \theta_s) \right] + n_{\text{f}}(t). \]  

(2-39)

\(^4\)This assumption can be relaxed by including a multiplier gain \( K_n \) in the output \( v(t) \).

\(^4\)This supposes an ideal multiplier. Physical realization of a multiplier may leave a carrier term and harmonics.
The bandpass filter \( F(s) \) following the mixer is centered at a frequency \( \omega_l \), which can either be \( \omega_0 + \omega_1 \) or \( \omega_0 - \omega_1 \). For convenience, say it is the latter, \( \omega_l = \omega_0 - \omega_1 \), with \( \omega_0 \leq \omega_1 \). One normally chooses \( B_F \leq B_H \), since there is nothing to be gained with \( B_F > B_H \). The new noise term \( n_s(t) \) is also split into two bands, and unless the bandwidths of \( F(s) \) and \( H(s) \) are chosen properly, overlapping of the noise spectra will allow unnecessary noise in the output of \( F(s) \). It is easy to see from Fig. 2-7 that this can be avoided by choosing\(^*\)

\[
2B_F \leq B_F + B_H < 4f_h
\]

\[
B_F < 2f_h \text{ or } W_F < 4f_h .
\]

When this is satisfied, the output \( z(t) \) from \( F(s) \) is

\[
z(t) = P^{\omega} |H(j\omega_0)|^2 |F(j\omega_1)| \cos(\omega_1t + \theta_1) + n_s(t)
\]

(2-41)

\(^*\)This may need to be modified by a factor of \( \frac{1}{2} \) if carrier is present in the multiplier output.

with \( \theta_1 = \theta_2 - \theta_3 + \arg F(j\omega_1) \). The output noise spectral density is, for positive values of \( \omega_0 \),

\[
S_{n_s} \left( j\omega \right) = \frac{1}{2} N_0 |H(j\omega_0)|^2 |F(j\omega)|^2
\]

(2-42)

We have tacitly assumed that \( F(s) \) is chosen in accordance with (2-40) to suppress completely the image term that would appear in \( S_{n_s} \left( j\omega \right) \) centered at \( \omega_0 + \omega_1 \).

---

**Fig. 2-6. The simple product-mixer**

---

**Fig. 2-7. Choosing filter bandwidths to avoid the image noise problem**
The equivalent noise bandwidth of the mixer is

\[
W_n = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{|H(\omega + \omega_0)|^2 |F(\omega)|^2 d\omega}{H(H(\omega + \omega_0)|F(\omega)|^2_{\text{max}}} = W_p. \tag{2-43}
\]

The output signal-to-noise ratio of the mixer is then given by

\[
\rho_s = \frac{P|H(j\omega_0)|^2 |F(\omega_0)|^2} {N_0 W_n |H[H(\omega + \omega_0)|F(\omega)|^2_{\text{max}}} . \tag{2-44}
\]

This attains its maximum value if \(H\) and \(F\) are built so that

\[
|H(j\omega)|^2_{\text{max}} = |H(j\omega_0)|^2
\]

and

\[
|F(\omega)|^2_{\text{max}} = |F(\omega_0)|^2 ,
\]

in which case

\[
\rho_s(\text{max}) = \frac{P}{N_0 W_n} \left\{ H(\omega_0)|F(\omega_0)|^2_{\text{max}} \right\} = \frac{P}{N_0 W_p}. \tag{2-45}
\]

This is one of the important features of a simple product mixer: the output signal-to-noise ratio is the same as that of a simple sinusoid passing through a linear filter having the same bandwidth as the mixer. Thus, even though mixing is a nonlinear operation on the incoming waveform, there is no degradation in the signal-to-noise ratio at the output. Also, there would be no nonlinear distortion if a bandlimited signal were put into the mixer, rather than a sine-wave. The heterodyne merely produces a translation in the frequency domain and may be treated as a linear device otherwise. Care must be exercised to avoid image noise, and this can be done by choosing \(B_r < 2f_0\) (balanced mixer) or \(B_r < f_s\) (non-balanced mixer).

2-G. Amplitude and Phase Detectors

When the heterodyne frequency is exactly equal to \(f_0\), the mixer becomes an amplitude detector or phase detector, depending upon the relative phases of the incoming wave and the mixing signal. In either case, the resulting output occurs at baseband (i.e., \(\omega_1 = 0\), and the post-multiplication filter \(F(s)\) is now a low-pass, rather than a band-pass, device. No longer are there four separate bands of frequencies in the multiplier output \(v(t)\) as shown in Fig. 2-6. Instead, the two low-frequency portions merge into one at zero frequency. Because of this, there is an apparent decrease, by a factor of 2, in the detector bandwidth, for we note that a band of frequencies \(W_n = 2B_H\) at carrier frequency (see Fig. 2-6) is heterodyned to a width \(W_{\text{bl}} = \frac{1}{2} W_H = B_H\) at baseband. At the same time, the merged noise spectra add together.

The sinusoidal signal components of the input are also heterodyned to zero frequency, and depending on the phase of the mixing frequency, these can add or subtract in varying degrees. With \(\omega_1 = 0\) in Eq. (2-41), we have

\[
p(t) = P^* H(j\omega_0)|F(0)\cos \theta_1 + n_0(t) \tag{2-46}
\]

where \(\theta_1 = \theta_2 - \theta_3\). Because of the merging of the heterodyned bands, the spectrum of \(n_0(t)\) becomes

\[
S_n \delta (f) = \frac{N_0}{2} \left\{ |H(j\omega_0)|^2 + |H(j\omega_0 + f)|^2 \right\} F(\omega_0) \tag{2-47}
\]

The latter approximation is valid when \(F(s)\) is much narrower in bandwidth than is \(H(s)\). The output signal component thus has a zero-frequency component with density

\[
S_{\text{r,sn}} (f) = P |H(j\omega_0)|^2 F^2(0) E(\cos^4 \theta_4, \delta(f)) \tag{2-48}
\]

The term \(E(\cos^4 \theta_4)\) above is referred to as the coherence factor. When \(\theta_1\) and \(\theta_3\) are completely unrelated, then \(\theta_4\) takes on any value between 0 and \(2\pi\) with equal likelihood, so the coherence factor in this case is

\[
E(\cos^4 \theta_4) = \frac{1}{2}. \tag{2-49}
\]

But when \(\theta_2\) and \(\theta_3\) are fully coherent, i.e., when \(\theta_4 = \theta_1\) with probability one, we have

\[
E(\cos^4 \theta_4) = 1. \tag{2-50}
\]

Thus, there is a 3-db advantage in signal-to-noise ratios to be gained by properly phased synchronous detection. In any event, the output signal-to-noise ratio is

\[
\rho_s = \frac{P}{N_0 W_D} |H(j\omega_0)|^2 F^2(0) E(\cos^4 \theta_4) \tag{2-51}
\]
where the detector bandwidth \( W_D \) is given by the same formula (2-43) as was the mixer bandwidth \( W_M \). By choosing \( H \) and \( F \) to have maximum response at \( f_0 \) and 0, respectively, the maximum SNR is achieved

\[
\rho_{\text{SNR max}} = \frac{P}{N_0W_D} E(\cos^2\theta_i) = \frac{P}{N_0W_F} E(\cos^2\theta_i) \,.
\]

(2-52)

The last two equations resemble those of the mixer, viz., (2-44) and (2-45), except for the coherence factor. Also, one should realize that the passband of frequencies around the carrier is \( 2W_D \) wide, whereas the detector bandwidth is only \( W_D \); hence, care must be taken not to confuse the carrier passband with the detector bandwidth.

For a synchronous phase detector, of course, \( \theta_2 \) and \( \theta_i \) are purposely made to differ by an angle of about \( \pi/2 \), so very little signal power is present in the output. The signal-to-noise ratio of a phase detector output is thus not a very meaningful quantity.

All of the expressions for signal-to-noise ratios above could have been more simply expressed in terms of fiducial bandwidths. The treatment here, however, illustrates the desirability of locating the signal power at the maximum response point of the filter.

2-H. Noise

Many sorts of electric signals are called noise. In the early days of radio, noise was familiar as the crash and crackle of static. Later, there was the rasp of ignition noise and the hiss of thermal and shot noises generated in radio circuits themselves. In the end, many engineers have come to regard any interfering signal of a more or less unpredictable nature as noise.

The study of noise began with the consideration of certain physical sources of noise and the sorts of noise that they generate. At first, only very simple properties of the noise signals so generated were understood and described. As the art has progressed, a mathematical theory of noise has grown up. This theory is a part of the general field of statistics, and it deals with signals that have an unpredictable, statistical, random element.

1. Thermal Noise

Perhaps the most fundamental noise is Johnson noise, the noise from a resistor. The engineering fact is that a resistor of resistance \( R \) acts as a noise generator with an equivalent circuit as in Fig. 2-8. In an ordinary resistor, it is a summation of the effects of the very short current pulses of many electrons as they travel between collisions, each pulse individually having a wide spectrum. In this case, the noise is a manifestation of the Brownian movement of the electrons in the resistor. In a resistor consisting of two opposed, close-spaced, hot, electron-emitting cathodes, it is a result of the current pulses of randomly emitted electrons passing from one cathode to the other. In a lossy dielectric, it is the result of random thermal excitations of polarizable molecules, forming little fluctuating dipoles.

For any particular sort of resistor, it should be possible to trace out the source and calculate the magnitude of the Johnson noise, and indeed, this approach has been used.

Any energy involved in thermal noise must clearly come from the surroundings in the form of heat transfer. Thus, derivations of the behavior are destined to involve thermodynamical arguments.

Consider a network containing many resistors. If we heat one hotter than the rest, energy tends to flow from the hot resistor to the cooler resistors. Johnson noise is such energy flowing as electric power. Even when the resistors are all at the same temperature, power will flow back and forth between them through the connecting network, always so that, on the average, a resistor receives just as much power as it sends out.

Statistical mechanics tells us how much energy must, on the average, be associated with each degree of freedom of a system when the system is in thermal equilibrium. In an electrical network of inductors, capacitors, and resistors, the number of degrees of freedom is the number of inductors plus the number of capacitors. (Inductors in series and capacitors in parallel are treated as a single inductance or capacitance, because in setting
up a signal on the network, we are free to specify arbitrary initial currents in all the inductors and arbitrary voltages across all capacitors.)

Classical statistical mechanics says that in a system (in our case, such an electrical network) that is in equilibrium (all at the same temperature $T$) there is, on the average, an energy $\frac{1}{2}kT$ joules associated with each degree of freedom:

$$k = 1.380 \times 10^{-23} \text{ joule/degree Kelvin.}$$

According to quantum mechanics, the energy is less than this at high frequencies; Nyquist and others have used the quantum-mechanical expression to get the correct result. However, even up to many thousands of megacycles, the classical expression is accurate. The only change that quantum mechanics makes is to say that the mean energy per degree of freedom is

$$\frac{1}{2} \frac{hf}{e^{hT/kT} - 1} = \frac{1}{2} kT + \ldots \quad (2-53)$$

where $h$ is Planck's constant

$$h = 6.625 \times 10^{-24} \text{ joule seconds}$$

and $f$ is the frequency. This reduces to the classical result when $f \ll kT/h$. At $300^\circ \text{K}$

$$\frac{kT}{h} = 6.25 \times 10^{15} \text{ cps}$$

Hence the theory holds, to engineering accuracy, up to 60 Gc. Even at very high frequencies, one may define an equivalent noise temperature $T_N$ as

$$T_N = \frac{hf/k}{e^{hT/kT} - 1} = T + \ldots$$

In this way the energy per degree of freedom is always $kT_N/2$ (anyway, $T_N$ is usually inferred from noise power measurements rather than calculated from $T$). Johnson noise thus often serves as a reference for the noisiness of radio receivers and amplifiers, and the effective noise temperature provides a useful way of specifying the noise output of any source.

Let us consider two simple circuits as particular examples, to see how things work out. In these circuits an inductance $L$ is in series with a resistance $R$ at temperature $T$, and a capacitance $C$ is in shunt with a conductance $G$ at a temperature $T$, as shown in Fig. 2-9.

**Fig. 2-9.** Total Johnson noise current- and voltage-squared relations in simple circuits

Let $E(i^2)$ be the total mean-square noise current in the inductor. We can write

$$\frac{1}{2} LE(i^2) = \frac{1}{2} kT. \quad (2-54)$$

On the left we have the average power in the inductance. On the right we have the average value this must have according to statistical mechanics. Accordingly,

$$E(i^2) = \frac{kT}{L}. \quad (2-55)$$

This must be true regardless of the value of $R$. If $R$ is low, we have a narrow-band circuit; if $R$ is high, we have a broad-band circuit.

In a similar way, in the case of the capacitance $C$ and the conductance $G$ in shunt, we easily find that

$$\frac{1}{2} CE(v^2) = \frac{1}{2} kT \quad (2-56)$$

$$E(v^2) = \frac{kT}{C}. \quad (2-56)$$

If the conductance is small, we have a narrow-band circuit with high low-frequency noise components. If the conductance is large, we have less low-frequency noise but more bandwidth.

The relations described above of course apply to capacitors and inductors not merely in the simple circuits we have considered, but to capacitors and inductors anywhere in all circuits, no matter how complicated they may be. In any case, we see that the noise voltage-or current-squared is proportional to the temperature $T$.

Among the circuits to which (2-55) and (2-56) apply is the simple RLC series resonant circuit of Fig. 2-10.
This circuit is characterized by a resonant frequency \( f_0 \), a \( Q_o \), and an impedance \( Z(j\omega) \), related by

\[
\omega_0 = \frac{1}{LC}^{1/2}
\]

\[
Q_o = \frac{\omega_0 L}{R}
\]  \hspace{2cm} (2-57)

\[
|Z(j\omega)|^2 = R^2 + \left(\frac{\omega L - \frac{1}{\omega C}}{\omega_0}\right)^2 = R^2 \left[ 1 + Q_o^2 \left( \frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \right)^2 \right].
\]

We can regard this circuit as one excited by the Thevenin equivalent Johnson noise voltage \( v_n(t) \). If we make the \( Q_o \) very high, so that the bandwidth of the circuit is very narrow, we can find out something about the spectral density of this voltage. Clearly, we must have the voltage and current spectra related by the equation

\[
S_{i,i}(j\omega) = \frac{S_{v,v}(j\omega)}{|Z(j\omega)|^2}.
\]  \hspace{2cm} (2-58)

The integral of \( S_{i,i}(j\omega) \) yields \( E^2 \), given by (2-55). Thus

\[
\frac{kT}{L} = \frac{1}{2\pi R^2} \int_{-\infty}^{\infty} S_{v,v}(j\omega) d\omega
\]  \hspace{2cm} (2-59)

At very high values of \( Q_o \), supposing \( S_{v,v}(j\omega) \) is fairly constant about \( \omega_0 \), we can bring \( S_{v,v}(j\omega) \) outside the integral:

\[
\frac{kT}{L} = \frac{S_{v,v}(j\omega)}{2\pi LR} \int_{-\infty}^{\infty} \frac{Q_o d(\omega/\omega_0)}{1 + Q_o^2 \left( \frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \right)}.
\]  \hspace{2cm} (2-60)

The value of the integral can be computed by substituting \( \omega/\omega_0 = e^\theta \) to reduce it to a standard form in integral tables; it is equal to

\[
\int_{-\infty}^{\infty} \frac{Q_o d(\omega/\omega_0)}{1 + Q_o^2 \left( \frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \right)} = \pi.
\]  \hspace{2cm} (2-61)

We thus have a white spectral density of the equivalent noise voltage in the circuit: For every \( \omega_0 \)

\[
S_{v,v}(j\omega) = 2kTR.
\]  \hspace{2cm} (2-62)

This is commonly referred to as Nyquist's Law.

The spectral density of the equivalent shunt noise current in a resistor can be computed similarly.

What happens if we connect two resistances in series or two conductances in parallel? In a given frequency range, the voltages or currents produced by different resistances are uncorrelated; they have random phases, and the mean square of the sum of the separate voltages or currents is equal to the sum of the mean-square voltages or currents of the separate resistors.

As we have noted, for a complex impedance the series noise voltage generator at any frequency can be calculated from the resistive component \( R \) of the impedance, and the shunt noise current generator from the conductive component \( G \) of the admittance. We can calculate the thermal noise spectral density for any network, simply by associating with each resistance a series voltage generator according to the relation above.

We may ask, what is the thermal noise power \( N \) available from a resistor? We will draw off the maximum power if we supply a matched load of the same resistance. Thus, the available noise power in the bandwidth \( W_x \) can be obtained by calculating the noise power flowing into a resistance \( R \) from a source with an internal resistance \( R \) and an open-circuit voltage spectral density given by (2-62). This power is

\[
N = \frac{kT}{2} W_x \text{ (watts)}.
\]  \hspace{2cm} (2-63)

This result is usually stated in the literature in single-sided notation:

\[
N = kT B_x \text{ (watts).}
\]  \hspace{2cm} (2-64)
The conclusion to be reached from this discussion is that when an amplifier is matched to its source, the amplifier input noise power spectral density $S_{nn}(f_0)$ is

$$S_{nn}(f_0) = N_0 = \frac{kT}{2} R \text{ (volts$^2$)}$$

or

$$G_{nn}(f_0) = \begin{cases} N_0 = kTR & f \geq 0 \\ 0 & f < 0 \end{cases} \quad (2-65)$$

One last point concerning thermal noise: By consideration of the entropy of a system in thermal equilibrium, it can be shown that individual electrons have a Gaussian velocity distribution of zero mean in each direction. The composite effect is a linear combination of Gaussian processes, and hence the total noise voltages in a circuit are independent white Gaussian processes.

2. Shot Noise and Other Noises in Electron Tubes

Electricity is not a smooth fluid; it comes in little pellets, that is, electrons. The flow of electrons in a vacuum tube is accompanied by a noise of the same nature as the patter of rain on a roof. Schottky, who first investigated this phenomenon, called it the Schrotteffekt (from shot); it is now usually called simply shot noise. Like Johnson noise, shot noise has a fairly flat spectrum. This is really what we should expect of a random collection of very short pulses (each of which has a wide spectrum). However, at frequencies for which the transit time from cathode to plate is comparable to the period, the noise induced in the plate circuit is a function of frequency.

Random processes other than the random emission of electrons can also give rise to noise. For example, noise can be introduced into an initially noiseless electron flow if the electrons randomly hit or miss the wires of a grid, with a certain average intercept... of current. Such noise is called partition noise or interception noise. If a small fraction only of the current is intercepted, the added noise is roughly equal to shot noise for the intercepted current.

At high frequencies and long transit times the excitation of a circuit may depend on the velocity of the entering electrons. In such a case, the random variation of velocity of emission from one electron to another, associated with the Maxwellian velocity distribution of electrons leaving a cathode, can give rise to noise. This fluctuation in velocity is also responsible for what is commonly called the modified or reduced shot noise in space-charge-limited flow of electrons from a cathode.

A very simple theory of noise in space-charge-limited diodes and triodes (at frequencies low enough so that transit time is not important) predicts that the noise can be represented by an impressed noise current $i(t)$ in the plate circuit with spectral density

$$S_{ii}(f_0) = (0.644)2kTg \cdot \quad (2-66)$$

Here $T_e$ is the temperature of the cathode and $g$ is the conductance of the diode or the transconductance of the triode. More elaborate theories lead to a factor that, in various circumstances, may be a little greater or a little less than 0.644.

At moderate frequencies, noises in triodes and diodes agree fairly well with (2-66), being perhaps a little higher. At low audio frequencies and below, flicker noise appears. This typically, but not always, has a $1/f$ spectrum (discussed below). Flicker noise is very variable from tube to tube. It has been ascribed to fluctuations in the work function of the cathode surface.

3. Noise With a $1/f$ Spectrum

Johnson noise is in a sense inherently white noise, in that the fundamental relation between the noise source (the resistance or conductance) and the amount of noise per unit bandwidth is independent of frequency. Shot noise is, in the same sense, inherently white noise too, although it can give rise to different spectra in circuits with different transfer admittances or in tubes of different transit times. In a close-spaced diode formed of opposed cathodes at the same temperature and with no average current, shot noise and Johnson noise are two names for the same thing.

Some important sorts of noise are generated only in nonequilibrium systems, for instance, in systems in which dc current flows. Among these are contact noise, such as is produced in a carbon microphone, and the noise produced in carbon resistors and in silicon and germanium diodes and transistors. Both contact noise and transistor noise have a spectrum that varies nearly as $1/f$ over a large frequency range, though it may be constant at high frequencies and at very low frequencies. One could obtain a $1/f$ spectrum down to any given frequency by a proper mixture of pulses of various lengths.
Transistor noise has been attributed to the trapping of the holes or electrons (carriers) that form the current 
flow. The trapping and subsequent release of a charge 
carrier is equivalent to a rectangular pulse in the current. 
The effect may be strengthened by the charge of the 
trapped carrier modulating the flow of other charges. By 
assuming a particular distribution of trapping times, or 
pulse lengths, a 1/f spectrum can be obtained. The matter 
of noise in semiconductors is by no means thoroughly 
understood, and somewhat different mechanisms have 
been suggested.

Actually, the power spectrum cannot vary as 1/f right 
down to \( f = 0 \), for this would imply an infinite noise 
power. Measurements do show a 1/f spectrum down to 
frequencies as low as \( 10^{-4} \) cps, however.

REFERENCES FOR CHAPTER 2

1. Davenport, W. B., Jr., and Root, W. L., Random Signals and Noise, McGraw-Hill 

CHAPTER 3
FORMULATION OF THE LOOP EQUATION AND BEHAVIOR IN THE ABSENCE OF NOISE

In this Chapter, we shall develop the basic equation governing the phase-locked loop. Assuming that noise is absent, we shall develop the acquisition and steady-state character of the loop. This essentially follows the work of Viterbi.

3-A. The Basic Integro-Differential Equation

The essentials of a phase-locked device are a multiplier, a loop filter, and a voltage-controlled oscillator (VCO), as shown in Fig. 3-1. Most of the more elaborate systems using double-heterodyne techniques, IF limiters, and acquisition aids reduce to this basic model, insofar as analysis of the behavior is concerned. The input is assumed to be a sinusoid of the form

\[ x(t) = A(2)^{\ast\ast} \sin \left[ \omega_0 t + \theta(t) \right] \]  

(3-1)
in which the quantities \( A, \omega_0, \) and \( \theta(t) \) are

- \( A \) = rms voltage amplitude of \( x(t) \)
- \( \omega_0 \) = frequency of the VCO when its input is shorted
- \( \theta(t) \) = the input signal phase process.

Normally, \( \theta(t) \) consists of an information-bearing term \( \psi(t) \) due to modulation and a term \( d(t) \) due either to doppler shift in the received signal, to drift in the VCO center frequency, or to some bias voltage appearing at the VCO input.

The incoming signal is multiplied by the VCO output

\[ v(t) = K_1 (2)^{\ast\ast} \cos \left[ \omega_0 t + \hat{\theta}(t) \right] \]  

(3-2)

\[ x(t) = A\sqrt{2} \sin \left( \omega_0 t + \theta(t) \right) \]

\[ y(t) = AK_1 K_m \sin \phi(t) \]

\[ v(t) = K_1 \sqrt{2} \cos \left( \omega_0 t + \hat{\theta}(t) \right) \]

The term \( \hat{\theta}(t) \) appearing here is the loop estimate of \( \theta(t) \), and \( K_1 \) is the rms output of the VCO.

The result of this multiplication, if perfect, would be

\[ y(t) = AK_1 (\sin \theta(t) - \hat{\theta}(t)) + \sin [2\omega_0 t + \theta(t) + \hat{\theta}(t)] \]

(3-3)

However, generally speaking, the multiplication is accomplished by a device unable to respond to the double-frequency term. Also, the multiplying device has some gain \( K_m \), and hence the actual output of the "phase-detector" is

\[ y(t) = AK_1 K_m \sin \phi(t) \]  

(3-4)

The quantity \( \phi(t) = \theta(t) - \hat{\theta}(t) \) is called the true phase error.

The multiplier output is then fed into the loop filter \( F(s) \) and emerges as \( z(t) \), which, in turn, supplies the input to the VCO.

In many actual implementations, the waveform \( y(t) \) is monitored (perhaps multiplied by a gain factor) and is called the "dynamic phase-error." Similarly, \( z(t) \) is often referred to as the "static phase-error." These are functions of \( \phi \), rather than the true phase error \( \phi \), and thus tend to generate some confusion. Little reference to these terms will appear in the material that follows.

**Fig. 3-1. The basic phase-locked loop**
The VCO output frequency is, as its name implies, a linear function of its input.

$$\omega_{\text{VCO}}(t) = \omega_0 + K_{\text{VCO}} z(t) + K_{\text{VCO}} e$$  \hspace{1cm} (3-5)

Consequently, the phase estimate \( \hat{\theta}(t) \) developed by the loop, being the integral of the VCO frequency, can be written, omitting \( e \) for the present,

$$\hat{\theta}(t) = \frac{K_{\text{VCO}}}{p} z(t)$$  \hspace{1cm} (3-6)

where \( p = \frac{d}{dt} \) is the Heaviside operator. Substitution for \( z(t) \) yields

$$\hat{\theta}(t) = AK_{\text{VCO}} \frac{F(p)}{p} \sin \phi(t)$$  \hspace{1cm} (3-7)

At this point it becomes convenient to define

$$K = K_{\text{VCO}} K_{\text{o}}$$  \hspace{1cm} (3-8)

as the open-loop gain of the loop, and to substitute \( \hat{\theta} = \theta - \phi \). This produces the fundamental equation that specifies the behavior of the loop in the absence of noise,

$$\theta(t) = \phi(t) + AK \frac{F(p)}{p} \sin \phi(t)$$  \hspace{1cm} (3-9)

When \( F(p) \) is a constant, (3-9) is a first-order integro-differential equation. Hence the configuration of Fig. 3-1 is called a first-order phase-locked loop. Similarly, when \( F(s) \) has \( n \) finite poles, the system equation is an \( (n+1) \)th-order one, and the device in Fig. 3-1 is said to be an \( (n+1) \)th-order loop.

**3-B. Tracking When the Loop Error is Small**

The phase-locked loop would be of little use if it were not possible to use it to reconstruct the input phase process with some degree of fidelity—that is, unless we are able to "lock" the loop in the first place. The term "lock" is somewhat subjective at this point, but it may generally be thought of as a condition in which \( \phi(t) \) never varies outside an interval of size \( 2\pi \). The mean-phase value is called the lock-in point; such lock-in points are located \( 2\pi \) radians apart. Whenever the phase error goes through \( 2\pi \) radians, we say that the loop has skipped a cycle. If the loop is capable of reducing the phase error to a small enough value, say \( |\phi| < \pi/6 \), we can approximate

$$\sin \phi \approx \phi$$

and, in such a case, the loop error is related to the input phase by a simple, linearized version of (3-9),

$$\theta(t) = \frac{p + AKF(p)}{p} \phi(t)$$  \hspace{1cm} (3-10)

Generally, the input process can be separated into an information-bearing part \( \phi(t) \) and a phase-offset term \( d(t) \):

$$\phi(t) = \frac{p}{p + AKF(p)} \psi(t) + \frac{p}{p + AKF(p)} d(t)$$  \hspace{1cm} (3-11)

We see that there are two kinds of error present in the loop. That part of the error due to the modulation \( \phi(t) \) is commonly called phase distortion. The remaining error is that produced, for example, by a doppler shift on the incoming carrier. To be effective, the loop must be designed to track whatever variations \( d(t) \) may have, so the filter \( F(s) \) must be properly chosen. Error arising from \( d(t) \) is called the tracking error, or transient distortion.

The steady-state tracking error can be determined by the final-value theorem of Laplace transform theory. Denoting \( D(s) = \mathcal{L}[d(t)] \), the Laplace transform of \( d(t) \), this theorem reads

$$\phi_{ss} = \lim_{s \to 0} \frac{s^2 D(s)}{s + AKF(s)}$$  \hspace{1cm} (3-12)

An example will elucidate the behavior:

**First-order loop**: Suppose \( F(s) = 1 \), and \( d(t) = \theta_0 + \Omega_0 t + \frac{1}{2} \Lambda_0 t^2 \), where \( \phi_0 \) is the initial phase offset, \( \Omega_0 \) the initial frequency offset, and \( \Lambda_0 \) the doppler rate,

$$D(s) = \frac{\phi_0}{s} + \frac{\Omega_0}{s^2} + \frac{\Lambda_0}{s^3}$$

This produces a steady-state tracking error given by

$$\phi_{ss} = \lim_{s \to 0} \frac{\theta_0 s + \Omega_0 + \Lambda_0 s}{s + AK}$$  \hspace{1cm} (3-13)

The error is not bounded if \( \Lambda_0 \neq 0 \); but when \( \Lambda_0 = 0 \),

$$\phi_{ss} = \frac{\Omega_0}{AK}$$  \hspace{1cm} (3-14)

In such a case, \( \phi_{ss} \) can be made small by adjustment of the loop gain. Note that the loop will track any initial phase offset \( \theta \), with no steady-state error in the absence of a frequency offset (i.e., \( \Omega_0 = 0 \)).
From this example, one can see that if \( d(t) = \frac{\lambda_e}{s^n/n!} \) (the \( n \)th derivative of \( d(t) \) is \( \lambda_e \), so \( D(s) = \lambda_e/s^{n+1} \)), and if \( F(s) \) has \( l \) poles at the origin, as
\[
F(s) = \frac{q(s)}{s^{l+1} p(s)}; \quad q(0) \neq 0, p(0) \neq 0. \tag{3-15}
\]
then the steady-state tracking error would be
\[
\phi_{ss} = \lim_{s \to 0} \frac{\lambda_e p(s) s^{-n+1}}{s^{l+n+1} p(s) + AKq(s)}. \tag{3-16}
\]
This is not finite if \( n > l + 1 \). When \( n = l + 1 \) the error is finite and has the value
\[
\phi_{ss} = \frac{\lambda_e}{AK} \frac{p(0)}{q(0)}. \tag{3-17}
\]
When \( n < l \), there is no steady-state error, \( \phi_{ss} = 0 \), and perfect tracking ultimately results.

The conclusion here is the following: In order for a phase-locked loop to track an \( n \)th-degree phase function, \( F(s) \) must have at least \( (n - 1) \) poles at the origin.

3-C. Acquiring Lock in the First-Order Loop

Let us concentrate now on the way a first-order loop (\( F(s) = 1 \)) behaves in a tracking mode only. We consider the case in which \( \psi(t) \) (modulation) is absent from the input phase function \( \theta(t) \); only the \( d(t) \) term due to input doppler, VCO drift, etc., is assumed to be present. From the discussion in the previous Section, it is clear that the first-order loop can track a function of the type (constant frequency offset)
\[
\theta(t) = \theta_0 + \Omega_o t \tag{3-18}
\]
whenever \( \Omega_o \) is small enough that \( \phi(t) \approx 0 \); that is, when
\[
\left| \frac{\Omega_o}{AK} \right| < \frac{\pi}{\delta}. \tag{3-19}
\]
To investigate the behavior for larger values of \( \Omega_o \), let us rewrite the loop equation, differentiating (3-9),
\[
\Omega_o = \phi + AK \sin \phi. \tag{3-20}
\]
We shall denote the frequency error by \( \psi(t) = \Omega(t) \), and plot \( \Omega \) versus \( \phi \) according to (3-20). The result is shown in Fig. 3-2. Whenever \( \Omega \) is positive, \( \phi \) tends to increase, and whenever \( \Omega \) is negative, \( \phi \) tends to decrease. Note that if \( \left| \Omega_o \right| < AK \), there are regular points at which \( \Omega = 0 \). Starting at \( \phi(0) = n\pi \), where \( n \) is an even integer, the system tends to move along the sinusoidal trajectory of Fig. 3-2 until it reaches the \( \phi \)-axis at
\[
\phi_{ss} = n\pi + \sin^{-1} \left( \frac{\Omega_o}{AK} \right). \tag{3-21}
\]
This is a stable point; \( \Omega \) cannot become negative because \( \phi \) would then tend to decrease and \( \phi \) turns the system to the \( \phi \)-axis. If \( n \) were an odd integer, the system would go through a larger part of the sinusoidal trajectory until it reached a stable point at
\[
\phi_{ss} = (n + 1)\pi + \sin^{-1} \left( \frac{\Omega_o}{AK} \right). \tag{3-22}
\]
If \( \Omega_o > AK \), however, the trajectory never crosses the \( \phi \)-axis, and phase lock is never achieved. The maximum pull-in range of a first-order loop is thus
\[
\Omega_{max} = AK. \tag{3-23}
\]
Whenever \( \Omega_o < AK \), the loop ultimately tracks the incoming function \( \psi(t) \) with no frequency error, but with a constant lagging phase error, given by (3-21) and (3-22). The loop never skips a cycle.

The multiples of \( \pi \) here are usually omitted, and the steady-state phase error is written merely
\[
\phi_{ss} = \sin^{-1} \left( \frac{\Omega_o}{AK} \right). \tag{3-24}
\]

![Fig. 3-2. First-order loop pull-in behavior](chart.png)
that is, the steady-state output of the VCO is a replica of the input, lagging by \( \phi_\infty \):

\[
v_\infty(t) = \left(\frac{1}{\omega_0 + \Omega_0} + \theta_0 - \phi_\infty\right) \sin \left(\omega_0 t + \theta_0 - \phi_\infty\right).
\]

(3-25)

The pull-in time may also be determined from (3-20) as follows: we can write

\[
\frac{d\phi}{dt} = \frac{1}{\Omega_0 - AK \sin \phi},
\]

(3-26)

This can be integrated to give the pull-in time to any particular value of \( \phi \), but, since the denominator above vanishes at the lock-in point, an infinite time is required before \( \Omega = 0 \). However, if we agree that the loop is in lock whenever \(| \phi - \phi_{lock} | < \delta_{lock} \), the corresponding time is finite. The maximum time required to achieve lock, designated the acquisition time, will be that time required for the system to pass from \(-\pi + \delta_{lock} - \sin^{-1}(\Omega_0/\Omega_{lock})\) to \(\sin^{-1}(\Omega_0/\Omega_{lock}) - \delta_{lock} \). When the value of \(\delta_{lock} \) is small, the answer reduces to approximately

\[
t_{acq} = \frac{2}{AK \cos \phi_\infty} \ln \frac{2}{\delta_{lock}}.
\]

(3-27)

If a loop is designed so that \( \phi_\infty = 5 \) deg when \( \Omega_0 \) is 200 rad/sec, and \( \delta_{lock} \) is taken to be 5 deg, then the required pull-in time is approximately 1 msec.

The trajectory described in Fig. 3-2 is called a phase-plane diagram of the loop behavior. It is particularly useful for analyzing the lock-in characteristics of lower-order loops.

3-D. Acquiring Lock in the Second-Order Loop With Passive Loop Filter

The system of interest in this Section is one in which the loop filter takes the form

\[
F(s) = \frac{1 + \tau_2 s}{1 + \tau_1 s}.
\]

(3-28)

The discussion in Section 3-B indicates that a loop with this filter can follow a phase function of the type

\[
d(t) = \theta_0 + \Omega_0 t \quad (\text{i.e., a constant frequency offset}),
\]

but that no higher-order terms may be considered. The equation governing the behavior, from (3-9), is

\[
\Omega_0 = \tau_1 \dot{\phi} + (1 + AK\tau_2 \cos \phi) \phi + AK \sin \phi.
\]

(3-29)

The steady-state phase error \( \phi_\infty \) is the same as for the first-order loop,

\[
\phi_\infty = \sin^{-1} \left( \frac{\Omega_0}{AK} \right),
\]

(3-30)

and, clearly, the loop never locks if \( \Omega_0 > AK \).

For values of \( \Omega_0 \) less than \( AK \), pull-in may possibly not occur, even though a stable point exists in the phase plane. This is a consequence of the fact that the trajectory that takes \( \phi \) to \( \phi_\infty \) is not the simple sinusoid of Section 3-C, but is rather the solution to Eq. (3-29).

To simplify things a bit, we can substitute for \( \phi \) the value

\[
\phi = \frac{d\Omega}{dt} = \frac{d\Omega}{d\phi} \cdot dt = \Omega \frac{dt}{d\phi},
\]

(3-31)

(note that \( \phi \) is proportional to the slope of the phase-plane trajectory). This produces the trajectory equation

\[
\Omega_0 = \left( \frac{\tau_1}{\delta_{lock}} + 1 + AK\tau_2 \cos \phi \right) \Omega + AK \sin \phi.
\]

(3-32)

A solution to (3-32) depends not only on \( \Omega_0 \), but also on \( \Omega(0) \) and \( \phi(0) \), the initial conditions of the VCO. For certain values of \( \Omega_0 \), the loop will lock regardless of what values \( \Omega(0) \) and \( \phi(0) \) take. See, for example, Fig. 3-3. In this Figure, \( \phi \) has been limited to the region \((-\pi, \pi\) by folding all the trajectories onto this region. When \( \Omega \) is positive, \( \phi \) increases, and when \( \Omega \) is negative, \( \phi \) decreases. Hence, motion as a function of time is from left to right in the upper half plane, and from right to left in the lower. Starting a trajectory, say in the upper half plane, one follows it to the right until \( \phi = \pi \), skipping back to \( \phi = -\pi \) at the same value of \( \Omega \) encountered at \( \phi = \pi \). This continues until the lock-in point is reached.

For values of \( \Omega_0 \) larger than those in Fig. 3-3, lock-in may occur for some initial conditions of the VCO but not for others, as illustrated in Fig. 3-4. This latter Figure shows that there is a limit cycle toward which all higher trajectories converge, as well as some of those from below. Even if \( \Omega(0) \) is negative, lock-in is not assured, since only those trajectories that ultimately pass through a strip determined by the asymptotes of the saddle points can converge to the lock-in point.

By integrating (3-32), Viterbi has derived a necessary condition on \( \Omega_0 \): If lock occurs for all initial conditions of the VCO, then \( \Omega_0 \) is bounded by

\[
|\Omega_0| < 2 \left( \frac{AK}{\tau_1} (1 + \frac{1}{2}AK\tau_2) \right)^{1/2}.
\]

(3-33)
Fig. 3-3. Lock-in behavior of a second-order loop with imperfect integrator, $F(s) = \frac{1 + \tau_2 s}{1 + \tau_1 s}$, for $\Omega_c/\Delta K = 0.4$ and $\Delta K \tau_2^2/\tau_1 = 2$. The upper bound of Eq. (3-33) is $\Omega_c/\Delta K < 0.693$. 
Fig. 3-4. Lock-in behavior of a second-order loop with imperfect integrator, \( F(s) = \frac{1 + \tau_1 s}{1 + \tau_2 s} \), for \( \Omega_0/AK = 0.9 \) and \( AK\tau_2 / r_1 = 2 \). The upper bound of Eq. (3-33) is \( \Omega_0/AK < 0.693 \). Note that lock-in occurs only when the trajectory happens to pass through the "slot." Otherwise, the trajectory enters the periodic frequency lag region shown.
Experimental evidence seems to indicate that this is a fairly sharp bound; that is, for \( \omega_s \) very close to the bound above, lock always ultimately occurs, although as the bound is approached, longer and longer times are required to pass through the region where the limit cycle would occur should \( \omega_s \) reach the bound.

Acquisition time for this loop is difficult to compute, but as an approximation, Viterbi has computed the time required for frequency lock (i.e., that time required before no more cycles are skipped); when \( \tau_i \) is large, it is

\[
t_{\text{acq freq}} \approx \frac{1}{\tau_i} \left( \frac{\Omega_0 \tau_i}{AK} \right)^2 \text{ sec.} \quad (3-34)
\]

The additional time to acquire phase lock is probably about the same as \( t_{\text{acq}} \) of the first-order loop, given in (3-27).

### 3-E. Tuning the VCO

In the first four Sections of this Chapter, we have neglected the effect of the VCO tuning voltage \( e \) shown in Fig. 3-1. Including it in the VCO equation (3-5), we can write

\[
\omega_{\text{VCO}}(t) = \omega_0 + K_{\text{VCO}}e + K_{\text{VCO}}z(t). \quad (3-35)
\]

The effect of \( e \) is that it determines the VCO frequency when \( z(t) \) is absent. The loop phase estimate is likewise changed to

\[
\hat{\theta}(t) = (\omega_0 + K_{\text{VCO}}e)t + \frac{K_{\text{VCO}}}{p} z(t). \quad (3-36)
\]

The basic integro-differential equation governing the loop is almost the same as (3-9), except for a term \( K_{\text{VCO}}et \) subtracted from \( \hat{\theta}(t) \). The result can be written

\[
\hat{\theta}(t) - K_{\text{VCO}}e = \Omega(t) + AKF(p) \sin \phi(t). \quad (3-37)
\]

The conclusion is now evident: All the answers we have obtained about lock-in, tracking, etc., are the same as before, except that we replace \( \hat{\theta} \) by \( \hat{\theta} - K_{\text{VCO}}e \). Alternatively, we can redefine \( \omega_0 \) as \( \omega_0 \equiv \omega_0 + K_{\text{VCO}}e \).

For example, the lock-in points now occur at

\[
\phi_{\text{st}} = \sin^{-1} \left( \frac{\Omega_0 - K_{\text{VCO}}e}{AK} \right). \quad (3-38)
\]

By choosing \( e \) properly, the steady-state error can be made zero. This value of \( e \) is clearly

\[
e = \frac{\Omega_0}{F_{\text{VCO}}}. \quad (3-39)
\]

In the first-order loop, the acquisition time with the foregoing value of \( e \) becomes

\[
t_{\text{acq}} = \left( \frac{2}{AK} \right) \ln \left( \frac{2}{\delta_{\text{lock}}} \right). \quad (3-40)
\]

### 3-F. Locking the Second-Order Loop with Perfect Integrator

When \( \tau_i \gg \tau_s \), the system whose loop filter is given by (3-28) behaves much the same as the one with

\[
F(s) = \frac{1 + r_s s}{\tau_i s}, \quad (3-41)
\]

except that the latter is able to track any constant frequency offset with zero steady-state error. However, there is a steady-state error when a doppler-rate \( \Lambda_s \) is present:

\[
\phi_{\text{ss}} = \sin^{-1} \left( \frac{\tau_i \Lambda_s}{AK} \right). \quad (3-42)
\]

and just as there was a limit on \( \Omega_s \) in Section 3-D for lock to occur for all initial conditions of the VCO in the passive-filter loop, there is a limit on \( \Lambda_s \) in the perfect-integrator loop.

Clearly, from (3-42), lock cannot occur at all if \( \Lambda_s > AK/\tau_i \). Figure 3-5 is a typical phase-plane diagram of the way such a loop behaves with a doppler-rate input. Almost all trajectories with \( \Omega(0) < 0 \) eventually lead to the lock-in point, while most of those with \( \Omega(0) > 0 \) diverge upward, never reaching lock. The Figure is drawn with the largest value of \( \Lambda_s \) (determined experimentally) such that all trajectories with \( \Omega(0) < 0 \) ultimately terminate at the lock-in point, while most of the trajectories with \( \Omega(0) > 0 \) fail to lock. This value of \( \Lambda_s \) is a function of the parameter \( AK \tau_i^2 / \tau_s = r \),

\[
\Lambda_s = k(r) \left( \frac{AK}{\tau_i} \right). \quad (3-43)
\]

Figure 3-6 illustrates the way \( k(r) \) varies as a function of \( r \).

When the VCO is being swept as in Section 3-E to lock onto a constant frequency offset \( \Omega_s \), the rate by which \( e \) is changed should not produce a \( \Lambda_s \) exceeding (3-43), and, as Fig. 3-5 shows for positive values of \( \Lambda_s \), the sweep should begin with a negative frequency error (a positive frequency error if \( \Lambda_s \) is negative). That is,

\[
K_{\text{VCO}} \frac{de}{dr} < k(r) \left( \frac{AK}{\tau_i} \right). \quad (3-44)
\]

In the passive-integrator loop, (3-43) and (3-44) hold approximately for short periods of time (see Section 6-C3).
Fig. 3-5. Phase-plane trajectory of a second-order loop with perfect integrator to a doppler-rate input $\Lambda_0$ for $AK\frac{\tau^2}{r_1} = 2$, and $\Lambda_0 = AK/2r_1$. 
Fig. 3-6. Normalized maximum doppler rate, \( k(r) = \frac{A_0 \tau_1}{AK} \), for which lock is guaranteed in the absence of noise, as function of the loop parameter \( r = AK \tau_2^2 / \tau_1 \).

REFERENCES FOR CHAPTER 3


Chapter 4

Behavior of Phase-Locked Loops with Stochastic Inputs

The consideration given to phase-locked devices in the preceding Chapter excluded both the possibility that the incoming phase process \( \theta(t) \) was a random, information-bearing process and the possibility that the input wave was a noisy one. We now wish to include these cases for consideration. Some comments are in order concerning the effect of noise on the pull-in and tracking behavior under such conditions. A small amount of phase jitter due to noise or \( \theta(t) \) will not affect the phase-plane trajectories significantly; it will result only in a small amount of hash being superimposed on the trajectories. Such a small amount of jitter on the trajectories will thus not appreciably change the number of cycles required to achieve lock. On the other hand, if the input noise is significant, the disturbance of the trajectory may easily be such that a drastically different number of cycles will be required before lock is achieved. The number of cycles required to achieve lock depends on the particular noise waveform observed at the input and is random. If there is a great amount of phase jitter, the response of the loop may be so erratic that lock never seems to occur. There are, as yet, no analytic results on the pull-in time when stochastic inputs are observed.

4-A. Development of a Mathematical Model and a Basic Loop Equation

We start as we did in Chapter 3, but now the input consists of a sinusoidal signal with power \( P = A^2 \) immersed in a noise \( n(t) \):

\[
x(t) = A 2^\sin \left[ \omega_0 t + \theta(t) \right] + n(t).
\]

(4-1)

The density of \( n(t) \), denoted \( S_n(n_0) \), will be arbitrary for a moment, although we shall restrict it later.

Following the same procedure as in Chapter 3, we derive the basic integro-differential equation, similar to (3-9):

\[
\dot{\theta}(t) = AK \frac{F(p)}{p} \sin \phi(t) + \frac{KF(p) n(t)}{p} + \frac{K_{vco} [e + n_e(t) - \dot{\theta}(t)]}{p}.
\]

(4-2)

Again, \( \dot{\phi}(t) = \dot{\theta}(t) - \ddot{\theta}(t) \), \( K = K_{n}K_{vco} \), \( e \) is a tuning bias applied to the VCO, and the term \( n_e(t) \) is the internal loop-noise referred to the VCO input. The noise \( n(t) \) appearing in (4-2) is

\[
n(t) = n_i(t) 2^{\alpha} \cos \left[ \omega_0 t + \dot{\theta}(t) \right].
\]

(4-3)

Except in special cases, the spectral density \( S_{nn}(f_0) \) of \( n(t) \) is not merely a heterodyned version of \( n_i(t) \), because of the correlation between \( n_i(t) \) and the phase estimate \( \dot{\theta}(t) \).

As we indicated in Section 3-E, the effect of \( e \) on the loop is that it changes the no-error VCO frequency from \( \omega_0 \) to \( \omega_0 + K_{vco} e \). We can thus merely redefine \( \omega_0 \) to include this offset and proceed with

\[
\dot{\theta}(t) = \frac{AKF(p)}{p} \sin \phi(t) + \frac{KF(p) n(t)}{p} + \frac{K_{vco} n_e(t)}{p}
\]

\[
= \frac{AKF(p)}{p} \sin \left[ \theta(t) - \dot{\theta}(t) \right]
\]

\[
+ \frac{KF(p)}{p} n(t) + \frac{K_{vco}}{p} n_e(t).
\]

(4-4)

This last equation gives us the basis of an exact mathematical model of the phase-locked loop. As shown in Fig. 4-1, there are three inputs to this model of the loop: the first, \( \theta(t) \), enters, is differenced with \( \dot{\theta}(t) \), then passes through a nonlinear, zero-memory amplifier whose characteristic is \( y_{out} = A \sin x_{in} \). The input system noise \( n(t) \), added to the output of the sine-amplifier, becomes the input to a filter with transfer function \( KF(s)/s \), whose output is then \( \dot{\theta}(t) \); in this way, \( \dot{\theta}(t) \) satisfies (4-4).

Solution of (4-4) is really made no easier by the model in Fig. 4-1; however, we are not generally interested in the solution to (4-4) for a specific sample function \( n(t) \) anyway. What we are interested in is some measure of the loop's average performance, such as rms phase error, mean rate of cycles skipped, etc. From the Figure, however, one is led to some natural approximations that aid in loop analysis, and these are developed in the ensuing chapters.
4-B. Discussion of Mathematical Models

As a first approximation, the output phase-jitter of a phase-locked loop can be obtained by replacing the sinusoidal nonlinearity of our mathematical model (Fig. 4-1) by a linear or polynomial approximation. However, the steady-state-error probability density \( p(\phi) \) is a periodic function because the loop is sensitive only to errors modulo \( 2\pi \); consequently, the steady-state phase variance in every phase-locked loop is infinite. This is borne out by the fact that, in actuality, the loop skips cycles at a certain mean rate, executing a random, nonstationary motion between lock-in points, much like a discrete random walk. The linear and polynomial approximations mentioned above do not exhibit this periodic lock-in behavior, and hence they not only give no information concerning loss of lock, but even come up with a finite-variance, stationary phase-process - a case that never actually occurs, as we have said.

But while the steady-state phase process itself does not possess a finite variance, it does when phase angles are reduced modulo \( 2\pi \). On this basis, Tikhonov and Viterbi were able to provide exact results for the first-order system.

Considering the exact equivalent model, there is certainly no loss in generality by the explicit inclusion of the "mod \( 2\pi \)" nonlinear function in the loop as in Fig. 4-2. However, there are several differences one can inject into the mathematical results: Although the steady-state \( \phi \)-process is nonstationary, the \( \phi(\mod 2\pi) \)-process usually is stationary (assuming \( \theta(t) \) is stationary); there then exists an equivalent stationary process (call it \( \Phi(t) \)) with finite variance, which, if reduced modulo \( 2\pi \), would result in a random process with the same probability density as \( \phi(t) \). And so long as our sole interest lies in \( \phi(\mod 2\pi) \), and not in \( \Phi(t) \) behavior of \( \phi \), we may assume that \( \phi \) is at work in the loop, and that stationarity prevails. In fact, the linear and polynomial approximations are direct attempts at calculating the behavior of \( \phi \).

4-C. Spectrum of the Loop Noise

The noise we deal with in our analysis is not that occurring at the input, but

\[
n(t) = n_1(t) \cos[\omega_0 t + \hat{\theta}(t)]
\]

in which \( \hat{\theta}(t) \) is derived, in part, from \( n_1(t) \). The non-independence of \( \hat{\theta} \) and \( n_1 \) means that \( S_{n_1}(f_0) \) is not generally the convolution of \( S_{\hat{\theta}}(f_0) \) with the spectrum of \( 2\pi \cos(\omega_0 t + \hat{\theta}) \). However, it has been reasoned that whenever \( n_1(t) \) has a much greater bandwidth than \( \hat{\theta}(t) \), and whenever \( S_{\hat{\theta}}(f_0) \) is symmetric about \( \omega_0 \), then the convolution formula is approximately correct. This gives

\[
S_{n_1}(f_0) = \frac{1}{2} S_{\hat{\theta}}(f_0) + \frac{1}{2} S_{\hat{\theta}}(f_0 - \omega_0) + \frac{1}{2} S_{\hat{\theta}}(f_0 + \omega_0).
\]

That is, the loop noise has the same spectral shape as \( n_1(t) \) does, except that it is heterodyned down to baseband. The spectral shape of \( n_1(t) \) usually comes from allowing white noise with density \( N_0 = kT/2 \) to be passed through a
predetection filter \( H(s) \), whose bandwidth is, say, \( W_H \). That is, \( n_i(t) \) has a spectrum of the form

\[
S_{n_i}(j\omega) = N_0|H(j\omega)|^2
\]

(4-7)

and, in this case, the loop noise \( n(t) \) has as its spectral density

\[
S_{nn}(j\omega) = \frac{1}{2} N_0 \{ |H[\omega-\omega_0]|^2 + |H[\omega+\omega_0]|^2 \}.
\]

(4-8)

In the zone about zero frequency, the region of interest, these two halves add:

\[
S_{nn}(0) = N_0|H(j\omega_0)|^2.
\]

(4-9)

(Again we note the change in bandwidth by a factor of one-half due to heterodyning \( n_i(t) \) to zero frequency, as in 2-6.)

**Fig. 4-2.** Equivalent exact mathematical model of phase-locked loop, with explicit reduction of \( \phi \mod 2\pi \). The VCO has been replaced by \( K_{\text{VCO}}/s \) (tuning bias and VCO noise omitted).

**REFERENCES FOR CHAPTER 4**


CHAPTER 5
THE LINEARIZED ANALYSIS OF PHASE-LOCKED SYSTEMS

Most of the important behavior of a phase-locked device can be predicted by replacing the loop phase-detector by an ideal phase-differencer. The distinction between the two is, of course, that there is a sinusoidal nonlinear characteristic for the phase detector, whereas the ideal phase-differencer is a linear error-detecting device. To justify this linearized analysis, one must assume that the loop error is very small, which usually means that a high signal-to-noise ratio prevails, and that modulation on the carrier is not excessive. The latter of these restrictions can be removed by subcarrier modulation. For a discussion of such techniques, see Volume II of this work.

5-A. Behavior of a Linear Loop

If the level of \( n(t) \) is sufficiently low, and if the loop is designed properly, the phase error \( \phi(t) \) should be very small. In such a case, the approximation

\[
\sin \phi = \phi , \tag{5-1}
\]

when inserted into (4-4), yields a linear equation relating the loop input and output

\[
\dot{\theta}(t) = \frac{AKF(p)}{p + AKF(p)} \left[ \theta(t) + \frac{n(t)}{A} \right], \tag{5-2}
\]

omitting VCO noise.

The output phase is thus the result of a linear filter acting upon the input phase process \( \theta(t) \) immersed in a normalized version of the loop noise, with the normalizing factor in this case being equal to the rms signal amplitude \( A \). An equivalent circuit of the linearized loop appears in Fig. 5-1. The overall loop transfer-function (call it \( L(s) \)) is related to the loop-filter \( F(s) \) by the relations

\[
L(s) = \frac{AKF(s)}{s + AKF(s)} \tag{5-3}
\]

As with any linear filter, \( L(s) \) has some effective noise bandwidth \( W_L \), which can be found by using (2-24),

\[
W_L = 2B_L = \frac{1}{2\pi} \int_{-\infty}^{\infty} |L(j\omega)|^2 d\omega \tag{5-4}
\]

where we use \( L^2 \) to denote \( |L(j\omega)|^2 \).

It is important to remember that \( W_L \) is the bandwidth of a transfer function at baseband and it is not the same as the width of the carrier passband produced by the loop. As we saw in Section 2-F, the carrier passband has width \( 2W_L \).

A slight rearrangement of (5-2) relates the phase error to the input processes:

\[
\phi(t) = [1 - L(p)] \theta(t) - L(p) \frac{n(t)}{A} \tag{5-5}
\]

(omitting for the present the VCO noise term \( n_e(t) \), to be considered later). The first term of (5-5) is an error due to the incoming phase function, and, hence, it represents distortion due to filtering. There are usually two effects comprising \( \theta(t) \); one is the "doppler" phase shift \( d(t) \), and the other is the information process \( \psi(t) \). The distinction between \( d(t) \) and \( \psi(t) \) is that \( d(t) \) is a non-stationary, more or less deterministic phase variation, whereas \( \psi(t) \) is a stationary, zero-mean random process. The last term in (5-5) is phase error due to the presence of noise at the input of the loop. The mean-square phase error, which we shall denote by \( \Sigma^2 \), is thus composed of three terms:

\[
\Sigma^2 = \mu^2(t) + \delta^2 + \sigma^2. \tag{5-6}
\]

The first term represents the transient distortion due to \( d(t) \), the second is modulation distortion, and the third is the mean-square phase-noise. The latter two, being stationary, can be computed by using (2-19):

\[
\delta^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} |1 - L(j\omega)|^2 S_{pp}(j\omega) d\omega \tag{5-7}
\]

\[
\sigma^2 = \frac{N_b W_LL^2}{A^2} = \frac{N_b B_e L^2}{A^2}. \tag{5-7}
\]
The expression for \( \sigma^2 \) is a very interesting one; it states that the mean-square phase-noise is precisely equal to the noise-to-signal power ratio in the loop effective noise bandwidth times the squared maximum loop response. This must not be given the wrong interpretation. It should be realized that \( N_\theta \) is the noise density at carrier frequency, that \( A^2 = P \) is the incoming signal power, also at carrier frequency, and yet \( W_L \) is the equivalent noise bandwidth at baseband, computed by (5-1). Therefore \( \sigma^2 \) is not the noise-to-signal power ratio in the passband about the carrier frequency, which has width \( 2W_L \), as discussed earlier.

Equation (5-8) can also be written in terms of fiducial bandwidth \( w_L \) (or \( b_L \)) as

\[
\sigma^2 = \frac{N_\theta w_L}{A^2} = \frac{N_\theta b_L}{A^2} ,
\]

and our previous statement can be amended to read: the mean-square phase-noise is precisely equal to the noise-to-signal ratio in the fiducial loop bandwidth.

The total transient distortion is defined by the equation

\[
\varepsilon_T^2 = \int_0^\infty \mu^2(t) \, dt . \tag{5-9}
\]

It is convenient to include in \( d(t) \) the initial phase offset \( \theta_0 \). There is usually a complete lack of knowledge about \( \theta_0 \), so it is assumed to be a uniformly distributed random variable over \((-\pi, \pi)\), thus with variance \( \pi^2/3 \). The remainder of \( d(t) \) (call it \( d_i(t) \)) is a time-varying phase function (usually due to doppler shift) whose form is fairly well known. By Parseval's theorem, the total transient distortion can be computed in the frequency domain by using the relation

\[
\varepsilon_T^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| 1 - L(j\omega) \right|^2 E \left[ |D(j\omega)|^2 \right] \, d\omega . \tag{5-10}
\]

Note the resemblance this bears to \( \delta^2 \) in (5-7). The term \( D(s) \) is the Laplace transform of \( d(t) \). Equation (5-10) results when one substitutes \( E [D(s)D(-s)] \) for \( S_{\delta\delta}(s) \) in (5-7).

We shall have occasion to investigate these quantities more fully in Section 5-C, to determine the best loop configuration.

5-8. Calculation of Loop Bandwidth

The loop bandwidth formula (5-4) is rather easy to apply to the loops with simple filters. For future reference, we shall tabulate a few of these.
1. First-Order Loop

The first-order loop is one with no loop filter, $F(s) = 1$. Hence (5.4) gives

$$w_L = W_L = \frac{1}{2} AK \quad \text{or} \quad b_L = B_L = \frac{AK}{4}.$$  (5.11)

This relation enables one to express the loop transfer function as

$$L(s) = \frac{2W_L}{s + 2W_L}.$$  (5.12)

2. Second-Order Loop, Passive Integrator

The second-order loop in most widespread use is one in which the loop filter takes the form (see Fig. 5-2)

$$F(s) = \frac{1 + \tau s}{1 + \tau_0 s}$$

$$L(s) = \frac{1 + \tau_0 s}{1 + (\tau_2 + 1/\tau_0 s) + (r \tau_0 / AK)s^2}$$  (5.13)

(again we have normalized $F(0) = 1$, as agreed to in Chapter 3). In the usual case, $F(s)$ is designed with $\tau_1 \gg \tau_0$, to make $F(s)$ appear to have a pole at the origin.

It is convenient to define a quantity $r$ as the following ratio:

$$r = \frac{AK\tau_0^2}{\tau_1}.$$  (5.14)

The corresponding value of $L^2$ is given by the equation

$$L^2 \approx \frac{r^2}{2 \left(1 + \frac{\tau_2}{\tau_1} + 2r\right)^2 \left(1 - \frac{\tau_2}{\tau_1}\right)^2}$$

$$+ \left(1 + \frac{\tau_2}{\tau_1} + 2r\right)^2 - 2(r + 1)$$

$$\approx \frac{r^2}{2(1 + 2r)^2 + r^2 - 2r - 2}.$$  (5.16)

This last function (Fig. 5-3) is monotone decreasing in $r$, approaching unity as $r \to \infty$, and becoming infinite as $r \to 0$.

Then, from (5.4), the equivalent loop noise bandwidth can be computed,

$$W_L = \frac{(r + 1)}{2\tau_2 (1 + \tau / \tau_1)} L^2$$

$$\approx \frac{(r + 1)\left[r^2 - 2r - 2 + 2(1 + 2r)^2\right]}{2\tau_2 r^2},$$  (5.17)

the latter approximation being valid when $\tau_1 \gg \tau_2$. The fiducial bandwidth (2-37) is a somewhat simpler quantity:

$$w_L = \frac{r + 1}{2\tau_2 (1 + \tau / \tau_1)} \approx \frac{r + 1}{2\tau_2}.$$  (5.18)

![Fig. 5-2. Passive integrator loop filter](image-url)

**Fig. 5-2.** Passive integrator loop filter

**Fig. 5-3.** Variation of maximum loop response $L^2$, noise bandwidth $W_L$, fiducial bandwidth $w_L$, and frequency at maximum loop response, for second-order phase-locked loop, as a function of the parameter $r = AK\tau_0^2/\tau_1$.
The loop transfer function is then easily expressed in fiducial bandwidth:

\[
L(s) = \frac{1 + \left(\frac{r + 1}{2w_L}\right) s}{1 + \left(\frac{r + 1}{2w_L}\right) s + \left(\frac{r + 1}{2w_Lr^2}\right) s^2}.
\] (5-19)

From (5-13), one may compute the system damping behavior. For \(r < 4\), the roots of the system are under-damped, with damping factor \(\zeta\) and loop natural frequency \(\beta\) given by

\[
\zeta = \frac{1}{2} (1 + \frac{\tau_d}{\tau_1}) = \frac{r\nu}{2}.
\]

\[
\beta = \left(\frac{\Delta K}{\tau_1}\right) = \frac{r\nu}{\tau_2}.
\] (5-20)

For \(r = 4\), critical damping \((\zeta = 1)\) occurs, and for \(r > 4\), \(L(s)\) has two real negative poles.

3. Second-Order Loop, Perfect Integrator

Whenever an operational amplifier is permissible in the loop, the loop filter takes the form

\[
F(s) = \frac{1 + \tau_d s}{\tau_1 s}.
\] (5-21)

We have kept the notation here the same as that in the previous example for later convenience. This loop filter gives the same response as that given in (5-18), (5-19), and (5-20) when the \(\tau_d/\tau_1\) term is dropped. This indicates that whenever \(\tau_d \gg \tau_1\), the loop with an imperfect integrating filter (5-13) performs very much the same as that with the perfect integrating filter (5-21).

5-C. Optimization of Loop Parameters

According to the preceding sections, we can perform a linear analysis to determine the behavior of a phase-locked device whenever there is sufficient justification to warrant the assumption that \(\sin \phi = \phi\). But whenever noise is present, there will be arbitrarily large phase deviations if only we wait long enough; because of this, the assumption above cannot be strictly valid at every instant of time.

It is natural, then, to design the loop transfer function to minimize the effect of noise, that is, to choose \(K\) and \(F(s)\) to minimize the mean-square phase error. In this way, not only does the loop operate in the most linear manner, but also \(\theta(t)\) is most faithfully reproduced by \(\hat{\theta}(t)\). As can be seen from (5-6), the mean-square phase error is not independent of time; to minimize \(\Sigma^2\) thus would require a time-varying loop filter.

Rather than derive an optimum time-varying filter, Jaffe and Rechtin modified the criterion somewhat, basing their optimization on minimizing a somewhat different quantity. They define the total phase error, denoted by \(\Sigma^2\), as

\[
\Sigma^2 = \lambda^2 (\phi^2 + \delta^2 + \sigma^2)
\] (5-22)

in which \(\lambda^2\) is a Lagrange multiplier, a design parameter related to the bandwidth of the loop.

Minimization of \(\Sigma^2\) by choice of the loop transfer function \(L(s)\) specifies both \(\Delta K\) and \(F(s)\). But the question now is, "How does one choose \(L(s)\) to minimize \(\Sigma^2\)"? Jaffe and Rechtin recognized that \(\lambda^2 \mathcal{L} + \delta^2\) could be computed by replacing \(S_{\phi\phi}(s)\) in (5-7) by \(\lambda^2 E[D(s)D(-s)] + S_{\phi\phi}(s)\). This led them to the conclusion that \(\Sigma^2\) is minimized whenever \(L(s)\) is chosen in accordance with the Weiner optimization technique, which yields the Yavits-Jackson formula

\[
L_{\text{opt}}(s) = 1 - \frac{N^p/A}{[S(s)]^2}.
\] (5-23)

\[
S(s) = \lambda^2 E[D(s)D(-s)] + S_{\phi\phi}(s) + N/A^2.
\]

Some explanation of the bracket terminology here is in order. The bracket \(\left[ \right]^p\) refers to a type of "square-root" factoring of the enclosed function, retaining in \(\left[ \right]\) the left-hand-plane poles and zeros of the enclosed function only; singularities on the imaginary axis are equally divided between \(\left[ \right]^p\) and its mirror image \(\left[ \right]^-\).

The mechanics involved in (5-23) will be made clearer in the next Chapter.

5-D. The Effects of VCO Noise

One of the limiting factors governing the design of narrow-band phase-locked devices is the phase noise inherent in the output of the VCO. This noise appears as a random fluctuation, drift, etc., and is sometimes called "oscillator instability." More often, however, the term "instability" is used in context to mean "frequency instability," whereas the quantity of concern to us here is the phase stability of an oscillator.
To represent this effect, let \( n_i(t) \) be that noise voltage applied at the input of a perfect noiseless VCO which produces the same noise as appears in the actual VCO output. That is, \( n_i(t) \) is the VCO phase noise referred to its input. The closed-loop behavior of the loop is now somewhat modified by the addition of this internal noise:

\[
\hat{\theta}(t) = L(p) \left[ \theta(t) + \frac{n_i(t)}{A} \right] + \left[ \frac{1 - L(p)}{p} \right] K_{\text{vco}} n_i(t).
\]

(5-24)

There are many factors that contribute to the spectral makeup of \( n_i(t) \). The two most significant terms that appear are (1) thermal (Johnson) noise generated in the resistances of the oscillator, and (2) noise with a 1/f spectrum, as discussed in Section 2-G, associated with the transistors, varactor diodes, carbon resistors, etc. Thus, the spectrum of \( K_n n_i(t) \) can be approximated by

\[
K_{\text{vco}} S_{n_i} (j\omega) = N_{ov} + 2\pi N_{1v}/|\omega|.
\]

As we have previously indicated, the 1/f law cannot extend all the way down to zero frequency, so the equation above is really valid only when \(|\omega|\) is greater than some small value \( \epsilon \). But if \(|1 - L(s)|/s \) has a zero at the origin, this \( \epsilon \) need not be known.

The amount of phase error in the closed-loop output due to VCO noise can then be found by integration:

\[
\sigma_{\text{vco}}^2 = \frac{1}{\pi} \int_{0}^{\infty} \left| \frac{1 - L(j\omega)}{\omega^2} \right|^2 \left[ N_{ov} + 2\pi N_{1v}/\omega \right] d\omega.
\]

(5-26)

At this point, for our treatment, we shall assume that \( L(s) \) is the passive-integrator loop of (5-13), and again set \( \tau = AK_{\text{r}}/\tau_c \). The total phase error due to input and VCO noise, \( \sigma \), in \( \text{rms} \) of the fiducial bandwidth \( \omega_m \), is

\[
\sigma^2 = \frac{N_o \omega_m}{A^2} + \left( \frac{r + 1}{4r} \right) \frac{N_{ov}}{\omega_m} + g(r) \frac{N_{1v}}{\omega_m^2}.
\]

(5-27)

The function \( g(r) \) is shown in Fig. 5-4; it is given by

\[
g(r) = \begin{cases} 
\frac{(r + 1)^2}{4[r^2(r - 4)]^\frac{3}{2}} & \text{for } r > 4 \\
\frac{25}{16} & \text{for } r = 4 \\
\frac{(r + 1)^2}{2[r^2(4 - r)]^\frac{3}{2}} \left[ \frac{\pi}{2} - \tan^{-1} \frac{r - 2}{\sqrt{4 - r}} \right] & \text{for } r < 4.
\end{cases}
\]

(5-28)

1. Optimization of \( \omega_L \) and \( r \)

Depending on the values of \( N_{ov}/A^2, N_{ov}, \) and \( N_{1v} \), there are optimum choices for both \( \omega_L \) and \( r \).

As shown in Fig. 5-4, there is quite a broad range of \( r \) for which \( g(r) \) is fairly constant and is nearly equal to its minimal value of 1.5491 at \( r = 5.22 \). A quite useful range to use this approximation for \( g(r) \) is from about \( r = 3 \) to \( r = 10 \). Outside this region, \( g(r) \) is increasing, drastically when \( r < 3 \), and more slowly for \( r > 10 \).

On the other hand, the coefficient of \( N_{ov} \), i.e., \( (r + 1)/4r \), is a monotone decreasing for all \( r \), asymptotically approaching the value 0.25 as \( r \) becomes infinite. If \( r \) is larger than 5 or 6, for most practical purposes, we can use the value 0.25 with little fear of producing any significant error.

As a conclusion, then, we see that the best value of \( r \) to minimize \( \sigma^2 \) lies to the right of 5.22 and probably is less than 10, if any \( N_{1v} \) is present at all. In fact, the difference between \( \sigma^2_{\text{min}} \) and \( \sigma^2 \) for any \( r \) between 5 and 10 is almost inconsequential insofar as \( \sigma^2 \) is concerned. The same statement cannot be made for \( r < 4 \) (an underdamped loop). Thus we may take a value of \( r = 7 \) as being practically as good as \( r_{\text{opt}} \), regardless of the other parameters. (This reasoning alleviates the necessity for differentiating \( \sigma^2 \) to find the exact value of \( r_{\text{opt}} \).)

We can now use this value of \( r \) with respect to \( \omega_L \) and \( r \), equate to zero, and solve for \( \omega_L \) to find its best value. Straightforwardly, \( \omega_L \) is the solution to the equation

\[
\omega_L^2 = \left( \frac{r + 1}{4r} \right) \left( \frac{A^2 N_{ov}}{N_o} \right) \omega_L - 2g(r) \left( \frac{A^2 N_{1v}}{N_o} \right) = 0.
\]

(5-29)
2. An Example

Just to see how a typical design should be made, let us assume that we have parameters

\[ \frac{A^2}{N_0} = 6 \times 10^4 \]

\[ N_{uv} = 0 \]

\[ N_{tr} = 0.08 \text{.} \] (5-30)

The optimum value of \( \omega_L \) is then

\[ \omega_L = \sqrt{3.098 \left( \frac{A^2 N_{tr}}{kT} \right)} \]

\[ = 26 \text{ cps} \]

and the optimum loop transfer function is

\[ L_{opt}(s) = \frac{1 + 0.154s}{1 + 0.154s + (3.38 \times 10^{-2}) s^2} \]

These parameters also produce a phase deviation of

\[ \sigma = 2.5 \times 10^{-2} \text{ rad} = 1.43 \text{ deg rms} \text{.} \]

3. Conclusion

It should perhaps be mentioned at this point that designs of this type are most valid for oscillator "clean-up" loops and spacecraft carrier tracking loops, or in situations where the tone to be tracked is spectrally a very pure one. Design of the ground receiver tracking loop would probably not use this analysis, since its loop must track phase deviations imposed on the carrier by the noise in the spacecraft system. There are other factors that must be considered, such as frequency acquisition interval, lock-in time, and doppler tracking rate. What has been presented here should be taken merely as a guide as to what the ideal bandwidth is from a minimal noise point of view.

REFERENCES FOR CHAPTER 5


CHAPTER 6

OPTIMIZED DESIGN OF TRACKING FILTERS (LINEAR ANALYSIS)

A phase-locked receiver can be used either to track the carrier component of the incoming signal, or to demodulate its information, or both simultaneously. In space communications, the amount of doppler shift on the carrier is always measured, because it provides information about the spacecraft velocity. A carrier tracking loop is thus always present in a space-communications receiver.

6-A. Tracking Loop Design

Suppose it is our aim to provide the best filter to track a given doppler-phase polynomial $d(t)$ of degree $N-1$, assuming for the present that modulation of the carrier is absent. The form of $D(s)$ is then

$$D(s) = \frac{\theta_0}{s} + \frac{\Omega_0}{s^2} + \cdots + \frac{Q(s)}{s^N} \quad (6-1)$$

in which the degree of $Q(s)$ is less than $N$. The filter specified by (5-23) is

$$L_{opt}(s) = 1 - \left[ (-1)^N s^{2N} + \left( \frac{\lambda A^2}{N_o} \right) E \left[ Q(s)Q(-s) \right] \right]^{-1} \quad (6-2)$$

In the examples worked by Jaffee and Rechtin, the $\theta_0$ term was always set equal to zero. This corresponds to the case in which the loop is initially tracking with no phase error.

6-B. Optimum Filter for Random Phase Offset

The simplest example of loop optimization occurs for $d(t) = \theta_0$, a uniformly distributed random phase-offset. To find $L_{opt}(s)$, we insert $N = 1$ into (6-2): The denominator in (6-2) is

$$s^2 + \frac{\lambda A^2 \omega^2}{N_o} s^2 + \frac{\lambda A^2 \omega^2}{3N_o} = s + \frac{\pi \lambda A}{(3N_o)\omega} \quad (6-3)$$

Thus, the optimum filter is given by

$$L_{opt}(s) = \frac{\lambda A \omega}{(3N_o)\omega} \left[ \frac{1}{s + \frac{\pi \lambda A}{(3N_o)\omega}} \right] \quad (6-4)$$

The loop bandwidth, specified by (5-11), is

$$\omega_c = \frac{\lambda A \omega}{2(3N_o)\omega} \quad (6-5)$$

We can now eliminate the Lagrange multiplier to give the optimum loop design equations in terms of the loop bandwidth, viz.,

$$L_{opt}(s) = \frac{2W_L}{s + 2W_L}$$

$$F(s) = 1$$

$$K = 2W_L/A \quad (6-6)$$

6-C. Optimum Filter for Frequency and Random Phase Offset

The next example of interest is the optimization of a loop with frequency offset $\Omega_0$:

$$d(t) = \theta_0 + \Omega_0 t \quad (6-7)$$

where again $\theta_0$ is random. With $N = 2$ and $Q(s) = \theta_0 + \Omega_0 s$ inserted into (6-2), the denominator of $L_{opt}(s)$ becomes

$$s^2 + \left( \frac{\lambda A^2 \omega^2}{3N_o} + \frac{2 \lambda A \Omega_0}{N_o} \right) s + \frac{\lambda \Omega_0}{N_o} \quad (6-8)$$

The loop natural frequency is thus

$$\beta^2 = \frac{\lambda \Omega_0}{N_o} \quad (6-9)$$

in which case the optimum filtering function is given by

$$L_{opt}(s) = \frac{s \left[ 2\beta^2 + \frac{\pi \beta}{3\beta_0} \right] + \beta^2}{s^2 + 2\beta^2 + \frac{\pi \beta}{3\beta_0}} \quad (6-10)$$

34
Comparing (6-10) with (5-13) and (5-14), we recognize
the optimum parameters:

$$\tau_2 = \frac{1}{\beta} \left[ 2 + \frac{\pi \beta}{3\Omega_o} \right]^{1/3} = \frac{\pi}{\Omega_o [3(r-2)]^{1/3}}$$

$$\frac{AK}{\tau_1} = \beta^2 = \frac{3\Omega_o^2}{\omega_o^2} (r-2)$$

$$r = 2 + \frac{\pi \beta^2}{3\Omega_o^2}$$

$$\omega_L = \frac{(r+1)\Omega_o}{2\pi r} \left[3(r-2)\right]^{1/3}.$$  

We have thus translated the effect of the Lagrange multiplier implicit in $\beta$ to the design parameter $r$. It is important here to note that $r$ must always exceed 2 in optimized loop design. This comes about because of the assumption that $\theta_o$ was a uniformly distributed random variable.

The Jaffee and Rechtin example using $\theta_o = 0$ produces a result somewhat different from that above:

$$\tau_2 = 2^{1/3}/\beta$$

$$\frac{AK}{\tau_1} = \beta^2$$

$$r = 2 \quad \text{(so $\zeta = 0.707$)}$$

$$\omega_L = \frac{3\beta}{2(2^{1/3})}.$$  

1. Choice of Parameters

In most cases, $\Omega_o$ is not known prior to locking the loop (or else the loop would have been pretuned so as not to encounter any frequency offset in the first place). Hence $\Omega_o$ must be treated as a random variable. An examination of the average transient error in (5-9) reveals that $\epsilon_2^2$ is minimized whenever $E(\Omega_o) = 0$ (that is, the loop should be initially tuned to the expected incoming frequency) and that the $\Omega_o$ of our previous calculations is replaced, in this case, by the rms frequency offset, $[E(\Omega_o)]^{1/2}$.

Once suitable values of $A/N_{\omega_o^2}$, $\omega_L$, and $\Omega_o$ have been established, $\beta$ (and hence $r$) can be found, and this specifies what $L_{opt}(s)$ is to be used. The optimum loop filter $F_{opt}(s)$ is related to $L_{opt}(s)$ by (5-3):

$$F_{opt}(s) = \frac{1 + \tau_2 s}{\tau_1 s}.$$  

But rather than synthesize this $F_{opt}(s)$, which would require an operational amplifier, it is quite usual to substitute a passive filter whose characteristic is approximately the same as (6-12); for example, whenever $r \tau_1 \gg \tau_2$,

$$F_{opt}(s) = \frac{1 + \tau_2 s}{1 + \tau_1 s}$$  

(6-14)

can be used. This will introduce a steady-state phase error (Fig. 6-1)

$$\phi_{ss} = \frac{\Omega_o}{AK} \frac{\Omega_o}{\omega_L} \left( \frac{r+1}{2r} \right) \left( \frac{\tau_2}{\tau_1} \right)$$  

(6-15)

which does not exist with the $F_{opt}(s)$ in (6-13). Once the loop is locked, however, the VCO can be retuned to eliminate $\phi_{ss}$.

As a further consideration, one cannot expect a very good lock-in behavior when $\Omega_o$ is so large that the carrier frequency falls outside the initial loop passband.

On the other hand, it is usually desirable to design a tracking loop with its bandwidth much narrower than the initial frequency uncertainty region. Such loops are usually frequency-swept (by controlling the VCO input voltage $v$) slowly through the uncertainty interval to acquire lock. It is thus reasonable to design the tracking loop with parameters to ensure that lock-in proceeds optimally whenever the carrier enters the fiducial loop passband, i.e., when $\Omega_o = 2\pi b_L$. Upon inserting this condition into (6-11), the proper value of $r$ can be computed numerically:

$$r = 2.28245$$  

(6-16)

$$\zeta = 0.755.$$  

This compares very favorably with the results obtained by Jaffee and Rechtin (6-12). The optimum loop parameters for $r = 2.282$ are

$$\tau_2 = 1.643/\omega_L$$

$$\tau_1/ AK = \frac{1.180}{\omega_2^2}.$$  

(6-17)

and the corresponding optimum loop transfer function is given by

$$L_{opt}(s) = \frac{1 + (1.643/\omega_2^2) s}{1 + (1.643/\omega_2^2) s + (1.18/\omega_2^2) s^2}.$$  

(6-18)

*Purely for computational convenience here, we have specified the fiducial bandwidth $b_L$ rather than the noise bandwidth $B_L$ as defining the edge of the passband.
Fig. 6-1. Response of first- and second-order loops to various inputs

- **(a) First-order loop, phase step**

- **(b) Ideal integrator second-order loop, frequency step**

- **(c) First-order loop, frequency step**

- **(d) Passive integrator second-order loop, frequency step**

- **(e) First-order loop, Doppler-rate input**

- **(f) Passive integrator second-order loop, Doppler-rate input**
2. Evaluation of Transient Error

The total transient error is given by the expression

\[ e^2_t = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left( -\frac{\pi}{3s^2} + \frac{\Omega_0^2}{s^2} \right) \times \left[ 1 - L_{\text{opt}}(s) \right] \left[ 1 - L_{\text{opt}}(-s) \right] ds . \]  

(6-19)

With the form of loop filter given in (5-21), the transient error is

\[ e^2_t = \frac{\tau_0 \Omega_0^2}{2AKr} \left( 1 + \frac{\pi AK}{3\Omega_0^2\tau_0} \right) . \]  

(6-20)

For the optimum loop, then

\[ e^2_{t\text{ (opt)}} = \frac{\pi (r - 1)}{60\pi^2 \left[ 3(r - 2)^2 \right]^{1/3}} . \]  

(6-21)

With the value \( r = 2.282 \), and \( \Omega_0 = \pi w_L \),

\[ e^2_{t\text{ (opt)}} = 5.37/w_L . \]  

(6-22)

The parameters in the Jaffee–Rocen example (6-12) produce almost the same transient behavior:

\[ e^2_{t\text{ (JRL)}} = 5.4/w_L . \]  

(6-23)

3. The Effects of Doppler Rates in Second-Order Loops

If we assume that there is a slow doppler rate \( \Lambda_0 \) (rad/sec²) superimposed on a simple doppler shift, the discussion in Chapter 4 indicates that the passive-filter (5-13) second-order loop cannot be expected to maintain a small steady-state error. The ultimate phase motion due to this doppler rate, \( \lambda_2 = \Lambda_0 \) in the notation of (3-17), is linearly expanding in time, as

\[ \phi_{\text{ex}(\text{doppler rate})} = \frac{\Lambda_0}{AK} t = \frac{\Lambda_0 (r + 1)^2}{4r w^2_L} \left( \frac{t}{\tau_0} \right) . \]  

(6-24)

On the other hand, when the perfect integrating filter second-order loop (5-21) is used, there is a finite value of the steady-state error, viz.,

\[ \phi_{\text{ex}(\text{doppler rate})} = \frac{\Lambda_0 \tau_0}{AK} = \frac{\Lambda_0 (r + 1)^2}{4r w^2_L} . \]  

(6-25)

There is naturally a great resemblance here, and at \( t = \tau_0 \), (6-24) and (6-25) are equal.

In either of the two cases above, it is evident that, with other things being equal, the value of \( r \) causing the least steady-state error is \( r = 1 \); that is, the loop is an underdamped one with damping factor \( \zeta = 1/2 \).

Another observation that can be made is the following: In using a passive-filter second-order loop to track a signal having a doppler rate, it is necessary to compensate periodically for the ever-growing steady-state phase error. This is conveniently done by manually retuning the VCO at least every \( \tau_0 \) sec (as discussed in 3-5) to zero the phase error due to doppler rate. In this way (6-24) never exceeds (6-25).

4. Comments on the Choice of \( r \) in Second-Order Tracking Loops

In Section 5-D, we concluded that a value of \( r \) between 6 and 10 is desirable to minimize the effects of VCO noise. In Section 6-C1, we concluded that an underdamped loop with \( r = 2.282 \) is needed to minimize the total phase error. And in Section 6-C2, an underdamped loop with \( r = 1 \) provides the best tracking of a signal with nonzero doppler rate.

Normally, in a communications system, the contribution of VCO noise is not a factor of utmost concern. On the other hand, minimizing VCO noise is of great importance in designing frequency-generator "clean-up" loops. Too, the doppler tracking capability of a loop is critical in the specification of many receivers. Some \( \gamma \)-choice is thus available to the design engineer in picking an \( r \) to suit his particular application.

5. Case With VCO Noise Included

Comparing (5-7), (5-10), and (5-26), one can see that VCO noise effects can be included in the optimizing analysis quite easily when \( S_{\text{in}}, \gamma(s) \) is a rational spectrum. However, for the case we have considered, the VCO (input) noise has \( r 1/f \) term, or more precisely in the \( s \)-plane, \( N_1K_{\text{vco}}(s)^{-1} \), and (5-27) and (6-2) are not valid when such a branch point appears in one of the spectra. A Wiener optimum \( L(s) \) certainly exists, but the resulting loop filter may not be a rational function of \( s \). It is more meaningful to limit the discussion here to optimization under the constraint that \( F(s) \) be rational.

However, we easily recognize that the optimum loop for an initial frequency-step offset again must be of second order, and that the loop filter must be of the same type as that given in (5-21). Then, too, we know that, depending on values of the various parameters, the optimum value of \( r \) lies between 2.282 and about 7. It remains only to find specific values of \( r \) and \( w_L \).
We have computed separately all of the terms appearing in the total phase error $\Sigma_p$ (except distortion due to carrier modulation). Hence

$$\Sigma_p = \lambda \omega_p^2 + \sigma^2$$

$$= \frac{\lambda^2 \omega_p \Omega_r}{2AK} \left( 1 + \frac{\pi^2 AK}{3 \Omega_p^2 r_1} \right) + \frac{N_r \omega_L}{A^2}$$

$$+ \frac{(r + 1)N_{os}}{4r \omega_L} + \frac{g(r)N_{iw}}{w_L^2}. \quad (6-26)$$

Applying the philosophy that $\Omega_r$ be set equal to $\pi \omega_L$ (so that lock-in proceeds optimally once the offset frequency enters the loop passband), we reduce this to

$$\Sigma_p = \frac{N_r \omega_L}{A^2} \left( 1 + \frac{1}{1 + r} + \frac{4r}{3(1 + r)^3} \right)$$

$$+ \frac{(r + 1)N_{os}}{4r \omega_L} + \frac{g(r)N_{iw}}{w_L^2}. \quad (6-27)$$

The value of $\lambda$, of course, is a parameter left unspecified. Previously (in 6-9), we related it to the loop natural frequency $\beta$ (and through $r$, to the loop bandwidth, as well), which then became the arbitrary parameter in the loop. We can also do this in the present case; inserting (6-9) into (6-27), we find

$$\Sigma_p = \frac{N_r \omega_L}{A^2} \left( 1 + \frac{1}{1 + r} + \frac{4r}{3(1 + r)^3} \right)$$

$$+ \frac{(r + 1)N_{os}}{4r \omega_L} + \frac{g(r)N_{iw}}{w_L^2}. \quad (6-28)$$

With this particular evaluation of $\lambda$, there is an optimum way to pick $r$ and $w_L$. However, we can note that none of the coefficients, as functions of $r$, varies drastically over the range 2.282 to 7; this means that $\Sigma_p$ is not extremely sensitive to $r$. Such is not the case with $w_L$, and the designer should, in all practicality, seek to determine that value of $w_L$ which will minimize $\Sigma_p$.

**REFERENCE FOR CHAPTER 6**

CHAPTER 7
THE DESIGN OF SYNCHRONOUS-DETECTOR AGC SYSTEMS

One of the fundamental circuits built into almost every receiver today is some sort of closed-loop regulating system that automatically adjusts the receiver gain to maintain a constant output level. The most common circuit used in broadcast receivers is one in which the IF signal is rectified, filtered, and applied to the grids of variable-μ tubes in the IF amplifier. Such a circuit (often referred to as Automatic Volume Control, or AVC) is equally as sensitive to noise voltages in the IF passband as it is to signal voltages, and, consequently, affords only moderate stabilization of the signal amplitude.

In phase-locked receivers, it is more usual to use a synchronous amplitude detector followed by a very narrow-band filter to derive the feedback voltage. The output of this filter, chiefly due to the presence of signal and relatively insensitive to noise, can be used to control the receiver gain very efficiently.

Such a loop we shall distinguish from the AVC loop mentioned above by the name Automatic Gain Control or AGC.

In what is to follow, we shall present a linear feedback theory for the design of AGC loops. As a result, one may predict performance analytically as a function of loop parameters within a certain measure of accuracy. The treatment here essentially follows the work of Victor and Brockman, with some minor extensions.

7-A. The Synchronous-Amplitude-Detector AGC Loop

The block diagram shown in Fig. 7-1 shows the principal elements involved in the design of an AGC System. Briefly, the input $2^n A(t) \cos (\omega_n t + \theta(t)) + n(t)$ enters a variable-gain amplifier whose gain is a function of a feedback control-voltage $c(t)$. The output (amplitude-stabilized), appearing as $2^n A \cos (\omega_n t + \theta(t)) + n(t)$, is passed to a phase-locked loop to derive a coherent reference, against which it is synchronously detected. At this point, an external gain-reference voltage $e_0$ is compared with the detector output; the resulting error voltage, filtered by $Y(s)$, then controls the receiver gain in a way that tends to null any differences between $e_0$ and the detector output.

There is a definite relation between AGC voltage $c(t)$ and the receiver gain (call it $1/A^*(c(t))$, or merely $1/A^*(t)$), for if the receiver output is held nearly constant at an rms value of $A$, it follows that $c(t)$ is also.

![Fig. 7-1. A synchronous-detector AGC loop, using a phase-locked loop to provide a coherent reference. (The actual phase-locked loop may be part of the gain-controlled receiver.)](image-url)
related to the rms incoming signal amplitude \( A(t) \). Thus, \( c(t) \) can be monitored by a voltmeter as a calibrated measure of input signal strength.

When double-frequency terms of the detector output are omitted, the input to \( Y(s) \) is

\[
Z_{AGC}(t) = -e_0 + AK_d \cos \phi(t) + K_d n_i(t)
\]

\[
= -e_0 + \frac{A(t)}{A^*(t)}K_d \cos \phi + \frac{K_d n_i(t)}{A^*(t)}
\]

The constant \( K_d \) refers to the gain of the synchronous detector, but otherwise the notation follows that previously used.

Victor and Brockman recognized that if the AGC loop is performing well, one may approximate\(^6\)

\[
-e_0 + \frac{A(t)}{A^*(t)}K_d \cos \phi \approx e_0 \ln \left[ \frac{A(t) K_d \cos \phi}{A^*(t) e_0} \right] = e_0 \ln A(t) - e_0 \ln \left[ \frac{A^*(t) e_0}{K_d} \right] + e_0 \ln \cos \phi(t).
\]  

(7.2)

The error here is less than half the square of the left-hand quantity.

The next step in the analysis is that of choosing a function for the variation of receiver gain with AGC voltage. Mathematically, it is convenient to express the gain in a logarithmic Taylor series:

\[
\frac{e_0 A^*(t)}{K_d} = 10^{[K_R e_0 K_c (t^2 + \cdots )/20]}
\]

\[
20 \log \left[ \frac{e_0 A^*(t)}{K_d} \right] = K_R + K_c c(t) + \cdots .
\]  

(7.3)

A characteristic with only the first two terms is nearly valid for many voltage-controlled-gain amplifiers, although \( K_R \) and \( K_c \) depend on the operating point to a certain degree. The quantity \( A^*(t)e_0/K_d \) is an adjusted attenuation factor. By retaining only the first two terms above, we assume the receiver has an adjusted attenuation in decibels which varies linearly with control voltage (over a limited range): there is an attenuation of \( K_R \) db in the absence of control, and an additional \( K_c \) db of attenuation when \( c(t) \) is applied. It should be stressed that \( K_R \) would be the adjusted value of receiver attenuation at a particular operating point if \( c(t) \) could be set equal to zero. The actual receiver attenuation \( K_{rec} \) corresponding to \( K_R \) is\(^15\)

\[
K_{rec} = K_R + 20 \log \left( \frac{e_0}{K_d} \right).
\]

Now define the quantities\(^11\)

\[
a(t) = 20 \log A(t)
\]

\[
a^*(t) = 20 \log \left( \frac{e_0 A^*(t)}{K_d} \right).
\]  

(7.4)

That is, \( a(t) \) and \( a^*(t) \) are the signal strength and adjusted receiver attenuation expressed in decibels. Since \( \ln x = \log x/\log e \),

\[
c(t) = K_R v(p) \left[ \frac{e_0}{20 \log e} \right] [a(t) - a^*(t)] + \left[ \frac{e_0}{\log e} \right] \log \cos \phi(t) + \frac{K_d n_i(t)}{A^*(t)}
\]

\[
= \frac{a^*(t) - K_R}{K_c}.
\]  

(7.5)

From this equation, we can draw the equivalent circuit shown in Fig. 7-2. Note, however, that the loop is linear in the logarithms of input and response, rather than these quantities themselves. The equivalent loop gain \( K_{AGC} \) and AGC-loop transfer function \( C(s) \) are given by

\[
K_{AGC} = \frac{K_R e_0 K_c}{20 \log e}
\]

\[
C(s) = \frac{K_{AGC} Y(s)}{[1 + K_{AGC} Y(s)] K_c}.
\]  

(7.6)

Straightforwardly, \( a(t) \) and \( a^*(t) \) are related by the linear operator equation

\[
a^*(t) = K_c C(p) \left[ a(t) + \left( \frac{20 \log e}{e_0} \right) K_d n_i(t) + 20 \log \cos \phi(t) \right] + [1 - K_c C(p)] K_R.
\]

(7.7)

\(^6\)Here we use "\( \ln x \)" to denote the natural logarithm of \( x \), and "\( \log x \)" to denote the base-ten logarithm of \( x \).

\(^{10}\)The value of \( K_{rec} \) can be set by a bias added to \( c(t) \), if desired. However, the equation for \( K_{rec} \) shows this is not necessary, as adjustment of \( e_0 \) will do the same thing.

\(^{11}\)Since the input signal power is related to \( A(t) \) by

\[
P = A^2/R_{in},
\]

where \( R_{in} \) is the input resistance of the receiver, \( a(t) \) differs from the value of \( P \) in \( \text{db} \) by the term \( 10 \log R_{in} \); \( 10 \log P = a(t) - 10 \log R_{in} \).
The AGC voltage is thus related to the input signal level by

\[ c(t) = C(p) \left[ a(t) + \left( \frac{20 \log e}{e_o} \right) K_o n_i(t) + 20 \log \cos \phi(t) - K_a \right]. \]  

(7-8)

7-B. AGC Stability

The steady-state mean and variance of \( a^*(t) \) due to input noise are readily found: if \( A(t) \) changes slowly with respect to the time constants in \( C(s) \),

\[ E[a^*(t)] = K_a C(0) E[a(t)] - K_a C(0)(10 \log e) \sigma^2 + [1-K_a C(0)] K_a \]  

\[ \text{var}[a^*(t)] = \left[ \left( \frac{20 \log e}{e_o} \right)^2 K_d \left( \frac{N_o w_o}{A^2} \right) A^2 \right] C(0) K_a^2 + 2(10 \log e)^2 \sigma^2 \left( \frac{w_o}{w_e} \right) C(0) K_a. \]  

(7-9)

In this calculation, we have approximated \( \ln \cos \phi \) by \( -\frac{1}{2} \phi^2 \), and we have assumed that \( \phi(t) \) is a Gaussian process. Both of these assumptions are somewhat in error, but behavioral indications are sufficient for our purposes. It should be noted that \( N_o \) is the spectral density of the stabilized noise \( n_i(t) \), and not the input noise \( n_i(t) \); for this reason, it is referred to \( A^2 \), the stabilized signal level, as \( N_o/A^2 \) (which is the same as \( N_o/A(t) \) at the receiver input, however). As a fair approximation, \( A = e_o/K_o \) can be used in evaluating the variance of \( a^*(t) \) above. When \( K_{100} \) is high, \( C(0) \approx 1/K_a \).

The fluctuation of \( a^*(t) \) depends not only on the input signal-to-noise ratios in the phase-locked loop bandwidth \( w_d \) and in the AGC loop bandwidth \( w_e \), but also upon the value of stabilized signal level at receiver output. The mean value of \( a^*(t) \) is less than that of \( a(t) \) by about 4.35 dB at \( \sigma^2 = 1 \).

7-C. Calibration Equation

Equation (7-7) shows how attenuation varies with signal strength and noise. Moreover, \( a(t) \) is related to \( c(t) \) by (7-8). Thus, it is possible to relate the steady-state mean value of \( a(t) \) (call it \( a \), to that of \( c(t) \), or merely \( c \), and thus, to relate \( c \) to the average input signal\(^{11} \) strength \( a \):

\[ a = K_a + \frac{1}{C(0)} c + (10 \log e) \sigma^2 \]

\[ = K_{rcs} + 20 \log(e_o/K_o) + \left( \frac{1}{C(0)} \right) c + (10 \log e) \sigma^2. \]  

(7-10)

This calibration curve is well demonstrated in actual practice, as attested to by the results shown in Fig. 7-3. The receiver whose characteristic is depicted is the S-band receiver at Pioneer Site, Goldstone Tracking Station, California.

The variance of this estimate of \( a \) for a particular measured value of \( c \) is the same as the variance of \( a^*(t)/K_a \) given in (7-8), because, as shown in (7-8), \( c(t) \) and \( a^*(t)/K_a \) have the same fluctuations due to noise.

\(^{11}\)If 10 log \( R_{in} \) is subtracted from the right-hand side of (7-10), the calibration curve relates \( c \) to signal power in dbw.
7-D. Dynamic Behavior of AGC Loops

In calibrating the AGC loop, we assumed that \( A(t) \) was more or less constant—or was at least changing very slowly with respect to time constants in the loop. But this may not be a realistic assumption during the early postlaunch phases of a spacecraft flight, or even later when the spacecraft is perhaps tumbling through space. In such cases, one must investigate the transient behavior of the loop in more detail and, if one can predict what types of transients are to be expected, choose parameters to optimize performance.

Fluctuations in signal amplitude can be of two types: first, there can be a nonstationary waxing and waning of the signal, such as is due to the effects of changes in range; and second, there can be superimposed on the first a stationary perturbation such as might occur as the spacecraft antenna seeks to retain its bearing toward Earth, or as a result of tumbling.

The modified Wiener formula (5-23) can be fitted to this problem by expressing

\[
A(t) = A_d(t)A_w(t) \\
a(t) = a_d(t) + a_w(t)
\]

where we have put

\[
a_d(t) = 20 \log A_d(t) \\
a_w(t) = 20 \log A_w(t)
\]

and in which we have designated \( A_d(t) \) and \( A_w(t) \) as the stationary and nonstationary parts of \( A(t) \), respectively.

For convenience, \( a_d(t) \) can be taken to be a zero-mean process by lumping the mean value of \( a(t) \) into \( a_d(t) \).

The Wiener optimum filter is

\[
C_{opt}(s) = \frac{1}{S'(s)} \cdot \left[ \frac{\lambda^2 E[\ddot a_d(s) \ddot a_d(-s)] + S_{a_d a_d}(s)}{[S'(s)]^2} \right]_{pr}
\]

\( \quad (7-13) \)

\[
S'(s) = \lambda^2 E[\ddot a_d(s) \ddot a_d(-s)] + S_{a_d a_d}(s) + (20 \log e)^2 \left( \frac{N_c}{\lambda^2} \right)
\]

\( + (10 \log e)^2 S_{a_d a_d}(s) .
\]

For convenience, we have approximated \( e_0/K_0 = A \) in the equation above and set \( \ddot a_d(s) = \ddot{L}[a_d(t)] \).

The bracket \( [\cdot] \) is the right-half-plane image of \( [\cdot] \), and \( [\cdot]_{pr} \) refers to the physically realizable part of the enclosed function.

The latter can be computed as \( \ddot{L}[\dddot{s}] \) (i.e., the Laplace transform of the inverse Fourier transform of the quantity). An equivalent method for computing \( \dddot{r} \), when the enclosed quantity is a rational function is obtained by expanding the function in a partial fraction expansion, but retaining only those terms having left-half-plane poles, poles at the origin, and poles at infinity.

For simple transients, much can be learned about the form of \( C_{opt}(s) \) without resorting to solution of (7-13). For example, if we are concerned with gain-tracking an input of the form \( a(t) = a_d(t) = \lambda t^k / n! \), then by a reasoning similar to that in Section 3-B, we find (neglecting the term due to noise)

\[
E [a_d(t) - a^*(t)]_t \sim \left\{ \begin{array}{ll}
\lambda e^{pt} & \text{for } n \geq 1 \\
K \log q(0)(n-l)! & \text{for } n < 1,
\end{array} \right.
\]

\( \quad (7-14) \)

We assumed here that \( Y(s) \) takes the form

\[
Y(s) = \frac{q_i + i u_i}{s^2 p(s), \quad a_i(0) \neq 0, \quad p(0) \neq 0. \quad (7-15) \)
\]

Thus, for stable AGC operation, \( Y(s) \) must have at least as many poles at the origin as the degree of \( a_d(t) \).
As a particular case, when \( a_d(t) = a \), \( Y(s) \) may take the form
\[
Y(s) = \frac{1}{1 + \tau_{AGC}} \quad (7-16)
\]
in which case there will be a constant gain-tracking transient error
\[
E [a_d(t) - a^*(t)]_{ss} = \frac{a}{K_{AGC}}. \quad (7-17)
\]

But if \( a_d(t) = a_t \), the loop filter (7-16) produces a steady-state transient error that grows linearly in time:
\[
E [a_d(t) - a^*(t)]_{ss} \to \frac{a_1}{K_{AGC}} t \approx \frac{a_1}{2\omega_c} \left( \frac{t}{\tau_{AGC}} \right). \quad (7-18)
\]

The AGC loop bandwidth in both cases is
\[
\omega_c = \frac{1 + K_{AGC}}{2\tau_{AGC}} \approx \frac{K_{AGC}}{2\tau_{AGC}}. \quad (7-19)
\]

By making \( K_{AGC} \) large enough, a signal-strength change of a few dB/sec can be tolerated for a short time.

A better \( Y(s) \) to follow \( a_t \) would, of course, be of the form \( Y(s) = 1/\tau_{AGC} s \), in which case there would be a steady-state gain-tracking error
\[
E [a(t) - a^*(t)] = \frac{a_1}{K_{AGC}} = \frac{a_1}{2\omega_c}. \quad (7-20)
\]

A comparison of (7-18) and (7-20) immediately shows that when \( K_{AGC} \) is large, the two AGC loops perform about the same for \( t < \tau_{AGC} \).

**REFERENCE FOR CHAPTER 7**

CHAPTER 8

THE DOUBLE-HETERODYNE PHASE-LOCKED RECEIVER

Communications requirements for deep-space missions require receiving systems whose capabilities at first sound so extreme as to be unrealizable. The development of the double-heterodyne receiver with phase-locked carrier reference, however, has made coherent communications the standard language of the systems engineer. Indeed, except for a few minor refinements, parameter optimizations, etc., there are few questions concerning the behavior of these receivers that cannot be readily answered.

A double-heterodyne receiver, briefly, is merely one in which there are two separate intermediate-frequency amplification stages. Such receivers can be designed to operate at very low signal levels with a great measure of stability, precision, and reliability.

8-A. Basic Configuration of the Receiver

The receiver shown in Fig. 8-1 combines the advantages of intermediate-frequency amplification to produce high gains (and there are two such stages) with those of the phase-locked loop: namely, coherent communications capability at low bandwidths and predictable stability. It is not immediately obvious that the concatenation of front-end mixing, subsequent IF stages, and ultimate phase detection constitutes a phase-locked loop as we have seen it in previous chapters. There are practical necessities, as well as subtle advantages, to the use of a limiter prior to phase detection, and although we have nowhere accounted for such a device in our previous theory, there is little, if any, deleterious effect due to its presence.

The tracking portion of the receiver operates as follows: The input function is $2^\alpha A(t) \cos [\omega_1 t + \theta(t)] + n_d(t)$, where $A(t)$ is a slowly varying rms signal amplitude and $n_d(t)$ is a wide-band noise with density $N_{no}$. The input is mixed with a multiplied-up version of the VCO output, so the first IF output process is

$$K_1 \ [2^\alpha A(t) \cos (\omega_1 t + \theta - \hat{\theta} + \theta_1) + n_i(t)].$$

The heterodyne operation followed by linear gain does not alter the signal power/noise-spectral-density ratio, as we have discussed in Chapter 2. The first IF frequency is

$$f_1 = \frac{\omega_1}{2\pi} = \frac{2\pi}{\omega_0 - \omega_1}.$$ 

(We assume $\omega_0 > \omega_1$, as a practical matter, although our analysis does not require it.) The signal is now heterodyned against a locally generated, free-running reference of the form $\cos (\omega_2(t) - \theta_2)$, so the output of the second IF amplifier is

$$K_{1s}K_{1s} \left\{ 2^\alpha A(t) \cos [(\omega_1 - \omega_2)t + \theta - \hat{\theta} + \theta_2] + n_i(t) \right\}.$$ 

Again, the signal-power/noise-spectral-density ratio is preserved, now centered at the second IF frequency

$$f_2 = \frac{\omega_2}{2\pi} = \frac{\omega_1 - \omega_2}{2\pi}.$$ 

(In general, $f_2$ may take a negative sign when $\omega_1 > \omega_2$. This is a purely mathematical distinction, and need cause us no great concern here.)

Each of the IF amplifiers has a known bandwidth, so the overall IF bandwidth or predetection bandwidth $W_N$ can be computed as in (2-43); however, the first IF normally has much wider bandwidth than does the second, so $W_N$ is essentially the equivalent noise bandwidth $W_H$ of the second IF filter $H(s)$.

As a matter of theory, either or both of the IF amplifiers can be gain-controlled by an AGC voltage, although it is practically necessary to control only one of them, say the first. Hence, in the terminology of Chapter 7, $K_{1s}K_{1s} = 1/A^*(t)$, so the second IF output is, in our previous notation,

$$x(t) = 2^\alpha A \cos (\omega_2 t + \theta - \hat{\theta} + \theta_2) + n_i(t).$$

This voltage feeds a band-pass limiter whose saturation limits are $\pm l$ v. The limiter mean-square output is, of course, constant at a value $l^2$, of which $2(2l^2/\pi)^2$ lies in the frequency zone about $f_2$. A signal component is present whose mean-square value is, let us say, $2(2la/\pi)^2$; the factor $a^2$ is then a signal-power suppression factor, with $0 \leq a \leq 1$. The remainder of the limiter output in the IF
zone is noise; it has total power $2(2l/n)^2\left(1 - a^2\right)$. The limiter output signal-to-noise ratio $\rho_i$ is thus a function of $a$ only:

$$\rho_i = \frac{a^2}{1 - a^2}.$$  \hfill (8-1)

We shall investigate the way $a$ behaves as a function of its input signal and noise a little later; for the present, let us express our answers in terms of $a$.

The limiter output feeds a phase detector whose reference input is rationally related to the free-running oscillator used to produce the second IF frequency; i.e., for some (relatively) fixed $\theta$, it may be taken as being of the form

$$\sin (M_1\omega_i t - M_1 \theta_1 - \theta).$$

and $M_1$, like $f_1$ and $f_2$, may take on a negative sign, if desired).

Correspondingly, the detector output is

$$y(t) = K_d \left\{ \alpha \sin \left[ \frac{\omega_0 - \omega_1 - (1 + M_1)\omega_{21}}{M_1} \right] t + \theta + \hat{\theta} \right. \left. + (1 + M_1) \theta_1 + \theta_2 \right\} + n(t).$$  \hfill (8-2)

The constant $K_d$ includes the gain of the detector, the limits $\pm l$, etc. When the loop is in lock, the multiplied-up VCO output must have as its frequency

$$\omega_{21} = \omega_0 - (1 + M_1)\omega_{21}.$$  \hfill (8-3)

The values of the IF frequencies are thus directly related to $f_{21}$:

$$f_1 = (1 + M_1)f_{21}$$  \hfill (8-4)

$$f_2 = M_1f_{21}$$

which may vary with any fluctuation $f_{21}$ may have.
The heterodyne signal coming into the first mixer from the VCO through a frequency multiplier will then be one of the form

\[ v(t) = \cos(w_{01}t + aMK_v v_{\text{vco}} F(p) + \frac{\sin(\theta - \hat{\theta} + (1 + M_1)\theta_1 + \theta_2)}{\alpha}) \]

\[ \times \left[ \frac{\sin(\theta - \hat{\theta} + (1 + M_1)\theta_1 + \theta_2)}{\alpha} + \frac{MK_v v_{\text{vco}} n_o(t)}{p} \right] \quad (8-5) \]

including the equivalent VCO input noise, \( n_o(t) \). The receiver estimate of the incoming phase function \( \hat{\theta}(t) \) thus satisfies the following relation:

\[ \hat{\theta} = aK_v v_{\text{vco}} M F(p) + \frac{\sin[\theta - \hat{\theta} + (1 + M_1)\theta_1 + \theta_2]}{\alpha} + \frac{MK_v v_{\text{vco}} n_o(t)}{p} \]

\[ = \frac{aK_v v_{\text{vco}} M F(p)}{p + aK_v v_{\text{vco}} M F(p)} \]

\[ \times \left[ \theta(t) + \frac{n_o(t)}{\alpha} + (1 + M_1)\theta_1(t) + \theta_2 + \frac{n_o(t)}{aK_v v_{\text{vco}} F(p)} \right] \]

\[ \quad (8-6) \]

If properly designed, the loop naturally adjusts itself, trying to zero the error; a linear analysis applies whenever the detector output error \( \phi = \theta - \hat{\theta} + (1 + M_1)\theta_1 + \theta_2 \) is small; in this case, the loop estimate is

\[ \hat{\theta}(t) = \frac{aK_v v_{\text{vco}} M F(p)}{p + aK_v v_{\text{vco}} M F(p)} \]

\[ \times \left[ \theta(t) + \frac{n_o(t)}{\alpha} + (1 + M_1)\theta_1(t) + \theta_2 + \frac{n_o(t)}{aK_v v_{\text{vco}} F(p)} \right] \]

\[ \quad (8-7) \]

This response is the same form as that given in (5-2), except for the term \((1 + M_1)\theta_1 + \theta_2\), which contributes to the loop noise, and which may also cause a varying phase lag between \( \theta \) and \( \hat{\theta} \) even in the absence of input noise. This term represents a heterodyne noise present in the loop.

The fact that \( \hat{\theta} \) does not follow \( \theta \) exactly (due to \((1 + M_1)\theta_1 + \theta_2\)) need cause little concern so long as synchronization detection of information processes, such as doppler shift, phase modulation, etc., are also accomplished using the same reference frequencies \( w_{01} \) and \( M_1 w_{01} \), combined in the same manner as above. For example, the AGC detector uses a 90-deg shifted version of \( M_1 w_{01} \) to derive the stabilized amplitude of the signal; hence the input to the AGC filter and amplifier is

\[ -e_o + AK_o \left[ \cos \phi + \frac{n_{\text{vco}}(t)}{A} \right] \]

So when the receiver is tracking, \( \phi = \theta_1 - \hat{\theta} + (1 + M_1)\theta_1 + \theta_2 \) is small, and the AGC detector works precisely as described in Chapter 7. Detection of phase modulation by a second synchronous detector also follows the same behavior exactly. Doppler-frequency extraction can also be implemented in a similar way.

There are, however, some other minor differences between the simple phase-locked loop and the double-heterodyne receiver, principally in the phase-detector outputs. It is necessary to distinguish between \( \epsilon_o = \hat{\theta} - \hat{\theta} \), the loop tracking error, and the quantity \( \phi = \theta - \hat{\theta} + (1 + M_1)\theta_1 + \theta_2 \), the loop detector error. The latter of these is the only one that need concern us if care is taken, as above, in proper information demodulation. In the simple no-heterodyne loops of Chapters 4, 5, and 6, \( \phi \) and \( \epsilon_o \) were the same, but they are not the same quantity in the receiver above.\(^13\)

Note that fluctuations in \( \phi \) occur (compare (8-7) with (5-24), for example) as if \((1 + M_1)\theta_1 \) were added to the input signal \( \theta(t) \). Of course, the input signal itself has been generated by an oscillator whose output frequency has been multiplied to carrier frequency by some factor \( M_2 \), and generally \( M_2 \gg 1 + M_1 \). Assuming that the oscillators associated with \( \theta(t) \) and \( \theta_1(t) \) are of equal quality, insofar as phase purity is concerned, we observe that unwanted fluctuations due to \((1 + M_1)\theta_1 \) are masked by those of \( \theta(t) \), the latter being roughly \( M_2/(1 + M_1) \) times as large. Hence, the contribution of \( \theta_1(t) \) to \( \phi \) can usually be neglected.

The loop transfer function \( L(s) \) is exactly the same form as (5-3) with the parameter \( AK \) replaced by

\[ AK = aK_v v_{\text{vco}} MF \]  

(8-8)

(Here we have made explicit inclusion of \( F \), the loop-filter dc gain, previously assumed to be unity, in the event that some loss is present. In this way, we may assume that \( F(s) \) takes the previously treated form.)

\( ^{13} \)There are ways of reducing the difference between \( \phi \) and \( \epsilon_o \). For example, \( w_{01} \) can be derived as \( w_{01} + w_o \), wherein a spectrally very pure standard reference frequency, such as one derived from an atomic-resonant oscillator, is mixed with a local free-running oscillator \( w_o \). Final phase detection then uses \( w_o \). The effect of this, in our terminology above, is that \( M_1 \) is set to \(-1 \), so that \( \phi \) differs from \( \epsilon_o \) only by \( \theta_1 \) plus the standard oscillator phase, \( \theta_o \). However, \( \theta_o \) is a constant, and the standard is very stable; hence for most practical purposes, \( \phi \) and \( \epsilon_o \) differ by a constant phase lag, tunable if desired.
8-B. Effects of Band-Pass Limiting

In Chapter 7 we discovered that AGC action may jitter the value of $A$ by several db even at a given constant-input amplitude $A(t)$. Without a limiter in the loop (A then replaces $a$ in (8-8) above), the loop bandwidth would jitter in the same manner as $A$. But with a limiter, $a$ is related directly to the predetection signal-to-noise ratio, generally a much more stable quantity. Further, $a$ is usually much larger than $A$, so the loop gain need not be so great to give a desired bandwidth.

Since the AGC alone is unable to maintain a constant receiver output level, limiting also tends to increase the effective dynamic range of the phase detector, and hence, that of the entire receiver. These reasons make it almost necessary to have IF limiting in a quality receiver.

The limiter output zonal signal-to-noise ratio varies as a function of its input SNR in a known manner, derived by Davenport in 1953. However, Davenport's result is a relation between the output signal strength and the total noise in the output zone. Generally speaking, a limiter tends to have a wider output noise bandwidth than does its input, whereas $L(s)$ is usually chosen to have a much narrower response than either of these bandwidths. Neglecting internal oscillator noises and assuming that $S_{nn}(jw)$ is fairly constant over the response region of the loop (i.e., $w_n > w_l$), one may derive a result similar to (5-8):

$$\sigma^2 = \frac{S_{nn}(0)}{\alpha^2} w_l. \quad (8-9)$$

The limiter output spectrum has some fiducial noise bandwidth $w_l$:

$$w_l = \frac{1}{S_{nn}(0)} \left[ \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_{nn}(j\omega) d\omega \right] = \frac{1}{S_{nn}(0)} (1 - \alpha^2) \quad (8-10)$$

since, of course, the integral represents the total normalized zonal noise power in accordance with the implicit definition of $n(t)$ in (8-2). Now as a result, the noise variance is

$$\sigma^2 = \left( \frac{1 - \alpha^2}{\alpha^2} \right) \frac{w_l}{w_1} = \frac{w_l}{\rho_l w_1}. \quad (8-11)$$

Davenport's result, which we mentioned earlier, is that $\rho_l = 2\rho_H$ when the predetection SNR, $\rho_H = A^2/\left( N_0 w_n \right)$, is large, and $\rho_l \approx (\pi/4) \rho_H$ when $\rho_H$ is very small. The actual formula for $\rho_l$ is very complicated. However, one approximation to this, which fits extremely well (see Fig. 8-2) over the entire range of $A^2/N_0 w_n$, is

$$\rho_l/\rho_H = \frac{0.7854 + 0.4768\rho_H}{1 + 0.2384\rho_H}. \quad (8-12)$$

The corresponding signal amplitude suppression factor is

$$\alpha = \left[ \frac{0.7854\rho_H + 0.4768\rho_H^2}{1 + 1.024\rho_H + 0.4768\rho_H^2} \right]^n. \quad (8-13)$$

As the loop bandwidth $w_l$ is a function of loop gain, it is thereby a function of $\alpha$. The bandwidth of a passive-integrator loop whose filter is that of (5-13) is related to the parameter $r$ of (5-14). In the double-heterodyne receiver with IF limiting, this $r$ is

$$r = \frac{aK_VK_BMF}{\tau_1} \quad (8-14)$$

and therefore the loop fiducial bandwidth is approximately

$$w_l = \frac{aK_VK_BMF \tau_1^2 + \tau_1}{2\tau_2 \tau_1}. \quad (8-15)$$

At very high input SNR's, $\alpha$ approaches unity, a result that causes the loop bandwidth to approach asymptotically the value

$$w_{l(\text{max})} = \frac{K_VK_BMF \tau_1^2 + \tau_1}{2\tau_2} \approx \frac{1}{2} K_VK_BMF \tau_2$$

$$\approx \frac{\tau_2 w_l}{\alpha_0 (\tau_a + 1)}. \quad (8-16)$$

The phase variance can be related to input signal and noise parameters by introducing $\rho_H = A^2/N_0 w_n$ into

![Fig. 8-2. Davenport's band-pass limiter zonal SNR curve](image-url)
(8-11). With this done (again assuming negligible oscillator noise), the following expression for \( \sigma^2 \) results:

\[
\sigma^2 = \frac{N_{awL}}{A^2} \left( \frac{\omega_H \rho_H}{\omega_I \rho_I} \right) \]

\[
= \frac{N_{awL}}{A^2} \Gamma. \quad (8-17)
\]

The factor \( \Gamma = \frac{\omega_H \rho_H}{\omega_I \rho_I} \) is the \textit{limiter performance factor}. One might at first presume, on the basis of Davenport's result, that \( \sigma^2 \) would be degraded by a factor of \( \pi/4 \) (about 1 dB) at very low predetection SNR's. But limiting tends to spread bandwidth, making \( \omega_I > \omega_H \). This effect tends to compensate somewhat for the \( \pi/4 \) expected loss, and measurements indicate that, in truth, this is the case.

As the predetection SNR becomes large, a different behavior results. There is a factor-of-2 improvement in \( \sigma^2 \) due merely to the Davenport phenomenon. But, in this case, it is not difficult to show that there is no further improvement, because the limiter's output noise spectrum has the same shape as that at its point. This makes \( \omega_I \) asymptotically equal to \( \omega_H \).

Going back to Davenport's original analysis, Springett has computed the factor that we would have denoted here by \( (\omega_I \rho_I/\omega_H \rho_H) \). His result\(^4\), shown in Fig. 8-3, is well approximated by

\[
\frac{\omega_I \rho_I}{\omega_H \rho_H} \approx \frac{0.862 + 0.690 \rho_H}{1 + 0.345 \rho_H}. \quad (8-18)
\]

\(^4\)Springett's result is based on the assumption that the predetection noise has an ideal band-pass characteristic. The results for an arbitrary unimodal noise density, however, to engineering accuracy, are the same.

\[\text{Fig. 8-3. The ratio of band-pass limiter output signal-to-noise spectral density to that at input. The reciprocal of this curve gives the limiter performance factor } \Gamma. \]

The expression in (8-18) is the reciprocal of the limiter performance factor, \( \Gamma \).

The final resulting value of \( \sigma^2 \) for a loop containing an IF limiter is therefore given by

\[
\sigma^2 = \frac{N_{awL}}{A^2} \left[ \frac{1 + 0.345 \left( \frac{A^2}{N_{awH}} \right)}{0.862 + 0.690 \left( \frac{A^2}{N_{awH}} \right)} \right].
\]

\[
= \frac{N_{awL}}{A^2}. \quad (8-19)
\]

This latter approximation is valid within \( \% \) db over the usual operating range of signal levels \( (\rho_H \leq 1) \). More accurate analyses should, of course, proceed under the more exact expression.

**REFERENCES FOR CHAPTER 8**


CHAPTER 9
BEHAVIOR OF PHASE-LOCKED LOOPS (NONLINEAR ANALYSIS)

Most of the important characteristics of phase-locked devices can be predicted by the linear analysis presented in earlier chapters. But since the linear representation of a loop is predicated on high SNR's in the loop passband, one can expect that, as this assumption begins to fail, significant accuracy is lost. At this point, it becomes apparent that another method of calculating loop performance is needed.

There have been many who have tried to perform a rigorous, exact analysis of the noisy loop, but most of these have failed. Develet modified the linear model to a "quasi-linear" model, and Van Trees was able to approximate the loop character by a Volterra-series expansion. Both of these techniques are capable of extending loop analysis past the range of linear theory, but they ultimately fail when the loop gets sufficiently noisy.

Tikhonov and Viterbi independently were able to solve for the exact steady-state behavior of the first-order loop by Fokker–Planck techniques. The second-order loop has been treated with some measure of success by Lindsey.

This Chapter presents still another method, due to the author and applicable only to the calculation of output variances and spectra, subject only to rather general assumptions. One assumption necessary to the analysis is that the joint-probability function of the phase error can be expressed as a diagonalyzed orthogonal expansion. Such an assumption is known to be valid for Gaussian processes, sine-wave processes, and the output process of a square-law device whose input is a narrow-band Gaussian process, and, in fact, it does not seem to be too restrictive. The second assumption is that the steady-state phase-error spectrum can be simply related to the spectrum of the detected phase error.

The analysis is not limited to Gaussian noise inputs or to wide-band processes, although such assumptions generally make calculations easier. The Gaussian assumption, in fact, yields results that agree excellently well with the known exact theoretical behavior of the first-order loop over the entire useful range of signal and noise values.

The advantages of this method are that it is simple in concept, and easy to perform and, further, that its accuracy improves as the loop-degree increases and as the loop filter becomes narrower-band.

9-A. The Spectral Equation

Since we are interested in the behavior of noisy loops, the small contribution of oscillator noise can be omitted from consideration. We focus our attention on the steady-state stationary equivalent phase process φ, discussed in Section 4-B, and the process ψ. A slight rearrangement of the exact loop equation (4-4) gives

\[
\frac{P}{F(s)} \phi + AK \sin \phi = \frac{P}{F(s)} \phi - K n(t). \tag{9-1}
\]

When we replace φ by its stationary equivalent φ, multiply (9-1) evaluated at \(t = t_1\), by its evaluation at \(t = t_2\), and average the resulting product, there results an equation involving various correlation functions in φ, sin φ, θ, n(t), etc., whose double-sided Laplace transform gives the analytic continuation of the power spectral density:

\[
\overline{\frac{s^2}{F(s)F(-s)} S_{\phi\phi}(s)} + AK \left[ \frac{s}{F(s)} S_{\phi, \sin \phi}(s) - \frac{s}{F(-s)} S_{\phi, \sin \phi}(-s) \right] \tag{9-2}
\]

\[
+ A^2K^2 S_{\sin \phi, \sin \phi}(s) = \frac{s^2}{F(s)F(-s)} S_{\sin \sin}(s) + K^2 S_{\sin \sin}(s). \tag{9-3}
\]

A result of Bussgang extended by Barrett and Lampard to a wide class of stochastic processes reveals

\[
R_{\sin \sin}(\tau) = \eta R_{\sin}(\tau) \tag{9-4}
\]

where \(\eta\) is a functional of \(p(\phi)\), given by

\[
\eta = \frac{1}{a^2} \int_{-\infty}^{\infty} (\phi - E\phi) \sin \phi p(\phi) d\phi. \tag{9-5}
\]

We have used \(a^2\) as the variance of \(\phi\).

The condition on \(\phi\) needed to justify (9-3) is that its two-dimensional frequency function be expressible in the form

\[
p[\phi(t_1), \phi(t_2)] = p[\phi(t_1)] p[\phi(t_2)] \sum_{n=0}^{\infty} a_n f_n(\phi(t_1)) f_n(\phi(t_2)) \tag{9-6}
\]

in terms of polynomials \(f_n(\phi)\) orthonormal with respect to the kernel \(p(\phi)\). When such an expansion is valid,

\[
S_{\phi, \sin \phi}(s) = S_{\phi, \sin \phi}(-s) = \eta S_{\sin \sin}(s). \tag{9-7}
\]
Even if (9-5) is not valid, the only modification needed in the theory is that \( \gamma \) in (9-6) must be replaced by \( \gamma(s) \).

Next, when \( \Phi \) is passed through the nonlinearity, \( \sin \phi \), there is a distortion of the zonal spectrum from \( S_{\Phi \Phi} \) into \( S_{\sin \phi \sin \phi}(s) \). Hence we define a function \( \gamma(s) \gamma(-s) \), which relates the way this occurs:

\[
\gamma(s)\gamma(-s) = \frac{S_{\sin \phi \sin \phi}(s)}{S_{\Phi \Phi}(s)}.
\]  

(9-7)

Substituting these latter relations into the spectral equation, we obtain an expression for the spectrum of \( \Phi \), namely

\[
S_{\Phi \Phi}(s) = \frac{-s^2S_{\Phi \Phi}(s) + K^2 F(s) F(-s) S_{\sin \phi \sin \phi}(s)}{-s^2 + \eta AK [s F(-s) - s F(s)] + A^2 K^2 F(s) F(-s) \gamma(s) \gamma(-s)}
\]  

(9-8)

from which the variance of \( \Phi \) can be computed:

\[
a^2 = \frac{1}{2\pi i} \int_{-\infty}^{\infty} S_{\Phi \Phi}(s) \, ds.
\]  

(9-9)

Generally (9-9) will be transcendental, since both \( \gamma \) and \( \gamma(s) \) are nontrivial functions of \( a^2 \).

If \( \gamma(s) \gamma(-s) \) is reasonably well-behaved, and if \( F(s) \) is a very narrow-band device, then \(| \gamma(\omega) F(\omega) |^2 \approx \gamma(0) | F(\omega) |^2 \). Similarly, under the same condition, \( S_{\sin \phi \sin \phi}(\omega) | F(\omega) |^2 \approx N_o | F(\omega) |^2 \), where \( N_o = S_{\sin \phi}(0) \). These relations, then, are particularly applicable when \( F(s) \) is a filter with a high cutoff rate. In fact, as long as the noise \( n(t) \) is wider-band than \( F(s) \), the system performance is essentially the same as if the noise were white, and the approximate \( S_{\Phi \Phi} \) in this case is given by

\[
S_{\Phi \Phi}(s) = \frac{-s^2S_{\Phi \Phi}(s) + K^2 N_o F(s) F(-s)}{-s^2 + \eta AK [s F(-s) - s F(s)] + A^2 K^2 F(s) F(-s)}
\]  

(9-10)

This method of analysis is thus called a linear spectral approximation.

As usual, there are two terms that contribute to the phase error: one due to the input signal component, and one due to noise. The term we wish to explore more fully here is that due to noise:

\[
S_{\Phi \Phi}(s) = \frac{K^2 N_o F(s) F(-s)}{-s^2 + \eta AK [s F(-s) - s F(s)] + (AK\gamma)^2 F(s) F(-s)}.
\]  

(9-11)

Concerning the spectral approximations, one further point can be made: In the linear and quasi-linear models of the loop, \( \phi \) is approximated by a linear function of \( \phi \), so in turn, \( S_{\sin \phi \sin \phi}(s) \) and \( S_{\Phi \Phi}(s) \) are proportional to \( S_{\Phi \Phi}(s) \). Hence, the method above yields the linear results when the value for \( \gamma \) and \( \gamma(s) \) are replaced by the common value given them by either the linear or the quasi-linear model. However, while including the linear cases, the linear spectral approximation is a more general concept, because \( \gamma \) and \( \gamma(s) \) are not necessarily equal. It is thus more apt to provide results closer to the exact behavior of the loop.

\*It must be remembered, however, that \( n(t) \) is the result of multiplying \( n(t) \) by \( \cos(\omega t + \theta) \), and hence will be somewhat wider-band than \( n(t) \).
9-B. Calculation of $\eta$

The calculation of $\eta$ as the constant relating $S_{3f}$ to $S_{\text{ojo}}$ involves first knowing $p(\phi)$, and second, knowing that $p[\Phi(t_1), \Phi(t_2)]$ has the diagonal form of (9-6).

At very low noise levels, assuming a Gaussian channel, the density for $p(\phi)$ certainly approaches a Gaussian density, as does $p(\Phi)$, and all analysis methods give compatible results; at high noise levels the density on $\phi$ (mod $2\pi$), viz., $p(\phi)$, approaches a uniform density of $1/2\pi$ on the interval ($-\pi, \pi$). In what follows, we denote $\phi$ (mod $2\pi$) merely by $\phi$.

As an approximation to this behavior, for definiteness, we may assume that $\phi$ is Gaussian, with variance $\sigma^2$; such an assumption satisfies both limiting conditions on the density of $\phi$, which is then given by

$$p(\phi) = \begin{cases} \frac{1}{\sigma^2 (2\pi)^{1/2}} \exp \left[ -\left( \phi + 2m_0 \right)^2 / 2\sigma^2 \right]; & |\phi| \leq \pi \\ 0; & |\phi| > \pi \end{cases}$$

(9-12)

The variance $\sigma^2$ of the $\phi$-process is related to the variance $\sigma^2$ by straightforward evaluation. The result is graphically presented in Fig. 9-1 along with the approximate formula

$$\sigma^2 = \frac{\sigma^2}{3} \left[ 1 - \exp \left( -\frac{3\sigma^2}{\pi^2} (1 + 0.13\sigma^2) \right) \right].$$

(9-13)

For the density in (9-12), the value of $\eta$ can be readily computed:

$$\eta = e^{-\sigma^2/2}.$$ (9-14)

9-C. Calculation of $\gamma^2$

For loops in which $F(s)$ is very narrow, $\gamma(s)\gamma(-s)$ can be replaced by its value at $s = 0$, viz., $\gamma^2 = \gamma^2(0)$. In the case that $F(s)$ has a very wide bandwidth, we can still use (9-11), but a better approximation to $\gamma(s)\gamma(-s)$, in this case, is the ratio of the total power in $\sin \phi$ to that in $\phi$, rather than $\gamma^2$, the ratio of powers around zero frequency. This total power value, $\gamma^2_{tp}$, is

$$\gamma^2_{tp} = \frac{R_{\sin \phi} \sin^2(0)}{R_{\phi}(0)}.$$ (9-15)

Since we have assumed that $\phi$ is a Gaussian process, Price's theorem can be used to compute the autocorrelation function of $\sin \phi$:

$$R_{\sin \phi}(r) = e^{-r^2} \sin R_\phi(r).$$ (9-16)

The relation above is certainly valid when $\sigma^2$ is small, because we are sure, in this case, that $\phi$ is nearly Gaussian. For larger $\sigma^2$, (9-16) is only approximate. But insofar as the Gaussian assumption is justified,

$$\gamma^2 = \frac{e^{-\sigma^2} \int_0^\infty \sin R_\phi(r) \, dr}{\int_0^\infty R_\phi(r) \, dr}.$$ (9-17)
for a loop with a narrow-band $F(s)$, and

$$\gamma_{tp}^2 = \sigma^2 e^{-\sigma^2} \sinh \sigma^2$$

(9-18)

for a loop with a wide-band $F(s)$.

The former quantity, $\gamma^2$, cannot be evaluated directly, since $R_s(t)$ is not known until $\gamma^2$ is, and vice versa. However, if loop noise is not completely overpowering, $\gamma^2$ is probably not as sensitive to the shape of $R_s(t)$ as it is to the value of $\sigma^2$. As a verification of this conjecture, the several different forms of $R_s(t)$ shown in Fig. 9-2 produce values of $\gamma$ that are in very close agreement over the range of interest. The upper curve is equivalent to $\gamma_{tp}$. The values of $\gamma$ are probably very well approximated by that in the middle:

$$\gamma = 2 \sigma^2 e^{-\sigma^2/2} \sinh \left( \frac{\sigma^2}{2} \right) = \sigma^2 (1 - e^{-\sigma^2}).$$

(9-19)

To see how well the spectral approximation method works, let us calibrate it by a few examples.

9-D. Behavior of the First-Order Loop

The spectrum of $\Phi$ for a first-order loop is approximately

$$S_{ss} = \frac{N_o K}{-\sigma^2 + (AK_{tp})}$$

(9-20)

for we merely insert $F(s) = 1$ into (9-11). We use $\gamma_{tp}$ to approximate $\gamma(s)$, because $F(s)$ is wide-band. Integration of (9-24) yields the estimate

$$\sigma^2 = \frac{N_o K}{2A_{tp}} = \frac{N_o}{A_{tp} (AK_{tp})}. \quad (9-21)$$

This equation compares in form with (5-8), except that the bandwidth of the loop has apparently changed to a new equivalent value because $A$ reduced to $A_{tp}$:

$$\omega_{(tq)} = \frac{AK_{tp}}{2}. \quad (9-22)$$

1. Linear Loop

Under the usual linear assumption, $\gamma_{tp} = 1$, so $\omega_{(tq)}$ reduces to $\omega_L$, the linear loop bandwidth, and, of course, the usual answer (5-8) results.

2. Quasi-linear Loop

When the constant gain $A$ in the linear loop is replaced by the equivalent statistical gain $ae^{-a^2/2}$ derived by Develet, i.e., when $\gamma_{tp} = e^{-a^2}$, the result is

$$\sigma^2 = \frac{N_o K}{2A} e^{-a^2/2}. \quad (9-23)$$

Notice that there is no solution for $\sigma^2$ here when $N_o K/2A \geq 1/2$. For $N_o K/2A < 1/2$, there is an equivalent bandwidth

$$\omega_{(tq)} = AK \quad e^{-a^2/2}. \quad (9-24)$$

3. Spectral Approximation

If we use the analysis presented above with $\gamma_{tp}$ as given in (9-18), the phase variance of the equivalent stationary process $\Phi$ is related to the linear-loop prediction by

$$ae^{-a^2/2} \sinh (a^2) = \frac{N_o K}{2A} \quad (9-25)$$

Here there is a solution $a^2$ for every value of $N_o K/2A$.

4. Reduction Modulo $2\pi$

The real indication of system performance, as we have agreed, is given not by the parameter $a^2$ (the variance of

![Graph](image-url)
\( \phi \), but by \( \sigma^2 \), the variance of \( \phi \pmod{2\pi} \). These are related as shown in Fig. 9-3. Each of the various performance approximations, reduced mod \( 2\pi \), is compared with the actual loop behavior (Fokker–Planck method) in Fig. 9-3; notice the closeness with which the spectral method predicts the actual behavior. We have also indicated that the spectral approximation method should yield results that improve as the filter \( F(s) \) becomes higher-order and narrower in bandwidth. Thus, the first-order system should be the one predicted with least fidelity — but, as shown in Fig. 9-3, even this proves to be very good.

The conclusion to be reached at this point, then, is that the linear spectral approximation method provides a very accurate means of predicting loop phase deviations (mod \( 2\pi \)) over the entire useful range of input signal and noise values. In addition, we have shown that the linear analysis is valid whenever \( A^2/N_0 \omega_L > 10 \).

9-E. Calculation of the Behavior of the Second-Order Loop

We now show the performance of a phase-locked loop whose loop filter is the passive integrator of (5-13). This filter function inserted into the spectral equation (9-11) produces the density

\[
S_{\phi}(x) = \frac{K^2 N_0 (1 - r^2 s^2)}{r^2 s^2 - s^4 [1 + 2 AK_l (r_2 - r_1) + (AK_l r_2)^2] + (AK_l)^2}.
\]

(9-26)

The density (9-26) is quite different from what one obtains from a simple linear model of the loop. Still, we can use the same types of ideas to express the output phase variance. For one thing, \( S_{\phi}(\omega) \) has a fiducial bandwidth

\[
\omega_L(\omega) = \frac{1}{2\pi S_{\phi}(0)} \int_{-\infty}^{\infty} S_{\phi}(\omega) d\omega.
\]

(9-27)

Note that this is computed in the same way as \( \omega_L \), except that we must use \( S_{\phi}(\omega) \) as given in (9-28) rather than \( |L(\omega)|^2 \), as one would do if the linear model were in effect. In these terms

\[
a^2 = \frac{N_0 \omega_L(\omega)}{(A \gamma)^2}.
\]

(9-28)

Again, it is evident that there are two effects in the second-order loop that deviate \( a^2 \) from the value it would take in a linear analysis: \( \omega_L(\omega) \) must be used rather than \( \omega_L \), and \( A \) is reduced to \( A\gamma \).

The value of \( \omega_L(\omega) \) can be found by integrating (9-27):

\[
\omega_L(\omega) = \frac{(1 + r_2 \gamma)}{2 r_2 \gamma} \left[ 1 + \frac{2(\gamma - 1)(1 - r_2 / r_2 \gamma)}{r_2 \gamma (1 + r_2 / r_2 \gamma)^2} \right]^{1/2}
\]

\[
\approx \frac{1 + r_2 \gamma}{2 r_2 \gamma} \left[ 1 + \frac{2(\gamma - 1)}{r_2 \gamma^2} \right]^{1/2}.
\]

(9-29)

The approximation follows the natural assumption \( r_2 \gamma \gg r_2 \).

A "damping factor" can be defined almost in the same way as it is for the linear loops, a way that reduces to the linear case for high input signal-to-noise ratios. We have already derived \( S_{\phi}^0 \); it is of the form

\[
S_{\phi}(\omega) = \frac{1 - a^2}{s^2 - b s^2 + c}.
\]

The roots of the denominator govern the "damping behavior" of \( \Phi(t) \). If these roots are \( r_1, r_2, -r_2, -r_1 \), then
we can define, for the left-half-plane factor \( s^2 + (r_1 + r_2) s + r_1 r_2 \), an equivalent damping coefficient

\[
\zeta_{eq} = \frac{(r_1 + r_2)}{2 (r_1 r_2)}.
\] (9-31)

We relate \( r_1 r_2 \) and \( (r_1 + r_2) \) to the parameters of \( S_{ee}(s) \) by noting

\[
(r_1 r_2)^2 = \left( \frac{A K_Y}{\tau_1} \right)^2
\] (9-32)

\[
(r_1 + r_2)^2 - 2r_1 r_2 = \tau_1^2 \left[ 1 + 2AK_Y(r_1 - \tau_1) + (AK_Y r_2)^2 \right].
\]

Hence, the equivalent output damping factor of the loop is

\[
\zeta_{eq} = \frac{(r r)^{16}}{2} \left[ (1 + \tau_2/r_{eq})^2 + \frac{2(y - \gamma)(1 - \tau_2/\tau_1)}{r_{eq}^2} \right]^{16}
\]

\[
\approx \frac{(r r)^{16}}{2} \left[ 1 + \frac{2(y - \gamma)}{r_{eq}^2} \right]^{16}.
\] (9-33)

1. Linear Analysis

It is easy to see when \( \eta = \gamma = 1 \) (the linear analysis parameters) that (9-28), (9-29), and (9-33) reduce to (5-8), (5-18), and (5-20), respectively.

2. Quasi-linear Analysis

The quasi-linear analysis replaces \( A \) everywhere by \( A e^{-2\eta^2} \) (i.e., \( \gamma = \eta = e^{-\eta^2/2} \)). Thus the behavior predicted in this case is given as

\[
a^2 e^{-a^2} = \frac{N_s w_{L(eq)}}{A^2}
\]

\[
w_{L(eq)} \approx \frac{1 + r e^{-\eta^2/2}}{2\tau_2}
\] (9-34)

\[
\zeta_{eq} \approx \frac{r_{eq}^4}{2} e^{-a^4/4}.
\]

From this it is important to note that while bandwidths and damping are less than the value predicted by the linear model, the overall value of \( a^2 \) is larger than its linearly predicted value.

3. Linear Spectral Analysis

In a still more general analysis, we set \( \eta = e^{-\eta^2/2} \) and \( \gamma = e^{-\gamma^2} \) \((1 - e^{-\gamma^2})\). The loop behavior is then approximated very closely by

\[
\frac{N_s w_{L}}{A^2} = \frac{(1 - e^{-\eta^2})^{16} w_L}{w_{L(eq)}}
\]

\[
\frac{w_{L(eq)}}{w_L} \approx \frac{1 + r (1 - e^{-\eta^2})/a^2}{1 + \frac{2a^2 (1 - e^{-\eta^2 - a^2 e^{-\gamma^2/2}})}{r(1 - e^{-\gamma^2})}}^{16}
\]

\[
\frac{\zeta_{eq}}{\zeta} \approx \frac{(1 - e^{-\eta^2})^{16}}{a} \left[ 1 + \frac{2a^2 (1 - e^{-\eta^2} - a^2 e^{-\gamma^2/2})}{r(1 - e^{-\gamma^2})} \right]^{16}.
\] (9-35)

Again, the same type of behavior results: \( w_{L(eq)} \) and \( \zeta_{eq} \) are less than their linearly predicted value, while \( \sigma^2 \) is greater. Although the quantities in (9-35) are generally transcendental in \( \sigma^2 \), nevertheless Newton’s rule is easily applied by a digital computer for solutions. Figure 9-4 relates various parameters above to corresponding values of \( \sigma^2 \), based on a constant value of \( r \) (viz., \( r = 2 \)). Figure 9-5 then shows how \( \sigma^2 \) compares with the value pre-
dicted by a linear analysis as a function of signal power for a fixed $N_0$; since $r$ varies with input signal level, we have arbitrarily chosen a value $r_o = 2$ at $A_o^2 = 3N_o/2\tau_2$.

Figure 9-6 illustrates the way $\omega_L_{(eq)}$, $\zeta_{(eq)}$, and $\zeta$ vary; parameters here are the same as those in Fig. 9-5.

4. Conclusions About the Second-Order Loop

The behavior depicted in Figs. 9-4, 9-5, and 9-6 indicates that a linear theory can be expected to yield satisfactory accuracy whenever $A^2/N_0\omega_h > 10$. Beyond this point, the linear spectral approximation probably agrees with actual results, if we may extrapolate the results obtained for the first-order TCP, for $\sigma^2 < 1.5 \text{ rad}^2$. This figure lies beyond the useful range of most receivers.

---

Fig. 9-5. Comparison of phase-noise variances by linear and linear-spectral approximations. The noise density is fixed, and the signal level is varied. The value of $r$ is taken as $r_o = 2$ at a reference signal level of $A_o^2 = 3N_o/2\tau_2 = N_0\omega_L$. Note that the ultimate roll-off is 5db/decade, rather than 10 db/decade, as in Fig. 9-4. Note also that even the linear approximation produces some curvature of $\sigma^2$ near $A_o^2$.

Fig. 9-6. Variation in bandwidth and damping parameters as a function of signal strength. The value of $r$ at a reference signal level $A_o^2 = 3N_o/2\tau_2 = N_0\omega_L$ was taken as $r_o = 2$. 
REFERENCES FOR CHAPTER 9


CHAPTER 10
DESIGNING A DOUBLE-HETERODYNE TRACKING LOOP

So far in this work, we have concerned ourselves, for the most part, in predicting the performance of a receiver when its parameters were specified. Exceptions to this dealt only with ways of choosing some of the receiver parameters, given the others. We may now pick a set of parameters, and, through the formulas that have been presented, predict quite accurately how the loop will operate. Once the receiver is built, it performs just so.

What we require for design, however, is an effective method for picking a nominal set of values that produces a desirable tracking function over a reasonable variation of the input signal power.

As long as the operation of the loop is to remain within the linear-analysis region, the task is much simplified, because there have been analyses put forth for extracting an optimum result from the linear loop. Nonlinear optimization in most cases is difficult, if not analytically impossible.

It is customary to optimize, insofar as possible, receiver performance at the "worst-case" parameter values; in this way one is sure that, while perhaps not operating optimally at any other set of values, the loop will do the best that it possibly can in those cases that require it most. This procedure is purely an arbitrary one, in that it is apt to change in accordance with the philosophy of the design engineer, and with the particular mission for which the receiver is intended.

What design rules are given here will be somewhat abstract, but perfectly general, so that the designer may issue his own philosophy in choosing values.

10-A. Definition of Receiver Threshold

The words "receiver threshold" conjure up a different image to each engineer: to one, it is that shaky, ill-defined signal level at which the receiver transits from operability to nonoperability; to another, it is some point at which the receiver operates with the least acceptable reliability; to still another, it may mean the signal level at which the receiver exits from its linear range. There are no absolutes when it comes to defining such a point; in all of the cases above, the engineer meant that the system hit "threshold" when its performance was in some way no longer tolerable, when the receiver no longer produced meaningful data. The trouble with such definitions tied to performance is that they tend to be subjective — what is tolerable in one case is not in another.

The word "threshold," as we shall use it here, is a precisely defined quantity, which can be subjectively interpreted as desired. Some give it the name "design-point threshold" and, in some cases, it has been called the "absolute receiver threshold." The concatenation of modifiers does not seem to be necessary at all, and we shall use the word "threshold" in only one sense.

Specifically, the loop threshold is defined as that input signal power $A_0^2$ given by

$$A_0^2 = \frac{N_0(r_0 + 1)}{2r_2 \left(1 + \frac{r_2}{r_{0r_1}}\right)}$$

$$= N_0 \omega_{L_0}, \quad (10-1)$$

Here, $\omega_{L_0}$ is the value of linear-loop fiducial bandwidth at threshold, i.e., computed by (5-18) using the threshold value of $r, r_0$. One must not be mistakenly led to believe that the loop is operating linearly; $\omega_{L_0}$ is merely the bandwidth a loop would have if $A_0$ were to lie in a linear-analysis region. In actuality, $A_0$ is a signal level at which the linear theory does not apply, for, in fact, $\sigma^2 > 1$.

In the notation that is to follow, all quantities subscripted by "0" refer to the value of that particular parameter at threshold.

10-B. Tracking Loop Performance of the Double-Heterodyne Receiver

Once the values of the receiver's physical parameters $K_0K_{tco}MF, r_1, r_2,$ and $W_H$ have been determined, performance proceeds as specified by the formulas\(^7\) in Chapter 8. But with a fixed receiver, the loop bandwidth changes as a function of signal or noise level. Rather than

\(^7\)Although $W_H$ is a measured physical quantity, it may not be equal to $\omega$ if the loop is not operating at the frequency producing $H(\omega)$ max. Care must often be taken to assure that $H(\omega)$ is properly tuned for maximum response at the loop operating frequency.
calculate each receiver characteristic separately, it is easier to evaluate the performance in terms of a reference signal level and to relate the remainder of the loop behavior to this point. One convenient reference point is the threshold signal level. Specifically, we define the ratio
\[ m = \frac{A_i}{A_{r_i}} \]  
(10-2)
as the loop performance margin. In terms of this \( m \),
\[ \rho_u = \frac{A_i^2}{N_i w_u} = \left( \frac{A_i}{A_o} \right) \left( \frac{w_{v_o}}{w_u} \right) = m \rho_{u_0}. \]  
(10-3)
As a result, all of the \( w \) p characteristics are expressible in terms of \( m, w_{v_o}, w_u, \) and \( \tau_2/\tau_1 \) (and this latter quantity is not needed if \( \tau_2 \gg \tau_1 \)). Thus a design consists really of specifying these three (or four) quantities.

However, it is often necessary to compute the performance of a receiver when \( w_{u}, K_{VCO}MF, \tau_2, \) and \( \tau_1 \) are given — quantities that can be measured. Most of the formulas we have given are in terms of \( r \), \( w_o \), etc. How does one proceed with the set of parameters above? We merely need to find \( r_o \), and the rest then follow. When \( 10w_{v_o} < w_u, a_o^2 \) is very well approximated by \( \pi w_{v_o}/4w_u \); from this, a slight manipulation of (8-14) produces the desired result:

\[ U = \frac{\pi(K_{VCO}MF)^{r_2}}{8r_1^{2}w_u} \]
\[ r_o = \left( \frac{U - 4r_2}{2} \right) \{ 1 + \left[ 1 + \frac{4U}{(U - 4r_2)^2} \right] \} \]
\[ \approx \frac{U}{2} \left[ 1 + \left( \frac{4}{U} \right) \right]^{\frac{1}{2}} \]  
(10-4)
We can also express \( U \) in terms of \( r_o \),
\[ U = \frac{r_o(r_2 + \tau_2/\tau_1)}{r_o + 1} \approx \frac{r_o^2}{r_o + 1}. \]  
(10-5)

**10-C. Nonlinear Behavior of the Double-Heterodyne Receiver**

The formulas in preceding chapters can be made to apply to loops with predetection bandpass limiters by making a few minor alterations, and we have seen how these changes come about in Chapter 8: first, \( AK \) is replaced by \( aK_{VCO}MF \), as in (8-8); and second, the expression (9-28) must be multiplied by the limiter performance factor \( \Gamma = w_u\rho_u/w_i\rho_i \), shown in Fig. 8-3.

These steps lead to a set of equations which characterize the double-heterodyne receiver:

\[ \rho_u = m \left( \frac{w_{v_o}}{w_u} \right) = m \rho_{u_0} \]  
(10-6a)
\[ \alpha = \left[ \frac{0.7854\rho_u + 0.4768\rho_u^2}{1 + 1.024\rho_u + 0.4768\rho_u^2} \right] \]  
(10-6b)
\[ \Gamma = \frac{1 + 0.345\rho_u}{0.862 + 0.690\rho_u} \]  
(10-6c)
\[ r = \frac{\alpha K_{VCO}MF(0)}{\tau_2} \approx \frac{\alpha}{\tau_2} \beta^2 = \tau_2 \left( \frac{\alpha}{\tau_2} \right) \beta^2 \]  
(10-6d)
\[ \sigma^2 = \frac{\pi^4}{3} \left( 1 - \exp \left[ \frac{-3a^2}{\pi^2} (1 + 0.13a^2) \right] \right) \]
\[ \approx a^2 \]  
(10-6e)
\[ \gamma = (1 - e^{-a^2})/a^2 \]  
(10-6f)
\[ \eta = e^{-a^2/2} \]  
(10-6g)
\[ w_u = \frac{1 + r}{2r_2(1 + \frac{r_2}{r_1})} \approx \frac{1 + r}{2r_2} = \frac{1 + \left( \frac{a}{\tau_2} \right) r_o}{1 + r_0} \]  
(10-6h)
\[ w_{u(eq)} = \frac{1 + ry}{2r_2 \left( 1 + \frac{r_2}{\eta \tau_1} \right) \left[ 1 + \frac{2(\gamma - \eta)(1 - r_2/\tau_1)}{\gamma^2 \left( 1 + \frac{r_2}{\eta \tau_1} \right)^2} \right] \]  
(10-6i)
\[ \zeta = \frac{1}{2} \left( 1 + \frac{r_2}{\eta \tau_1} \right) r_2 = \frac{1}{2} \left( 1 + \frac{1}{\tau_2 \beta^2} \right) r_2 \beta \]
\[ \approx \frac{r_2 \beta}{2} \]  
(10-6j)
\[ \zeta_{eq} = \frac{1}{2} \left( 1 + \frac{r_2}{\eta \tau_1} \right) \left( \gamma \left[ 1 + \frac{2(\gamma - \eta)(1 - r_2/\tau_1)}{\gamma^2 \left( 1 + \frac{r_2}{\eta \tau_1} \right)^2} \right] \right) \]
\[ \approx \frac{1}{2} (\gamma)^{\alpha} \approx \zeta \gamma^{\alpha} \]  
(10-6k)
\[
\beta_{eq} = \frac{(r_0)_{18}}{r_0} = \beta \gamma_{18}
\]  
(10-6)

\[
a^2 = \frac{N_1 \omega_{(eq)} \Gamma}{A^2 \gamma^2} = \frac{1}{m} \left( \frac{\omega_{(eq)} \Gamma}{\omega_{no} \gamma} \right)
\]  
(10-6m)

\[
\beta \approx \frac{2 \omega_{o}}{r + 1} = \left( \frac{\alpha}{\alpha_o} \right)^{\frac{2 \omega_{o}}{r_o}}
\]  
(10-6n)

The use of \( \gamma \) as given in (10.6f) requires \( \tau_1 \gg \tau_1; \) otherwise, \( \gamma_{1p} \) of (9.18) should be substituted in its place. Notice that the only quantities needed to specify everything in (10.6) are the margin \( m \), the predetection SNR \( \rho_{\eta_0} = \omega_{\eta_0}/\omega_{\eta}, r_0 \), and \( \tau_0/\tau_1 \).

The performance observed here is very much the same as that exhibited in loops without limiters, except that the value of \( K, K_{\text{vco}} MF \) required is drastically less than the value of \( K \) required when there is no limiter in the loop. This, of course, is due to the fact that while \( \Delta_1 \) is very small (the threshold rms signal level), \( \alpha_o \) is many orders of magnitude greater, approximately equal to the predetection SNR.

10-D. The Signal Level Producing \( a^2 = 1 \)

Since it is doubtful that one would ever expect a loop to operate usefully with a value of \( a^2 \) greater than unity, the specific value of \( a^2 \) is more a point of academic interest than anything else. Furthermore, if we may judge from the first-order loop result, the threshold value of phase noise predicted by the linear spectral method would probably not be a very accurate one, for the theory begins to fail somewhere near this point.

Holding the rest of the loop parameters constant, we can solve for the value of margin (call it \( m_1 \)) that produces \( a^2 = a^2_1 = 1 \). This is a more meaningful quantity, and certainly much more accurately calculable by the methods in Chapter 9. It follows from (10.6) that when \( a^2 = 1 \), the value of \( \gamma_1/\gamma_0 = m_1^{18} \) satisfies the equation

\[
A_1 = \frac{\Gamma_1 (r_0 + \tau_1/\tau_1) (A_1/A_0 + 1/r_0 \gamma_1)}{\gamma_1 (r_0 + 1) (A_1/A_0 + \tau_1/\tau_1 \gamma_1/\gamma_0)} \left[ 1 + \frac{2(\gamma_1 - \gamma_0) (1 - \gamma_1/\gamma_0) (A_1/A_0)}{\gamma_1^2 r_0 (A_1/A_0 + \tau_1/\tau_1 \gamma_1/\gamma_0)^2} \right]^{18}
\]

\[
r_1/a_0 = a_0 A_1 A_0
\]

(10-7)

The subscript "1" on \( \Gamma_1, \gamma_1, \eta_1, \) etc., refers to evaluation at \( a^2 = 1 \). Under the usual assumption \( \gamma_1 r_0 \tau_1 \gg \tau_0 \), the approximate value for \( m_1 \) is

\[
m_1 = \frac{r_0 \Gamma_1}{2 \gamma_1 (\tau_0 + 1)} \left[ 1 + \frac{4(r_0 + 1)}{\Gamma_1} \frac{1}{r_0^2} \right]^{18}.
\]

(10-8)

The value of \( a^2 \) is within 0.5% of unity when \( a^2 = 1 \); hence to the accuracy we need, the values of \( m_1 \) and \( r_1 \) above give \( a^2 = 1 \). The curves in Fig. 10-1 show how \( m_1 \) varies as a function of \( r_0 \) and \( \tau_1/\tau_1 \). As \( \tau_1/\tau_1 \) nears unity, it is necessary to replace \( \gamma_1 \) by \( \gamma_{1p} \) (see 9.18). The values given in the Figure result by assuming that \( \gamma_1 \) in (10.3) takes the value

\[
\gamma_1 = \left( 1 - \frac{\tau_0}{\tau_1} \right) (1 - e^{-1}) + \left( \frac{\tau_0}{\tau_1} \right) e^{-18}
\]

\[
= 0.6321 \left( 1 - \frac{\tau_0}{\tau_1} \right) + 0.60653 \left( \frac{\tau_0}{\tau_1} \right) .
\]

(10-9)

Fig. 10-1. Variation in margin producing \( a^2 = 1 \) as a function of threshold design parameter,

\[
r_0 = \alpha_0 K_{\text{vco}} MF \tau_1 r_1
\]
10-E. Choice of Receiver Parameters

Choosing values of \( N_n, \omega_{v0}, r_0, K_dK_{vco}MF \), and \( w_H \) specifies a typical receiver design. Values for \( r_1, r_2, \rho_H, \xi_0 \), and \( \Gamma_0 \) follow directly from (10-1) and the first four parts of (10-6). The value of \( \frac{r_0}{r_1} \) is a measure of how well the imperfect integrating loop filter performs in the loop; with \( \frac{r_2}{r_1} = 1 \), for example, the loop filter is merely \( F(s) = 1 \), and all the results in (10-6) reduce to the first-order loop equations. On the other hand, the results in (10-6) are approximately the same for all values of \( \frac{r_0}{r_1} \) less than about \( r_0/10 \). There is some flexibility in choosing loop gain parameters because \( K_d, K_{vco}, M \), and \( F \) appear as a single factor in the theory.

However, the only thing an engineer needs to know in order to know how well any receiver will perform are \( m, r_0, \rho_H, \) and perhaps \( \frac{r_2}{r_1} \). This latter set of quantitiries depends on several things: \( r_0 \) sets the loop damping factor, to be chosen in accordance with the discussion in Section 6-C4; the value of \( \rho_H \) depends on \( \omega_{v0} \) and \( \omega_H \); \( \omega_{v0} \) is determined by considering the desired acquisition ease, the doppler tracking capability, the state of the art in VCO design, the spectral purity of the received carrier, the expected incoming \( A^2/N_0 \), and the amount of phase jitter that can be tolerated at that operating level. In addition, \( r_1 \) should be chosen (Section 6-C3) so that the phase error due to the expected doppler rate is increasing slowly enough that only infrequent retunings of the VCO are required.

Once \( r_0 \) and \( \rho_H \) (and perhaps \( \frac{r_2}{r_1} \)) are given, the receiver performance in terms of margin can be found from (10-6); a typical plot of \( \sigma \) vs \( m \) is given in Fig. 10-2, wherein linear approximations are compared with actual behavior; Fig. 10-3 shows how bandwidth and damping factor change as a function of margin. With assumed values of \( N_n \) and \( \omega_{v0} \), threshold is specified (10-1), so the curves are easily referenced to the actual incoming signal power.

![Fig. 10-2. Comparison of linear and nonlinear approximations to loop rms phase error, as a function of loop margin](image1)

![Fig. 10-3. Variation of loop bandwidths and damping factors, as a function of loop margin](image2)
PART II
SUMMARY OF PHASE-LOCKED LOOP DESIGN CONCEPTS

In this second part, the important definitions, concepts, and formulas are collected chapter by chapter. The notation used in these formulas is listed in the Appendix with names, units, and text references. Many of the formulas here are only approximations of more accurate ones in Part I.

Since not all of the subdivisions in Part I are referred to in Part II, there are some discontinuities both in the headings and in the figure and equation numbers of Part II.
CHAPTER 1
INTRODUCTION AND HISTORY OF THE PHASE-LOCKED LOOP

The first serious application of the phase-lock principle was as a synchronizing device for television in the late 1940's. Since then it has evolved as the heart of the most sensitive, versatile, and flexible receiver in existence today. There are many approximate analyses for predicting the behavior of the loop (Fig. II-1-1); some of these are set forth in the ensuing chapters.

![Diagram of a simple phase-locked loop](image)

**Fig. II-1-1.** Basic configuration of a simple phase-locked loop. The mixer output, filtered by $F(s)$, is used to control the frequency of the voltage-controlled oscillator (VCO).
CHAPTER 2
FUNDAMENTAL CONCEPTS

2-A. Statistics

Process: a time function, \( x(t) \).

Deterministic process: a process \( x(t) \) for which \( x(t) \) is specified at every value of \( t \).

Stochastic (or random) process: a set, or ensemble, of time functions \( \{x(t)\} \), any particular member of which is observed according to some probability law.

Stationary process: a stochastic process whose statistical behavior does not depend on its time origin.

Ergodic process: a stochastic process wherein time averages produce the same results as statistical averages.

Time-autocorrelation of the function \( x(t) \): given a function \( x(t) \)

\[
\mathbb{R}(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x(t)x(t+\tau) \, dt. \tag{2-3}
\]

Statistical autocorrelation of a stochastic process \( \{x(t)\} \):

\[
\mathbb{R}(t_1, t_2) = E[x(t_1)x(t_2)]
= \mathbb{R}(t_1 - t_2) \quad \text{(if } x(t) \text{ is stationary)} \tag{2-5}
= \mathbb{R}(t_1, t_2) \quad \text{(if } x(t) \text{ is ergodic).} \tag{2-6}
\]

Cross-correlation \( \mathbb{R}(t_1-t_2) \): obtained by replacing \( x(t_1) \) by \( y(t_1) \) above.

Spectral density of a function \( x(t) \):

\[
S_x(\omega) = \int_{-\infty}^{\infty} \mathbb{R}(\tau) e^{-i\omega \tau} \, d\tau. \tag{2-7}
\]

Spectral density of a stationary process \( \{x(t)\} \):

\[
S_x(\omega) = \int_{-\infty}^{\infty} \mathbb{R}(\tau) e^{-i\omega \tau} \, d\tau. \tag{2-8}
\]

Cross-spectral density \( S_{xy}(\omega) \): similarly defined as the Fourier transform of \( R_{xy}(\tau) \).

Mean value of \( x(t) \):

\[
\mu(t) = E[x(t)]
= \mu \quad \text{(if } x(t) \text{ is stationary).} \tag{2-10}
\]

Variance of \( x(t) \):

\[
\sigma_x^2(t) = E[x(t) - \mu(t)]^2
= E[x^2(t)] - \mu^2(t). \tag{2-11}
\]

2-B. Linear Filtering

Unit impulse response of filter \( H \) (Fig. II-2-1): The output \( h(t) \) of \( H \) when the input is \( \delta(t) \), the unit-weight Dirac impulse function (see Fig. II-2-2).

Linear filter: any filter whose input \( x(t) \) and output \( y(t) \) are related by

\[
y(t) = \int_{-\infty}^{\infty} x(t_1)h(t - t_1) \, dt_1. \tag{2-14}
\]

![Fig. II-2-1. Filtering device](image)

![Fig. II-2-2. Response of filter to a unit-impulse function](image)
When \( x(t), y(t), \) and \( h(t) \) have Laplace transforms \( X(s), Y(s), \) and \( H(s) \), respectively, linearity reduces to

\[
Y(s) = X(s) H(s).
\]

When \( x(t) \) is a stationary process,

\[
S_{xy}(f) = |H(f)|^2 S_{xx}(f).
\]

2-C. Noise Bandwidth

White noise: a random process \( n(t) \) such that

\[
R_{nn}(\tau) = N_o \delta(\tau) \quad \text{(volts)}
\]

\[
S_{nn}(f) = N_o \quad \text{for all } -\infty < \omega < +\infty \quad \text{(volts/cps)}.
\]

Effective noise bandwidth of a linear filter \( H(s) \):

\[
W_H = 2B_H = \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(f)|^2 df \quad \text{(cps)}.
\]

Note that \( W_H \) is the two-sided bandwidth and \( B_H \) is the single-sided bandwidth of \( H(s) \) in cps (see Figs. 11-2-3, 11-2-4, and 11-2-5). The output noise power with a white noise input is

\[
N = N_o W_H |H(f)|^2_{\text{max}} \quad \text{(volts)}.
\]

2-D. Sinusoidal Filter Inputs

If \( 2\pi P \sin(\omega_0 t + \theta) + n(t) \) is put into \( H(s) \), where \( n(t) \) is white with spectral density \( N_o \), the output SNR of \( H(s) \) is

\[
\rho_H = \frac{P}{N_o W_H} \cdot \frac{|H(f)|^2_{\text{max}}}{|H(f)|^2_{\text{max}}} = \frac{P}{N_o \omega_H}.
\]

2-E. Fiducial Bandwidth

Fiducial bandwidth: equivalent (2-sided) bandwidth referred to a specific frequency \( \omega_0 \), such as the carrier above, rather than that \( \omega \) producing \( |H(f)|^2_{\text{max}} \),

\[
\omega_H = 2b_H : \omega_H \frac{|H(f)|^2_{\text{max}}}{|H(f)|^2_{\text{max}}} = \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(f)|^2 df \quad \text{(cps)}.
\]

It is this fiducial bandwidth that is needed in most of phase-locked loop theory, rather than the equivalent noise bandwidth \( W_H \).

2-F. Band-Pass Mixers

Heterodyning (Fig. 11-2-6) does not affect signal-to-noise ratios so long as the heterodyning frequency and the carrier frequency are independent. However, when the two are the same, the output signal-to-noise ratio \( \rho \) is

\[
\rho = \frac{P}{N_o w_D} \cos^2 \phi
\]
Fig. II-2-6. The simple product-mixer

where \( P \) = input signal power (watts), \( N_0 = 2 \)-sided input noise density (watts/cps), \( \omega_p \) is the fiducial bandwidth of the detector (cps), and \( \phi \) is the phase angle between the carrier and the detecting sinusoid. The term \( E(\cos^2 \phi) \) is the coherence factor of the detector. Image noise must be eliminated by post-heterodyne filtering (Fig. II-2-7).

2-H. Noise

Johnson (or thermal) noise: noise due to thermal agitation of electrons in a resistor. There is a noise voltage \( v_n(t) \) in series (see Fig. II-2-8) with each such resistance \( R \), with

\[
S_{v_n} (j\omega) = 2kTR \quad \text{(volts}^2/\text{cps}) \quad (2-62)
\]

where \( k \) is Boltzmann's constant and \( T \) is the resistor temperature in degrees Kelvin. This is Nyquist's Law.

Maximum noise power obtainable from such a source in a bandwidth \( W_x \) is

\[
N = \frac{kT}{2} W_x = \frac{1}{2} kTR_e \quad \text{(watts).} \quad (2-63)
\]

(Compare this with (2-25), for a filter with bandwidth \( W_x \) and gain \( |H(j\omega)|^2_{\text{max}} \).

Hence for a white thermal noise the matched-load noise voltage appearing across the load resistance \( R \) has uniform spectral density \( N_0 = kT R \) volts\(^2/\text{cps}.

Fig. II-2-7. Choosing filter bandwidths to avoid the image noise problem

Fig. II-2-8. Thevenin and Norton equivalent circuits of noisy resistors
CHAPTER 3
FORMULATION OF THE LOOP EQUATION AND BEHAVIOR IN THE ABSENCE OF NOISE

3-A. The Basic Integro-Differential Equation

The basic equation relating to the device in Fig. II-3-1 is

\[ \theta(t) = \phi(t) + AK \frac{F(p)}{p} \sin \phi(t) \]  

(3-9)

where \( A = \text{rms signal amplitude, volts} \)

\( K = \text{total open-loop amplitude, volts}^{-1} \)

\( \theta(t) = \text{input phase function, rad} \)

\( \phi(t) = \text{phase-error function, rad} \)

\( p = \frac{d}{dt}, \text{the Heaviside operator.} \)

3-B. Tracking When the Loop Error Is Small

If \( |\phi| < \pi/6 \), then \( \sin \phi(t) \approx \phi(t) \), so

\[ \theta(t) \approx p + AKF(p) \phi(t) . \]  

(3-10)

Considering \( \theta(t) = d(t) \) radians, i.e., a doppler-shift function only, the steady-state phase-error is

\[ \phi_{ss} = \lim_{t \to \infty} \frac{\omega D(s)}{s + AKF(s)} \]  

(3-12)

where \( D(s) = L[d(t)] \). If \( d(t) = \lambda \omega t^2 / n! \), and if \( F(s) \) has \( l \) poles at the origin, as \( F(s) = q(s)/s^l p(s) \), then for \( n \geq l + 1 \)

\[ \phi_{ss} = \frac{\lambda \omega p(0)}{AK q(0)} \frac{t^{n-l-1}}{(n-l-1)!} \text{ as } t \to \infty . \]  

(3-18)

For \( n < l + 1, \phi_{ss} = 0 \), so perfect doppler tracking ultimately results. When \( n = l + 1 \), there is a constant steady-state error

\[ \phi_{ss} = \frac{\lambda \omega p(0)}{AK q(0)} \]  

(3-17)

3-C. Acquiring Lock in the First-Order Loop

A phase-plane diagram of first-order (\( F(s) = 1 \)) lock-in is given in Fig. II-3-2. Lock-in proceeds as

\[ \Omega_{\text{max}} \text{ (rad/sec)} = AK = 2W_c \text{ (cps)} \]  

(3-23)

\[ \phi_{ss} = \sin^{-1} \left( \frac{\Omega_{\text{max}}}{AK} \right) = \sin^{-1} \left( \frac{\Omega_{\text{max}}}{2W_c} \right) \text{ (rad)} \]  

(3-24)

\[ t_{\text{sec}} \approx \frac{1}{W_c \cos \phi_{ss}} \ln \frac{2}{\delta_{\text{lock}}} \text{ (sec)} \]  

(2-37)

where \( \Omega_{\text{max}} \) is the maximum value of \( \Omega_{\text{max}} \), the initial loop frequency offset in rad/sec, to which the loop ultimately locks, \( t_{\text{sec}} \) is the maximum time in seconds required to achieve lock within \( \delta_{\text{lock}} \) radians, and \( W_c \) is the 2-sided loop bandwidth in cps.

3-D. Acquiring Lock in the Second-Order Loop with Passive Loop Filter

With a loop filter of the form

\[ F(s) = \frac{1 + \tau_2 s}{1 + \tau_1 s} \]  

(3-28)

where \( \tau_1, \tau_2 \) are in seconds, for \( d(t) = \phi_0 + \Omega_d t \), there results

\[ \phi_{ss} = \sin^{-1} \left( \frac{\Omega_{\text{max}}}{AK} \right) = \sin^{-1} \left( \frac{\Omega_{\text{max}} \tau_2^2}{\tau_1} \right) \text{ (rad)} . \]  

(3-30)

But for lock-in to occur for all initial conditions of the VCO,

\[ \Omega_{\text{max}} < \frac{2}{\tau_2} \left[ r \left( 1 + \frac{r_1}{2\tau_2} \right)^{16} = 4r \left( 1 + \frac{r_1}{2\tau_2} \right)^{16} \right] \text{ (rad/sec)} . \]  

(3-33)

The time to reach frequency lock (not phase lock) is

\[ t_{\text{freq lock}} \approx \frac{(r + 1)^2}{8r^2 \omega c} \left( \frac{\Omega_{\text{max}}}{\omega c} \right)^2 \text{ (sec)} \]  

(3-34)

where \( r = AK \tau_2^2/\tau_1 \) and \( \omega c \) is the loop fiducial bandwidth (see Chapter 5). We assume \( r_1 \gg \tau_1 \). Typical plots of lock-in behavior are illustrated in Figs. II-3-3 and II-3-4.
\[ x(t) = A \sqrt{2} \sin(w_0 t + \theta(t)) \]
\[ y(t) = AK_m \sin \phi(t) \]
\[ v(t) = K_i \sqrt{2} \cos(w_0 t + \dot{\theta}(t)) \]

**Fig. II-3-1.** The basic phase-locked loop

**Fig. II-3-2.** First-order loop pull-in behavior
Fig. II-3-3. Lock-in behavior of a second-order loop with imperfect integrator, \( F(s) = \frac{1 + \tau_2 s}{1 + \tau_1 s} \), for \( \Omega_0/\Delta K = 0.4 \) and \( \Delta K^2/\tau_1 = 2 \). The upper bound of Eq. (3-33) is \( \Omega_0/\Delta K < 0.693 \).
Fig. II-3-4. Lock-in behavior of a second-order loop with imperfect integrator, $F(s) = (1 + \tau s)/(1 + \tau_1 s)$, for $\omega_0/\Delta K = 0.9$ and $\Delta K \tau_2^2/\tau_1 = 2$. The upper bound of Eq. (3-33) is $\omega_0/\Delta K < 0.693$. Note that lock-in occurs only when the trajectory happens to pass through the "slot." Otherwise, the trajectory enters the periodic frequency lag region shown.
Fig. II-3-5. Phase-plane trajectory of a second-order loop with perfect integrator to a doppler-rate input $A_0$ for $AK \tau_2/\tau_1 = 2$, and $A_0 = AK/2\tau_1$. 
3-F. Locking the Second-Order Loop with Perfect Integrator

With a loop filter of the form

\[ F(s) = \frac{1 + \tau_2 s}{\tau_1 s} \quad (3-41) \]

where \( \tau_1 \) and \( \tau_2 \) are in seconds, the doppler polynomial \( d(t) = \theta_s + \Omega_s t + \frac{1}{2} \Lambda_o t^2 \) produces a constant steady-state error given by

\[ \phi_{ss} = \sin^{-1} \left( \frac{\Lambda_o}{\beta^2} \right) = \sin^{-1} \left( \frac{\Lambda_o (r + 1)^3}{4r W_L^2} \right) \quad (3-42) \]

where \( \beta \) is the loop natural frequency (see 5-20). To guarantee lock-in for \( r = 2 \), it is necessary to have (see Fig. II-3-5)

\[ | \Lambda_o | < k(r) \beta^2 \quad (3-43) \]

\[ \text{sgn} \left[ \Omega(0) \right] = - \text{sgn} (\Lambda_o) \]

When \( r = 2 \), the value \( k(r) \) in (3-43) is about 0.5; for other values of \( r \), factors are given in Fig. II-3-6.

When the VCO is swept for lock-in with \( \Omega_s \neq 0 \) and \( \Lambda_o = 0 \), the equations above restrict the VCO input sweep \( e \) by

\[ \left| K_{v_{r0}} \frac{de}{dt} \right| < k(r) \beta^2 \quad (3-44) \]

These relations are approximately true for the passive loop of Section 3-D when \( \tau_1 \gg \tau_2 \).

Fig. II-3-6. Normalized doppler rates \( \Lambda_o/\beta^2 \) and \( \Lambda_o/w_L^2 \) for which lock is guaranteed in the absence of noise, as a function of the loop parameter \( v = AK \tau_2^2/\tau_1 \). Note that for a fixed bandwidth \( w_L \), the optimum value of \( r \) is about two (\( \zeta = 0.707 \)).
CHAPTER 4

BEHAVIOR OF PHASE-LOCKED LOOPS WITH STOCHASTIC INPUTS

With input and VCO noises \( n(t), n_v(t) \), the basic equation is

\[
\dot{\theta}(t) = \frac{\Delta K F(p)}{p} \sin [\theta(t) - \hat{\theta}(t)] + \frac{K F(p)}{p} n(t) + \frac{K_v}{p} n_v(t)
\]

and equivalent mathematical models are shown in Figs. II-4-1 and II-4-2.

Owing to noise, the loop skips cycles at a certain rate, executing a nonstationary random walk between lock-in points. The actual phase-error variances are thus nonstationary, growing without bound as time goes on. When phase angles are reduced modulo \( 2\pi \), however, the resulting steady-state phase error process is stationary.

---

**Fig. II-4-1.** Exact mathematical equivalent of the phase-locked loop. Sources of external, as well as internal, noises are shown; VCO tuning voltage \( e \) is also indicated.

**Fig. II-4-2.** Equivalent exact mathematical model of phase-locked loop, with explicit reduction of \( \phi (\text{mod } 2\pi) \). The VCO has been replaced by \( K_v/s \) (tuning bias and VCO noise omitted).
CHAPTER 5
THE LINEARIZED ANALYSIS OF PHASE-LOCKED SYSTEMS

5-A. Behavior of a Linear Loop

When \( \phi(t) \) is very small, we can approximate \( \sin \phi \approx \phi \). In such a case there is a linear equation relating the input phase function \( \theta(t) \) to the loop estimate \( \hat{\theta}(t) \),

\[
\hat{\theta}(t) = \frac{AKF(p)}{p + AKF(p)} \left[ \theta(t) + \frac{n(t)}{A} \right]
\]

(5.2)

omitting VCO noise. An equivalent circuit appears in Fig. II-5-1. Hence we can define a loop transfer function

\[
L(s) = \frac{AKF(s)}{s + AKF(s)}
\]

(5.3)

\[
F(s) = \frac{sL(s)}{AK[1 - L(s)]}
\]

and \( L(s) \) has some fiducial bandwidth \( \omega_L \) (see (2.37)).

There are two effects in \( \hat{\theta}(t) \): doppler \( d(t) \) and modulation \( \psi(t) \). As a result, there are three kinds of phase errors in the loop; the mean-square phase-error \( \Sigma^2 \) is of the form

\[
\Sigma^2 = \mu^2(t) + \delta^2 + \sigma^2 \text{ (rad}^2\text{)}
\]

(5.6)

where \( \mu(t) \) is transient (or doppler) distortion due to \( d(t) \), \( \delta^2 \) is modulation distortion, and \( \sigma^2 \) is phase noise:

\[
\delta^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} |L(j\omega)|^2 S_{\psi\psi}(j\omega) \, d\omega \text{ (rad}^2\text{)}
\]

(5.7)

\[
\sigma^2 = \frac{N_v \omega_L}{A^2} \text{ (rad}^2\text{)}.
\]

Total transient distortion \( \epsilon^2_t \) is defined by the integral

\[
\epsilon^2_t = \int_{-\infty}^{\infty} \mu^2(t) \, dt \text{ (rad-sec)²}
\]

(5.9)

Fig. II-5.1. The linearized model of the phase-locked loop
and the total phase-error $\Sigma^2$ is defined as

$$\Sigma^2 = \lambda^2 \epsilon^2 + \beta^2 + \sigma^2 \text{ (rad}^2) \text{.}$$  \hspace{1cm} (5-22)$$

Here $\lambda^2$ is a Lagrange multiplier, in units of seconds$^{-2}$, which can be evaluated in terms of loop bandwidth.

5-B. Calculation of Loop Bandwidths

1. First-Order Loop

If $F(s) = 1$,

$$w_r = W_r = \frac{AK}{2} \text{ (cps)} \quad (5-11)$$

$$L(s) = \frac{2w_r}{s + 2w_r} \text{.} \quad (5-12)$$

2. Second-Order Loop, Passive Integrator

With a loop filter of the form (Fig. II-5-2),

$$F(s) = \frac{1 + r_2 s}{1 + r_1 s} \text{,}$$

$$L(s) = \frac{1 + r_2 s}{1 + r_2 \left(1 + \frac{r_2}{r_1} \right) s + (r_2/r_1)s^2} \text{.} \quad (5-13)$$

where $r$ is the loop parameter ratio,

$$r = \frac{AKr_2^2}{r_1}. \quad (5-14)$$

We shall assume $r_1 \gg r_2$ in the formulas that follow. The loop fiducial bandwidth is

$$w_f = \frac{r + 1}{2r_2} \text{ (cps)} . \quad (5-18)$$

The loop has a damping coefficient $\zeta$ and natural frequency $\beta$ given by

$$\zeta = \frac{r_{\phi}^2}{2} \quad (5-20)$$

$$\beta = \frac{2r_{\phi}}{r + 1} w_f \text{ (rad/sec)} . \quad (5-20)$$

The maximum gain of the loop occurs at $\omega_{\max}$

$$\omega_{\max} = \frac{1}{r_2} \left[ (1 + 2r)^{\phi} - 1 \right]^{\phi}$$

$$= \beta \frac{1 + 2r)^{\phi} - 1}{r} \text{ (rad/sec)} \quad (5-15)$$

$$\rightarrow 0 \quad \text{as} \ r \rightarrow \infty$$

$$\rightarrow \beta \quad \text{as} \ r \rightarrow 0.$$

The maximum value of $L^2$ of loop power gain is

$$L^2 = \left| L(i\omega) \right|_{\max}^2 = \frac{r^2}{2(1 + 2r)^{\phi} + r^2 - 2r - 2}. \quad (5-16)$$

These effects are shown in Fig. II-5-3.
3. Second-Order Loop, Perfect Integrator

With a loop filter of the form

\[ F(s) = \frac{1 + \tau s}{\tau s} \]  

(5-21)

the values of \( \omega_c, \zeta, \beta, \omega_{\text{max}}, \) and \( L^2 \) are the same as those given in Section 5-B3 when \( \tau_r/\tau_r \to 0 \) in such a way that \( r \) in (5-14) remains fixed.

5-C. Optimization of Loop Parameters

The Weiner optimization of the total phase error \( \Sigma^2_{\phi} = \lambda^2 \epsilon^2 + \delta^2 + \sigma^2 \) is achieved by choosing

\[ L_{\text{opt}}(s) = 1 - \frac{N_{\text{vco}}/A}{[S(s)]^*} \]  

(5-23)

\[ S(s) = \lambda^2 E[D(s)D(-s)] + S_{\phi\phi}(s) + N_\phi/A^2. \]

5-D. Effects of VCO Noise

The total phase noise in the linear loop due to input, as well as to VCO noise, is

\[ \sigma^2 = \frac{N_{\text{vco}}}{A^2} + \left( \frac{\tau + \frac{1}{4r}}{4r} \right) \frac{N_{\text{vco}}}{\omega_c} + g\left(r\right) \frac{N_{\text{vco}}}{\omega_c^2} \]  

(5-27)

where \( N_{\text{vco}} \) is the white-noise-spectrum portion of \( \dot{K}_{\text{vco}}(t) \), and \( N_{\text{vco}} \) is the coefficient of the \( 1/f \) noise component of \( \dot{K}_{\text{vco}}(t) \):

\[ K_{\text{vco}} S_{\phi\phi}\left(j\omega\right) = N_{\text{vco}} + 2\pi N_{\text{vco}}/\omega^3 \]  

(5-25)

The function \( g\left(r\right) \) is given in Fig. II-5-4, but to good accuracy we can use \( g\left(r\right) = 1.55 \) for \( 3 < r < 10 \).

Optimum values of \( r \) and \( \omega_c \) that minimize \( \sigma^2 \) can be found: \( \sigma^2 \) is fairly insensitive to \( r \), and any value in the range \( 3 < r < 10 \) is acceptable; \( \omega_c \) can be found as the solution to

\[ \omega_c^2 - 0.286 \left( \frac{A^2 N_{\text{vco}}}{N_\phi} \right) \omega_c - 3.1 \left( \frac{A^2 N_{\text{vco}}}{N_\phi} \right) = 0. \]  

(5-29)

Fig. II-5-4. Factors governing relative contributions of VCO noise to output phase noise
CHAPTER 6

OPTIMIZED DESIGN OF TRACKING FILTERS (LINEAR ANALYSIS)

6-A. Tracking Loop Design

The filter $L_{opt}(s)$ that minimizes $\Sigma \frac{e^2}{T}$ when the doppler polynomial $d(t)$ has a Laplace transform $D(s)$ given by

$$D(s) = \frac{\theta_n}{s} + \frac{\Omega_n}{s^2} + \frac{L_n}{s^3} + \cdots = \frac{Q(s)}{s^N}$$

(6-1)

(the degree of $Q(s)$ is less than $N$) is given by the Wiener formula

$$L_{opt}(s) = 1 - \left[ (-1)^N \frac{s^N}{\theta_n} + \frac{\lambda A \pi}{N_0 \pi^2} E[Q(s)Q(-s)] \right]$$

(6-2)

6-B. Optimum Filter for Random Phase Offset

When $d(t) = \theta_n$, where $\theta_n$ is a uniformly distributed random phase offset, we find

$$L_{opt}(s) = \frac{\lambda A \pi}{(3N_0)\pi^2} \left[ \frac{1}{s + \frac{\lambda A \pi}{(3N_0)\pi^2}} \right]$$

(6-6)

$$W_L = w_L = \frac{\lambda A \pi}{2(3N_0)\pi^2} \text{ (cps).}$$

(6-7)

6-C. Optimum Filter for Frequency and Random Phase Offset

When $d(t) = \theta_n + \Omega_n t$, in which $\theta_n$ is uniformly distributed over $(-\pi, \pi)$, it follows that

$$L_{opt}(s) = \frac{s \left( 2\beta^2 + \frac{\pi^2 \beta^4}{3N_0^2} \right) + \beta^2}{s^2 + \left( 2\beta^2 + \frac{\pi^2 \beta^4}{3N_0^2} \right) s + \beta^2}$$

(6-13)

where the loop natural frequency $\beta$ is given by

$$\beta = \left( \frac{\lambda A \Omega_n}{N_0 \pi^2} \right)^{1/3}.$$ 

(6-12)

This $L_{opt}(s)$ can be realized with $F(s) = (1 + \tau_2 s)/\tau_2 s$ by setting

$$\tau_2 = \frac{1}{\beta} \left( 2 + \frac{\pi^2 \beta^4}{3N_0^2} \right)^{1/3}$$

(6-14)

$$AK \frac{\lambda A \pi}{\tau_1} = \frac{3\Omega_n}{\pi^2} (r - 2),$$

$$r = 2 + \frac{\pi^2 \beta^4}{3N_0^2}$$

When $\theta_n = 0$ above (not random), the value of $\beta$ given in (6-12) remains unchanged, but the loop parameters are somewhat different:

$$\tau_2 = \frac{2\beta}{\pi^2}$$

$$AK \frac{\lambda A \pi}{\tau_1} = \beta$$

$$r = 2 \text{ (so } \zeta = 0.707)$$

$$w_L = \frac{3\beta}{2(2)^{1/3}}$$

1. Choice of Parameters

If we design the loop optimally to lock onto an $\Omega_n$ appearing at the bandwidth edge (i.e., $\Omega_n = 2\pi b_L$), the proper values of $r$ and $\zeta$ for $\theta_n$ random are

$$r = 2.282$$

$$\zeta = 0.755.$$ 

(6-19)

Hence the loop configuration is described by

$$\tau_2 = 1.643/w_L$$

(6-20)

$$AK \frac{\lambda A \pi}{\tau_1} = \frac{w_L^2}{1.180}$$

$$L_{opt}(s) = \frac{1 + 1.643/w_L s}{1 + (1.643/w_L s)(1.180/w_L^2)s^2}.$$ 

(6-21)
The resulting transient error is
\[
\Sigma^2_T^{(opt)} = \frac{\pi^2}{6\Omega_0 (3r(r-2))^{1/2}} \tag{6.25}
\]
\[
= 0.537/w_L \text{ for } r = 2.282 \text{ and } \Omega_0 = \pi w_L.
\]

3. The Effects of Doppler Rates in Second-Order Loops

If there is a small doppler-rate term \( \Lambda_0 \text{ (rad/sec)} \) in \( d(t) \), the loop with \( F(s) = (1 + \tau s)/\tau s \) has a steady-state error
\[
\phi_{ss} \text{ (doppler rate)} = \frac{\Lambda_0 \tau_1}{AK} = \frac{\Lambda_0 (r + 1)^2}{4r w_L^2} \text{ (rad)}. \tag{6.28}
\]

When the imperfect-integrator loop filter \( F(s) = (1 + \tau_1 s)/(1 + \tau_1 s) \) is used, there is a steady-state growth in phase error.

\[
\phi_{ss} \text{ (loop rate)} \rightarrow \frac{\Lambda_0}{AK} t = \frac{\Lambda_0 (r + 1)^2}{4r w_L^2} \left( \frac{t}{\tau_1} \right) \text{ (rad)}. \tag{6.27}
\]

These values are minimized, for a fixed \( w_L \) and \( \tau_1 \), by choosing \( r = 1 \). Parameters must be chosen to maintain less than about 30 deg error in the loop, and the loop must be periodically retuned. Representative types of transient behavior are illustrated in Fig. II-6-1.

4. Comments on the Choice of \( r \) in Second-Order Tracking Loops

An \( r \) between 6 and 10 minimizes the effects of VCO noise, whereas a value of 2.282 (or 2 if \( \phi_0 = 0 \)) minimizes the total phase error in tracking a frequency offset, and, further, \( r = 1 \) provides the best doppler-rate tracking capability. However, loop performance is generally rather insensitive to changing values of \( r \).
Fig. II-6-1. Response of first- and second-order loops to various inputs.
CHAPTER 7
DESIGN OF SYNCHRONOUS DETECTOR AGC SYSTEMS

7-A. Synchronous Detector AGC Loop

With the AGC device in Fig. II-7-1, the AGC voltage \( c(t) \) is given by

\[
c(t) = C(p) \left[ a(t) + \left( \frac{20 \log e}{e_o} \right) K_D n_i(t) + 20 \log \cos \phi(t) - K_e \right]
\]

(7-8)

where

\[
a(t) = 20 \log A(t) \quad \text{(db-volts\textsuperscript{2})}
\]

\[
C(p) = \frac{K_{AGC}}{K_A \left[ 1 + K_{10C} Y(s) \right]}
\]

\[
K_{AGC} = \frac{K_i e_o K_c}{20 \log e}
\]

and \( Y(s) \) is the AGC loop filter. The other quantities are given in the Appendix with units. To convert \( a \) to db-watts, subtract \( 10 \log R_{in} \) from \( a \), with \( R_{in} \) in ohms. An equivalent circuit is shown in Fig. II-7-2.

7-C. Calibration Equation

The steady-state mean value of AGC voltage \( c \) is related to the input signal strength \( a \) by

\[
a = K_{rec} + 20 \log \left( \frac{K_D}{e_o} \right) + \left( \frac{1}{C(0)} \right) c + (10 \log e) \sigma^2
\]

\[\text{(db-volts\textsuperscript{2})}\]

(7-10)

Thus, given \( c \), one can infer a value of \( a \). A typical calibration is given in Fig. II-7-3. The AGC voltage fluctuates with noise; its steady-state variance is approximately

\[
\text{var} [c] = \left[ (20 \log e)^2 \left( \frac{N_{AT} w_c}{A^2} \right) \right]
\]

\[+ 2(10 \log e)^2 \sigma^4 \frac{w_c}{w_c} \] \( C^2(0) \).

(7-9)

where \( \sigma^2 \) is the phase noise variance (5-7).

Fig. II-7-1. A synchronous-detector AGC loop, using a phase-locked loop to provide a coherent reference. (The actual phase-locked loop may be part of the gain-controlled receiver.)
Fig. II-7-2. Equivalent diagram of an AGC loop. Linearized analysis follows the assumptions that $A^*(t)$ is exponentially related to $c(t)$ and that $a(t) = a^*(t)$. The input is $a(t) = 20 \log |A(t)|$. Adjusted loop gain $a^*(t) = 20 \log \left( A^*(t) e_{in}/K_i \right)$.

Fig. II-7-3. Measured AGC curve showing departure from linear behavior.
CHAPTER 8
THE DOUBLE-HETERODYNE PHASE-LOCKED RECEIVER

8-A. Basic Configuration of the Receiver

A block diagram is shown in Fig. II-8-1; the input and output phases are related by

\[
\dot{\theta} = aK_vK_{VCO}M \frac{F(p)}{p} \sin \left[ \theta - \hat{\theta} + (1 + M_1)\theta_1 + \theta_2 \right] + \frac{n(t)}{a} + \frac{MK_{VCO}}{p} n_r(t).
\]  

(8-6)

This is the same form that governs the behavior of a simple loop when we set

\[
AK = aK_vK_{VCO}M F \quad \text{sec}^{-1}
\]

\[
\phi = \theta - \hat{\theta} + (1 + M_1)\theta_1 + \theta_2 \quad \text{rad}.
\]  

(8-8)

The term \( F \) is the dc gain of \( F(s) \), viz., \( F(0) \).

8-B. Effects of Band-Pass Limiting

The parameter \( a \) above is the signal suppression factor derived by Davenport (Fig. II-8-2). In addition to this,

---

**Fig. II-8-1.** The double-heterodyne receiver, with equations for signals at each point.
the ratio $A^2/N_0$ at the limiter input comes out of the limiter as $A^2/N_0 \Gamma$, where $\Gamma$ is a factor shown in Fig. II-8-3. With the loop filter $F(s) = F \cdot (1 + r_s)/s(1 + r)$, we thus have (linear theoretic approximation):

$$\sigma^2 = \frac{N_o \omega_0}{A^2} \Gamma = \frac{\Gamma}{m(1 + r)\omega_0} \left(1 + \frac{\alpha}{\alpha_0} r_0\right) \text{ (rad)}$$

$$r = \frac{\alpha K_v K_{VCO} M \omega_r^2}{r_1} = \frac{\alpha}{\alpha_0} r_0$$  \hspace{1cm} (8-14)

$$\omega_L = \frac{1 + r}{2r_2} \omega_0 \left[1 + \left(\frac{\alpha}{\alpha_0}\right)^2 r_0\right] \text{ (cps)}$$

$$\zeta = \frac{r_{\nu}}{2} = \left(\frac{\alpha}{\alpha_0}\right)^{\nu} \zeta_0$$

$$r_0 = \frac{U}{2} \left[1 + \left(1 + \frac{4}{U}\right)^{\nu} \right] \hspace{1cm} (10-4)$$

$$A_o^2 = N_o \omega_0 \omega$$ (watts)

$$m = \left(\frac{A}{A_o}\right)^2$$  \hspace{1cm} (10-2)

$$\rho_H = m \left(\frac{\omega_{L,0}}{\omega_H}\right)$$  \hspace{1cm} (10-3)

$$\alpha = \frac{0.7854\rho_H + 0.4768\rho_H^2}{1 + 1.024\rho_H + 0.4768\rho_H^2}$$  \hspace{1cm} (8-13)

$$U = \frac{\pi (K_v K_{VCO} M \omega)^2 r_2}{8r_1^2 \omega_H} = \frac{r_0^2}{r_0 + 1/ \omega_H}$$  \hspace{1cm} (10-5)
CHAPTER 9

BEHAVIOR OF PHASE-LOCKED LOOPS (NONLINEAR ANALYSIS)

9-A. The Spectral Equation

The phase-noise spectrum can be approximated by

\[ S_{\text{ee}}(s) = \frac{K^2 N_0 F(s) F(-s)}{-s^2 + \gamma AK [sF(s) - sF(-s)] + (AK\gamma)^2 F(s) F(-s)} \]

\[ (\text{rad}^2/\text{cps}) \quad (9-11) \]

\[ \eta = e^{-\sigma^2/2} \]

\[ \gamma = \frac{1 - e^{-\sigma^2}}{\sigma^2} \quad (9-19) \]

when \( \sigma^2 \leq 1 \) (rad\(^2\)). Then (9-11) can be integrated to give \( \sigma^2 \):

\[ \sigma^2 = \frac{N_0 \omega_{\text{eq}}}{A^2 \gamma^2} \quad (\text{rad}^2) \quad (9-28) \]

where \( \omega_{\text{eq}} \) is the loop equivalent bandwidth,

\[ \omega_{\text{eq}} = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{\text{ee}}(j\omega) \, d\omega \quad (\text{cps}) \]

\[ (9-27) \]

9-D. Behavior of First-Order Loop

For \( F(s) = 1 \), a better value of \( \gamma \) than (9-19) is (see Fig. II-9-2)

\[ \gamma_{fp} = \left( \frac{\sinh \sigma^2}{\sigma^2 e^{\sigma^2}} \right)^{1/\sigma} \quad (9-18) \]

for \( \sigma^2 \leq 1 \). The phase-noise variance is thus a transcendental equation in \( \sigma^2 \), whose results are given in Fig. II-9-3,

\[ \sigma e^{-\sigma^2/2} \left( \sinh \sigma^2 \right)^{1/\sigma} = \frac{N_0 K}{2A} = \frac{N_0 \omega_{\text{eq}}}{A^2} \quad (9-25) \]

At \( \sigma^2 = 1 \), the value of \( N_0 \omega_{\text{eq}}/A^2 \) is 0.057.

---

Fig. II-9-2. Variation of the parameter \( \gamma \) as a function of the Gaussian variance \( \sigma^2 \), for various forms of \( R_{\text{eq}}[\tau] \).

Note that, for \( \sigma^2 < 1 \), there is not a significant dependence on the form of \( R_{\text{eq}}[\tau] \).

Fig. II-9-3. Comparison of linear, quasi-linear, and linear-spectral approximate methods with the actual behavior of the first-order loop.
Fig. II-9-4. Comparison of linear and nonlinear theories for second-order, constant linear bandwidth loop, i.e., the value of $r$ is kept constant at $r_0 = 2$. 

Signal strength, dB above $A_0^2 = N_0w_0$.
Fig. II-9-5. Comparison of phase-noise variances by linear and linear-spectral approximations. The noise density is fixed, and the signal level is varied. The value of $r$ is taken as $r_0 = 2$ at a reference signal level of $A_0^2 = 3N_0/2r_0 = N_0w_{L_0}$. Note that the ultimate roll-off is 5db/decade, rather than 10db/decade, as in Fig. 9-4. Note also that even the linear approximation produces some curvature of $\sigma^2$ near $A_0^2$. 
9-E. Calculation of Behavior of the Second-Order Loop

With the usual passive-integrator filter second-order loop,

\[
\frac{\omega_{l(eq)}}{\omega_L} = \frac{1 + ry}{(1 + \gamma) \left[ 1 + \frac{2(\gamma - \eta)}{ry^2} \right]^{1/4}} \quad \text{(9-29)}
\]

\[
\zeta_{\text{eq}} = \gamma^{1/4} \left[ 1 + \frac{2(\gamma - \eta)}{ry^2} \right]^{1/4} \quad \text{(9-33)}
\]

These, with (9-28) above, specify the loop behavior.

Figures II-9-4, II-9-5, and II-9-6 indicate the true loop behavior compared with that predicted by linear theory.

Fig. II-9-6. Variation in bandwidth and damping parameters as a function of signal strength. The value of \( r \) at a reference signal level \( A_n^2 = 3N_0/2r_0 \) was taken as \( r_0 = 2 \).
CHAPTER 10

DESIGNING A DOUBLE-HETERODYNE TRACKING LOOP

10-A. Definition of Receiver Threshold

Threshold is defined as a signal level $A^2_n$ such that

$$A^2_n = N_0\omega_{-6} = \frac{N_0}{2\tau_2}$$ \hspace{1cm} (10-1)

(assuming $r_\tau \gg \tau_\gamma$). The loop is not acting linearly at this signal level; nevertheless $\omega_{-6}$ is the bandwidth a linear loop would have at $A = A_n$.

10-B. Tracking Loop Performance of the Double-Heterodyne Receiver

The ratio $(A/A_n)^2$ is the receiver margin:

$$m = (A/A_n)^2.$$ \hspace{1cm} (10-2)

With measured parameters $W_c K_v K_{vo} M F, \tau_\gamma$, and $\gamma$, the value of $r_\gamma$ can be calculated (when $10\omega_{-6} < \omega_\mu$) as

$$r_\gamma = \frac{U}{2} \left[ 1 + \left( 1 + \frac{4}{U} \right)^{1/2} \right]$$

$$U = \frac{\alpha(K_v K_{vo} M F)^2 \tau^3_1}{8 \tau^4_1 \omega_\mu}$$ \hspace{1cm} (10-4)

$$= \frac{r^2_\gamma}{r_\gamma + 1}.$$

The only modifications to the linear theory in Chapter 8 that need concern us for $\sigma^2 \leq 1$ are that $A$ is replaced by $A_n$, $r$ by $r_\gamma$, and $\omega_{-6}$ by $\omega_{-6}(\gamma)$. The resulting equations that specify loop behavior are

$$\rho_\mu = m \left( \frac{\omega_{-6}}{\omega_\mu} \right)$$ \hspace{1cm} (10-6a)

$$\alpha = \left[ \frac{0.7854\rho_\mu + 0.4768\rho^2_\mu}{1 + 1.024\rho_\mu + 0.4768\rho^2_\mu} \right]^{1/4}$$ \hspace{1cm} (10-6b)

$$\Gamma = \frac{1 + 0.345\rho_\mu}{0.862 + 0.696\rho_\mu}$$ \hspace{1cm} (10-6c)

$$r = \frac{a K_v K_{vo} M F \tau^3_1}{\tau_1} = \left( \frac{\alpha}{a_0} \right) r_\gamma = r^2_\gamma \left( \frac{\alpha}{a_0} \right) \beta^2_\gamma$$ \hspace{1cm} (10-6d)

$$\gamma = \left( 1 - e^{-\sigma^2} \right) / \sigma^2$$ \hspace{1cm} (10-6f)

$$\eta = e^{-\sigma^2} / \sigma^2$$ \hspace{1cm} (10-6g)

$$\omega_{-6} = \frac{1 + r}{2\tau_2} = \frac{\omega_{-6}}{1 + r_\gamma}$$ \hspace{1cm} (cps) \hspace{1cm} (10-6h)

$$\omega_{-6}(\gamma) = \frac{1 + r_\gamma}{2\tau_2} \left[ 1 + \frac{2(\gamma - \eta)}{\gamma r_\gamma} \right]^{1/2}$$ \hspace{1cm} (10-6i)

$$\xi = \frac{r_\gamma}{2} = \left( \frac{\alpha}{a_0} \right)^{1/2} \zeta_\gamma = \left( \frac{\alpha}{a_0} \right)^{1/2} \frac{r_\gamma}{2}$$ \hspace{1cm} (10-6j)

$$\zeta_{\gamma} = \frac{r^{1/2}_{-6}(\gamma)}{2} \left[ 1 + \frac{2(\gamma - \eta)}{\gamma r_\gamma} \right]^{1/2} = \left( \frac{\alpha}{a_0} \gamma \right)^{1/2} \zeta_\gamma$$ \hspace{1cm} (10-6k)

$$\sigma^2 = \frac{N_0 \omega_{-6}(\gamma)}{A^2 \tau^2_\gamma} \approx \frac{1 + \left( \frac{\alpha}{a_0} \gamma \right)^{1/2}}{1 + r_\gamma}$$ \hspace{1cm} (10-6m)

$$\beta = \frac{2\omega_{-6}}{r + 1} = \left( \frac{\alpha}{a_0} \right)^{1/2} \frac{2\omega_{-6}}{r_\gamma + 1}$$ \hspace{1cm} (10-6n)

---

*S*Since $N_0 = kT\theta/2$ (volts$^2$/cps), the threshold signal power is $P_\theta = A^2_n/R = kT\omega_\mu/2 = kT\theta$ (watts).

---

Fig. 10-1. Variation in margin producing $\sigma^2 = 1$ as a function of threshold design parameter,

$r_\gamma = a_0 K_v K_{vo} M F \tau^3_1 / \tau_1$. 

---

87
10-D. The Signal Level Producing $\sigma^2 = 1$

The signal level $A^2$, producing $\sigma^2 = 1$, corresponds to a margin $m$, approximately equal to

$$m = \left( \frac{r_0 \Gamma_o}{2 \Gamma_o (r_0 + 1)} \right)^2 \left[ 1 + \left( 1 + \frac{4(r_0 + 1)}{\Gamma_o r_0^2} \right)^{\frac{3}{2}} \right]$$

$$= 3.13 \text{ (for } r_0 = 2). \quad (10-8)$$

Hence the $\sigma^2 = 1$ margin is about 5 db at $r_0 = 2$. (This result is shown in Fig. II-10-1.)

10-E. Choice of Receiver Parameters

Choosing values of $N_0$, $\omega_n$, $K_0$, $K_{VCO}$, $MF$, and $\omega_H$ specifies a typical receiver design. However, the only parameters required to plot performance are $m$, $\gamma_0$, $\rho_{he}$, and perhaps $\tau_o/\tau_1$ ($\tau_o/\tau_1$ is not needed if $\tau_1 \gg \tau_2$). Such plots are given in Figs. II-10-2 and II-10-3.

---

Fig. II-10-2. Comparison of linear and nonlinear approximations to loop rms phase error, as a function of loop margin
Fig. II-10-3. Variation of loop bandwidths and damping factors, as a function of loop margin.
APPENDIX

Nomenclature

Those symbols used throughout the text are listed here, along with assigned names, proper units, and the equation number that either defines the quantity, gives its first use, or else refers to its approximate point of first introduction. Quantities labeled "abs" are dimensionless.

- $a$: either mean value of $a(t)$ below, or square root of $a^2$ below.
- $a_k$: coefficients in expansion of $p(\phi_1, \phi_2)$, (9-5).
- $a^2$: variance of $\phi(t)$, in rad$^2$, (9-4).
- $a(t)$: input rms signal $A(t)$ expressed in db-volts$^2$, (7-4).
- $a_d(t)$: that part of $a(t)$ due to deterministic and non-stationary changes in signal, in db-volts$^2$, (7-11).
- $a_d(t)$: ha' part of $a(t)$ due to stationary random signal fluctuations, in db, (7-11).
- $\mathcal{A}(s)$: the Laplace transform of $a_d(t)$, (7-13).
- $a^*(t)$: adjusted receiver attenuation $A^*(t)$ expressed in db-volts$^2$, (7-4).
- $A$: rms signal amplitude into loop, volts, (3-1).
- $A(t)$: rms signal amplitude into receiver at time $t$, in volts, (7-1).
- $A^*(t)$: receiver attenuation factor at time $t$, abs, (7-1).
- $A_d(t)$: nonstationary part of $A(t)$, in rms volts, (7-12).
- $A_d(t)$: stationary gain fluctuation part of $A(t)$, abs, (7-12).
- $A_t$: receiver rms signal level producing $\sigma = 1$ rad, volts, (10-7).
- $b_n$: fiducial (one-sided) bandwidth of the linear transfer function $H(s)$, cps, (2-37).
- $B_n$: equivalent noise (one-sided) bandwidth of the linear transfer function $H(s)$, cps, (2-27).
- $c, c(t)$: AGC control signal, volts, (7-3).
- $C$: capacitance, farads.
- $C(s)$: AGC closed-loop transfer function, (7-5).
- $d(t)$: doppler phase function on input signal, rad, (3-11).
- $d(t)$: time-varying part of doppler phase function on input signal, rad, (5-9).
- $D(s)$: doppler phase function in s-domain, $D(s) = \mathcal{L}[d(t)]$, (3-12).
- $e$: VCO tuning bias, volts, (3-5).
- $e_0$: AGC gain-adjust bias, volts, (7-1).
- $E(\cdot)$: statistical expectation operator (2-1).
- $f$: frequency variable, cps,$^1$\textsuperscript{10}
- $f_1$: first IF frequency, cps, (8-1),$^1$\textsuperscript{19}
- $f_2$: second IF frequency, cps, (8-1),$^1$\textsuperscript{19}
- $f_A$: first-mixer heterodyne frequency, $M$ times VCO output frequency, cps, (8-1),$^1$\textsuperscript{19}
- $f_x$: second-mixer heterodyne frequency, from internal oscillator, cps, (8-1),$^1$\textsuperscript{19}
- $F$: (finite) dc gain of loop filter, abs, (8-8).
- $F(s)$: loop low-pass filter transfer function, (3-7).
- $\mathcal{F}$: Fourier transformation operator.
- $g$: conductance of diode, or transconductance of triode, mhos, (2-66).
- $g(r)$: VCO noise term coefficients, abs, (5-28).
- $G$: conductance, mhos, Fig. 2-9.
- $G_{ce}(f_w)$: single-sided spectral density of $x(t)$, volts$^2$/cps, (2-26).
- $h$: Planck's constant, $6.625 \times 10^{-34}$ joule-sec, (2-53).
- $h(t)$: unit-impulse response of linear filter $H$, volts, (2-14).

$^1$May take on a negative value.
H a filter operator, Section 2-B.

\( H(s) \) linear filter transfer function, \( L[h(t)] \), (2-28).

\( i(t) \) current, amperes, (2-54).

\( j = \sqrt{-1} \), the imaginary unit.

\( k \) Boltzmann’s constant, \( 1.38 \times 10^{-23} \) joule/°K, (2-53).

\( k(t) \) maximum normalized lock-on doppler rate, abs, (3-43).

\( K \), \( K_m, K_{VCO} \) equivalent simple-loop gain, volts^{-1}sec^{-1}, (3-8).

\( K_1 \) rms VCO signal output, volts, (3-2).

\( K_4 \) receiver AGC attenuation, db/volt, (7-3).

\( K_{AC} \) AGC-loop equivalent loop gain, abs, (7-6).

\( K_C \) AGC amplifier gain, abs, (7-6).

\( K_p \) phase detector gain, volts/rad, (8-2).

\( K_B \) AGC detector gain, (volts peak out)/(volts rms in), (7-3).

\( K_{11} \) first IF gain, abs, (8-1).

\( K_{12} \) second IF gain, abs, (8-1).

\( K_m \) mixer gain (simple loop), volts^{-1}, (3-4).

\( K_s \) adjusted receiver attenuation with no AGC, db, (7-3).

\( K_{VCO} \) VCO gain constant, rad/sec-volt, (3-5).

\( K_{rec} \) actual receiver attenuation with no AGC, db, (7-3).

\( l \) limiter rms output level, volts, (8-2).

\( L \) inductance, henries, Fig. 2-9.

\( L^* \) maximum value of \( |L(j\omega)| \), abs, (5-4).

\( L(s) \) linear loop phase-transfer function, (5-3).

\( \mathcal{L} \) Laplace transformation operator.

\( m \) receiver margin above threshold, abs, (10-2).

\( M \) VCO output frequency multiplication factor, abs, (8-1).

\( M \) internal oscillator frequency multiplication ratio, \( (f_o/f_{ox}) - 1 \), abs, (8-4). It must be a rational number, and may be negative.

\( n_0(t) \) receiver input noise waveform, volts, Section 8-A.

\( n_1(t) \) loop input noise waveform, volts, (4-1).

\( n(t) \) simple loop baseband noise waveform, volts, (4-3).

\( N_i(t) \) equivalent VCO noise referred to VCO input, volts, (5-24).

\( N \) noise power out of linear filter, volts^2, (2-23).

\( N_o \) noise (two-sided) spectral density, volts^2/cps, (2-20).

\( N_{oo} \) receiver input noise (two-sided) spectral density of \( n_o(t) \), volts^2/cps, Section 8-A.

\( N_s \) noise (one-sided) spectral density, volts^2/cps, (2-27).

\( N_{ov} \) VCO adjusted input white-noise density, (rad/sec)^2/cps, (5-25).

\( N_{ov} \) VCO adjusted input 1/f noise density, (rad/sec)^2/cps, (5-25).

\( p \) Heaviside operator, \( d/dt \), (3-6).

\( p(A) \) probability density on \( A \), abs, (2-1).

\( p(s) \) denominator of \( F(s) \), (3-15).

\( P \) input signal power, \( A^2 \), volts^2, (2-29).

\( P_y \) power in \( y(t) \), volts^2, (2-19).

\( q(s) \) numerator of \( F(s) \), (3-15).

\( Q_o \) resonance quality factor, \( w_o L/R \), abs, (2-57).

\( Q(s) \) numerator polynomial of \( D(s) \), (6-1).

\( r \) second-order loop-parameter ratio \( AKr^2_o / \tau_i \), abs, (5-14).

\( R_1 \) resistance, ohms, Fig. 2-9.

\( R_{xy}(t_1, t_2) \) statistical cross-correlation between \( x(t_1) \) and \( y(t_2) \), volts^2, (2-4).

\( R_x(r) \) statistical autocorrelation of stationary process \( x(t) \), volts^2, (2-5).

\( R_x(r) \) time autocorrelation of the function \( x(t) \), volts^2, (2-6).

\( s \) complex frequency variable, rad/sec.

\( S \) output signal power of linear filter, volts^2, (2-33).

\( S_{xy}(j\omega) \) two-sided spectral density of stationary process \( x(t) \); \( \mathcal{F}[R_{xy}(r)] \), volts^2/cps, (2-8).

\( S_{xy}(j\omega) \) two-sided spectral density of function \( x(t) \); \( \mathcal{F}[R_x(r)] \), volts^2/cps, (2-9).
\[ S(s) \text{ spectral density of total phase-error, rad}^2/\text{cps, (5-23)}. \]
\[ S'(s) \text{ spectral density of total AGC error, (db-volts)}^2/\text{cps, (7-13)}. \]
\[ t \text{ time, sec.} \]
\[ t_1, t_2 \text{ specific instants of time, sec.} \]
\[ t_{\text{seq}} \text{ first-order loop phase-acquisition time, sec, (3-27)} \]
\[ t_{\text{freq seq}} \text{ second-order loop frequency-acquisition time, sec, (3-34)} \]
\[ T, T_\infty \text{ equivalent noise temperature of receiver, } ^\circ \text{K, (2-53).} \]
\[ T_c \text{ cathode temperature of diode or triode, } ^\circ \text{K, (2-66).} \]
\[ U \text{ receiver measured-parameter ratio, ab, (10-4).} \]
\[ v(t) \text{ VCO output, volts, (3-2).} \]
\[ w_\text{H} \text{ two-sided fiducial bandwidth of linear filter } H(s), \text{ cps, (2-37).} \]
\[ w_\text{L} \text{ two-sided fiducial linear loop bandwidth of phase transfer function } L(s), \text{ cps, (5-3).} \]
\[ w_{\text{L(eq)}} \text{ equivalent two-sided fiducial bandwidth of loop output spectrum, cps, (9-29).} \]
\[ W_\text{H} \text{ two-sided noise bandwidth of a linear filter } H(s), \text{ cps, (2-24).} \]
\[ W_\text{L} \text{ two-sided noise bandwidth of a linear loop whose phase transfer function is } L(s), \text{ cps, (5-4).} \]
\[ x(t) \text{ arbitrary time function.} \]
\[ X(s) \text{ Laplace transform of } x(t), \mathcal{L}[x(t)]. \]
\[ y(t) \text{ time function.} \]
\[ Y(s) \text{ AGC loop-filter response, (7-5).} \]
\[ z(t) \text{ time function.} \]
\[ Z(jw) \text{ complex impedance function, ohms.} \]
\[ \alpha \text{ signal voltage suppression factor, ab, (8-1).} \]
\[ \alpha_0 \text{ value of } \alpha \text{ at threshold, (10-6).} \]
\[ \beta \text{ second-order loop natural frequency } (AK/\tau_c)^{0.5}, \text{ rad/sec, (6-12).} \]
\[ \gamma(s) \text{ sinusoidal-error spectral ratio function, ab, (9-7).} \]
\[ \gamma(0). \]
\[ \gamma_{\text{tp}} \text{ sinusoidal-error function total-power transfer ratio, ab, (9-15).} \]
\[ \delta^2 \text{ modulation distortion, rad}^2, (5-6). \]
\[ \delta_{\text{lock}} \text{ first-order loop "in-lock" constant, rad, (3-27).} \]
\[ \delta_\phi(t) \text{ rectangular unit-pulse function, volts, (2-13).} \]
\[ \delta(t) \text{ Dirac delta function, lim } \delta(t) \text{ as } t \to \infty, (2-13). \]
\[ \epsilon_\phi \text{ total transient distortion, rad}^2, (5-9). \]
\[ \epsilon \text{ loop tracking error, rad, (8-7).} \]
\[ \xi \text{ linear loop damping factor, ab, (5-20).} \]
\[ \xi_{\text{eq}} \text{ nonlinear loop equivalent damping factor, ab, (9-33).} \]
\[ \eta \text{ Bussgang coefficient, ab, (9-3).} \]
\[ \phi \text{ phase angle, rad.} \]
\[ \theta_0(0) \text{ initial value of loop phase offset, rad, (3-13).} \]
\[ \theta(t) \text{ or } \theta \text{ input signal phase function, rad, (3-1).} \]
\[ \hat{\theta}(t) \text{ or } \hat{\theta} \text{ loop estimate of } \theta(t), \text{ rad, (3-2).} \]
\[ \lambda^2 \text{ Lagrange multiplier, sec}^{-1}, (5-22). \]
\[ \lambda_n \text{ } d^{n+1}(0 +), \text{ the } n\text{th doppler moment, rad/ sec}^n, (3-15). \]
\[ \mu \text{ mean of a stationary random variable, volts, (2-12).} \]
\[ \mu(t) \text{ mean of a nonstationary random variable, volts, (2-10).} \]
\[ \pi \text{ 3.14169} \ldots \]
\[ \rho_{y} \text{ signal-to-noise ratio of } y(t), \text{ ab, (2-35).} \]
\[ \sigma^2_x \text{ variance of stationary random variable } x, (x \text{ units})^2, (2-12). \]
\[ \sigma^2_{\varphi}(t) \text{ variance of nonstationary random process } x(t) \text{ at time } t, (x \text{ units})^2, (2-11). \]
\[ \sigma^2_{\phi} \text{ variance of loop phase-noise, rad}^2, (5-7). \]
\[ \sigma^2_{\text{vco}} \text{ variance of loop phase-noise due to noise in VCO, rad}^2, (5-26). \]
\[ \tau \text{ } t_1 - t_2, \text{ time difference, variable of } R_{\text{av}}(\tau), \text{ sec, (2-4).} \]
\[ \tau_1 \text{ second-order loop denominator time constant of } F(s), \text{ sec, (3-28).} \]
\[ \tau_2 \text{ second-order loop numerator time constant of } F(s), \text{ sec, (3-28).} \]
\( \phi(t) = \phi \) two-heterodyne loop detector error (same as \( \phi_p \) for simple loop), rad, (5-8), (8-6).

\( \phi_s \) steady-state loop detector phase error, rad, (3-12).

\( \phi_0 \) initial value of loop detection phase error, rad, (3-32).

\( \phi \) stationary equivalent phase-error process, rad, Section 4-B.

\( \omega \) angular frequency variable, rad/sec.

\( \omega_0 \) VCO short-circuit output frequency, rad/sec, (3-1).

\( \omega_{\text{max}} \) frequency at which \( |L(j\omega)| \) is maximum for second-order loops, rad/sec, (5-15).

\( \omega_1 \) first IF frequency, rad/sec, (8-1).\(^1\)

\( \omega_2 \) second IF frequency, rad/sec, (8-1).\(^1\)

\( \omega_{\text{h1}} \) first mixer heterodyne frequency, \( M \) times VCO output frequency, rad/sec, (8-1).\(^1\)

\( \omega_{\text{h2}} \) second mixer heterodyne frequency, from internal oscillator, rad/sec, (8-1).\(^1\)

\( \Gamma \) limiter performance factor, \( w_l a_t / w_l a_t n \), abs, (8-17).

\( \Lambda_0 \) initial value of doppler phase-rate, rad/sec\(^2\), (3-13).

\( \Sigma^2 \) mean-square phase error, rad\(^2\), (5-6).

\( \Sigma^2_f \) total mean-square phase error, rad\(^2\), (5-22).

\( \psi(t) \) modulation phase function, rad, (3-11).

\( \Omega = \Omega(t) \) loop frequency error, rad/sec, (3-20).

\( \Omega_0 \) initial value of loop frequency offset rad/sec, (3-12).

\( \Omega_{\text{max}} \) maximum value of \( \Omega \) for which loop locks in absence of noise, rad/sec, (3-23).

\(^1\) May take on a negative value.