NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Technical Report No. 32-819

Theory and Practical Design of Phase-Locked Receivers

Volume I

GPO PRICE $ _____

CFSTI PRICE(S) $ _____

Hard copy (HC) 4.00

Microfiche (MF) .75

ff 653 July 65

Robert C. Tausworthe

JET PROPULSION LABORATORY

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA

February 15, 1966

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Technical Report No. 32-819

Theory and Practical Design of Phase-Locked Receivers Volume I

Robert C. Tausworthe

Richard Goldstein

Richard M. Goldstein, *Manager*
Communications Systems Research Section

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

February 15, 1966

BLANK PAGE

CONTENTS

PART I

THE THEORY OF PHASE-LOCKED LOOPS

CONTENTS (Cont'd)

CONTENTS (Cont'd)

PART II

SUMMARY OF PHASE-LOCKED LOOP DESIGN

CONTENTS (Cont'd)

FIGURES

PART I

FIGURES (Cont'd)

PART II

FIGURES (Cont'd)

FIGURES (Cont'd)

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

17323

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

BLANK PAGE

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.

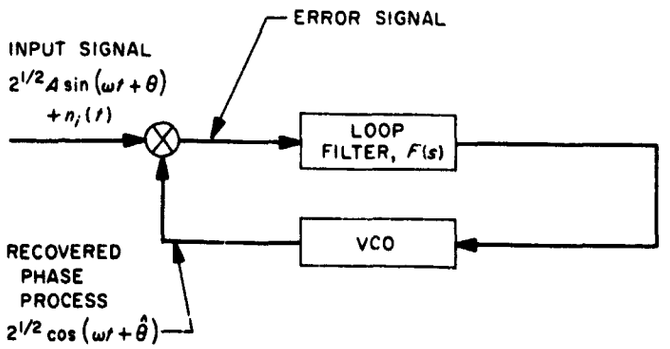


Fig. 1-1. Basic configuration of a simple phase-locked loop. The mixer output, filtered by F (s), is used to control the frequency of the voltage-controlled oscillator (VCO).

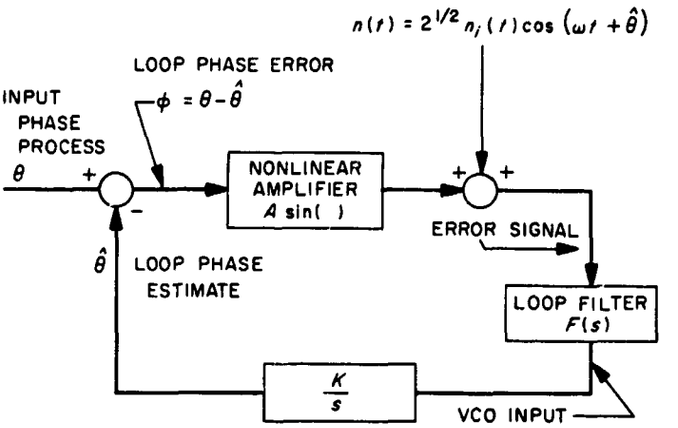


Fig. 1-2. A mathematically equivalent model of the simple phase-locked loop. Here there are two separate inputs, θ and $n(t)$, whereas the two are combined in the actual input.

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

1. Jaffe, R. M., and Rechtin, E., "Design and Performance of Phase-Lock Circuits Capable of Near Optimum Performance Over a Wide Range of Input Signals and Noise Levels," *IRE Transactions on Information Theory*, Vol. IT-1, pp. 66–76, March 1955.
2. Develet, J. A., Jr., "A Threshold Criterion for Phase-Lock Demodulation," *Proceedings of the IRE*, Vol. 51, pp. 349–356, February 1956.
3. Gruen, W. J., "Theory of AFC Synchronization," *Proceedings of the IRE*, Vol. 41, pp. 1043–1048, August 1953.
4. Viterbi, A. J., "Acquisition and Tracking Behavior of Phase-Locked Loops," *Proceedings of the Symposium on Active Networks and Feedback Systems*, Polytechnic Institute of Brooklyn, N. Y., Vol. 10, pp. 583–619, April 1960.
5. Margolis, S. G., "The Response of a Phase-Locked Loop to a Sinusoid Plus Noise," *IRE Transactions on Information Theory*, Vol. IT-3, pp. 135–144, March 1957.
6. Booton, R. C., Jr., "The Analysis of Nonlinear Control Systems with Random Inputs," *Proceedings of the Symposium on Nonlinear Circuit Analysis*, Polytechnic Institute of Brooklyn, N. Y., pp. 369–391, April 1953.
7. Van Trees, H. L., "Functional Techniques for the Analysis of the Nonlinear Behavior of Phase-Locked Loops," presented at WESCON, San Francisco, Calif., August 20–23, 1963.
8. Tikhonov, V. I., "The Effects of Noise on Phase-Lock Oscillation Operation," *Automatika i Telemakhanika*, Vol. 22, No. 9, 1959.
9. Tikhonov, V. I., "Phase-Lock Automatic Frequency Control Application in the Presence of Noise," *Automatika i Telemekhanika*, Vol. 23, No. 3, 1960.
10. Viterbi, A. J., "Phase-Locked Loop Dynamics in the Presence of Noise by Fokker-Planck Techniques," *Proceedings of the IEEE*, Vol. 51, No. 12, pp. 1737–1753, December 1963.
11. Tausworthe, Robert C., "A New Method for Calculating Phase-Locked Loop Performance," *Space Programs Summary, No. 37-31*, Vol. IV, pp. 292–300, Jet Propulsion Laboratory, Pasadena, California, February 1965.

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 θ can assume any value over $[0, 2\pi]$ with equal likelihood. Again, θ 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:

$$E(A) = \int A\, p(A) dA \tag{2-1}$$

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

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

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[1] stationary, but, as our example shows, the converse need not be true.

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

$$\mathcal{R}_{xx}(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x(t) x(t+\tau) dt \tag{2-3}$$

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

$$R_{xx}(t_1, t_2) = E\left[x(t_1)\, x(t_2)\right] \tag{2-4}$$

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

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

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

$$R_{xx}(\tau) = \mathcal{R}_{xx}(\tau). \tag{2-6}$$

[1]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,

$$\mathcal{S}_{xx}(j\omega) = \int_{-\infty}^{+\infty} \mathcal{R}_{xx}(\tau)\, e^{-j\omega\tau}\, dt \tag{2-7}$$

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

$$S_{xx}(j\omega) = \int_{-\infty}^{+\infty} R_{xx}(\tau)\, e^{-j\omega\tau}\, dt. \tag{2-8}$$

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

$$\mathcal{S}_{xx}(j\omega) = S_{xx}(j\omega). \tag{2-9}$$

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 $S_{xx}(j\omega)$, we may replace $j\omega$ by the complex frequency variable s; the resulting function $S_{xx}(s)$ is the analytic continuation of $S_{xx}(j\omega)$ 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., $|\omega| \to (-s^2)^{1/2}$.)

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

$$\mu(t) = E[x(t)] \tag{2-10}$$

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

$$\begin{aligned} \sigma_x^2(t) &= E\{[x(t) - \mu(t)]^2\} \\ &= E[x^2(t)] - \mu^2(t) \end{aligned} \tag{2-11}$$

For stationary processes, both σ_x^2 and μ are independent of t, and, if $R_{xx}(\infty)$ exists,

$$\begin{aligned} \mu_x^2 &= R_{xx}(\infty) \\ \sigma_x^2 &= R_{xx}(0) - R_{xx}(\infty) \end{aligned} \tag{2-12}$$

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 $t_2 = t_1 - \tau$:

$$\begin{aligned} R_{xy}(\tau) &= E[x(t)\, y(t-\tau)] \\ &= R_{yx}(-\tau). \end{aligned}$$

The Fourier transform $S_{xy}(j\omega)$ 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 $\delta_T(t)$:

$$\delta_T(t) = \begin{cases} \dfrac{1}{T} & \text{for } 0 \leqslant t \leqslant T \\ 0 & \text{for all other values of } t. \end{cases} \tag{2-13}$$

As T approaches zero, $\delta_T(t)$ 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 $\delta_T(t)$ as $T \to 0$.

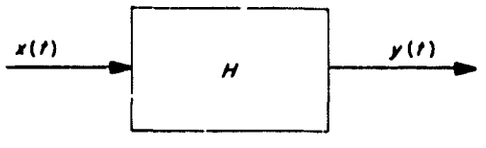


Fig. 2-1. Filtering device

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:

$$\begin{aligned} y(t) &= \int_{-\infty}^{+\infty} x(t_1)\, h(t-t_1)\, dt_1 \\ &= \int_{-\infty}^{+\infty} x(t-t_1)\, h(t_1)\, dt_1 . \end{aligned} \tag{2-14}$$

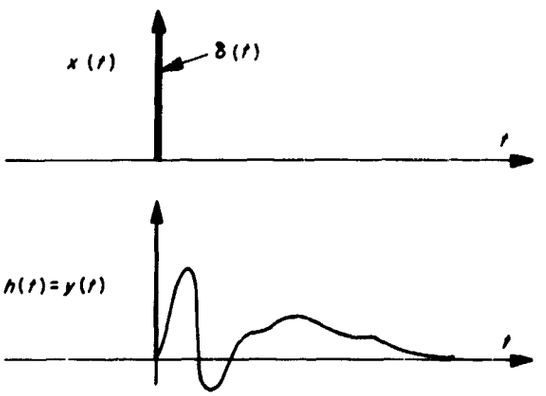


Fig. 2-2. Response of a realizable filter to a unit-impulse function

This law states that the output y at time t is the superposition of the input t_1 seconds ago, weighted by the decay response of the filter after t_1 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,

$$R_{yy}(\tau) = \int_{-\infty}^{+\infty}\int_{\infty}^{\infty} h(t_1)\, h(t_2)\, R_{xx}(\tau + t_1 - t_2)dt_1 dt_2 \tag{2-16}$$

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

$$\begin{aligned} S_{yy}(j\omega) &= H(j\omega)\, H(-j\omega)\, S_{xx}(j\omega) \\ &= |H(j\omega)|^2\, S_{xx}(j\omega). \end{aligned} \tag{2-17}$$

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[2] appearing at the filter output can be found by inverse transformation

$$P_y = R_{yy}(0) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} |H(j\omega)|^2\, S_{xx}(j\omega)d\omega. \tag{2-19}$$

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, $R_{nn}(\tau) = 0$ for every $\tau \neq 0$. Such a process is often called a white noise process. It is often the character of such noises that

$$R_{nn}(\tau) = N_0\delta(\tau). \tag{2-20}$$

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

$$S_{nn}(j\omega) = N_0 \quad \text{for all } -\infty < \omega < \infty. \tag{2-21}$$

It should be pointed out that $R_{nn}(0)$ indicates $n(t)$ has infinite power, while $S_{nn}(j\omega)$ 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

$$S_{yy}(j\omega) = N_0|H(j\omega)|^2 \tag{2-22}$$

and the output noise power N from the filter is

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

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

$$W_H = \frac{\frac{1}{2\pi}\int_{-\infty}^{+\infty} |H(j\omega)|^2\, d\omega}{|H(j\omega)|^2_{\max}}. \tag{2-24}$$

[2]Actually $R_{yy}(0)$ is the mean-square voltage of the process $y(t)$, so the unit of power we are speaking of here is volts^2; 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 signal-to-noise, so the units are unimportant, as long as they are consistent.

In these terms, the output noise power is

$$N = N_0 W_H |H(j\omega)|^2_{\max} . \qquad (2\text{-}25)$$

Viewed graphically (Fig. 2-3), W_H 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 folds the negative frequency response onto the positive. The result is the single-sided spectrum, which we give a different notation:

$$G_{zz}(j\omega) = \begin{cases} 2\,S_{zz}(j\omega) & \omega \geqslant 0 \\ 0 & \omega < 0 . \end{cases} \qquad (2\text{-}26)$$

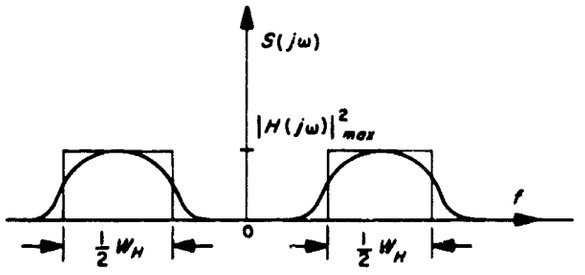


Fig. 2-3. Equivalent noise bandwidth

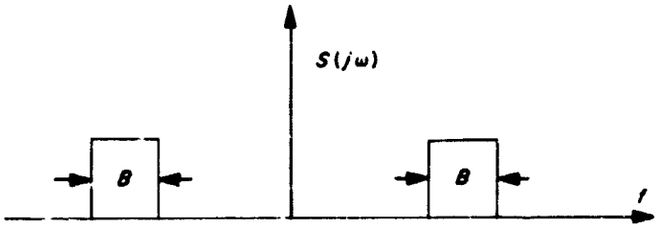


Fig. 2-4. Double-sided frequency response

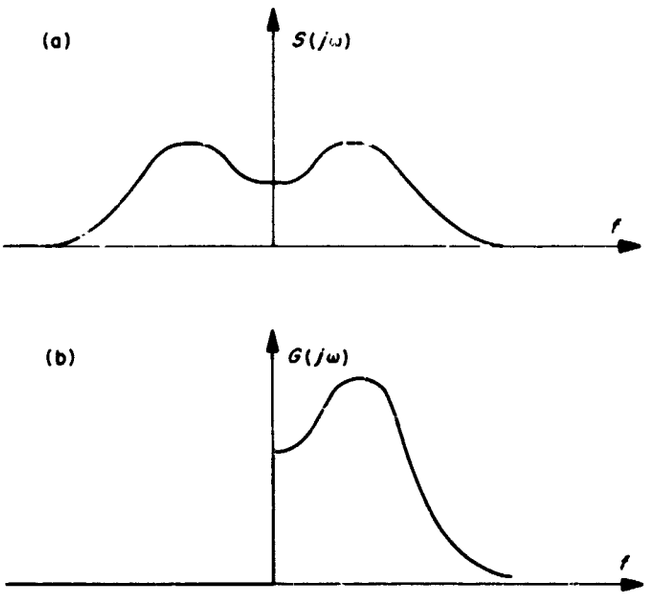


Fig. 2-5. Double- and single-sided spectra

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

$$B_H = \frac{\frac{1}{2\pi}\int_0^\infty |H(j\omega)|^2\, d\omega}{|H(j\omega)|^2_{\max}} = \frac{1}{2} W_H \qquad (2\text{-}27)$$

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

$$G_{nn}(j\omega) = \begin{cases} N_+ \\ 0 \end{cases} = \begin{cases} 2N_0 \text{ for } \omega \geqslant 0 \\ 0 \qquad \omega < 0 . \end{cases} \qquad (2\text{-}28)$$

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

$$N = N_+ B_H |H(j\omega)|^2_{\max} .$$

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 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 $G(j\omega)$ 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

$$x(t) = (2P)^{1/2}\cos(\omega_0 t + \theta) \tag{2-29}$$

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

$$\begin{aligned} R_{xx}(\tau) &= P\cos\omega_0\tau \\ S_{xx}(j\omega) &= \pi P\left[\delta(\omega-\omega_0) + \delta(\omega+\omega_0)\right]. \end{aligned} \tag{2-30}$$

The factor of π is present in S_{xx} because we must satisfy

$$R_{xx}(0) = P = \frac{1}{2\pi}\int_{-\infty}^{+\infty} S_{xx}(j\omega)d\omega. \tag{2-31}$$

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

$$S_{yy}(j\omega) = \pi P\,|H(j\omega)|^2\left[\delta(\omega-\omega_0) + \delta(\omega+\omega_0)\right] \tag{2-32}$$

and the output signal power (call it S) is

$$S = R_{yy}(0) = P\,|H(j\omega_0)|^2. \tag{2-33}$$

If white noise is also present at the input, so that $x(t) = (2P)^{1/2}\cos(\omega_0 t + \theta) + n(t)$, then the output consists of a signal component with power S and noise with power N:

$$R_{yy}(0) = S + N. \tag{2-34}$$

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

$$\rho_y = \frac{S}{N} = \frac{P}{N_0 W_H}\,\frac{|H(j\omega_0)|^2}{|H(j\omega)|^2_{\max}}. \tag{2-35}$$

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

$$\rho_{y(\max)} = \frac{P}{N_0 W_H}. \tag{2-36}$$

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

2-E. Fiducial Bandwidth

The ratio $W_H|H(j\omega)|^2_{\max}/|H(j\omega_0)|^2$ 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

$$w_H = 2b_H = \frac{W_H|H(j\omega)|^2_{\max}}{|H(j\omega_0)|^2} = \frac{\frac{1}{2\pi}\int_{-\infty}^{+\infty}|H(j\omega)|^2 d\omega}{|H(j\omega_0)|^2}. \tag{2-37}$$

This quantity is much like the noise bandwidth of H, and, in fact, reduces to W_H when the maximum filter response occurs at ω_0. It is the same form of definition as that given to W_H, except that it is referenced to an arbitrary frequency ω_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 $W_H = 2B_H$, as shown in the Figure. The filter input, denoted by $x(t)$, is

$$x(t) = (2P)^{1/2}\sin(\omega_0 t + \theta_1) + n_1(t),$$

where $n_1(t)$ 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

$$y(t) = (2P)^{1/2}|H(j\omega_0)|\sin(\omega_0 t + \theta_2) + n_2(t). \tag{2-38}$$

Here $\theta_2 = \theta_1 + \arg H(j\omega_0)$, the new noise $n_2(t)$ is band-limited by $H(s)$ to the filter width W_H, and $S_{n_2 n_2}(j\omega) = N_0|H(j\omega)|^2$.

We shall assume now that $y(t)$ is multiplied by a unit-power[3] sinusoid, $2^{1/2}\sin(\omega_h t + \theta_3)$. Thus, the mixer output[4] is

$$\begin{aligned} v(t) = {} & P^{1/2}|H(j\omega_0)|\cos\left[(\omega_0-\omega_h)t + (\theta_2-\theta_3)\right] \\ & - P^{1/2}|H(j\omega_0)|\cos\left[(\omega_0+\omega_h)t + (\theta_2+\theta_3)\right] + n_3(t). \end{aligned} \tag{2-39}$$

[3] This assumption can be relaxed by including a multiplier gain K_m in the output $v(t)$.

[4] This supposes an ideal multiplier. Physical realization of a multiplier may leave a carrier term and harmonics.

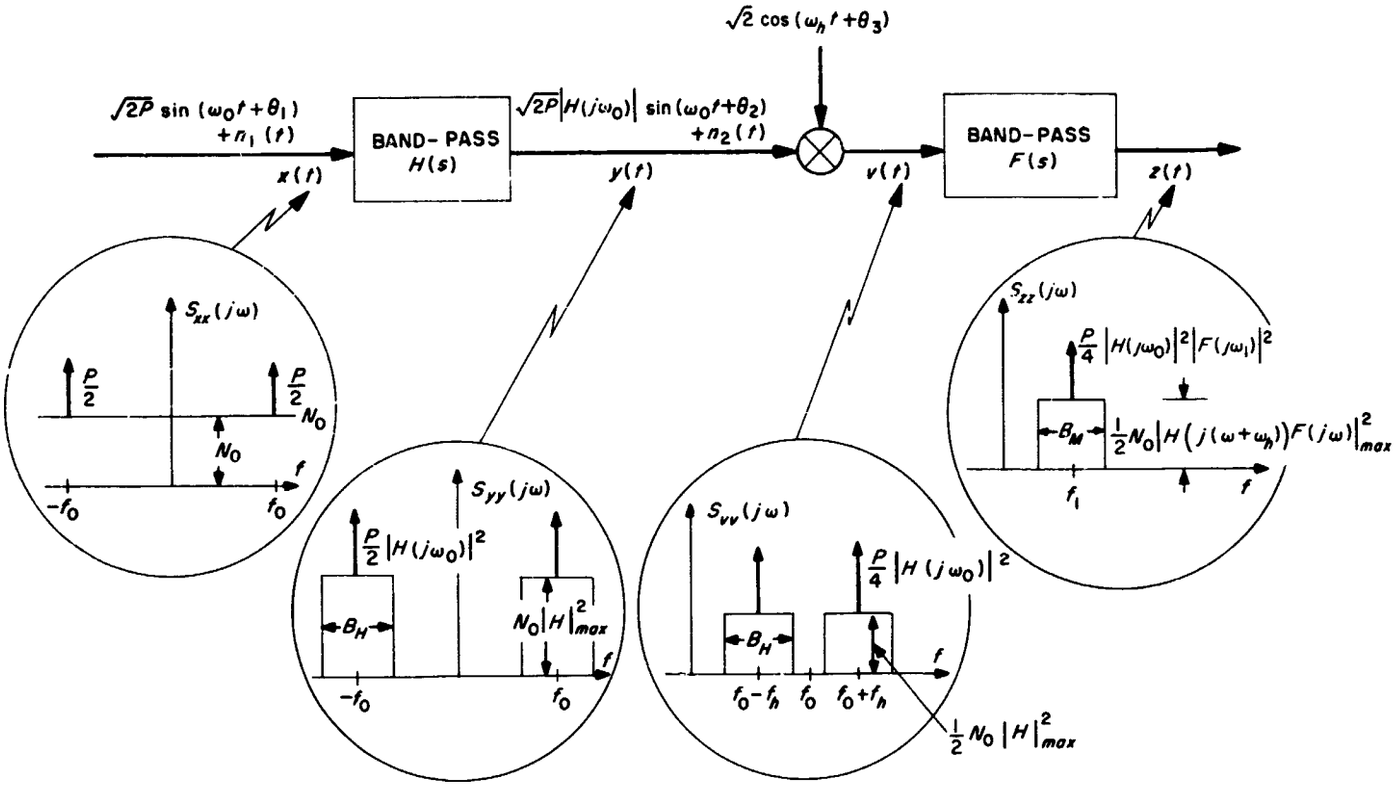


Fig. 2-6. The simple product-mixer

The bandpass filter $F(s)$ following the mixer is centered at a frequency ω_1, which can either be $\omega_0 + \omega_h$ or $\omega_0 - \omega_h$. For convenience, say it is the latter, $\omega_1 = \omega_0 - \omega_h$, with $\omega_0 \leqslant \omega_h$. One normally chooses $B_F \leqslant B_H$, since there is nothing to be gained with $B_F > B_H$. The new noise term $n_3(t)$ 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[5]

$$2B_F \leqslant B_F + B_H < 4f_h$$

$$B_F < 2f_h \text{ or } W_F < 4f_h \ . \qquad (2\text{-}40)$$

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

$$z(t) = P^{1/2} \, |H(j\omega_0)| \cdot |F(j\omega_1)| \cos(\omega_1 t + \theta_4) + n_3(t) \qquad (2\text{-}41)$$

with $\theta_4 = \theta_2 - \theta_3 + \arg F(j\omega_1)$. The output noise spectral density is, for positive values of ω,

$$S_{n_4 n_4}(j\omega) = \frac{1}{2} N_0 \, |H[j(\omega+\omega_h)]|^2 \cdot |F(j\omega)|^2 \ . \qquad (2\text{-}42)$$

We have tacitly assumed that $F(s)$ is chosen in accordance with (2-40) to suppress completely the image term that would appear in $S_{n_4 n_4}(j\omega)$ centered at $\omega_0 + \omega_h$.

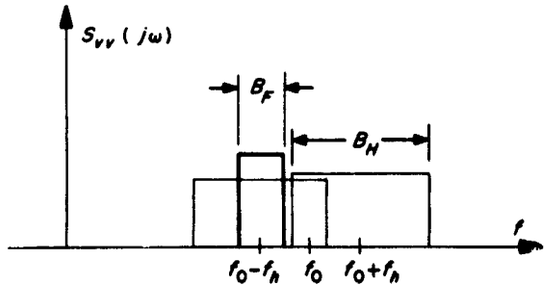


Fig. 2-7. Choosing filter bandwidths to avoid the image noise problem

[5]This may need to be modified by a factor of ½ if carrier is present in the multiplier output.

The equivalent noise bandwidth of the mixer is

$$W_M = \frac{\frac{1}{\pi}\int_0^\infty |H[j(\omega+\omega_h)]|^2 \cdot |F(j\omega)|^2 d\omega}{|H[j(\omega+\omega_h)]\,F(j\omega)|^2_{\max}} \approx W_F. \tag{2-43}$$

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

$$\rho_z = \frac{P|H(j\omega_0)|^2 \cdot |F(j\omega_1)|^2}{N_0 W_M |H[j(\omega+\omega_h)]\,F(j\omega)|^2_{\max}}. \tag{2-44}$$

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

$$|H(j\omega)|^2_{\max} = |H(j\omega_0)|^2$$

and

$$|F(\omega)|^2_{\max} = |F(j\omega_1)|^2,$$

in which case

$$\rho_{z(\max)} = \frac{P}{N_0 W_M}\left[\frac{|H(j\omega)|^2_{\max}\,|F(j\omega)|^2_{\max}}{|H[j(\omega+\omega_h)]F(j\omega)|^2_{\max}}\right] = \frac{P}{N_0 W_M} \approx \frac{P}{N_0 W_F}. \tag{2-45}$$

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 $B_F < 2f_h$ (balanced mixer) or $B_F < f_h$ (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 $W_H = 2B_H$ at carrier frequency (see Fig. 2-6) is heterodyned to a width $W_{het} = \frac{1}{2} W_H = B_H$ 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

$$z(t) = P^{1/2}\,|H(j\omega_0)|\,F(0)\cos\theta_4 + n_4(t) \tag{2-46}$$

where $\theta_4 = \theta_2 - \theta_3$. Because of the merging of the heterodyned bands, the spectrum of $n_4(t)$ becomes

$$S_{n_4 n_4}(j\omega) = \frac{N_0}{2}\{|H[j(\omega-\omega_0)]|^2 + |H(j\omega+\omega_0)|^2\}|F(j\omega)|^2 \approx N_0\,|H[j(\omega-\omega_0]|^2\,|F(j\omega)|^2. \tag{2-47}$$

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

$$S_{zz(sig)}(j\omega) = P\,|H(j\omega_0)|^2 F^2(0)\,E(\cos^2\theta_4)\delta(f). \tag{2-48}$$

The term $E(\cos^2\theta_4)$ above is referred to as the *coherence factor*. When θ_2 and θ_3 are completely unrelated, then θ_4 takes on any value between 0 and 2π with equal likelihood, so the coherence factor in this case is

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

But when θ_2 and θ_3 are fully coherent, i.e., when $\theta_2 = \theta_3$ with probability one, we have

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

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

$$\rho_z \approx \frac{P}{N_0 W_D}\frac{|H(j\omega_0)|^2 F^2(0)}{|H[j(\omega-\omega_0)]F[j(\omega)]|^2_{\max}} E(\cos^2\theta_4) \tag{2-51}$$

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

$$\rho_{z(\max)} = \frac{P}{N_0 W_D} E(\cos^2\theta_4) \approx \frac{P}{N_0 W_F} E(\cos^2\theta_4) \ . \tag{2-52}$$

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, θ_2 and θ_3 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,

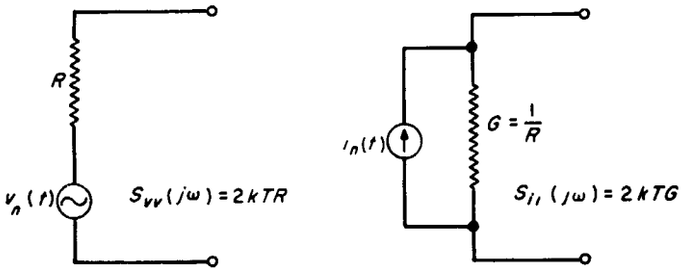


Fig. 2-8. Thevenin and Norton equivalent circuits of noisy resistors

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

$$\frac{1}{2}\frac{hf}{e^{hf/kT}-1} = \frac{1}{2}kT + \ldots \tag{2-53}$$

where h is Planck's constant

$$h = 6.625 \times 10^{-34} \text{ joule seconds}$$

and f is the frequency. This reduces to the classical result when $f \ll kT/h$. At 300°K

$$\frac{kT}{h} = 6.25 \times 10^{12} \text{ cps}$$

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

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

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.

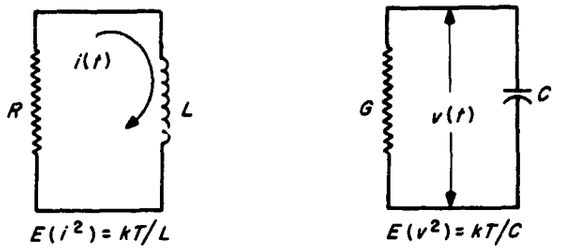


Fig. 2-9. Total Johnson noise current- and voltage-squared relations in simple circuits

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

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

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,

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

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

$$\frac{1}{2}CE(v^2) = \frac{1}{2}kT$$
$$E(v^2) = \frac{kT}{C}. \tag{2-56}$$

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.

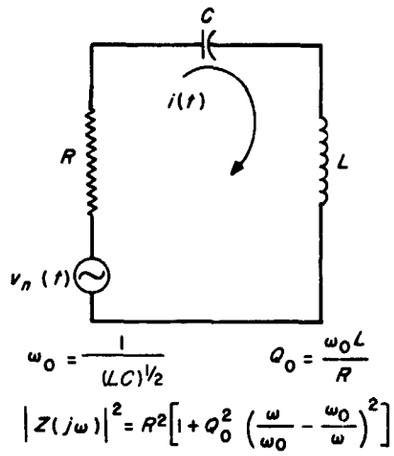


Fig. 2-10. Tuned circuit for investigating the spectral density of $v_n(t)$

This circuit is characterized by a resonant frequency f_0, a Q_0, and an impedance $Z(j\omega)$, related by

$$\omega_0 = \frac{1}{(LC)^{1/2}}$$

$$Q_0 = \frac{\omega_0 L}{R} \tag{2-57}$$

$$|Z|^2 = R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2 = R^2\left[1 + Q_0^2\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)^2\right].$$

We can regard this circuit as one excited by the Thevenin equivalent Johnson noise voltage $v_n(t)$. If we make the Q_0 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

$$S_{ii}(j\omega) = \frac{S_{v_n v_n}(j\omega)}{|Z(j\omega)|^2}. \tag{2-58}$$

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

$$\frac{kT}{L} = \frac{1}{2\pi R^2}\int_{-\infty}^{+\infty} \frac{S_{v_n v_n}(j\omega)d\omega}{1 + Q_0^2\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)^2}. \tag{2-59}$$

At very high values of Q_0, supposing $S_{v_n v_n}(j\omega)$ is fairly constant about ω_0, we can bring $S_{v_n v_n}(j\omega)$ outside the integral:

$$\frac{kT}{L} = \frac{S_{v_n v_n}(j\omega_0)}{2\pi LR}\int_{-\infty}^{+\infty} \frac{Q_0 d(\omega/\omega_0)}{1 + Q_0^2\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)^2}. \tag{2-60}$$

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

$$\int_{-\infty}^{+\infty} \frac{Q_0 d(\omega/\omega_0)}{1 + Q_0^2\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)} = \pi. \tag{2-61}$$

We thus have a white spectral density of the equivalent noise voltage in the circuit: For every ω_0

$$S_{v_n v_n}(j\omega_0) = 2kTR. \tag{2-62}$$

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_N 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

$$N = \frac{kT}{2} W_N \text{ (watts)}. \tag{2-63}$$

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

$$N = kTB_N \text{ (watts)}. \tag{2-64}$$

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 $S_{nn}(j\omega)$ is

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

or

$$G_{nn}(j\omega) = \begin{cases} N_+ = kTR \text{ for } f \geq 0 \\ 0 \qquad\qquad f < 0 \,. \end{cases} \tag{2-65}$$

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 *Schroteffekt* (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 interception 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

$$S_{ii}(j\omega) = (0.644)2kT_c g \,. \tag{2-66}$$

Here T_c 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 Book Co., Inc., New York, N. Y., 1958.

2. Pierce, J. R., "Physical Sources of Noise," *Proceedings of the IRE*, Vol. 44, pp. 601—608, May 1956.

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

$$x(t) = A(2)^{1/2} \sin [\omega_0 t + \theta(t)] \tag{3-1}$$

in which the quantities A, ω_0, and $\theta(t)$ are

A = rms voltage amplitude of $x(t)$

ω_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

$$v(t) = K_1 (2)^{1/2} \cos [\omega_0 t + \hat{\theta}(t)] . \tag{3-2}$$

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

$$y_1(t) = AK_1 \{\sin [\theta(t) - \hat{\theta}(t)] + \sin [2\omega_0 t + \theta(t) + \hat{\theta}(t)]\}. \tag{3-3}$$

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

$$y(t) = AK_1K_m \sin \phi(t) . \tag{3-4}$$

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 sin ϕ, rather than the true phase error ϕ, and thus tend to generate some confusion. Little reference to these terms will appear in the material that follows.

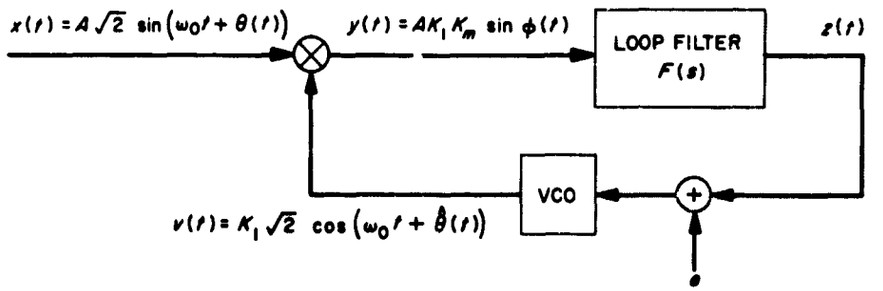


Fig. 3-1. The basic phase-locked loop

The VCO output frequency is, as its name implies, a linear fu[illegible]n of its input.

$$\omega_{VCO}(t) = \omega_0 + K_{VCO}\, z(t) + K_{VCO}\, e \ . \tag{3-5}$$

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,

$$\hat{\theta}(t) = \frac{K_{VCO}}{p}\, z(t) \tag{3-6}$$

where $p = d/dt$ is the Heaviside operator. Substitution for $z(t)$ yields

$$\hat{\theta}(t) = AK_1K_mK_{VCO}\,\frac{F(p)}{p}\sin\phi(t) \ . \tag{3-7}$$

At this point it becomes convenient to define

$$K = K_1K_mK_{VCO} \tag{3-8}$$

as the *open-loop gain*[6] 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,

$$\theta(t) = \phi(t) + AK\,\frac{F(p)}{p}\sin\phi(t) \ . \tag{3-9}$$

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 2π. The mean-phase value is called the *lock-in point*; such lock-in points are located 2π radians apart. Whenever the phase error goes through 2π 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),

$$\theta(t) = \frac{p + AKF(p)}{p}\,\phi(t) \ . \tag{3-10}$$

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

$$\phi(t) = \frac{p}{p + AKF(p)}\,\psi(t) + \frac{p}{p + AKF(p)}\,d(t) \ . \tag{3-11}$$

We see that there are two kinds of error present in the loop. That part of the error due to the modulation $\psi(t)$ 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

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

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 ϕ_0 is the initial phase offset, Ω_0 the initial frequency offset, and Λ_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

$$\phi_{ss} = \lim_{s\to 0}\frac{\theta_0 s + \Omega_0 + \Lambda_0/s}{s + AK} \ . \tag{3-13}$$

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

$$\phi_{ss} = \frac{\Omega_0}{AK} \ . \tag{3-14}$$

In such a case, ϕ_{ss} can be made small by adjustment of the loop gain. Note that the loop will track any initial phase offset θ_0 with no steady-state error in the absence of a frequency offset (i.e., $\Omega_0 = 0$).

[6]Often the open-loop gain K is defined to contain the dc gain of $F(s)$. However, $F(0)$ may not always be finite (e.g., if $F(s)$ has an integration), so we do not include it here. In cases where it is a finite value, we shall be careful to design $F(0) = 1$.

From this example, one can see that if $d(t) = \lambda_n t^n/n!$ (the nth derivative of $d(t)$ is λ_n, so $D(s) = \lambda_n/s^{n+1}$), and if $F(s)$ has l poles at the origin, as

$$F(s) = \frac{q(s)}{s^l p(s)}, \quad q(0) \neq 0, \; p(0) \neq 0, \tag{3-15}$$

then the steady-state tracking error would be

$$\phi_{ss} = \lim_{s \to 0} \frac{\lambda_n p(s) s^{l-n+1}}{s^{l+1} p(s) + AKq(s)} \; . \tag{3-16}$$

This is not finite if $n > l + 1$. When $n = l + 1$ the error is finite and has the value

$$\phi_{ss} = \frac{\lambda_n}{AK} \frac{p(0)}{q(0)} \; . \tag{3-17}$$

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 nth-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)

$$\theta(t) = \theta_0 + \Omega_0 t \tag{3-18}$$

whenever Ω_0 is small enough that $\phi(t) \sim 0$; that is, when

$$\left| \frac{\Omega_0}{AK} \right| < \frac{\pi}{6} \; . \tag{3-19}$$

To investigate the behavior for larger values of Ω_0, let us rewrite the loop equation, differentiating (3-9),

$$\Omega_0 = \dot{\phi} + AK \sin \phi \; . \tag{3-20}$$

We shall denote the *frequency error* by $\dot{\phi}(t) = \Omega(t)$, and plot Ω versus ϕ according to (3-20). The result is shown in Fig. 3-2. Whenever Ω is positive, ϕ tends to increase, and whenever Ω is negative, ϕ tends to decrease. Note that if $|\Omega_0| < AK$, 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 ϕ-axis at

$$\phi_{ss} = n\pi + \mathrm{Sin}^{-1}(\Omega_0/AK) \; . \tag{3-21}$$

This is a stable point; Ω cannot become negative because ϕ would then tend to decrease and return the system to the ϕ-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

$$\phi_{ss} = (n + 1)\pi + \mathrm{Sin}^{-1}\left(\frac{\Omega_0}{AK}\right) \; . \tag{3-22}$$

If $\Omega_0 > AK$, however, the trajectory never crosses the ϕ-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 $\Omega_0 < AK$, the loop ultimately tracks the incoming function $\theta(t)$ 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 π here are usually omitted, and the steady-state phase error is written merely

$$\phi_{ss} = \mathrm{Sin}^{-1}\left(\frac{\Omega_0}{AK}\right) \tag{3-24}$$

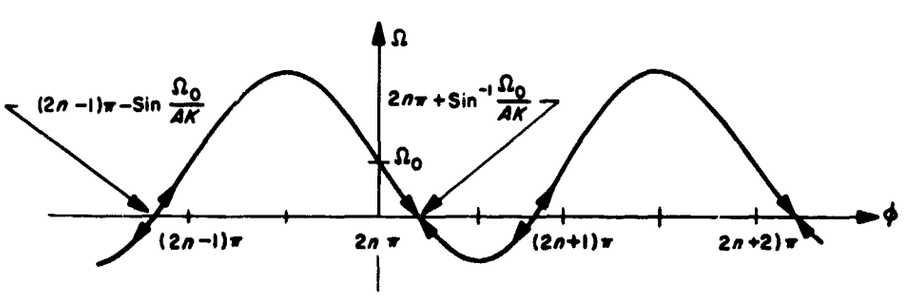


Fig. 3-2. First-order loop pull-in behavior

that is, the steady-state output of the VCO is a replica of the input, lagging by ϕ_{ss}:

$$v_{ss}(t) = 2^{1/2} \cos\left[(\omega_0 + \Omega_0)t + \theta_0 - \phi_{ss}\right]. \tag{3-25}$$

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

$$\frac{dt}{d\phi} = \frac{1}{\Omega_0 - AK \sin\phi}. \tag{3-26}$$

This can be integrated to give the pull-in time to any particular value of ϕ, 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 $-\pi + \delta_{lock} - \mathrm{Sin}^{-1}(\Omega_0/AK)$ to $\mathrm{Sin}^{-1}(\Omega_0/AK) - \delta_{lock}$. When the value of δ_{lock} is small, the answer reduces to approximately

$$t_{acq} = \frac{2}{AK \cos\phi_{ss}} \ln\frac{2}{\delta_{lock}}. \tag{3-27}$$

Hence, if a loop is designed so that $\phi_{ss} = 5$ deg when Ω_0 is 200π rad/sec, and δ_{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 low-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

$$F(s) = \frac{1 + \tau_2 s}{1 + \tau_1 s}. \tag{3-28}$$

The discussion in Section 3-B indicates that a loop with this filter can follow a phase function of the type $d(t) = \theta_0 + \Omega_0 t$ (i.e., a constant frequency offset), but that no higher-order terms may be considered. The equation governing the behavior, from (3-9), is

$$\Omega_0 = \tau_1 \ddot{\phi} + (1 + AK\tau_2 \cos\phi)\dot{\phi} + AK \sin\phi. \tag{3-29}$$

The steady-state phase error ϕ_{ss} is the same as for the first-order loop,

$$\phi_{ss} = \sin^{-1}\left(\frac{\Omega_0}{AK}\right), \tag{3-30}$$

and, clearly, the loop never locks if $\Omega_0 > AK$. Even for values of Ω_0 less than AK, pull-in possibly may not occur, even though a stable point exists in the phase plane. This is a consequence of the fact that the trajectory that takes ϕ to ϕ_{ss} 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 ϕ the value

$$\ddot{\phi} = \frac{d\Omega}{dt} = \frac{d\Omega}{d\phi} \cdot \frac{d\phi}{dt} = \Omega \frac{d\Omega}{d\phi} \tag{3-31}$$

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

$$\Omega_0 = \left(\tau_1 \frac{d\Omega}{d\phi} + 1 + AK\tau_2 \cos\phi\right)\Omega + AK \sin\phi. \tag{3-32}$$

A solution to (3-32) depends not only on Ω_0, but also on $\Omega(0)$ and ϕ_0, the initial conditions of the VCO. For certain values of Ω_0, the loop will lock regardless of what values $\Omega(0)$ and ϕ_0 take. See, for example, Fig. 3-3. In this Figure, ϕ has been limited to the region $(-\pi, \pi)$ by folding all the trajectories onto this region. When Ω is positive, ϕ increases, and when Ω is negative, ϕ 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 Ω encountered at $\phi = \pi$. This continues until the lock-in point is reached.

For values of Ω_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 Ω_0: If lock-in occurs for all initial conditions of the VCO, then Ω_0 is bounded by

$$|\Omega_0| < 2\left[\left(\frac{AK}{\tau_1}\right)(1 + \tfrac{1}{2}AK\tau_2)\right]^{1/2}. \tag{3-33}$$

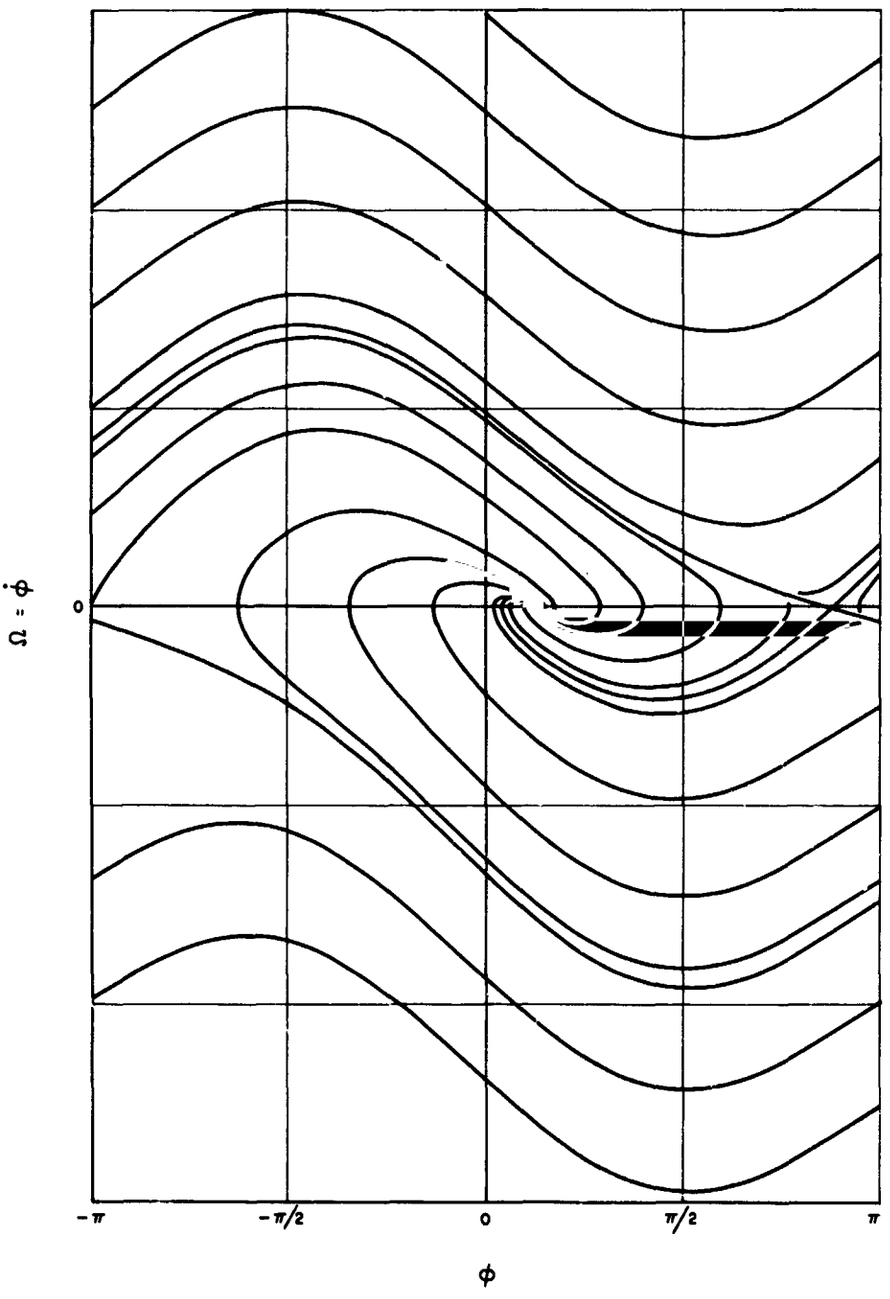


Fig. 3-3. Lock-in behavior of a second-order loop with imperfect integrator, $F(s) = (1+\tau_2 s)/(1+\tau_1 s)$, for $\Omega_0/AK = 0.4$ and $AK\tau_2^2/\tau_1 = 2$. The upper bound of Eq. (3-33) is $\Omega_0/AK < 0.693$.

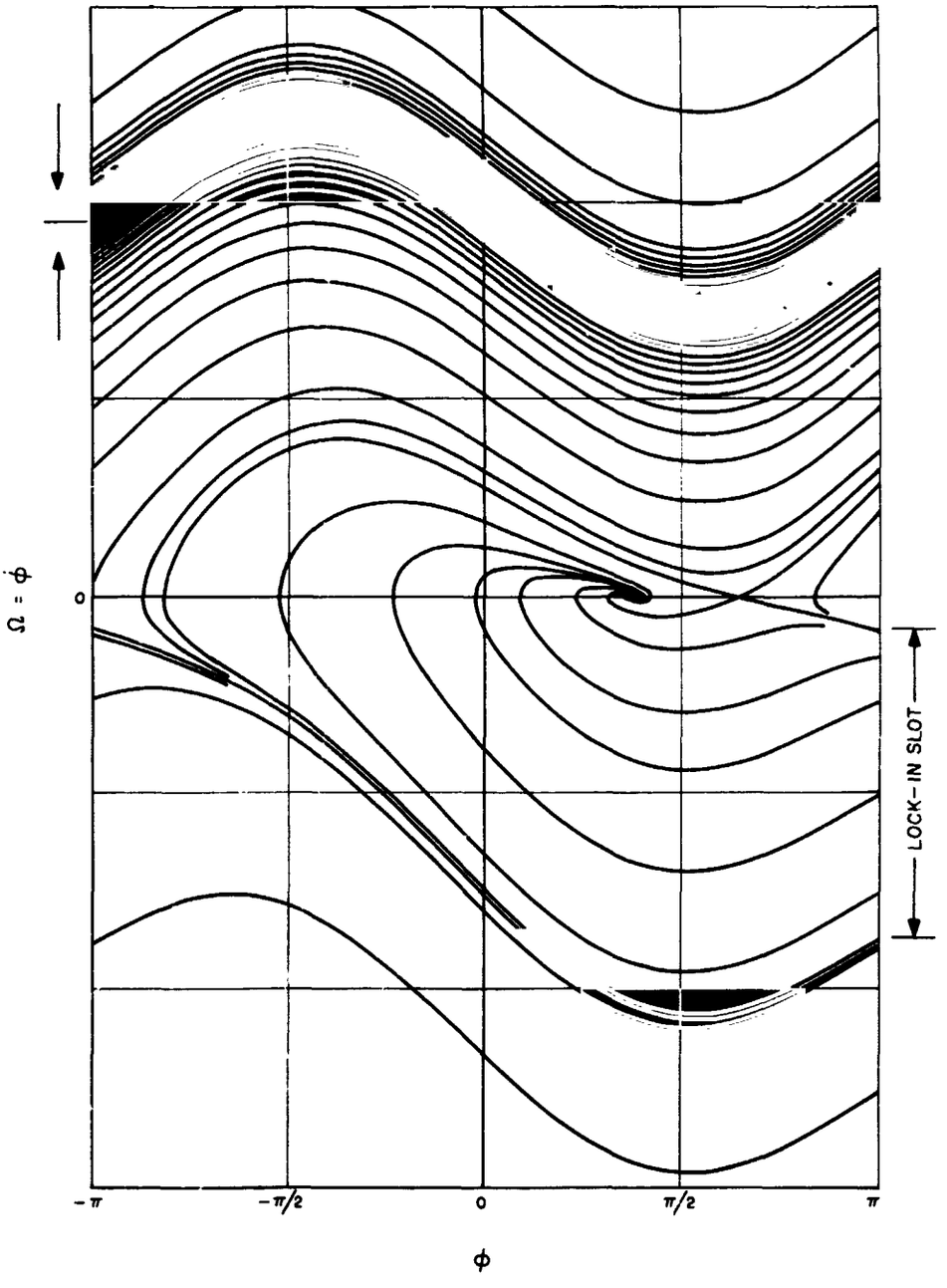


Fig. 3-4. Lock-in behavior of a second-order loop with imperfect integrator, $F(s) = (1+\tau_2 s)/(1+\tau_1 s)$, for $\Omega_0/AK = 0.9$ and $AK\tau_2^2/\tau_1 = 2$. The upper bound of Eq. (3-33) is $\Omega_0/AK < 0.693$. Note that lock-in occurs only when the trajectory happens to pass through the "slot." 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 Ω_0 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 Ω_0 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 τ_1 is large, it is

$$t_{\text{freq acq}} \approx \frac{1}{\tau_2}\left(\frac{\Omega_0\tau_1}{AK}\right)^2 \text{sec.} \tag{3-34}$$

The additional time to acquire phase lock is probably about the same as t_{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

$$\omega_{VCO}(t) = \omega_0 + K_{VCO}e + K_{VCO}\,z(t)\,. \tag{3-35}$$

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

$$\hat{\theta}(t) = (\omega_0 + K_{VCO}e)t + \frac{K_{VCO}}{p}\,z(t)\,. \tag{3-36}$$

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

$$\dot{\theta}(t) - K_{VCO}e = \Omega(t) + AKF(p)\sin\phi(t)\,. \tag{3-37}$$

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 $\dot{\theta}$ by $\dot{\theta} - K_{VCO}e$. Alternatively, we can redefine ω_0 as $\omega_{0p} = \omega_0 + K_{VCO}e$.

For example, the lock-in points now occur at

$$\phi_{ss} = \sin^{-1}\left(\frac{\Omega_0 - K_{VCO}e}{AK}\right). \tag{3-38}$$

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

$$e = \frac{\Omega_0}{K_{VCO}}\,. \tag{3-39}$$

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

$$t_{acq} = \left(\frac{2}{AK}\right)\ln\left(\frac{2}{\delta_{lock}}\right). \tag{3-40}$$

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

When $\tau_1 \gg \tau_2$, the system whose loop filter is given by (3-28) behaves much the same as the one with

$$F(s) = \frac{1 + \tau_2 s}{\tau_1 s}\,, \tag{3-41}$$

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 Λ_0 is present:

$$\phi_{ss} = \text{Sin}^{-1}\left(\frac{\tau_1\Lambda_0}{AK}\right) \tag{3-42}$$

and just as there was a limit on Ω_0 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 Λ_0 in the perfect-integrator loop.

Clearly, from (3-42), lock cannot occur at all if $\Lambda_0 > AK/\tau_1$. 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 Λ_0 (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 Λ_0 is a function of the parameter $AK\tau_2^2/\tau_1 = r$;

$$\Lambda_0 = k(r)\left(\frac{AK}{\tau_1}\right). \tag{3-43}$$

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 Ω_0, the rate by which e is changed should not produce a Λ_0 exceeding (3-43), and, as Fig. 3-5 shows for positive values of Λ_0, the sweep should begin with a negative frequency error (a positive frequency error if Λ_0 is negative). That is,

$$\left|K_{VCO}\frac{de}{dt}\right| < k(r)\left(\frac{AK}{\tau_1}\right). \tag{3-44}$$

In the passive-integrator loop, (3-43) and (3-44) hold approximately for short periods of time (see Section 6-C3).

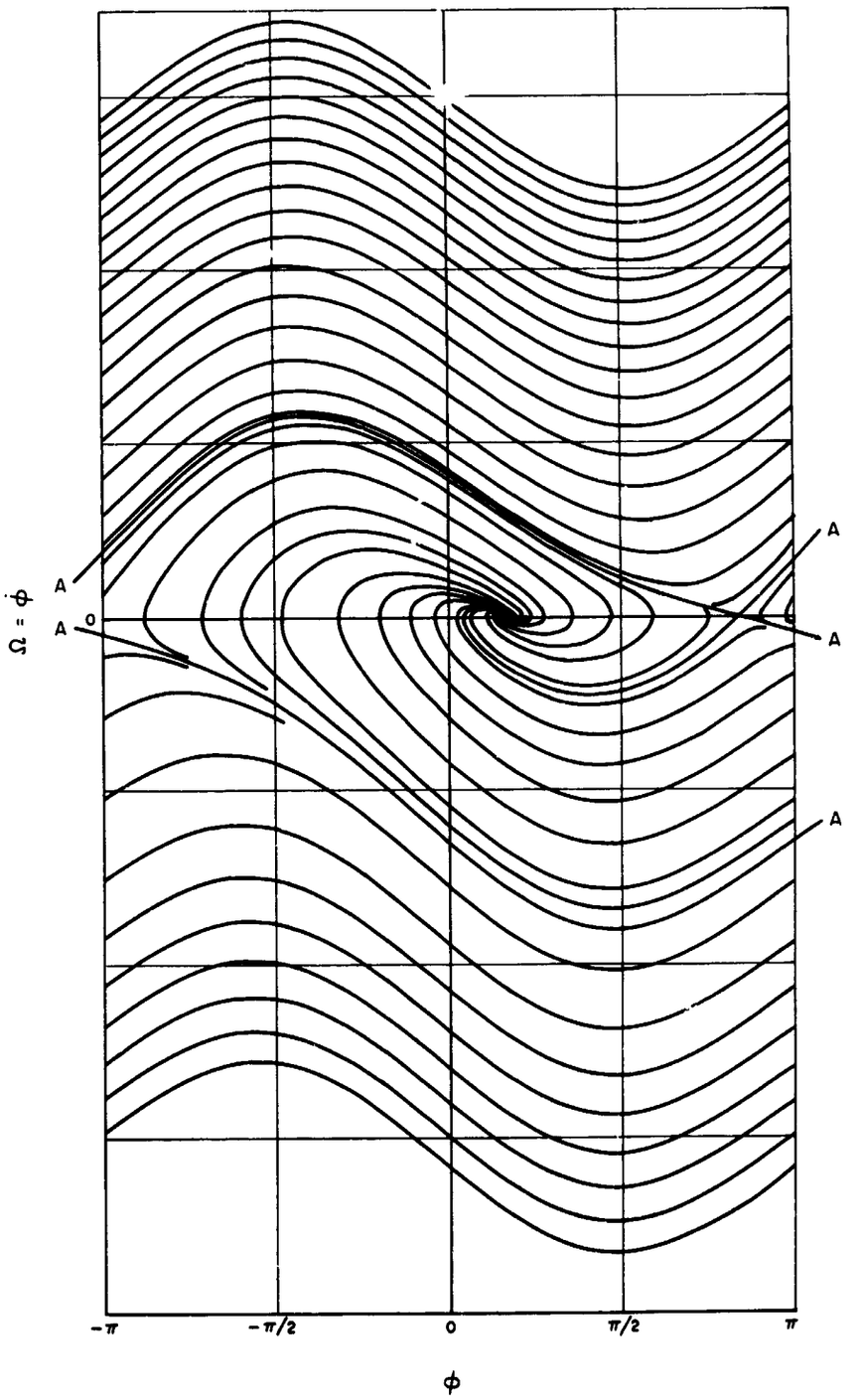


Fig. 3-5. Phase-plane trajectory of a second-order loop with perfect integrator to a doppler-rate input Λ_0 for $AK\,\tau_2^2/\tau_1 = 2$, and $\Lambda_0 = AK/2\tau_1$

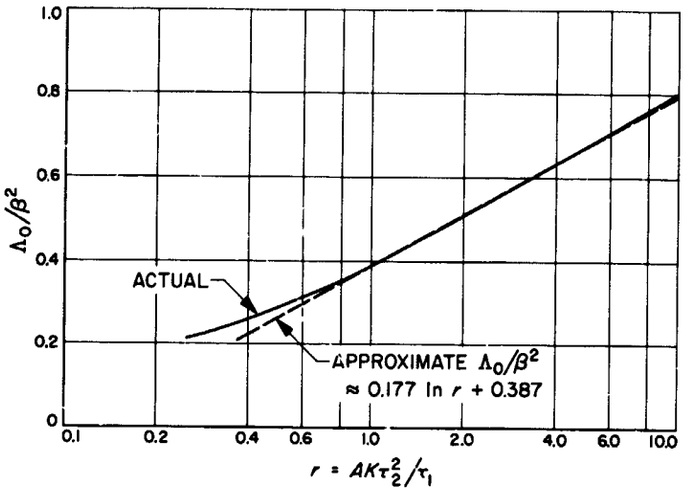


Fig. 3-6. Normalized maximum doppler rate, $k(r) = \Lambda_0 \tau_1/AK$, for which lock is guaranteed in the absence of noise, as function of the loop parameter $r = AK \tau_2^2/\tau_1$

REFERENCES FOR CHAPTER 3

1. Viterbi, A. J., "Acquisition and Tracking Behavior of Phase-Locked Loops," *Proceedings of the Symposium on Active Networks and Feedback Systems,* April 19–21, 1960, Polytechnic Institute of Brooklyn, N. Y., Vol. 10, p. 583.

2. Dye, R.A., "Phase Lock Loop Swept-Frequency Synchronization Analysis," 1965 *International Space Electronics Symposium Record,* IEEE Space Electronics and Telemetry Group, pp. 7-D1–7-D6.

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 $n_i(t)$:

$$x(t) = A\,2^{1/2} \sin\left[\omega_0 t + \theta(t)\right] + n_i(t)\,. \qquad (4\text{-}1)$$

The density of $n_i(t)$, denoted $S_{n_i n_i}(j\omega)$, 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):

$$\hat{\theta}(t) = AK\frac{F(p)}{p}\sin\phi(t) + \frac{KF(p)n(t)}{p} + \frac{K_{VCO}\left[e + n_v(t)\right]}{p}\,. \qquad (4\text{-}2)$$

Again, $\phi(t) = \theta(t) - \hat{\theta}(t)$, $K = K_1 K_m K_{VCO}$, e is a tuning bias applied to the VCO, and the term $n_v(t)$ is the internal loop-noise referred to the VCO input. The noise $n(t)$ appearing in (4-2) is

$$n(t) = n_i(t)\,2^{1/2}\cos\left[\omega_0 t + \hat{\theta}(t)\right]\,. \qquad (4\text{-}3)$$

Except in special cases, the spectral density $S_{nn}(j\omega)$ 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 $\hat{\theta}(t)$.

As we indicated in Section 3-E, the effect of e on the loop is that it changes the no-error VCO frequency from ω_0 to $\omega_{0p} = \omega_0 + K_{VCO}e$. We can thus merely redefine ω_0 to include this offset and proceed with

$$\begin{aligned}\hat{\theta}(t) &= \frac{AKF(p)}{p}\sin\phi(t) + \frac{KF(p)}{p}n(t) + \frac{K_{VCO}}{p}n_v(t)\\ &= \frac{AKF(p)}{p}\sin\left[\theta(t) - \hat{\theta}(t)\right]\\ &\quad + \frac{KF(p)}{p}n(t) + \frac{K_{VCO}}{p}n_v(t)\,. \qquad (4\text{-}4)\end{aligned}$$

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 $\hat{\theta}(t)$, 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 $\hat{\theta}(t)$; in this way, $\hat{\theta}(t)$ 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.

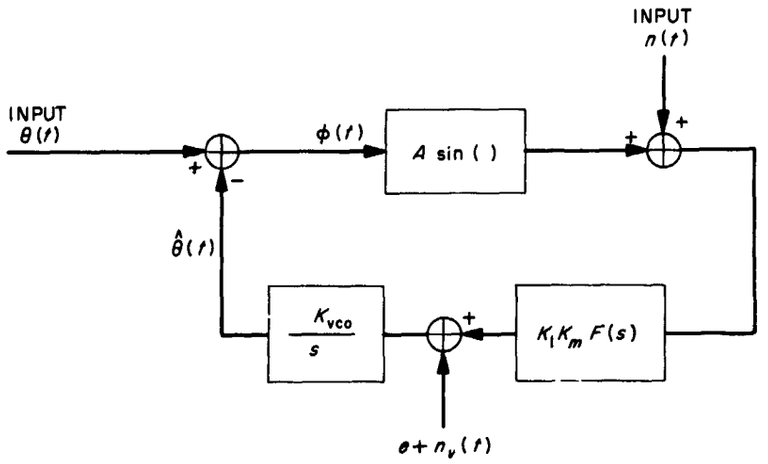


Fig. 4-1. Exact mathematical equivalent of the phase-locked loop. Sources of external, as well as internal, noises are shown; VCO tuning voltage e is also indicated.

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 2π; 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 2π. 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 2π" 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 ϕ-process is nonstationary, the ϕ(mod 2π)-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 2π, would result in a random process with the same probability density as $\phi(t)$. And so long as our sole interest lies in ϕ(mod 2π), and not in the behavior of ϕ, we may assume that Φ 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 Φ.

4-C. Spectrum of the Loop Noise

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

$$n(t) = n_i(t)\, 2^{1/2} \cos\left[\omega_0 t + \hat{\theta}(t)\right] \tag{4-5}$$

in which $\hat{\theta}(t)$ is derived, in part, from $n_i(t)$. The non-independence of $\hat{\theta}$ and n_i means that $S_{nn}(j\omega)$ is not generally the convolution of $S_{n_i n_i}(j\omega)$ with the spectrum of $2^{1/2} \cos(\omega_0 t + \hat{\theta})$. However, it has been reasoned that whenever $n_i(t)$ has a much greater bandwidth than $\hat{\theta}(t)$, and whenever $S_{n_i n_i}(j\omega)$ is symmetric[7] about ω_0, then the convolution formula is approximately correct. This gives

$$S_{nn}(j\omega) = \frac{1}{2} S_{n_i n_i}\left[j(\omega-\omega_0)\right] + \frac{1}{2} S_{n_i n_i}\left[j(\omega+\omega_0)\right]. \tag{4-6}$$

That is, the loop noise has the same spectral shape as $n_i(t)$ does, except that it is heterodyned down to baseband. The spectral shape of $n_i(t)$ usually comes from allowing white noise with density $N_0 = kT/2$ to be passed through a

[7]This condition is stronger than it needs to be. See Davenport and Root, p. 162.

predetection filter $H(s)$, whose bandwidth is, say, W_H. That is, $n_i(t)$ has a spectrum of the form

$$S_{n_i n_i}(j\omega) = N_0|H(j\omega)|^2 \tag{4-7}$$

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

$$S_{nn}(j\omega) = \frac{1}{2} N_0\{|H[j(\omega-\omega_0)]|^2 + |H[j(\omega+\omega_0)]|^2\}. \tag{4-8}$$

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

$$S_{nn}(0) = N_0|H(j\omega_0)|^2. \tag{4-9}$$

(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.)

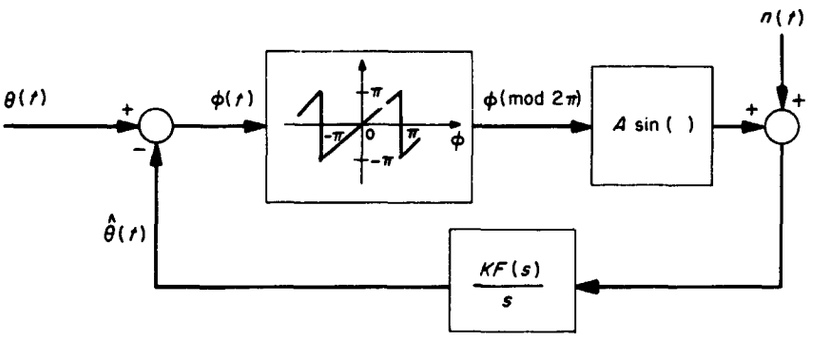


Fig. 4-2. Equivalent exact mathematical model of phase-locked loop, with explicit reduction of ϕ (mod 2π). The VCO has been replaced by K_{VCO}/s (tuning bias and VCO noise omitted).

REFERENCES FOR CHAPTER 4

1. Jaffe, R. M., and Rechtin, E., "Design and Performance of Phase-Lock Circuits Capable of Near-Optimum Performance Over a Wide Range of Input Signal and Noise Levels," *IRE Transactions on Information Theory*, Vol. IT-1, pp. 66–76, March 1955.

2. Davenport, W. B., Jr., and Root, W. L., *Random Signals and Noise*, McGraw-Hill Book Co., Inc., New York, N. Y., 1958.

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

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

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

$$L(s) = \frac{AKF(s)}{s + AKF(s)}$$

$$F(s) = \frac{sL(s)}{AK[1 - L(s)]} . \tag{5-3}$$

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

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

where we use L^2 to denote $| L(j\omega) |^2_{\max}$.

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 $n_v(t)$, 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 Σ^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):

$$\delta^2 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} | 1 - L(j\omega) |^2 \, S_{\psi\psi}(j\omega) \, d\omega$$

$$\sigma^2 = \frac{N_0 W_L L^2}{A^2} = \frac{N_0 B_L L^2}{A^2} . \tag{5-7}$$

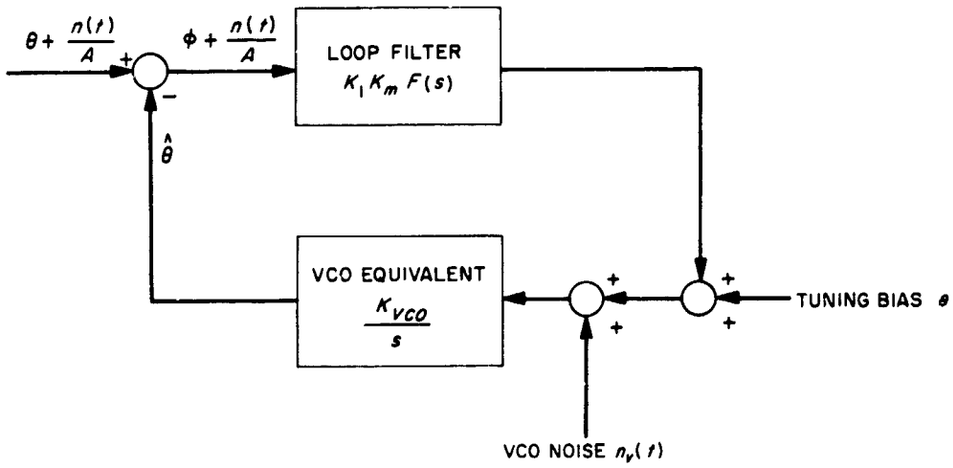


Fig. 5-1. The linearized model of the phase-locked loop

The expression for σ^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_0 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 (5-4)! Therefore σ^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

$$\sigma^2 = \frac{N_0 w_L}{A^2} = \frac{N_0 b_L}{A^2}, \tag{5-8}$$

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

$$\epsilon_T^2 = \int_0^\infty \mu^2(t)\, dt. \tag{5-9}$$

It is convenient to include in $d(t)$ the initial phase offset θ_0. There is usually a complete lack of knowledge about θ_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 $d_1(t)$) 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

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

Note the resemblance this bears to δ^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 $S_{\psi\psi}(s)$ in (5-7).

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

5-B. 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

$$w_L = W_L = \frac{1}{2} AK \qquad \text{or} \qquad b_L = B_L = \frac{AK}{4} . \quad (5\text{-}11)$$

This relation enables one to express the loop transfer function as

$$L(s) = \frac{2W_L}{s + 2W_L} . \quad (5\text{-}12)$$

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)

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

$$L(s) = \frac{1 + \tau_2 s}{1 + (\tau_2 + 1/AK)s + (\tau_1/AK)s^2} \quad (5\text{-}13)$$

(again we have normalized $F(0) = 1$, as agreed to in Chapter 3). In the usual case, $F(s)$ is designed with $\tau_1 \gg \tau_2$, to make $F(s)$ appear to have a pole at the origin.

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

$$r = \frac{AK\tau_2^2}{\tau_1} . \quad (5\text{-}14)$$

The value of ω producing maximum loop response, $L^2 = |L(j\omega)|^2_{\max}$, can be found by differentiation; it is

$$\omega_{\max} = \frac{1}{\tau_2}\left\{\left[\left(1 - \frac{\tau_2}{\tau_1}\right)\left(1 + \frac{\tau_2}{\tau_1} + 2r\right)\right]^{1/2} - 1\right\}^{1/2}$$

$$\approx \frac{1}{\tau_2}\left[(1 + 2r)^{1/2} - 1\right]^{1/2} . \quad (5\text{-}15)$$

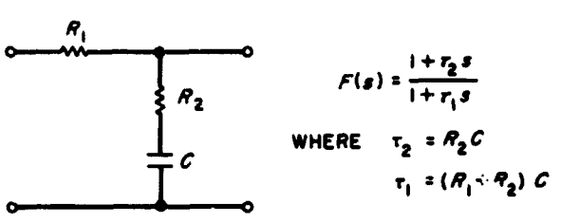


Fig. 5-2. Passive integrator loop filter

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

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

$$\approx \frac{r^2}{2(1+2r)^{1/2} + r^2 - 2r - 2} . \quad (5\text{-}16)$$

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,

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

$$\approx \frac{(r + 1)\left[r^2 - 2r - 2 + 2(1 + 2r)^{1/2}\right]}{2\tau_2 r^2} , \quad (5\text{-}17)$$

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

$$w_L = \frac{r + 1}{2\tau_2 (1 + \tau_2/r\tau_1)} \approx \frac{r + 1}{2\tau_2} . \quad (5\text{-}18)$$

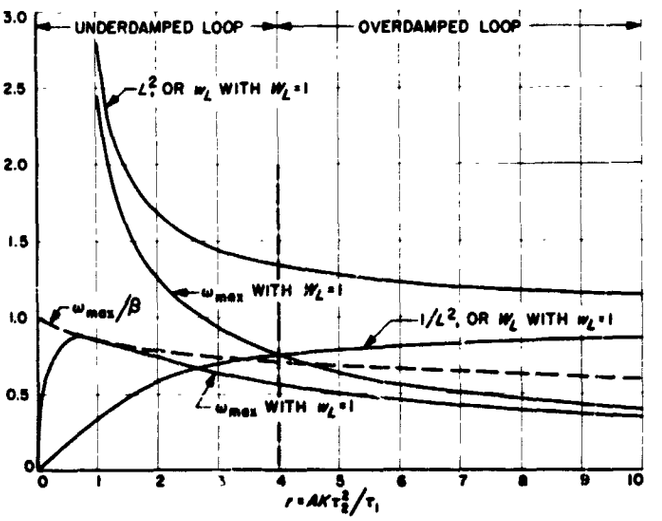


Fig. 5-3. Variation of maximum loop response L^2, noise bandwidth W_L, fiducial bandwidth w_L, and frequency at maximum loop response, for second-order phase-locked loop, as a function of the parameter $r = AK\tau_2^2/\tau_1$

The loop transfer function is then easily expressed in fiducial bandwidth:

$$L(s) \approx \frac{1 + \left(\frac{r+1}{2w_L}\right) s}{1 + \left(\frac{r+1}{2w_L}\right) s + \left(\frac{r + \frac{\tau_1}{\tau_2}}{2w_L r^{1/2}}\right)^2 s^2} . \qquad (5\text{-}19)$$

From (5-13), one may compute the system damping behavior. For $r < 4$, the roots of the system are underdamped, with damping factor ζ and loop natural frequency β given by

$$\zeta = \frac{1}{2}(1 + \tau_2/r\tau_1)\, r^{1/2} \approx \frac{r^{1/2}}{2}$$

$$\beta = \left(\frac{AK}{\tau_1}\right)^{1/2} = \frac{r^{1/2}}{\tau_2} . \qquad (5\text{-}20)$$

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

$$F(s) = \frac{1 + \tau_2 s}{\tau_1 s} . \qquad (5\text{-}21)$$

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 $\tau_2/r\tau_1$ term is dropped. This indicates that whenever $r\tau_1 \gg \tau_2$, 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 loop 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 Σ^2 thus would require a time-varying loop filter.

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

$$\Sigma_T^2 = \lambda^2 \epsilon_T^2 + \delta^2 + \sigma^2 \qquad (5\text{-}22)$$

in which λ^2 is a Lagrange multiplier, a design parameter related to the bandwidth of the loop.

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

$$L_{opt}(s) = 1 - \frac{N_0^{1/2}/A}{[S(s)]^+} \qquad (5\text{-}23)$$

$$S(s) = \lambda^2 E\,[D(s)\,D(-s)] + S_{\psi\psi}(s) + N_0/A^2 .$$

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

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 $n_v(t)$ 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, $n_v(t)$ 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:

$$\hat{\theta}(t) = L(p)\left[\theta(t) + \frac{n(t)}{A}\right] + \left[\frac{1 - L(p)}{p}\right] K_{VCO}\, n_v(t)\,. \tag{5-24}$$

There are many factors that contribute to the spectral makeup of $n_v(t)$. 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 $K_{VCO}n_v(t)$ can be approximated by

$$K_{VCO}^2\, S_{n_v n_v}(j\omega) = N_{0v} + 2\pi N_{1v}/|\,\omega\,|\,. \tag{5-25}$$

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 ϵ. But if $[1 - L(s)]/s$ has a zero at the origin, this ϵ need not be known.

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

$$\sigma_{VCO}^2 = \frac{1}{\pi}\int_0^{+\infty} \frac{|\,1 - L(j\omega)\,|^2}{\omega^2}\left[N_{0v} + 2\pi N_{1v}/\omega\right] d\omega\,. \tag{5-26}$$

At this point, for our treatment, we shall assume that $L(s)$ is the passive-integrator loop of (5-13), and again set $r = AK\tau_2^2/\tau_1$. The total phase error due to input and VCO noise, in terms of the fiducial bandwidth w_L, is

$$\sigma^2 = \frac{N_0 w_L}{A^2} + \left(\frac{r+1}{4r}\right)\frac{N_{1v}}{w_L} + g(r)\,\frac{N_{1v}}{w_L^2}\,. \tag{5-27}$$

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

$$g(r) = \begin{cases} \dfrac{(r+1)^2}{4[r^3(r-4)]^{1/2}} \ln\left[\dfrac{r - 2 + \sqrt{r(r-4)}}{r - 2 - \sqrt{r(r-4)}}\right] & \text{for } r > 4 \\ \dfrac{25}{16} & \text{for } r = 4 \\ \dfrac{(r+1)^2}{2[r^3(4-r)]^{1/2}}\left[\dfrac{\pi}{2} - \tan^{-1}\dfrac{r-2}{\sqrt{r(4-r)}}\right] & \text{for } r < 4. \end{cases} \tag{5-28}$$

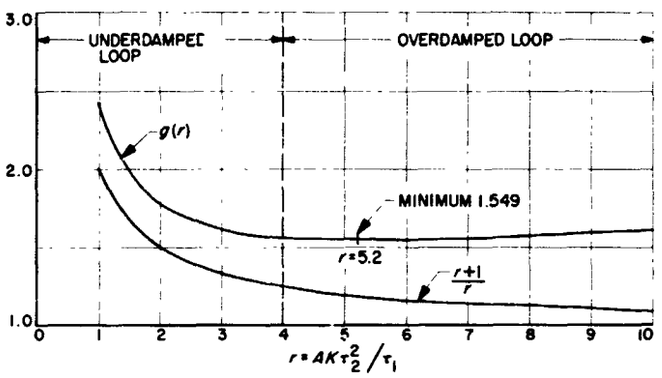


Fig. 5-4. Factors governing relative contributions of VCO noise to output phase noise

1. Optimization of w_L and r

Depending on the values of N_0/A^2, N_{0v}, and N_{1v}, there are optimum choices for both w_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_{0v}, 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 σ^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 σ_{min}^2 and σ^2 for any r between 5 and 10 is almost inconsequential insofar as σ^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_{opt}, regardless of the other parameters. (This reasoning alleviates the necessity for differentiating σ^2 to find the exact value of r_{opt}.)

We can now use this value of r, differentiate σ^2 with respect to w_L, equate to zero, and solve for w_L to find its best value. Straightforwardly, w_L is the solution to the equation

$$w_L^3 - \left(\frac{r+1}{4r}\right)\left(\frac{A^2 N_{0v}}{N_0}\right) w_L - 2g(r)\left(\frac{A^2 N_{1v}}{N_0}\right) = 0\,. \tag{5-29}$$

2. An Example

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

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

$$N_{0r} = 0 \qquad (5\text{-}30)$$

$$N_{1r} = 0.08\,.$$

The first of these is typical for an Earth–spacecraft link at a distance of about 3 million kilometers, while the values of N_{0r} and N_{1r} have been approximated from plots of σ^2 under very high signal-to-noise conditions.

The optimum value of w_L is then

$$w_L = \sqrt[3]{3.098\left(\frac{A^2 N_{1r}}{kT}\right)}$$

$$= 26 \text{ cps}$$

and the optimum loop transfer function is

$$L_{\text{opt}}(s) = \frac{1 + 0.154s}{1 + 0.154s + (3.38 \times 10^{-3})\, s^2}\,.$$

These parameters also produce a phase deviation of

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

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

1. Jaffee, R. M., and Rechtin, E., "Design and Performance of Phase-Locked Circuits Capable of Near-Optimum Performance Over a Wide Range of Input Signal and Noise Levels," *IRE Transactions on Information Theory*, Vol. IT-1, pp. 66–76, March 1955.

2. Bode, H. W., and Shannon, C. E., "A Simplified Derivation of Linear Least-Square Smoothing and Prediction Theory," *Proceedings of the IRE*, Vol. 38, pp. 417–426, April 1950.

3. Wiener, N., *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, John Wiley, New York, 1949.

4. Yavits, M. C., and Jackson, J. L., "Linear Filter Optimization With Game-Theoretic Considerations," *1955 IRE National Convention Record*, Part 4, pp. 193–199.

5. Malling, L., "Phase-Stable Oscillators," *Space Communications*, Chap. 13, McGraw-Hill, New York, 1963.

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

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

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

$$L_{opt}(s) = 1 - \frac{s^N}{\left[(-1)^N s^{2N} + \left(\frac{A^2\lambda^2}{N_0}\right) E\,[Q(s)Q(-s)]\right]^+}. \tag{6-2}$$

In the examples worked by Jaffee and Rechtin, the θ_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

$$\left[-s^2 + \frac{\lambda^2 A^2}{N_0}\frac{\pi^2}{3}\right]^+ = s + \frac{\pi\lambda A}{(3N_0)^{1/2}}. \tag{6-3}$$

Thus, the optimum filter is given by

$$L_{opt}(s) = \frac{\lambda A\pi}{(3N_0)^{1/2}}\left[\frac{1}{s + \frac{\lambda A\pi}{(3N_0)^{1/2}}}\right]. \tag{6-4}$$

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

$$w_L = W_L = \frac{\lambda A\pi}{2(3N_0)^{1/2}}. \tag{6-5}$$

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

$$\begin{aligned} L_{opt}(s) &= \frac{2W_L}{s + 2W_L} \\ F(s) &= 1 \\ K &= 2W_L/A\,. \end{aligned} \tag{6-6}$$

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

The next example of interest is the optimization of a loop with frequency offset Ω_0:

$$d(t) = \theta_0 + \Omega_0 t \tag{6-7}$$

where again θ_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

$$\left[s^4 - \frac{\lambda^2 A^2\pi^2}{3N_0}s^2 + \frac{\lambda^2 A^2\Omega_0^2}{N_0}\right]^+ = s^2 + \left(\frac{A^2\lambda^2\pi^2}{3N_0} + \frac{2\lambda A\Omega_0}{N_0^{1/2}}\right)^{1/2} s + \frac{\lambda A\Omega_0}{N_0^{1/2}}. \tag{6-8}$$

The loop natural frequency is thus

$$\beta^2 = \frac{A\lambda\Omega_0}{N_0^{1/2}} \tag{6-9}$$

in which case the optimum filtering function is given by

$$L_{opt}(s) = \frac{s\left[2\beta^2 + \frac{\pi^2\beta^4}{3\Omega_0^2}\right]^{1/2} + \beta^2}{s^2 + \left[2\beta^2 + \frac{\pi^2\beta^4}{3\Omega_0^2}\right]^{1/2} s + \beta^2}. \tag{6-10}$$

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

$$\begin{aligned} \tau_2 &= \frac{1}{\beta}\left[2 + \frac{\pi^2\beta^2}{3\Omega_0^2}\right]^{1/2} = \frac{\pi r}{\Omega_0\left[3r(r-2)\right]^{1/2}} \\ \frac{AK}{\tau_1} &= \beta^2 = \frac{3\Omega_0^2}{\pi^2}(r-2) \\ r &= 2 + \frac{\pi^2\beta^2}{3\Omega_0^2} \\ w_L &= \frac{(r+1)\Omega_0}{2\pi r}\left[3r(r-2)\right]^{1/2}. \end{aligned} \tag{6-11}$$

We have thus translated the effect of the Lagrange multiplier implicit in β 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 θ_0 was a uniformly distributed random variable.

The Jaffee and Rechtin example using $\theta_0 = 0$ produces a result somewhat different from that above:

$$\left.\begin{aligned} \tau_2 &= 2^{1/2}/\beta \\ \frac{AK}{\tau_1} &= \beta^2 \\ r &= 2 \quad (\text{so } \zeta = 0.707) \\ w_L &= \frac{3\beta}{2(2^{1/2})} \end{aligned}\right\} \text{Jaffee–Rechtin.} \tag{6-12}$$

1. Choice of Parameters

In most cases, Ω_0 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 Ω_0 must be treated as a random variable. An examination of the average transient error in (5-9) reveals that ϵ_T^2 is minimized whenever $E(\Omega_0) = 0$ (that is, the loop should be initially tuned to the expected incoming frequency) and that the Ω_0 of our previous calculations is replaced, in this case, by the rms frequency offset, $[E(\Omega_0^2)]^{1/2}$.

Once suitable values of $A/N_0^{1/2}$, w_L, and Ω_0 have been established, β (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):

$$F_{opt}(s) = \frac{1 + \tau_2 s}{\tau_1 s}. \tag{6-13}$$

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 $r\tau_1 \gg \tau_2$,

$$F_{opt}(s) \approx \frac{1 + \tau_2 s}{1 + \tau_1 s} \tag{6-14}$$

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

$$\phi_{ss} = \frac{\Omega_0}{AK} \approx \frac{\Omega_0}{w_L}\left(\frac{r+1}{2r}\right)\left(\frac{\tau_2}{\tau_1}\right) \tag{6-15}$$

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 ϕ_{ss}.

As a further consideration, one cannot expect a very good lock-in behavior when Ω_0 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 e) 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[8] loop passband, i.e., when $\Omega_0 = 2\pi b_L$. Upon inserting this condition into (6-11), the proper value of r can be computed numerically:

$$\begin{aligned} r &= 2.28245 \\ \zeta &= 0.755. \end{aligned} \tag{6-16}$$

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

$$\begin{aligned} \tau_2 &= 1.643/w_L \\ \tau_1/AK &= \frac{1.180}{w_L^2}. \end{aligned} \tag{6-17}$$

and the corresponding optimum loop transfer function is given by

$$L_{opt}(s) = \frac{1 + (1.643/w_L)s}{1 + (1.643/w_L)s + (1.18/w_L^2)s^2}. \tag{6-18}$$

[8]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.

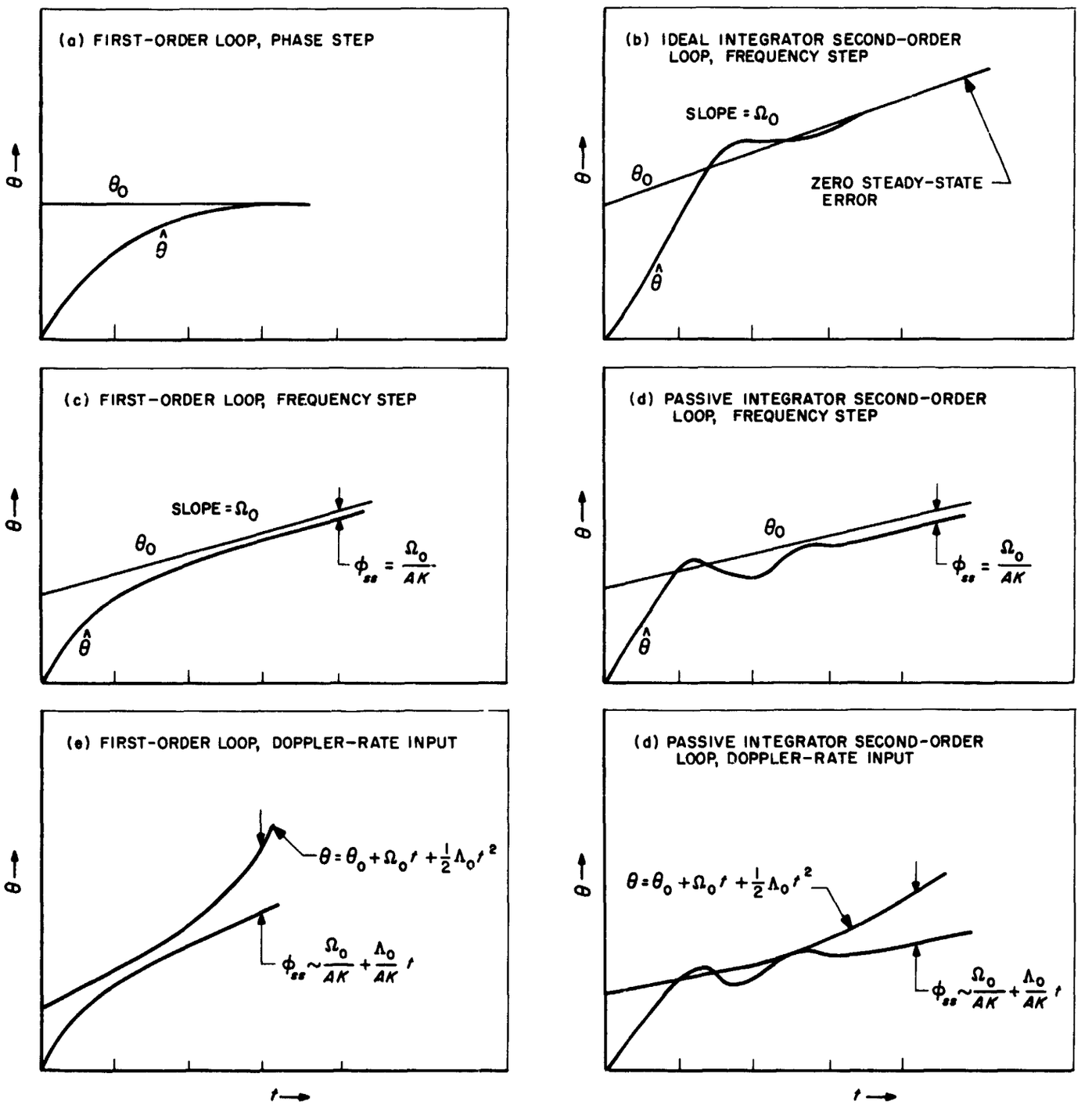


Fig. 6-1. Response of first- and second-order loops to various inputs

2. Evaluation of Transient Error

The total transient error is given by the expression

$$\epsilon_T^2 = \frac{1}{2\pi j}\int_{-j\infty}^{+j\infty}\left(-\frac{\pi^2}{3s^2}+\frac{\Omega_0^2}{s^4}\right) \times [1 - L_{\text{opt}}(s)][1 - L_{\text{opt}}(-s)]\,ds\,. \tag{6-19}$$

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

$$\epsilon_T^2 = \frac{\tau_1\tau_2\Omega_0^2}{2AK\tau}\left(1+\frac{\pi^2 AK}{3\Omega_0^2\tau_1}\right)\,. \tag{6-20}$$

For the optimum loop, then

$$\epsilon_{T\,(\text{opt})}^2 = \frac{\pi^3(r-1)}{6\Omega_0\,[3r\,(r-2)^3]^{1/2}}\,. \tag{6-21}$$

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

$$\epsilon_{T\,(\text{opt})}^2 = 5.37/w_L\,. \tag{6-22}$$

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

$$\epsilon_{T(\text{JR})}^2 = 5.4/w_L\,. \tag{6-23}$$

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

If we assume that there is a slow doppler rate Λ_0 (rad/sec^2) 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

$$\phi_{ss(\text{dopp rate})} = \frac{\Lambda_0}{AK}\,t \approx \frac{\Lambda_0(r+1)^2}{4r\,w_L^2}\left(\frac{t}{\tau_1}\right)\,. \tag{6-24}$$

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.,

$$\phi_{ss(\text{dopp rate})} = \frac{\Lambda_0\tau_1}{AK} = \frac{\Lambda_0(r+1)^2}{4r\,w_L^2}\,. \tag{6-25}$$

There is naturally a great resemblance here, and at $t = \tau_1$, (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 τ_1 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-C3, 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 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 $S_{n_{1v}n_{1v}}(s)$ is a rational spectrum. However, for the case we have considered, the VCO (input) noise has a $1/f$ term, or more precisely in the s-plane, $N_{1v}/K_{\text{VCO}}^2(-s^2)^{1/2}$, and (5-22) 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 Σ_T^2 (except distortion due to carrier modulation). Hence

$$\Sigma_T^2 = \lambda^2\epsilon_T^2 + \sigma^2$$
$$= \frac{\lambda^2\tau_1\tau_2\Omega_0^2}{2AKr}\left(1 + \frac{\pi^2AK}{3\Omega_0^2\tau_1}\right) + \frac{N_0w_L}{A^2}$$
$$+ \frac{(r+1)N_{0v}}{4r\,w_L} + \frac{g(r)N_{1v}}{w_L^2}\,. \qquad (6\text{-}26)$$

Applying the philosophy that Ω_0 be set equal to πw_L (so that lock-in proceeds optimally once the offset frequency enters the loop passband), we reduce this to

$$\Sigma_T^2 = \frac{N_0w_L}{A^2} + \frac{\lambda^2\pi^2(r+1)^3}{16r^2w_L}$$
$$+ \frac{(r+1)}{4rw_L}\left(\frac{\lambda^2\pi^2}{3} + N_{0v}\right) + \frac{g(r)N_{1v}}{w_L^2}\,. \qquad (6\text{-}27)$$

The value of λ, of course, is a parameter left unspecified. Previously (in 6-9), we related it to the loop natural frequency β (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

$$\Sigma_T^2 = \frac{N_0\,w_L}{A^2}\left(1 + \frac{1}{1+r} + \frac{4r}{3(1+r)^3}\right)$$
$$+ \frac{(r+1\,N_{0r}}{4r\,w_L} + \frac{g\,(r)\,N_{1r}}{w_L^2} \qquad (6\text{-}28)$$

With this particular evaluation of λ, 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 Σ_T^2 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 Σ_T^2.

REFERENCE FOR CHAPTER 6

1. Jaffee, R. M., and Rechtin, E., "Design and Performance of Phase-Locked Circuits Capable of Near-Optimum Performance Over a Wide Range of Input Signals and Noise Levels," *IRE Transactions on Information Theory*, Vol. IT-1, pp. 66–76, March 1955.

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 $2^{1/2} A(t) \cos [\omega_0 t + \theta(t)] + n_0(t)$ enters a variable-gain amplifier whose gain is a function of a feedback control-voltage $c(t)$. The output (amplitude-stabilized), appearing as $2^{1/2} A \cos [\omega_0 t + \theta(t)] + n_i(t)$, 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_G 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_G 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

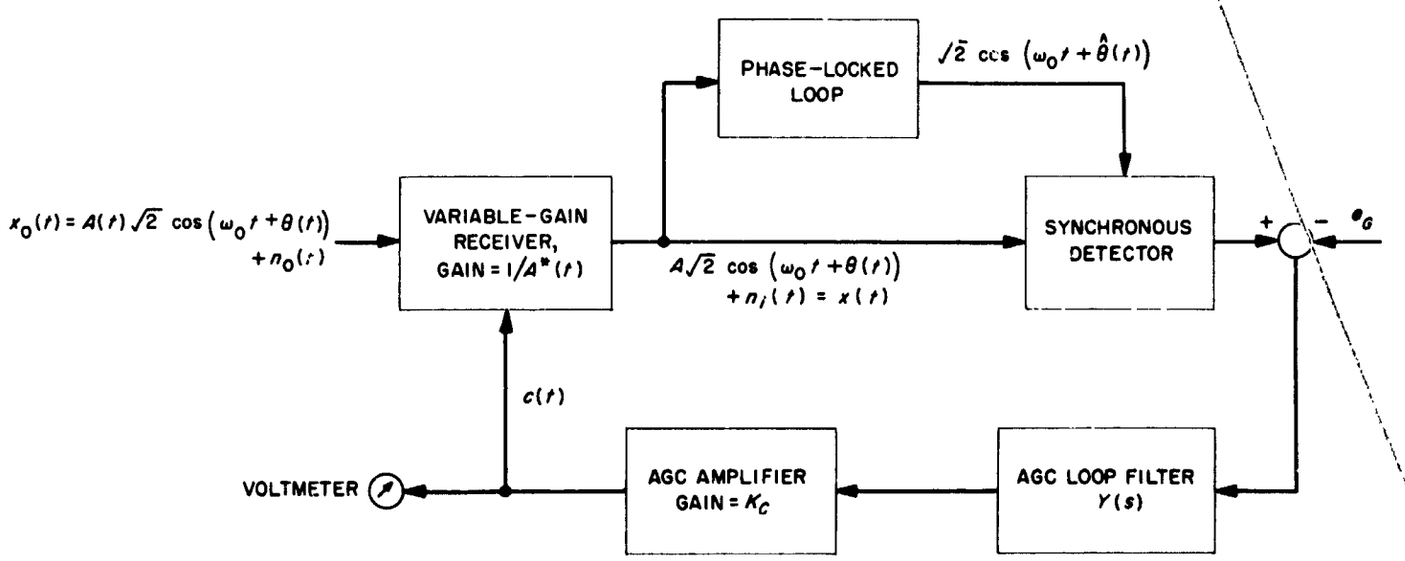


Fig. 7-1. A synchronous-detector AGC loop, using a phase-locked loop to provide a coherent reference. (The actual phase-locked loop may be part of the gain-controlled receiver.)

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

$$z_{AGC}(t) = -e_G + AK_D \cos\phi(t) + K_D n_i(t)$$
$$= -e_G + \frac{A(t)}{A^*(t)} K_D \cos\phi(t) + \frac{K_D n_0(t)}{A^*(t)} .$$

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[9]

$$-e_G + \frac{A(t)}{A^*(t)} K_D \cos\phi \approx e_G \ln\left[\frac{A(t) K_D \cos\phi}{A^*(t) e_G}\right]$$
$$\approx e_G \ln A(t) - e_G \ln\left[\frac{A^*(t) e_G}{K_D}\right] + e_G \ln\cos\phi(t) . \quad (7\text{-}2)$$

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:

$$\frac{e_G A^*(t)}{K_D} = 10^{[K_R + K_A c(t) + \cdots]/20}$$
$$20 \log\left[\frac{e_G A^*(t)}{K_D}\right] = K_R + K_A c(t) + \cdots . \quad (7\text{-}3)$$

A characteristic with only the first two terms is nearly valid for many voltage-controlled-gain amplifiers, although K_R and K_A depend on the operating point to a certain degree. The quantity $A^*(t)e_G/K_D$ 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_A db/v 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[10]

$$K_{rec} = K_R + 20 \log\left(\frac{K_D}{e_G}\right) .$$

Now define the quantities[11]

$$a(t) = 20 \log A(t)$$
$$a^*(t) = 20 \log\left[\frac{e_G A^*(t)}{K_D}\right] . \quad (7\text{-}4)$$

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

$$c(t) = K_C Y(p)\left[\left(\frac{e_G}{20 \log e}\right)[a(t) - a^*(t)] + \left(\frac{e_G}{\log e}\right)\log\cos\phi(t) + \frac{K_D n_0(t)}{A^*(t)}\right]$$
$$= \frac{a^*(t) - K_R}{K_A} . \quad (7\text{-}5)$$

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

$$K_{AGC} = \frac{K_A e_G K_C}{20 \log e}$$
$$C(s) = \frac{K_{AGC} Y(s)}{[1 + K_{AGC} Y(s)] K_A} . \quad (7\text{-}6)$$

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

$$a^*(t) = K_A C(p) \times \left[a(t) + \left(\frac{20 \log e}{e_G}\right) K_D n_i(t) + 20 \log\cos\phi(t)\right] + [1 - K_A C(p)] K_R . \quad (7\text{-}7)$$

[9]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_G 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 dbw by the term 10 log R_{in}: 10 log $P = a(t) - 10 \log R_{in}$.

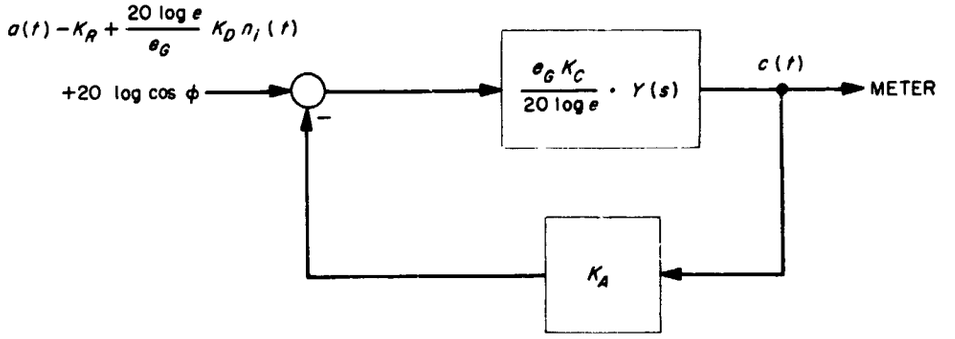


Fig. 7-2. Equivalent diagram of an AGC loop. Linearized analysis follows the assumptions that $A^*(t)$ is exponentially related to $c(t)$ and that $a(t) \approx a^*(t)$. The input is $a(t) = 20 \log A(t)$. Adjusted loop gain $a^*(t) = 20 \log [A^*(t)\, e_G/K_D]$

The AGC voltage is thus related to the input signal level by

$$c(t) = C(p)\left[a(t) + \left(\frac{20 \log e}{e_G}\right) K_D\, n_i(t) + 20 \log \cos \phi(t) - K_R\right]. \tag{7-8}$$

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 $C(s)$,

$$\begin{aligned} E\,[a^*(t)] &= K_A C(0) E[a(t)] - K_A C(0)(10 \log e)\,\sigma^2 \\ &\quad + [1 - K_A C(0)] K_R \\ \text{var}\,[a^*(t)] &= \left[\left(\frac{20 \log e}{e_G}\right)^2 K_D^2 \left(\frac{N_0 w_C}{A^2}\right) A^2 \right. \\ &\quad \left. + 2(10 \log e)^2\, \sigma^4 \left(\frac{w_C}{w_L}\right)\right] C^2(0) K_A^2. \end{aligned} \tag{7-9}$$

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_0 is the spectral density of the stabilized noise $n_i(t)$, and not the input noise $n_0(t)$; for this reason, it is referred in (7-8) to A^2, the stabilized signal level, as N_0/A^2 (which is the same as $N_{00}/A^2(t)$ at the receiver input, however). As a fair approximation, $A = e_G/K_D$ can be used in evaluating the variance of $a^*(t)$ above. When K_{AGC} is high, $C(0) \sim 1/K_A$.

The fluctuation of $a^*(t)$ depends not only on the input signal-to-noise ratios in the phase-locked loop bandwidth w_L and in the AGC loop bandwidth w_C, 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[12] strength a:

$$\begin{aligned} a &= K_R + \frac{1}{C(0)} c + (10 \log e)\,\sigma^2 \\ &= K_{rec} + 20 \log(e_G/K_D) + \left(\frac{1}{C(0)}\right) c + (10 \log e)\,\sigma^2. \end{aligned} \tag{7-10}$$

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 (7-9), because, as shown in (7-8), $c(t)$ and $a^*(t)/K_A$ have the same fluctuations due to noise.

[12] 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.

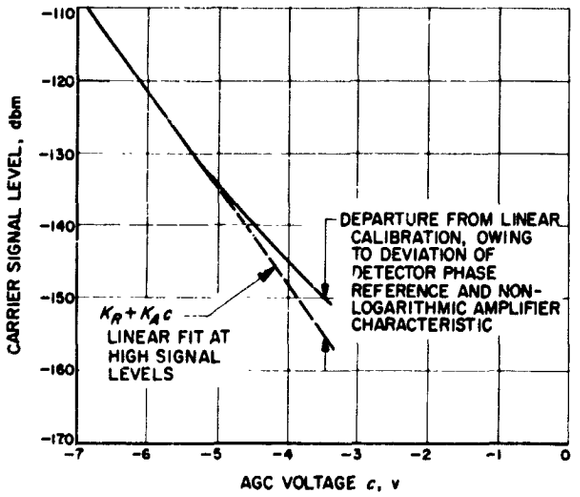


Fig. 7-3. Measured AGC curve showing departure from linear behavior

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

$$\begin{aligned} A(t) &= A_d(t)A_\psi(t) \\ a(t) &= a_d(t) + a_\psi(t) \end{aligned} \tag{7-11}$$

where we have put

$$\begin{aligned} a_d(t) &= 20 \log A_d(t) \\ a_\psi(t) &= 20 \log A_\psi(t) \end{aligned} \tag{7-12}$$

and in which we have designated $A_\psi(t)$ and $A_d(t)$ as the stationary and nonstationary parts of $A(t)$, respectively. For convenience, $a_\psi(t)$ 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

$$C_{\text{opt}}(s) = \frac{1}{[S'(s)]^+} \cdot \left[\frac{\lambda_1^2 E[\tilde{a}_d(s)\tilde{a}_d(-s)] + S_{a_\psi a_\psi}(s)}{[S'(s)]^-} \right]_{pr} \tag{7-13}$$

$$S'(s) = \lambda_1^2 E[\tilde{a}_d(s)\tilde{a}_d(-s)] + S_{a_\psi a_\psi}(s) + (20 \log e)^2 \left(\frac{N_0}{A^2}\right) + (10 \log e)^2 S_{\phi^2\phi^2}(s) .$$

For convenience, we have approximated $e_G/K_D = A$ in the equation above and set $\tilde{a}_d(s) = \mathcal{L}[a_d(t)]$.

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

The latter can be computed as $\mathcal{L}\mathcal{F}^{-1}[\ ]$ (i.e., the Laplace transform of the inverse Fourier transform of the quantity). An equivalent method for computing it 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 $C_{\text{opt}}(s)$ without resorting to solution of (7-13). For example, if we are concerned with gain-tracking an input of the form $a(t) = a_d(t) = \lambda_n t^n/n!$, then by a reasoning similar to that in Section 3-B, we find (neglecting the term due to noise)

$$E[a_d(t) - a^*(t)]_{ss} \sim \begin{cases} \dfrac{\lambda_n p(0) t^{n-l}}{K_{AGC} q(0)(n-l)!} & \text{for } n \geq l \\ 0 & \text{for } n < l. \end{cases} \tag{7-14}$$

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

$$Y(s) = \frac{q(s)}{s^l p(s)}, \quad q(0) \neq 0, \quad p(0) \neq 0. \tag{7-15}$$

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

$$Y(s) = \frac{1}{1 + \tau_{AGC}s} \tag{7-16}$$

in which case there will be a constant gain-tracking transient error

$$E\left[a_d(t) - a^*(t)\right]_{ss} = \frac{a}{K_{AGC}} . \tag{7-17}$$

But if $a_d(t) = a_1 t$, the loop filter (7-16) produces a steady-state transient error that grows linearly in time:

$$E\left[a_d(t) - a^*(t)\right]_{ss} \to \frac{a_1}{K_{AGC}} t \approx \frac{a_1}{2w_c}\left(\frac{t}{\tau_{AGC}}\right). \tag{7-18}$$

The AGC loop bandwidth in both cases is

$$w_c = \frac{1 + K_{AGC}}{2\tau_{AGC}} \approx \frac{K_{AGC}}{2\tau_{AGC}} . \tag{7-19}$$

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 $a_1 t$ 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

$$E\left[a(t) - a^*(t)\right] = \frac{a_1\tau_{AGC}}{K_{AGC}} = \frac{a_1}{2w_c} . \tag{7-20}$$

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

1. Victor, W. K., and Brockman, M. H., "The Application of Linear Servo Theory to the Design of AGC Loop," *Proceedings of the IRE*, Vol. 48, No. 2, pp. 234–238, February 1960.

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 $2^{1/2} A(t) \cos[\omega_0 t + \theta(t)] + n_0(t)$, where $A(t)$ is a slowly varying rms signal amplitude and $n_0(t)$ is a wide-band noise with density N_{00}. The input is mixed with a multiplied-up version of the VCO output, so the first IF output process is

$$K_{I1}\left[2^{1/2} A(t)\cos(\omega_1 t + \theta - \hat{\theta}) + n_1(t)\right].$$

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

$$f_1 = \frac{\omega_1}{2\pi} = \frac{2\pi}{\omega_0 - \omega_{h1}}.$$

(We assume $\omega_0 > \omega_{h1}$ 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 $\cos(\omega_{h2}(t) - \theta_1)$, so the output of the second IF amplifier is

$$K_{I1}K_{I2}\left\{2^{1/2} A(t)\cos\left[(\omega_1 - \omega_{h2})t + \theta - \hat{\theta} + \theta_1\right] + n_2(t)\right\}.$$

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

$$f_2 = \frac{\omega_2}{2\pi} = \frac{\omega_1 - \omega_{h2}}{2\pi}.$$

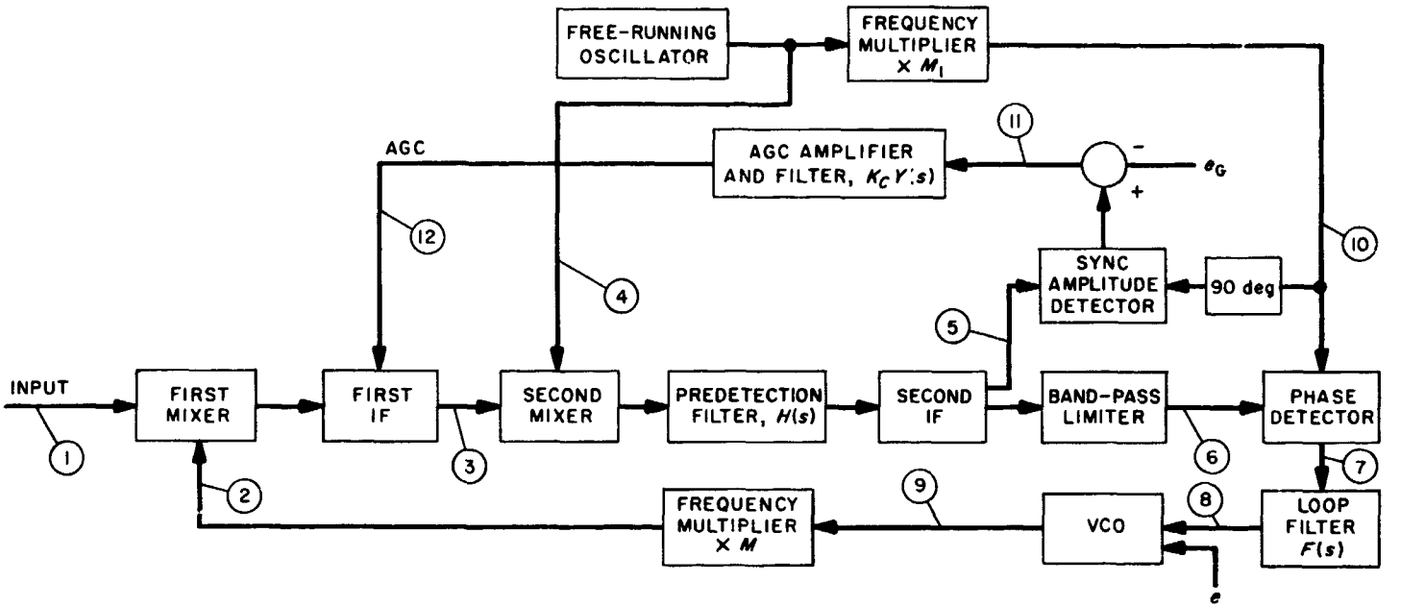
(In general, f_2 may take a negative sign when $\omega_{h2} > \omega_1$. This is a purely mathematical distinction, and need cause us no great concern here.)

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, $K_{I1}K_{I2} = 1/A^*(t)$, so the second IF output is, in our previous notation,

$$x(t) = 2^{1/2} A\cos(\omega_2 t + \theta - \hat{\theta} + \theta_1) + n_i(t).$$

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 $2(2l/\pi)^2$ lies in the frequency zone about f_2. A signal component is present whose mean-square value is, let us say, $2(2l\alpha/\pi)^2$; the factor α^2 is then a signal-power suppression factor, with $0 \leq \alpha \leq 1$. The remainder of the limiter output in the IF



(1) $x_0(t) = \sqrt{2}\, A(t) \cos\left[\omega_0 t + \theta(t)\right] + n_0(t)$

(2) $v(t) \sim \sqrt{2} \cos\left[\omega_{h1} t + \hat{\theta}(t)\right]$

(3) $x_{I1}(t) = K_{I1}\left\{\sqrt{2} \cos\left[(\omega_0 - \omega_{h1})\, t + \theta - \hat{\theta}\right] + n_{I1}(t)\right\}$

(4) $v_2(t) \sim \sqrt{2} \cos(\omega_{h2} t - \theta_1)$

(5) $x(t) = \sqrt{2}\, A \cos\left[(\omega_0 - \omega_{h1} - \omega_{h2})\, t + \theta - \hat{\theta} + \theta_1\right] + n_i(t)$

(6) $x_\ell(t) = \frac{2\ell}{\pi}\left\{\sqrt{2}\, \alpha \cos\left[(\omega_0 - \omega_{h1} - \omega_{h2})\, t + \theta - \hat{\theta} + \theta_1\right] + n_\ell(t)\right\}$

(7) $y(t) = K_d\left\{\alpha \sin\left[(\omega_0 - \omega_{h1} - (1 + M_1)\,\omega_{h2})\, t + \theta - \hat{\theta} + (1 + M_1)\,\theta_1 + \theta_2\right] + n(t)\right\}$

(8) $z(t) = F(\rho)\; y(t)$

(9) $v_1(t) \sim \sqrt{2} \cos\left[\frac{\omega_{h1} t}{M} + \frac{K_{VCO} F(\rho)}{\rho}\, y(t)\right]$

(10) $v_3(t) \sim \sqrt{2} \sin(M_1 \omega_{h2} t + M_1 \theta_1 - \theta_2)$

(11) $z_{AGC}(t) = -e_G + K_D\left\{A \cos\left[(\omega_0 - \omega_{h1} - (1 + M_1)\,\omega_{h2})\, t + \theta - \hat{\theta} + (1 + M_1)\,\theta_1 + \theta_2\right] + n(t)\right\}$

(12) $c(t) = K_C\, Y(\rho)\, z_{AGC}(t)$

Fig. 8-1. The double-heterodyne receiver, with equations for signals at each point

zone is noise; it has total power $2\,(2l/\pi)^2\,(1 - \alpha^2)$. The limiter output signal-to-noise ratio ρ_l is thus a function of α only:

$$\rho_l = \frac{\alpha^2}{1 - \alpha^2}. \tag{8-1}$$

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

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 θ_2, it may be taken as being of the form

$$\sin(M_1 \omega_{h2} t - M_1 \theta_1 - \theta_2).$$

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

Correspondingly, the detector output is

$$y(t) = K_d\left(\alpha \sin\left\{\left[\omega_0 - \omega_{h1} - (1 + M_1)\,\omega_{h2}\right] t + \theta - \hat{\theta} + (1 + M_1)\,\theta_1 + \theta_2\right\} + n(t)\right). \tag{8-2}$$

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

$$\omega_{h1} = \omega_0 - (1 + M_1)\,\omega_{h2}. \tag{8-3}$$

The values of the IF frequencies are thus directly related to f_{h2}:

$$\begin{aligned} f_1 &= (1 + M_1)\, f_{h2} \\ f_2 &= M_1 f_{h2} \end{aligned} \tag{8-4}$$

which may vary with any fluctuation f_{h2} may have.

The heterodyne signal coming into the first mixer from the VCO through a frequency multiplier will then be one of the form

$$v(t) = \cos\left\{\omega_{h1}t + \alpha M K_d K_{VCO}\frac{F(p)}{p} \times\left[\sin\left(\theta - \hat{\theta} + (1 + M_1)\theta_1 + \theta_2\right) + \frac{n(t)}{\alpha}\right] + \frac{MK_{VCO}n_v(t)}{p}\right\} \tag{8-5}$$

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

$$\hat{\theta} = \alpha K_d K_{VCO} M \frac{F(p)}{p}\left\{\sin\left[\theta - \hat{\theta} + (1 + M_1)\theta_1(t) + \theta_2\right] + \frac{n(t)}{\alpha}\right\} + \frac{MK_{VCO}}{p}n_v(t). \tag{8-6}$$

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

$$\hat{\theta}(t) = \frac{\alpha K_d K_{VCO} M F(p)}{p + \alpha K_d K_{VCO} M F(p)} \times\left[\theta(t) + \frac{n(t)}{\alpha} + (1 + M_1)\theta_1(t) + \theta_2 + \frac{n_v(t)}{\alpha K_d F(p)}\right]. \tag{8-7}$$

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 θ 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 θ exactly (due to $(1 + M_1)\theta_1 + \theta_2$) need cause little concern so long as synchronous detection of information processes, such as doppler shift, phase modulation, etc., are also accomplished using the same reference frequencies ω_{h2} and $M_1\omega_{h2}$ combined in the same manner as above. For example, the AGC detector uses a 90-deg shifted version of $M_1\omega_{h2}$ to derive the stabilized amplitude of the signal; hence the input to the AGC filter and amplifier is

$$-e_G + AK_D\left[\cos\phi + \frac{n_{AGC}(t)}{A}\right].$$

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 $\epsilon_\phi = \theta - \hat{\theta}$, 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, ϕ and ϵ_ϕ were the same, but they are not the same quantity in the receiver above.[13]

Note that fluctuations in ϕ occur (compare (8-7) with (5-24), for example) as if $(1 + M_1)\theta_1(t)$ 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 $(1 + M_1)\theta_1(t)$ 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 ϕ can usually be neglected.

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

$$AK = \alpha K_d K_{VCO} M F. \tag{8-8}$$

(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 ϕ and ϵ_ϕ; For example, ω_{h2} can be derived as $\omega_{std} + \omega_3$, 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 ω_3; final phase detection then uses ω_3. The effect of this, in our terminology above, is that M_1 is set to -1, so that ϕ differs from ϵ_ϕ only by θ_2 plus the standard oscillator phase, θ_3. However, θ_2 is a constant, and the standard is very stable; hence for most practical purposes, θ and $\hat{\theta}$ 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 α in (8-8) above), the loop bandwidth would jitter in the same manner as A. But with a limiter, α is related directly to the predetection signal-to-noise ratio, generally a much more stable quantity. Further, α 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 $S_{nn}(j\omega)$ is fairly constant over the response region of the loop (i.e., $w_H \gg w_L$), one may derive a result similar to (5-8):

$$\sigma^2 = \frac{S_{nn}(0)\, w_L}{\alpha^2} . \tag{8-9}$$

The limiter output spectrum has some fiducial noise bandwidth w_l:

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

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,

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

Davenport's result, which we mentioned earlier, is that $\rho_l \approx 2\rho_H$ when the predetection SNR, $\rho_H = A^2/(N_0 w_H)$, is large, and $\rho_l \approx (\pi/4)\, \rho_H$ when ρ_H is very small. The actual formula for ρ_l is very complicated. However, one

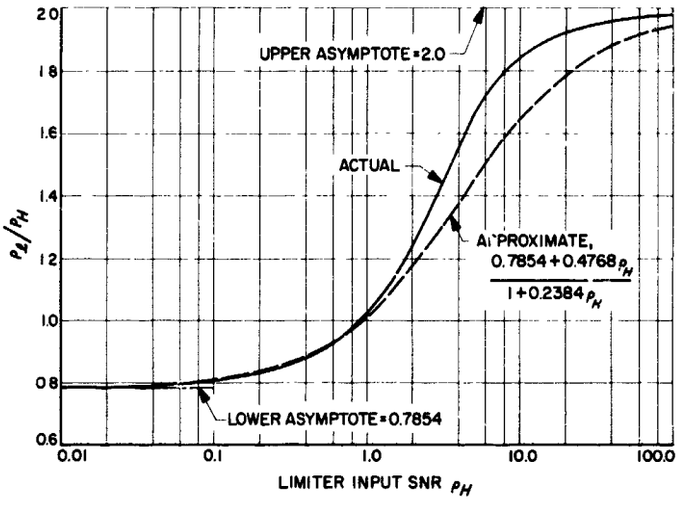


Fig. 8-2. Davenport's band-pass limiter zonal SNR curve

approximation to this, which fits extremely well (see Fig. 8-2) over the entire range of $A^2/N_0 w_H$, is

$$\rho_l/\rho_H \approx \frac{0.7854 + 0.4768\rho_H}{1 + 0.2384\rho_H} . \tag{8-12}$$

The corresponding signal amplitude suppression factor is

$$\alpha = \left[\frac{0.7854\rho_H + 0.4768\rho_H^2}{1 + 1.024\rho_H + 0.4768\rho_H^2} \right]^{1/2} . \tag{8-13}$$

As the loop bandwidth w_L is a function of loop gain, it is thereby a function of α. 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

$$r = \frac{\alpha K_d K_{VCO} M F\, \tau_2^2}{\tau_1} \tag{8-14}$$

and therefore the loop fiducial bandwidth is approximately

$$w_L = \frac{\alpha K_d K_{VCO} M F\, \tau_2^2 + \tau_1}{2\tau_2 \tau_1} . \tag{8-15}$$

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

$$w_{L(\max)} = \frac{K_d K_{VCO} M F\, \tau_2^2 + \tau_1}{2\tau_2} \approx \frac{1}{2} K_d K_{VCO} M F\, \tau_2$$

$$\approx \frac{r_0 w_{L_0}}{\alpha_0 (r_0 + 1)} . \tag{8-16}$$

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

(8-11). With this done (again assuming negligible oscillator noise), the following expression for σ^2 results:

$$\sigma^2 = \frac{N_0 w_L}{A^2}\left(\frac{w_H \rho_H}{w_l \rho_l}\right)$$

$$= \frac{N_0 w_L}{A^2}\,\Gamma\,. \qquad (8\text{-}17)$$

The factor $\Gamma = w_H \rho_H / w_l \rho_l$ is the *limiter performance factor*. One might at first presume, on the basis of Davenport's result, that σ^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 $w_l > w_H$. 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 σ^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 w_l asymptotically equal to w_H.

Going back to Davenport's original analysis, Springett has computed the factor that we would have denoted here by $(w_l \rho_l / w_H \rho_H)$. His result[14], shown in Fig. 8-3, is well approximated by

$$\frac{w_l \rho_l}{w_H \rho_H} \approx \frac{0.862 + 0.690\rho_H}{1 + 0.345\rho_H}\,. \qquad (8\text{-}18)$$

[14]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.

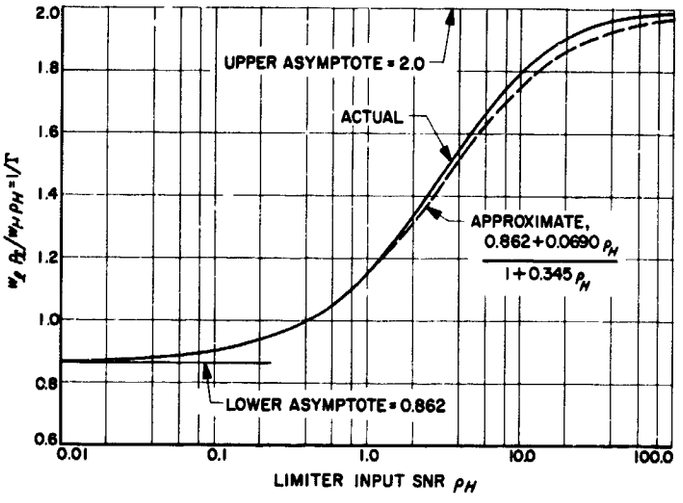


Fig. 8-3. The ratio of band-pass limiter output signal-to-noise spectral density to that at input. The reciprocal of this curve gives the limiter performance factor Γ.

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

The final resulting value of σ^2 for a loop containing an IF limiter is therefore given by

$$\sigma^2 = \frac{N_0 w_L}{A^2}\left[\frac{1.0 + 0.345\left(\frac{A^2}{N_0 w_H}\right)}{0.862 + 0.690\left(\frac{A^2}{N_0 w_H}\right)}\right]$$

$$\approx \frac{N_0 w_L}{A^2}\,. \qquad (8\text{-}19)$$

This latter approximation is valid within ¾ 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

1. Davenport, W. B., "Signal-to-Noise Ratios in Band-Pass Limiters," *Journal of Applied Physics*, Vol. 24, No. 6, pp. 720–727, June 1953.

2. Springett, J. C., "Signal-to-Noise and Signal-to-Noise Spectral-Density Ratios at the Output of a Filter–Limiter Combination," to be published in the *Space Programs Summary*, Vol. IV, Jet Propulsion Laboratory, Pasadena, Calif.

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 diagonalized 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 ϕ. A slight rearrangement of the exact loop equation (4-4) gives

$$\frac{p}{F(p)}\,\phi + AK\sin\phi = \frac{p}{F(p)}\,\theta - K\,n(t)\,. \qquad (9\text{-}1)$$

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, $\sin\Phi$, θ, $n(t)$, etc., whose double-sided Laplace transform gives the analytic continuation of the power spectral density:

$$\frac{-s^2}{F(s)F(-s)}\,S_{\Phi\Phi}(s) + AK\left[\frac{s}{F(s)}\,S_{\Phi,\sin\Phi}(s) - \frac{s}{F(-s)}\,S_{\Phi,\sin\Phi}(-s)\right] \qquad (9\text{-}2)$$

$$+ A^2K^2\,S_{\sin\Phi,\sin\Phi}(s) = \frac{-s^2}{F(s)F(-s)}\,S_{\theta\theta}(s) + K^2S_{nn}(s)\,.$$

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

$$R_{\Phi,\sin\Phi}(\tau) = \eta\,R_{\Phi\Phi}(\tau) \qquad (9\text{-}3)$$

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

$$\eta = \frac{1}{a^2}\int_{-\infty}^{+\infty}(\Phi - E\Phi)\sin\Phi\,p(\Phi)\,d\Phi\,. \qquad (9\text{-}4)$$

We have used a^2 as the variance of Φ.

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

$$p[\Phi(t_1),\Phi(t_2)] = p[\Phi(t_1)]\,p[\Phi(t_2)]\sum_{n=0}^{\infty}a_nf_n\,[\Phi(t_1)]\,f_n[\Phi(t_2)] \qquad (9\text{-}5)$$

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

$$S_{\Phi,\sin\Phi}(s) = S_{\Phi,\sin\Phi}(-s) = \eta S_{\Phi\Phi}(s). \qquad (9\text{-}6)$$

Even if (9-5) is not valid, the only modification needed in the theory is that η in (9-6) must be replaced by $\eta(s)$.

Next, when Φ is passed through the nonlinearity, sin Φ, 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:

$$\gamma(s)\gamma(-s) = \frac{S_{\sin\phi\,\sin\phi}(s)}{S_{\phi\phi}(s)}. \tag{9-7}$$

Substituting these latter relations into the spectral equation, we obtain an expression for the spectrum of Φ, namely

$$S_{\phi\phi}(s) = \frac{-s^2 S_{\theta\theta}(s) + K^2 F(s)\,F(-s)\,S_{nn}(s)}{-s^2 + \eta AK\,[sF(-s) - sF(s)] + A^2K^2\,F(s)\,F(-s)\,\gamma(s)\,\gamma(-s)} \tag{9-8}$$

from which the variance of Φ can be computed:

$$a^2 = \frac{1}{2\pi j}\int_{-j\infty}^{+j\infty} S_{\phi\phi}(s)\,ds. \tag{9-9}$$

Generally (9-9) will be transcendental, since both η 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 $|\gamma(j\omega)F(j\omega)|^2 \approx \gamma^2(0)\,|F(j\omega)|^2$. Similarly, under the same condition, $S_{nn}(j\omega)\,|F(j\omega)|^2 \approx N_0\,|F(j\omega)|^2$, where $N_0 = S_{nn}(0)$. 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[15] 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

$$S_{\phi\phi}(s) = \frac{-s^2 S_{\theta\theta}(s) + K^2 N_0 F(s)\,F(-s)}{-s^2 + \eta AK\,[sF(-s) - sF(s)] + A^2K^2\gamma^2\,F(s)\,F(-s)}. \tag{9-10}$$

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:

$$S_{\phi\phi}(s) = \frac{K^2 N_0 F(s)\,F(-s)}{-s^2 + \eta AK\,[sF(-s) - sF(s)] + (AK\gamma)^2\,F(s)\,F(-s)}. \tag{9-11}$$

Concerning the spectral approximations, one further point can be made: In the linear and quasi-linear models of the loop, ϕ is approximated by a linear function of Φ, so in turn, $S_{\sin\phi\,\sin\phi}(s)$ and $S_{\phi\,\sin\phi}(s)$ are proportional to $S_{\phi\phi}(s)$. Hence, the method above yields the linear results when the value for η 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 η and γ are not necessarily equal. It is thus more apt to provide results closer to the exact behavior of the loop.

[15]It must be remembered, however, that $n(t)$ is the result of multiplying $n_i(t)$ by cos $(\omega t+\theta)$, and hence will be somewhat wider-band than $n_i(t)$.

9-B. Calculation of η

The calculation of η as the constant relating $S_{\Phi\Phi}$ to $S_{\Phi \sin \Phi}$ 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 ϕ (mod 2π), viz., $p(\phi)$, approaches a uniform density of $1/2\pi$ on the interval $(-\pi, \pi)$. In what follows, we denote ϕ (mod 2π) merely by ϕ.

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

$$p(\phi) = \begin{cases} \dfrac{1}{a(2\pi)^{1/2}} \displaystyle\sum_{m=-\infty}^{+\infty} \exp\left[-(\phi + 2m\pi)^2/2a^2\right]; \\ \qquad\qquad \text{for } |\phi| \leq \pi \\ 0; \qquad\qquad \text{for } |\phi| > \pi\,. \end{cases} \tag{9-12}$$

The variance σ^2 of the ϕ-process is related to the variance a^2 by straightforward evaluation. The result is graphically presented in Fig. 9-1 along with the approximate formula

$$\sigma^2 = \frac{\pi^2}{3}\left\{1 - \exp\left[-\frac{3a^2}{\pi^2}(1 + 0.13a^2)\right]\right\}. \tag{9-13}$$

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

$$\eta = e^{-a^2/2}. \tag{9-14}$$

9-C. Calculation of γ^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 Φ, rather than γ^2, the ratio of powers around zero frequency. This total power value, γ_{tp}^2, is

$$\gamma_{tp}^2 = \frac{R_{\sin \Phi \sin \Phi}(0)}{R_{\Phi\Phi}(0)}. \tag{9-15}$$

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

$$R_{\sin \Phi}(\tau) = e^{-a^2} \sinh R_{\Phi}(\tau). \tag{9-16}$$

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

$$\gamma^2 = \frac{e^{-a^2}\displaystyle\int_0^\infty \sinh R_{\Phi}(\tau)\, d\tau}{\displaystyle\int_0^\infty R_{\Phi}(\tau)\, d\tau} \tag{9-17}$$

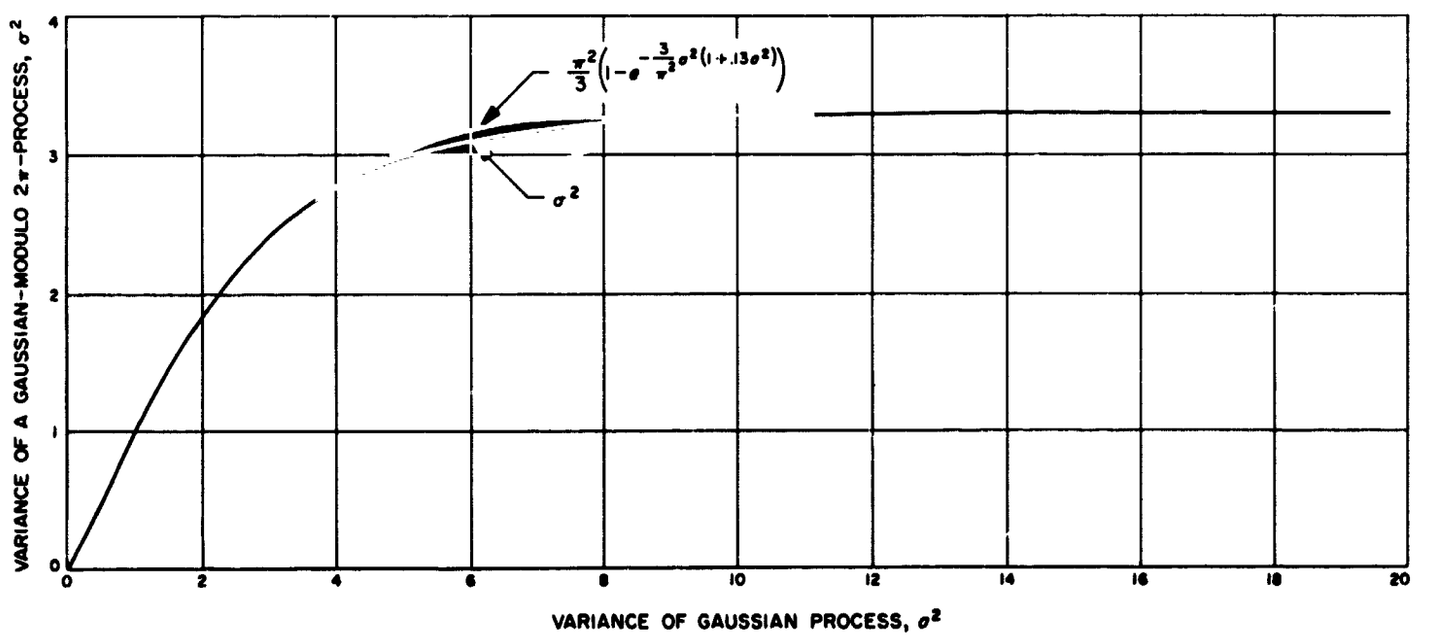


Fig. 9-1. Relationship between the variance of ϕ (mod 2π) and that of a Gaussian stationary process Φ

for a loop with a narrow-band $F(s)$, and

$$\gamma_{tp}^2 = a^{-2}\, e^{-a^2} \sinh a^2 \tag{9-18}$$

for a loop with a wide-band $F(s)$.

The former quantity, γ^2, cannot be evaluated directly, since $R_\phi(\tau)$ is not known until γ^2 is, and vice versa. However, if loop noise is not completely overpowering, γ^2 is probably not as sensitive to the shape of $R_\phi(\tau)$ as it is to the value of a^2. As a verification of this conjecture, the several different forms of $R_\phi(\tau)$ shown in Fig. 9-2 produce values of γ that are in very close agreement over the range of interest. The upper curve is equivalent to γ_{tp}. The values of γ are probably very well approximated by that in the middle:

$$\gamma = 2\, a^{-2}\, e^{-a^2/2} \sinh\left(\frac{a^2}{2}\right) = a^{-2}\left(1 - e^{-a^2}\right). \tag{9-19}$$

To see how well the spectral approximation method works, let us calibrate it by a few examples.

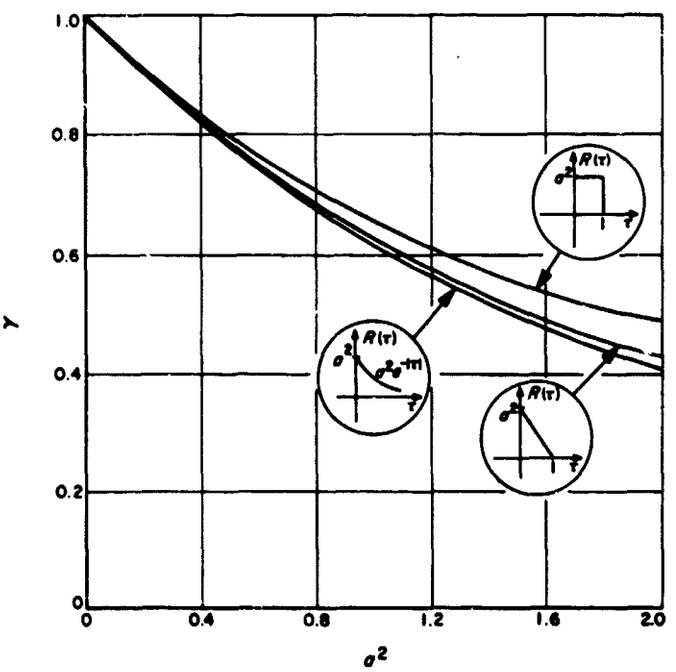


Fig. 9-2. Variation of the parameter γ as a function of the Gaussian variance a^2, for various forms of $R_{\phi\phi}(\tau)$. Note that, for $a^2 < 1$, there is not a significant dependence on the form of $R_{\phi\phi}(\tau)$.

9-D. Behavior of the First-Order Loop

The spectrum of Φ for a first-order loop is approximately

$$S_{\phi\phi} = \frac{N_0 K^2}{-s^2 + (AK\gamma_{tp})^2} \tag{9-20}$$

for we merely insert $F(s) = 1$ into (9-11). We use γ_{tp} to approximate $\gamma(s)$, because $F(s)$ is wide-band. Integration of (9-24) yields the estimate

$$a^2 = \frac{N_0 K}{2A\gamma_{tp}} = \frac{N_0}{A^2\gamma_{tp}^2}\left(\frac{A\gamma_{tp}K}{2}\right). \tag{9-21}$$

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 $A\gamma_{tp}$:

$$w_{L(eq)} = \frac{A\gamma_{tp}K}{2}. \tag{9-22}$$

1. Linear Loop

Under the usual linear assumption, $\gamma_{tp} = 1$, so $w_{L(eq)}$ reduces to w_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 $\gamma_{tp}^2 = e^{-a^2}$, the result is

$$a^2 = \frac{N_0 K}{2A}\, e^{a^2/2}. \tag{9-23}$$

Notice that there is no solution for a^2 here when $N_0K/2A \geq 1/2$. For $N_0K/2A < 1/2$, there is an equivalent bandwidth

$$w_{L(eq)} = \frac{AK}{2}\, e^{-a^2/2}. \tag{9-24}$$

3. Spectral Approximation

If we use the analysis presented above with γ_{tp} as given in (9-18), the phase variance of the equivalent stationary process Φ is related to the linear-loop prediction by

$$ae^{-a^2/2}\,(\sinh a^2)^{1/2} = \frac{N_0 K}{2A}. \tag{9-25}$$

Here there is a solution a^2 for every value of $N_0K/2A$.

4. Reduction Modulo 2π

The real indication of system performance, as we have agreed, is given not by the parameter a^2 (the variance of

Φ), but by σ^2, the variance of ϕ (mod 2π). These are related as shown in Fig. 9-2. Each of the various performance approximations, reduced mod 2π, 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 2π) over the entire useful range of input signal and noise values. In addition, we have shown that the linear analysis is valid whenever $A^2/N_0 w_L > 10$.

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

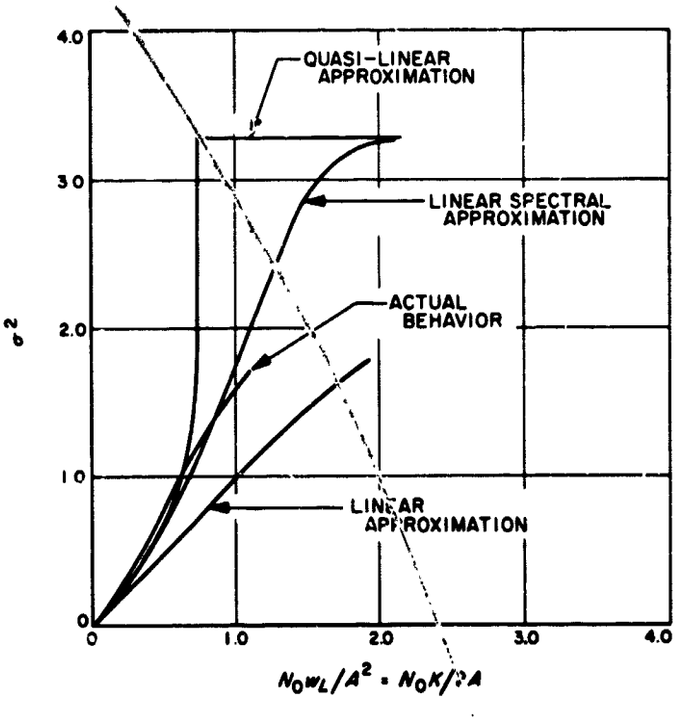


Fig. 9-3. Comparison of linear, quasi-linear, and linear-spectral approximate methods with the actual behavior of the first-order loop

filter function inserted into the spectral equation (9-11) produces the density

$$S_{\phi\phi}(s) = \frac{K^2 N_0 (1 - \tau_2^2 s^2)}{\tau_1^2 s^4 - s^2 [1 + 2AK\eta (\tau_2 - \tau_1) + (AK\gamma \tau_2)^2] + (AK\gamma)^2} \tag{9-26}$$

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, $S_{\phi\phi}(j\omega)$ has a fiducial bandwidth

$$w_{L(\text{eq})} = \frac{1}{2\pi S_{\phi\phi}(0)} \int_{-\infty}^{+\infty} S_{\phi\phi}(j\omega)\, d\omega . \tag{9-27}$$

Note that this is computed in the same way as w_L, except that we must use $S_{\phi\phi}(j\omega)$ as given in (9-26) rather than $|L(j\omega)|^2$, as one would do if the linear model were in effect. In these terms

$$a^2 = \frac{N_0 w_{L(\text{eq})}}{(A\gamma)^2} . \tag{9-28}$$

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: $w_{L(\text{eq})}$ must be used rather than w_L, and A is reduced to $A\gamma$.

The value of $w_{L(\text{eq})}$ can be found by integrating (9-27):

$$w_{L(\text{eq})} = \frac{(1 + \tau\gamma)}{2\tau_2 (1 + \tau_2/\tau\tau_1\gamma) \left[1 + \dfrac{2(\gamma - \eta)(1 - \tau_2/\tau_1)}{\tau\gamma^2 (1 + \tau_2/\tau\tau_1\gamma)^2}\right]^{1/2}}$$

$$\approx \frac{1 + \tau\gamma}{2\tau_2 \left[1 + \dfrac{2(\gamma - \eta)}{\tau\gamma^2}\right]^{1/2}} . \tag{9-29}$$

The approximation follows the natural assumption $\tau\tau_1\gamma \gg \tau_2$.

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 $S_{\phi\phi}$; it is of the form

$$S_{\phi\phi}(s) = \frac{1 - as^2}{s^4 - bs^2 + c} .$$

The roots of the denominator govern the "damping behavior" of $\Phi(t)$. If these roots are r_1, r_2, $-r_1$, $-r_2$, 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*

$$\zeta_{eq} = \frac{(r_1 + r_2)}{2(r_1 r_2)^{1/2}}. \tag{9-31}$$

We relate $r_1 r_2$ and $(r_1 + r_2)$ to the parameters of $S_{\phi\phi}(s)$ by noting

$$(r_1 r_2)^2 = \left(\frac{AK\gamma}{\tau_1}\right)^2 \tag{9-32}$$

$$(r_1 + r_2)^2 - 2r_1 r_2 = \tau_1^{-2}\left[1 + 2AK\eta(\tau_2 - \tau_1) + (AK\gamma\tau_2)^2\right].$$

Hence, the equivalent output damping factor of the loop is

$$\zeta_{eq} = \frac{(\gamma r)^{1/2}}{2}\left[(1 + \tau_2/r\tau_1\gamma)^2 + \frac{2(\gamma - \eta)(1 - \tau_2/\tau_1)}{r\gamma^2}\right]^{1/2}$$

$$\approx \frac{(\gamma r)^{1/2}}{2}\left[1 + \frac{2(\gamma - \eta)}{r\gamma^2}\right]^{1/2}. \tag{9-33}$$

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 $Ae^{-a^2/2}$ (i.e., $\gamma = \eta = e^{-a^2/2}$). Thus the behavior predicted in this case is given as

$$a^2 e^{-a^2} = \frac{N_0 w_{L(eq)}}{A^2}$$

$$w_{L(eq)} \approx \frac{1 + r\,e^{-a^2/2}}{2\tau_2} \tag{9-34}$$

$$\zeta_{eq} \approx \frac{r^{1/2}}{2}\,e^{-a^2/4}.$$

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 $\eta = e^{-a^2/2}$ and $\gamma = a^{-2}\,(1 - e^{-a^2})$. The loop behavior is then approximated very closely by

$$\frac{N_0 w_L}{A^2} = \frac{(1 - e^{-a^2})^2}{a^2}\,\frac{w_L}{w_{L(eq)}}$$

$$\frac{w_{L(eq)}}{w_L} \approx \frac{1 + r(1 - e^{-a^2})/a^2}{(1 + r)\left[1 + \dfrac{2a^2(1 - e^{-a^2} - a^2 e^{-a^2/2})}{r(1 - e^{-a^2})^2}\right]^{1/2}} \tag{9-35}$$

$$\frac{\zeta_{eq}}{\zeta} \approx \frac{(1 - e^{-a^2})^{1/2}}{a}\left[1 + \frac{2a^2(1 - e^{-a^2} - a^2 e^{-a^2/2})}{r(1 - e^{-a^2})^2}\right]^{1/2}.$$

Again, the same type of behavior results: $w_{L(eq)}$ and ζ_{eq} are less than their linearly predicted value, while a^2 is greater. Although the quantities in (9-35) are generally transcendental in a^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 σ^2, based on a constant value of r (viz., $r = 2$). Figure 9-5 then shows how σ^2 compares with the value pre-

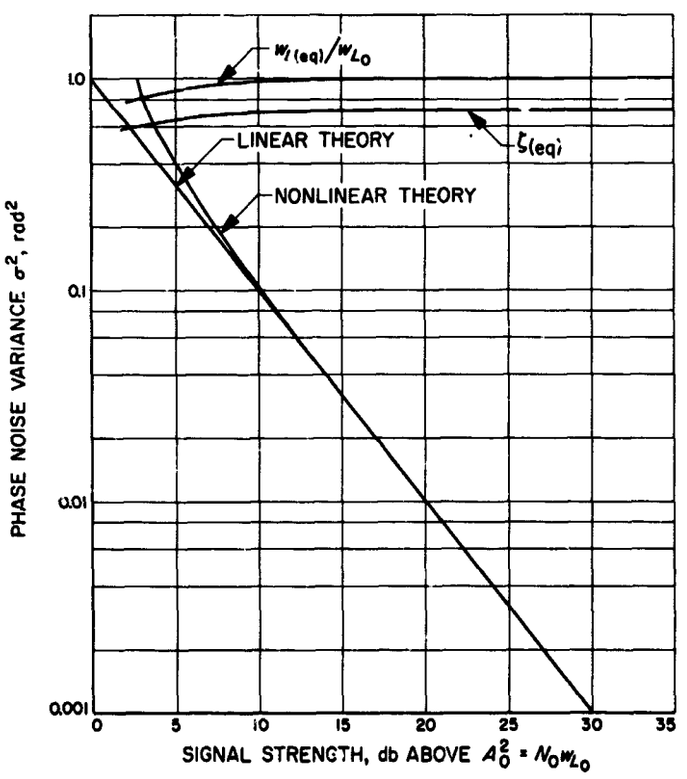


Fig. 9-4. Comparison of linear and nonlinear theories for second-order, constant linear bandwidth loop, i.e., the value of r is kept constant at $r_0 = 2$.

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[16] chosen a value $r_0 = 2$ at $A_0^2 = 3N_0/2\tau_2$.

[16]Reasons for this choice are given in Chapter 10.

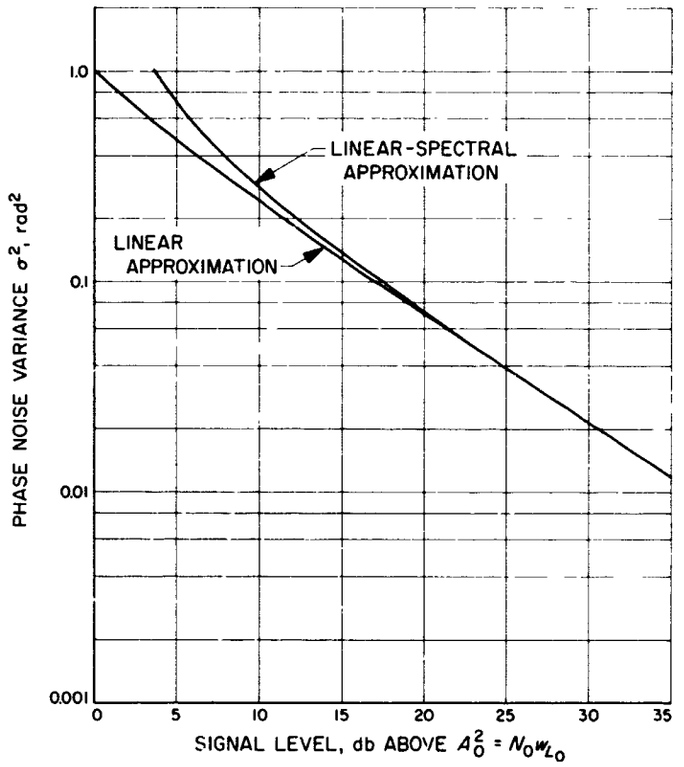


Fig. 9-5. Comparison of phase-noise variances by linear and linear-spectral approximations. The noise density is fixed, and the signal level is varied. The value of r is taken as $r_0 = 2$ at a reference signal level of $A_0^2 = 3N_0/2\tau_2 = N_0 w_{L_0}$. Note that the ultimate roll-off is 5db/decade, rather than 10 db/decade, as in Fig. 9-4. Note also that even the linear approximation produces some curvature of σ^2 near A_0^2.

Figure 9-6 illustrates the way $w_{L(eq)}$, w_L, $\zeta_{(eq)}$, and ζ 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 $A^2/N_0 w_L > 10$. Beyond this point, the linear spectral approximation probably agrees with actual results, if we may extrapolate the results obtained for the first-order loop, for $\sigma^2 < 1.5$ rad^2. This figure lies beyond the useful range of most receivers.

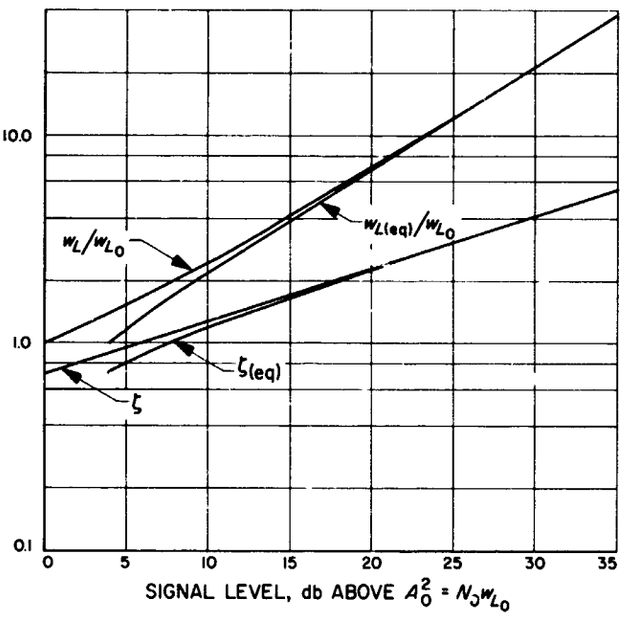


Fig. 9-6. Variation in bandwidth and damping parameters as a function of signal strength. The value of r at a reference signal level $A_0^2 = 3N_0/2\tau_2 = N_0 w_{L_0}$ was taken as $r_0 = 2$.

REFERENCES FOR CHAPTER 9

1. Tausworthe, R. C., "A New Method for Calculating Phase-Locked Loop Behavior," *Space Programs Summary No. 37-31*, Vol. IV, pp. 292–300, Jet Propulsion Laboratory, Pasadena, Calif., February 1965.

2. Bussgang, J. J., "Cross-Correlation Functions of Amplitude-Distorted Gaussian Signals," *MIT Research Laboratory of Electronics Technical Report*, No. 216, The M.I.T. Press, Cambridge, Mass., March 1952.

3. Barrett, J. F., and Lampard, D. G., "An Expansion for Some Second-Order Probability Distributions and its Applications to Noise Problems," *IRE Transactions of the Professional Group on Information Theory*, Vol. IT-1, March 1955, pp. 10–15.

4. Price, Robert, "A Useful Theorem for Nonlinear Devices Having Gaussian Inputs," *IRE Transactions of the Professional Group on Information Theory*, Vol. IT-4, June 1958, pp. 69–72.

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 each 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

$$A_0^2 = \frac{N_0(r_0 + 1)}{2\tau_2\left(1 + \frac{\tau_2}{r_0\tau_1}\right)} = N_0 w_{L_0} . \qquad (10\text{-}1)$$

Here, w_{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; w_{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 $K_d K_{VCO} MF$, τ_1, τ_2, and W_H have been determined, performance proceeds as specified by the formulas[17] in Chapter 8. But with a fixed receiver, the loop bandwidth changes as a function of signal or noise level. Rather than

[17]Although W_H is a measured physical quantity, it may not be equal to w_H if the loop is not operating at the frequency producing $H(j\omega)_{max}$. Care must often be taken to assure that $H(s)$ 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

$$m = \frac{A^2}{A_0^2} \tag{10-2}$$

as the loop *performance margin*. In terms of this m,

$$\rho_H = \frac{A^2}{N_0 w_H} = \left(\frac{A^2}{A_0}\right)\left(\frac{w_{L_0}}{w_H}\right) = m\rho_{H_0} . \tag{10-3}$$

As a result, all of the loop characteristics are expressible in terms of m, w_{L_0}, w_H, and τ_2/τ_1 (and this latter quantity is not needed if $\tau_1 \gg \tau_2$). 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_H, $K_dK_{VCO}MF$, τ_2, and τ_1 are given — quantities that can be measured. Most of the formulas we have given are in terms of r's, w_L's, etc. How does one proceed with the set of parameters above? We merely need to find r_0, and the rest then follow. When $10w_{L_0} < w_H$, α_0^2 is very well approximated by $\pi w_{L_0}/4w_H$; from this, a slight manipulation of (8-14) produces the desired result:

$$U = \frac{\pi(K_dK_{VCO}MF)^2\tau_2^3}{8\tau_1^2 w_H}$$

$$r_0 = \left(\frac{U - \tau_2/\tau_1}{2}\right)\left\{1 + \left[1 + \frac{4U}{(U - \tau_2/\tau_1)^2}\right]^{1/2}\right\}$$

$$\approx \frac{U}{2}\left[1 + \left(1 + \frac{4}{U}\right)^{1/2}\right] . \tag{10-4}$$

We can also express U in terms of r_0,

$$U = \frac{r_0(r_0 + \tau_2/\tau_1)}{r_0 + 1} \approx \frac{r_0^2}{r_0 + 1} . \tag{10-5}$$

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 $\alpha K_dK_{VCO}MF$, as in (8-8); and second, the expression (9-28) must be multiplied by the limiter performance factor $\Gamma = w_H\rho_H/w_i\rho_i$, shown in Fig. 8-3. These steps lead to a set of equations which characterize the double-heterodyne receiver:

$$\rho_H = m\left(\frac{w_{L_0}}{w_H}\right) = m\rho_{H_0} \tag{10-6a}$$

$$\alpha = \left[\frac{0.7854\rho_H + 0.4768\rho_H^2}{1 + 1.024\rho_H + 0.4768\rho_H^2}\right]^{1/2} \tag{10-6b}$$

$$\Gamma = \frac{1 + 0.345\rho_H}{0.862 + 0.690\rho_H} \tag{10-6c}$$

$$r = \frac{\alpha K_dK_{VCO}MF(0)\tau_2^2}{\tau_1} = \frac{\alpha}{\alpha_0} r_0 = \tau_2^2\beta^2 = \tau_2^2\left(\frac{\alpha}{\alpha_0}\right)\beta_0^2 \tag{10-6d}$$

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

$$\approx a^2 \tag{10-6e}$$

$$\gamma = (1 - e^{-a^2})/a^2 \tag{10-6f}$$

$$\eta = e^{-a^2/2} \tag{10-6g}$$

$$w_L = \frac{1 + r}{2\tau_2\left(1 + \frac{\tau_2}{r\tau_1}\right)} \approx \frac{1 + r}{2\tau_2} = \left[\frac{1 + \left(\frac{\alpha}{\alpha_0}\right)r_0}{1 + r_0}\right] w_{L_0} \tag{10-6h}$$

$$w_{L(eq)} = \frac{1 + r\gamma}{2\tau_2\left(1 + \frac{\tau_2}{r\gamma\tau_1}\right)\left[1 + \frac{2(\gamma - \eta)(1 - \tau_2/\tau_1)}{\gamma^2 r\left(1 + \frac{\tau_2}{r\gamma\tau_1}\right)^2}\right]^{1/2}} \tag{10-6i}$$

$$\zeta = \frac{1}{2}\left(1 + \frac{\tau_2}{r\tau_1}\right) r^{1/2} = \frac{1}{2}\left(1 + \frac{1}{\tau_1\tau_2\beta^2}\right)\tau_2\beta$$

$$\approx \frac{r^{1/2}}{2} = \frac{\tau_2\beta}{2} \tag{10-6j}$$

$$\zeta_{eq} = \frac{1}{2}\left(1 + \frac{\tau_2}{r\gamma\tau_1}\right)\left\{r\gamma\left[1 + \frac{2(\gamma - \eta)(1 - \tau_2/\tau_1)}{r\gamma^2(1 + \tau_2/r\gamma\tau_1)^2}\right]\right\}^{1/2}$$

$$\approx \frac{1}{2}(r\gamma)^{1/2} \approx \zeta\gamma^{1/2} \tag{10-6k}$$

$$\beta_{eq} = \frac{(r\gamma)^{1/2}}{\tau_2} = \beta\gamma^{1/2} \tag{10-6l}$$

$$a^2 = \frac{N_0 w_{L(eq)}\Gamma}{A^2\gamma^2} = \frac{1}{m}\left(\frac{w_{L(eq)}\Gamma}{w_{L_0}\gamma^2}\right) \tag{10-6m}$$

$$\beta \approx \frac{2w_L}{r+1} = \left(\frac{\alpha}{\alpha_0}\right)^{1/2}\frac{2w_{L0}}{r_0+1}\,. \tag{10-6n}$$

The use of γ as given in (10-6f) requires $\tau_1 \gg \tau_2$; otherwise, γ_{tp} 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 $\rho_{H_0} = w_{H_0}/w_H$, r_0, and τ_2/τ_1.

The performance observed here is very much the same as that exhibited in loops without limiters, except that the value of $K_d K_{VCO} MF$ 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 A_0 is very small (the threshold rms signal level) α_0 is many orders of magnitude greater, approximately equal to the predetection SNR.

10-D. The Signal Level Producing $\sigma^2 = 1$

Since it is doubtful that one would ever expect a loop to operate usefully with a value of σ^2 greater than unity, the specific value of a_0^2 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 theory 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 $a^2 = a_1^2 = 1$. 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 $A_1/A_0 = m_1^{1/2}$ satisfies the equation

$$\frac{A_1}{A_0} = \frac{\Gamma_1(r_0 + \tau_2/\tau_1)(A_1/A_0 + 1/r_0\gamma_1)}{\gamma_1(r_0+1)(A_1/A_0 + \tau_2/r_0\gamma_1\tau_1)\left[1 + \dfrac{2(\gamma_1-\eta_1)(1-\tau_2/\tau_1)(A_1/A_0)}{\gamma_1^2 r_0(A_1/A_0 + \tau_2/\gamma_1 r_0\tau_1)^2}\right]^{1/2}}$$

$$r_1 = \frac{\alpha_1}{\alpha_0}r_0 \approx \frac{A_1}{A_0}r_0\,. \tag{10-7}$$

The subscript "1" on Γ, γ, η, etc., refers to evaluation at $a^2 = 1$. Under the usual assumption $\gamma_1 r_0 \tau_1 \gg \tau_2$, the approximate value for m_1 is

$$m_1 = \left[\frac{r_0\Gamma_1}{2\gamma_1(r_0+1)}\right]^2\left\{1 + \left[1 + \frac{4(r_0+1)}{\Gamma_1 r_0^2}\right]^{1/2}\right\}^2\,. \tag{10-8}$$

The value of σ^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 $\sigma^2 = 1$. The curves in Fig. 10-1 show how m_1 varies as a function of r_0 and τ_2/τ_1. As τ_2/τ_1 nears unity, it is necessary to replace γ by γ_{tp} (see 9-18). The values given in the Figure result by assuming that γ_1 in (10-3) takes the value

$$\gamma_1 = \left(1 - \frac{\tau_2}{\tau_1}\right)(1 - e^{-1}) + \left(\frac{\tau_2}{\tau_1}\right)e^{-1/2}$$

$$= 0.6321\left(1 - \frac{\tau_2}{\tau_1}\right) + 0.60653\left(\frac{\tau_2}{\tau_1}\right)\,. \tag{10-9}$$

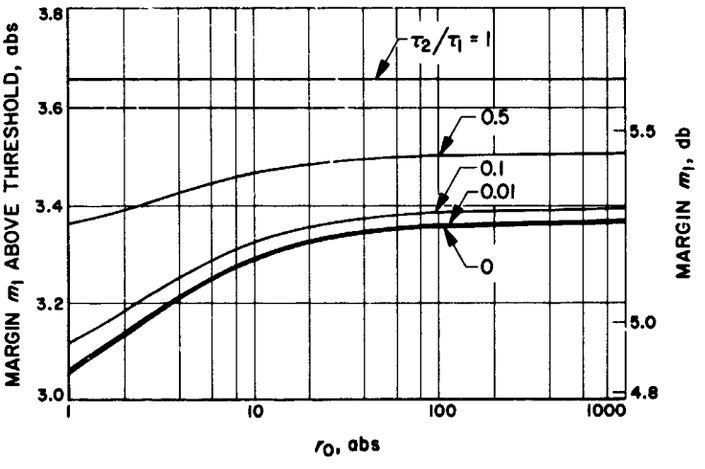


Fig. 10-1. Variation in margin producing $\sigma^2 = 1$ as a function of threshold design parameter, $r_0 = \alpha_0 K_d K_{VCO} MF\tau_2^2/\tau_1$

10-E. Choice of Receiver Parameters

Choosing values of N_0, w_{L_0}, r_0, $K_dK_{VCO}MF$, and w_H specifies a typical receiver design. Values for τ_1, τ_2, ρ_{H_0}, α_0, and Γ_0 follow directly from (10-1) and the first four parts of (10-6). The value of τ_2/τ_1 is a measure of how well the imperfect integrating loop filter performs in the loop; with $\tau_2/\tau_1 = 1$, 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 τ_2/τ_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 things an engineer needs to know in order to know how well any receiver will perform are m, r_0, ρ_{H_0}, and perhaps τ_2/τ_1. This latter set of quantities 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 ρ_{H_0} depends on w_{L_0} and w_H; w_{L_0} 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, τ_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 ρ_{H_0} (and perhaps τ_2/τ_1) are given, the receiver performance in terms of margin can be found from (10-6); a typical plot of σ 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_0 and w_{L_0}, threshold is specified (10-1), so the curves are easily referenced to the actual incoming signal power.

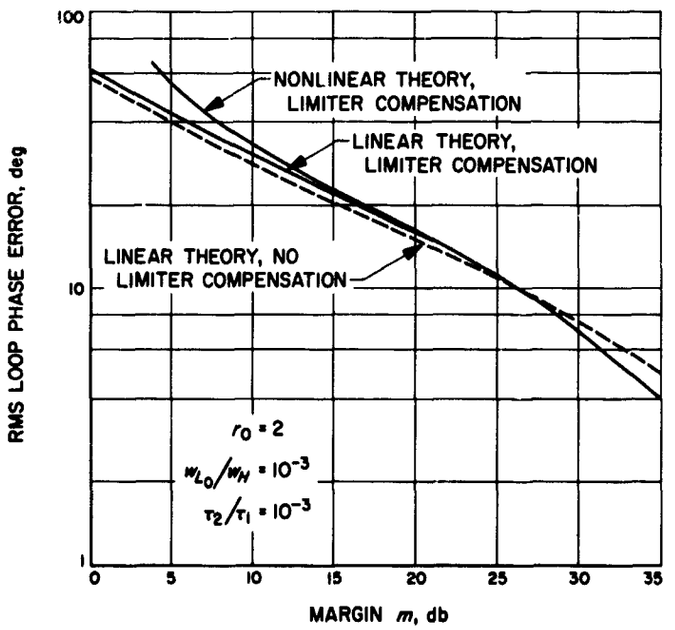


Fig. 10-2. Comparison of linear and nonlinear approximations to loop rms phase error, as a function of loop margin

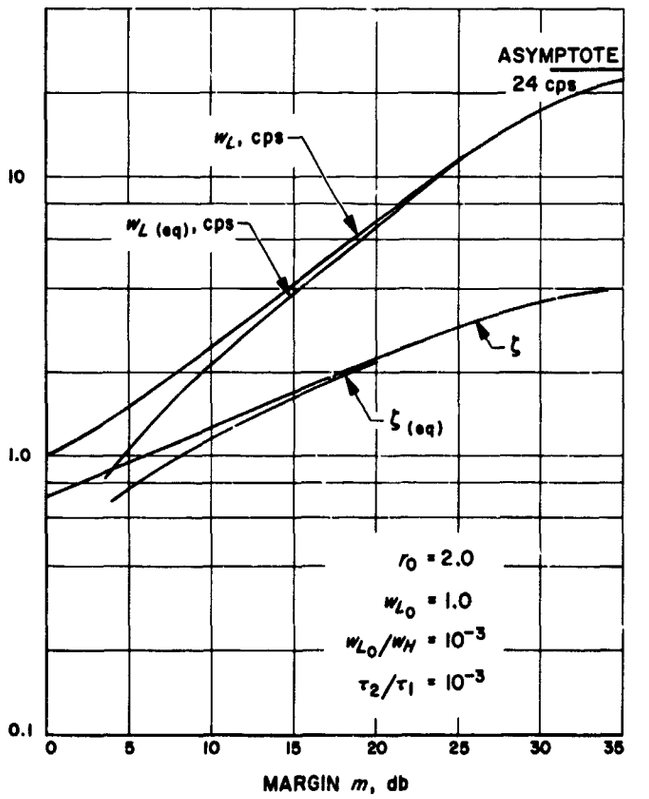


Fig. 10-3. Variation of loop bandwidths and damping factors, as a function of loop margin

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.

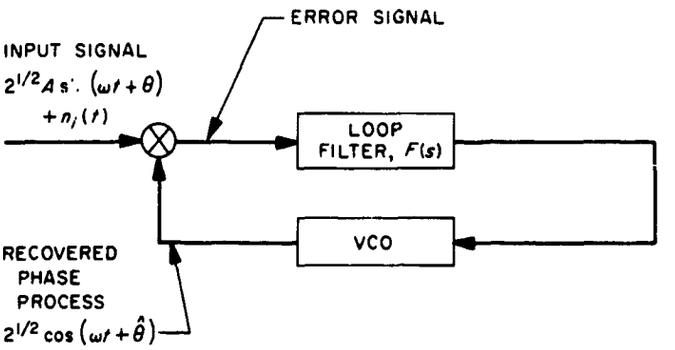


Fig. II-1-1. Basic configuration of a simple phase-locked loop. The mixer output, filtered by $F(s)$, is used to control the frequency of the voltage-controlled oscillator (VCO).

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 $x(t)$

$$\mathcal{R}_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)\, x(t+\tau)\, dt . \qquad (2\text{-}3)$$

Statistical autocorrelation of a stochastic process $\{x(t)\}$:

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

$$= R_{xx}(t_1 - t_2) \quad \text{(if } x(t) \text{ is stationary)} \qquad (2\text{-}5)$$

$$= \mathcal{R}_{xx}(t_1 - t_2) \quad \text{(if } x(t) \text{ is ergodic).} \qquad (2\text{-}6)$$

Cross-correlation $R_{xy}(t_1 - t_2)$: obtained by replacing $x(t_2)$ by $y(t_2)$ above.

Spectral density of a function $x(t)$:

$$\mathcal{S}_{xx}(j\omega) = \int_{-\infty}^{+\infty} \mathcal{R}_{xx}(\tau)\, e^{-j\omega\tau}\, d\tau . \qquad (2\text{-}7)$$

Spectral density of a stationary process $\{x(t)\}$:

$$S_{xx}(j\omega) = \int_{-\infty}^{+\infty} R_{xx}(\tau)\, e^{-j\omega\tau}\, d\tau . \qquad (2\text{-}8)$$

Cross-spectral density $S_{xy}(j\omega)$: similarly defined as the Fourier transform of $R_{xy}(\tau)$.

Mean value of $x(t)$:

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

$$= \mu \quad \text{(if } x(t) \text{ is stationary).} \qquad (2\text{-}10)$$

Variance of $x(t)$:

$$\sigma_x^2(t) = E\,[x(t) - \mu(t)]^2$$

$$= E\,[x^2(t)] - \mu^2(t) . \qquad (2\text{-}11)$$

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

$$y(t) = \int_{-\infty}^{+\infty} x(t_1) h(t - t_1)\, dt_1 . \qquad (2\text{-}14)$$

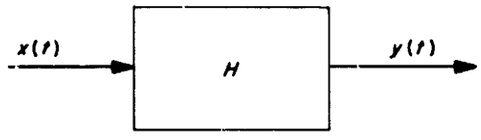


Fig. II-2-1. Filtering device

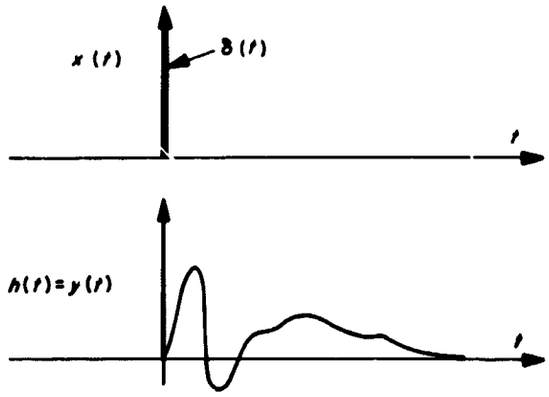


Fig. II-2-2. Response of filter to a unit-impulse function

When $x(t)$, $y(t)$, and $h(t)$ have Laplace transforms $X(s)$, $Y(s)$, and $H(s)$, respectively, linearity reduces to

$$Y(s) = X(s)\,H(s)\,. \tag{2-15}$$

When $x(t)$ is a stationary process,

$$S_{yy}(j\omega) = |H(j\omega)|^2\, S_{xx}(j\omega)\,. \tag{2-17}$$

2-C. Noise Bandwidth

White noise: a random process $n(t)$ such that

$$R_{nn}(\tau) = N_0\,\delta(\tau) \quad \text{(volts}^2\text{)} \tag{2-20}$$

$$S_{nn}(j\omega) = N_0 \quad \text{for all } -\infty < \omega < +\infty \quad \text{(volts}^2\text{/cps)}\,. \tag{2-21}$$

Effective noise bandwidth of a linear filter $H(s)$:

$$W_H = 2B_H = \frac{\frac{1}{2\pi}\int_{-\infty}^{+\infty} |H(j\omega)|^2\, d\omega}{|H(j\omega)|_{\max}} \quad \text{(cps)}\,. \tag{2-24}$$

Note that W_H is the *two-sided bandwidth* and B_H is the *single-sided bandwidth* of $H(s)$ in cps (see Figs. II-2-3, II-2-4, and II-2-5). The output noise power with a white noise input is

$$N = N_0 W_H\, |H(j\omega)|^2_{\max} \quad \text{(volts}^2\text{)}\,. \tag{2-25}$$

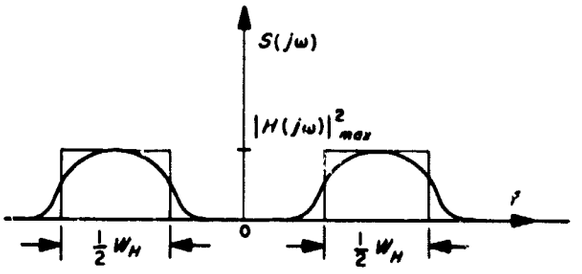


Fig. II-2-3. Equivalent noise bandwidth

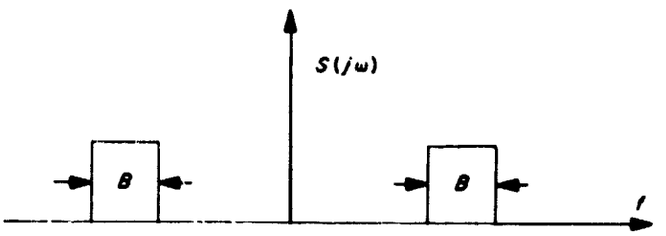


Fig. II-2-4. Double-sided frequency response

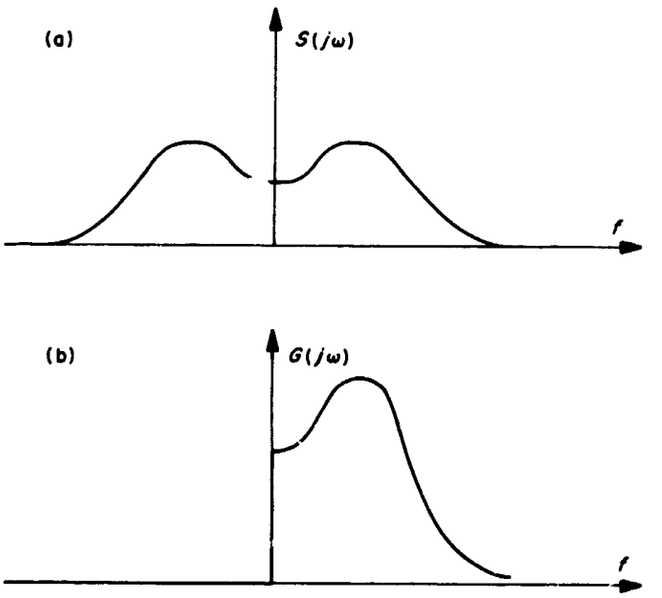


Fig. II-2-5. Double- and single-sided spectra

2-D. Sinusoidal Filter Inputs

If $2^{1/2} P \sin(\omega_0 t + \theta) + n(t)$ is put into $H(s)$, where $n(t)$ is white with spectral density N_0, the output SNR of $H(s)$ is

$$\rho_H = \frac{P}{N_0 W_H} \cdot \frac{|H(j\omega_0)|^2}{|H(j\omega)|^2_{\max}} = \frac{P}{N_0 w_H}\,. \tag{2-35}$$

2-E. Fiducial Bandwidth

Fiducial bandwidth: equivalent (2-sided) bandwidth referred to a specific frequency ω_0, such as the carrier above, rather than that ω producing $|H(j\omega)|^2_{\max}$,

$$w_H = 2b_H = W_H \frac{|H(j\omega)|^2_{\max}}{|H(j\omega_0)|^2} = \frac{\frac{1}{2\pi}\int_{-\infty}^{+\infty} |H(j\omega)|^2\, d\omega}{|H(j\omega_0)|^2} \quad \text{(cps)}\,. \tag{2-37}$$

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. II-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 ρ is

$$\rho = \frac{P}{N_0 w_D}\, E(\cos^2\phi) \tag{2-52}$$

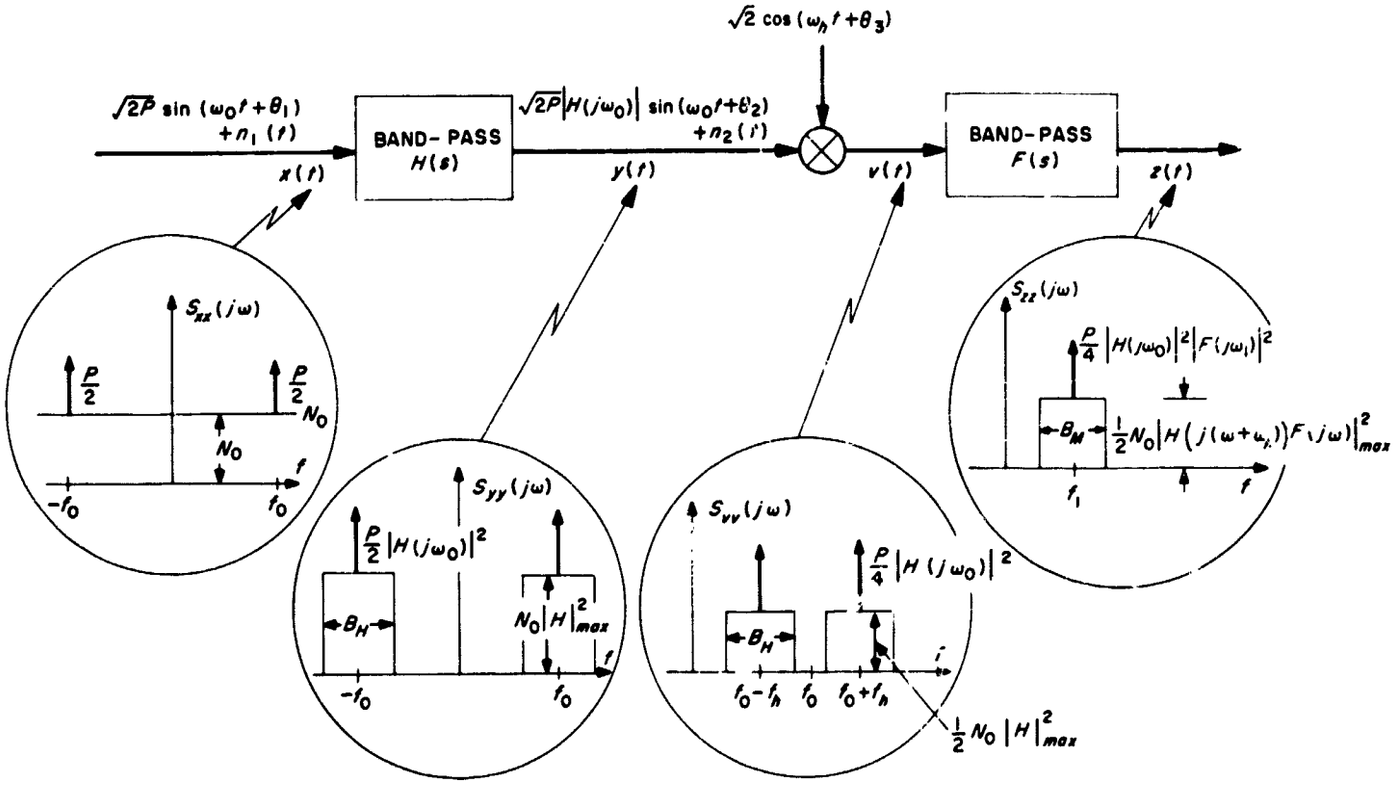


Fig. II-2-6. The simple product-mixer

where P = input signal power (watts), N_0 = 2-sided input noise density (watts/cps), ω_D is the fiducial bandwidth of the detector (cps), and ϕ 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

$$S_{v_n v_n}(j\omega) = 2kTR \quad \text{(volts}^2\text{/cps)} \qquad (2\text{-}62)$$

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_N is

$$N = \frac{kT}{2} W_N = kTB_N \quad \text{(watts)}. \qquad (2\text{-}63)$$

(Compare this with (2-25), for a filter with bandwidth W_N and gain $|H(j\omega)|^2_{max}$.)

Hence for a white thermal noise the matched-load noise voltage appearing across the load resistance R has uniform spectral density $N_0 = \frac{1}{2}kTR$ volts²/cps.

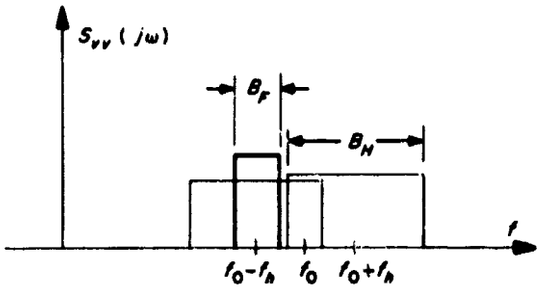


Fig. II-2-7. Choosing filter bandwidths to avoid the image noise problem

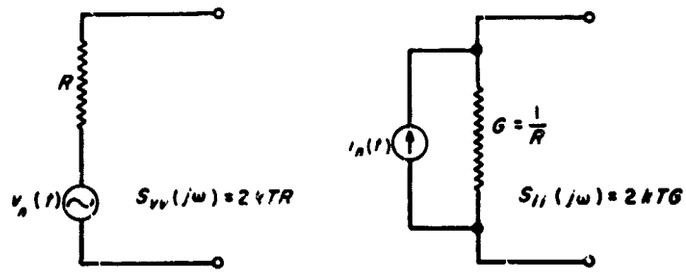


Fig. II-2-8. Thevenin and Norton equivalent circuits of noisy resistors

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

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

where A = rms signal amplitude, volts

K = total open-loop gain, $\sec^{-1}$ volt^{-1}

$\theta(t)$ = input phase function, rad

$\phi(t)$ = phase-error function, rad

$p = d/dt$, the Heaviside operator.

3-B. Tracking When the Loop Error Is Small

If $|\phi| < \pi/6$, then $\sin \phi(t) \approx \phi(t)$, so

$$\theta(t) \approx \frac{p + AKF(p)}{p} \phi(t). \tag{3-10}$$

Considering $\theta(t) = d(t)$ radians, i.e., a doppler-shift function only, the steady-state phase-error is

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

where $D(s) = \mathcal{L}[d(t)]$. If $d(t) = \lambda_n t^n/n!$, and if $F(s)$ has l poles at the origin, as $F(s) = q(s)/s^l p(s)$, then for $n \geq l+1$

$$\phi_{ss} \to \frac{\lambda_n p(0)}{AKq(0)} \frac{t^{n-l-1}}{(n-l-1)!} \quad \text{as } t \to \infty. \tag{3-16}$$

For $n < l + 1$, $\phi_{ss} = 0$, so perfect doppler tracking ultimately results. When $n = l + 1$, there is a constant steady-state error

$$\phi_{ss} = \frac{\lambda_n p(0)}{AKq(0)} \quad \text{(rad)}. \tag{3-17}$$

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

$$\Omega_{max} \text{ (rad/sec)} = AK = 2W_L \text{ (cps)} \tag{3-23}$$

$$\phi_{ss} = \sin^{-1}\left(\frac{\Omega_0}{AK}\right) = \sin^{-1}\left(\frac{\Omega_0}{2W_L}\right) \text{ (rad)} \tag{3-24}$$

$$t_{acq} \approx \frac{1}{W_L \cos \phi_{ss}} \ln \frac{2}{\delta_{lock}} \text{ (sec)} \tag{2-37}$$

where Ω_{max} is the maximum value of Ω_0, the initial loop frequency offset in rad/sec for which the loop ultimately locks, t_{acq} is the maximum time in seconds required to achieve lock within δ_{lock} radians, and W_L 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

$$F(s) = \frac{1 + \tau_2 s}{1 + \tau_1 s} \tag{3-28}$$

where τ_1, τ_2 are in seconds, for $d(t) = \theta_0 + \Omega_0 t$, there results

$$\phi_{ss} = \sin^{-1}\left(\frac{\Omega_0}{AK}\right) = \sin^{-1}\left(\frac{\Omega_0 \tau_2^2}{r\tau_1}\right) \text{ (rad)}. \tag{3-30}$$

But for lock-in to occur for all initial conditions of the VCO,

$$\Omega_0 < \frac{2}{\tau_2}\left[r\left(1 + \frac{r\tau_1}{2\tau_2}\right)\right]^{1/2} \approx 4w_L\left(\frac{r}{r+1}\right)\left(\frac{\tau_1}{2\tau_2}\right)^{1/2} \text{ (rad/sec)}. \tag{3-33}$$

The time to reach frequency lock (not phase lock) is

$$t_{\text{freq acq}} \approx \frac{(r+1)^3}{8r^2 w_L}\left(\frac{\Omega_0}{w_L}\right)^2 \text{ (sec)} \tag{3-34}$$

where $r = AK\tau_2^2/\tau_1$ and w_L is the loop fiducial bandwidth (see Chapter 5). We assume $r\tau_1 \gg \tau_2$. Typical plots of lock-in behavior are illustrated in Figs. II-3-3 and II-3-4.

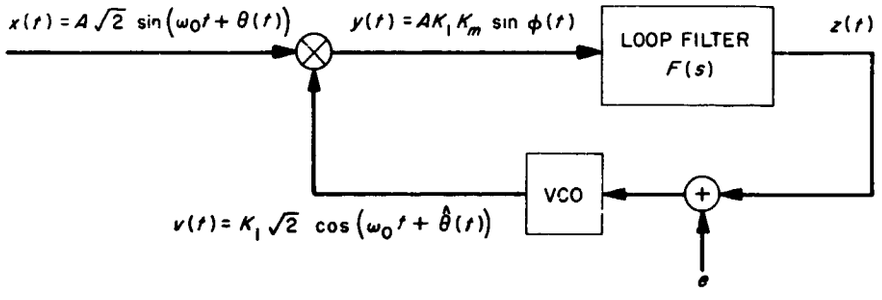


Fig. II-3-1. The basic phase-locked loop

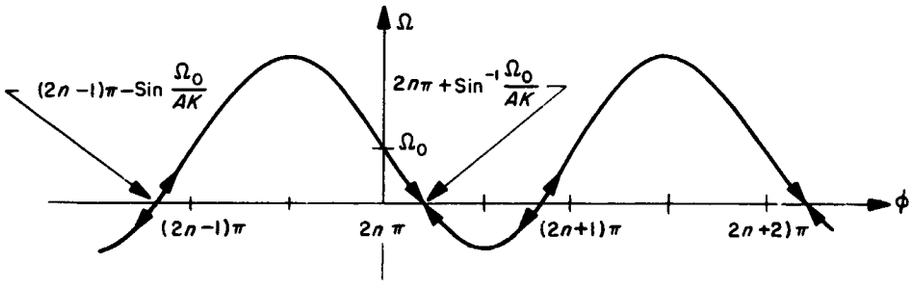


Fig. II-3-2. First-order loop pull-in behavior

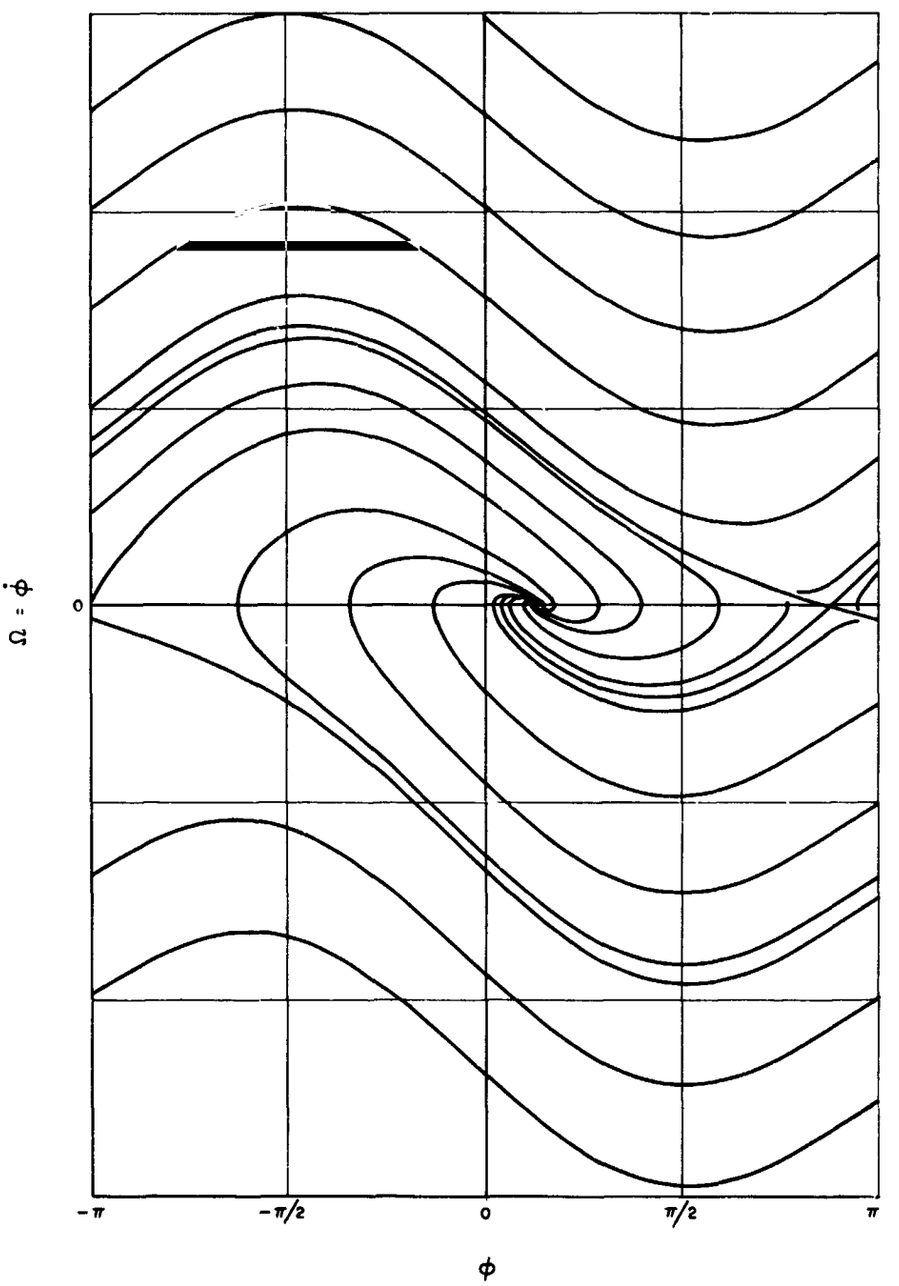


Fig. II-3-3. Lock-in behavior of a second-order loop with imperfect integrator, $F(s) = (1 + \tau_2 s)/(1 + \tau_1 s)$, for $\Omega_0/AK = 0.4$ and $AK\tau_2^2/\tau_1 = 2$. The upper bound of Eq. (3-33) is $\Omega_0/AK < 0.693$.

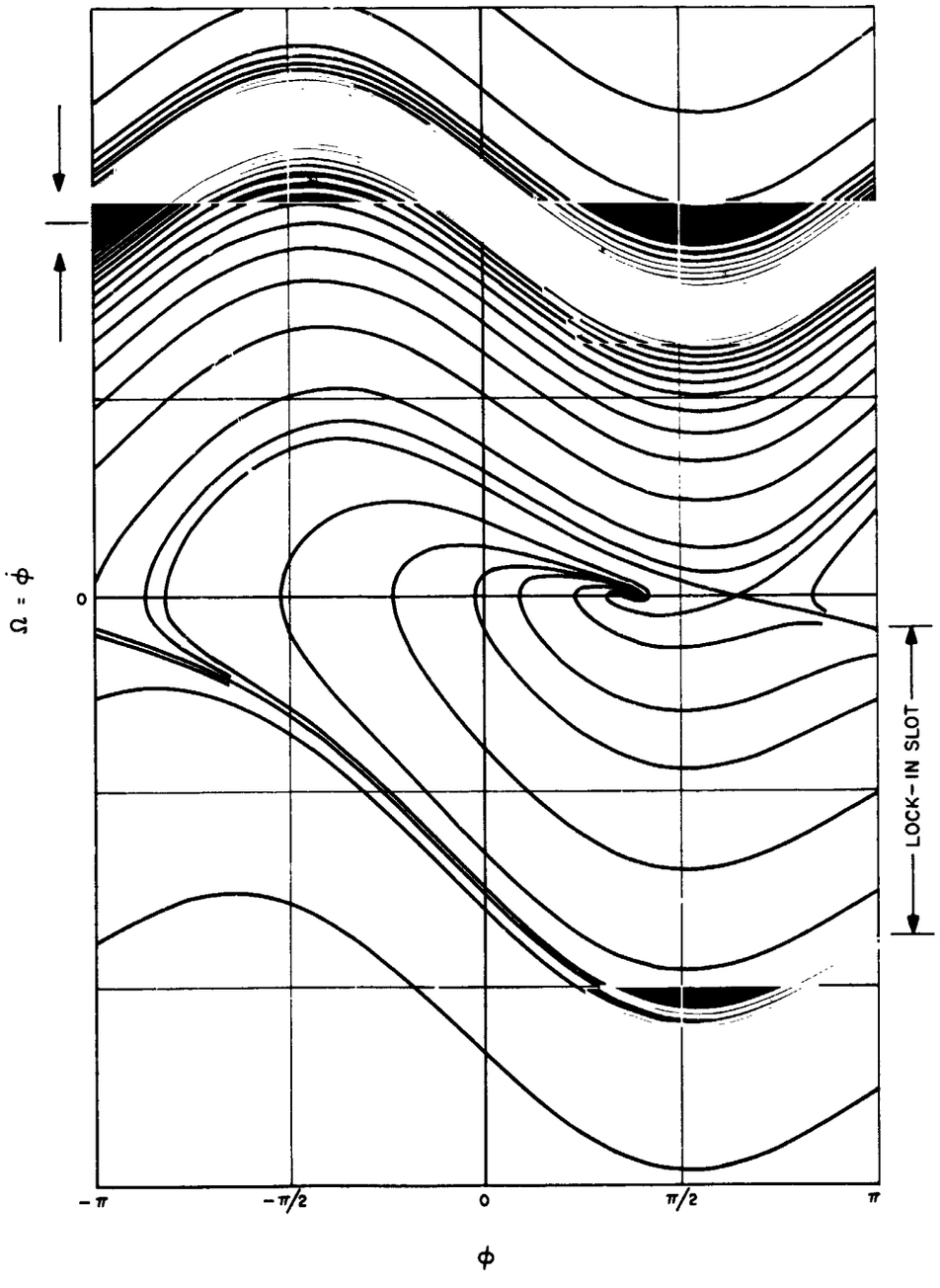


Fig. II-3-4. Lock-in behavior of a second-order loop with imperfect integrator, $F(s) = (1 + \tau_2 s)/(1 + \tau_1 s)$, for $\Omega_0/AK = 0.9$ and $AK\tau_2^2/\tau_1 = 2$. The upper bound of Eq. (3-33) is $\Omega_0/AK < 0.693$. Note that lock-in occurs only when the trajectory happens to pass through the "slot." Otherwise, the trajectory enters the periodic frequency lag region shown.

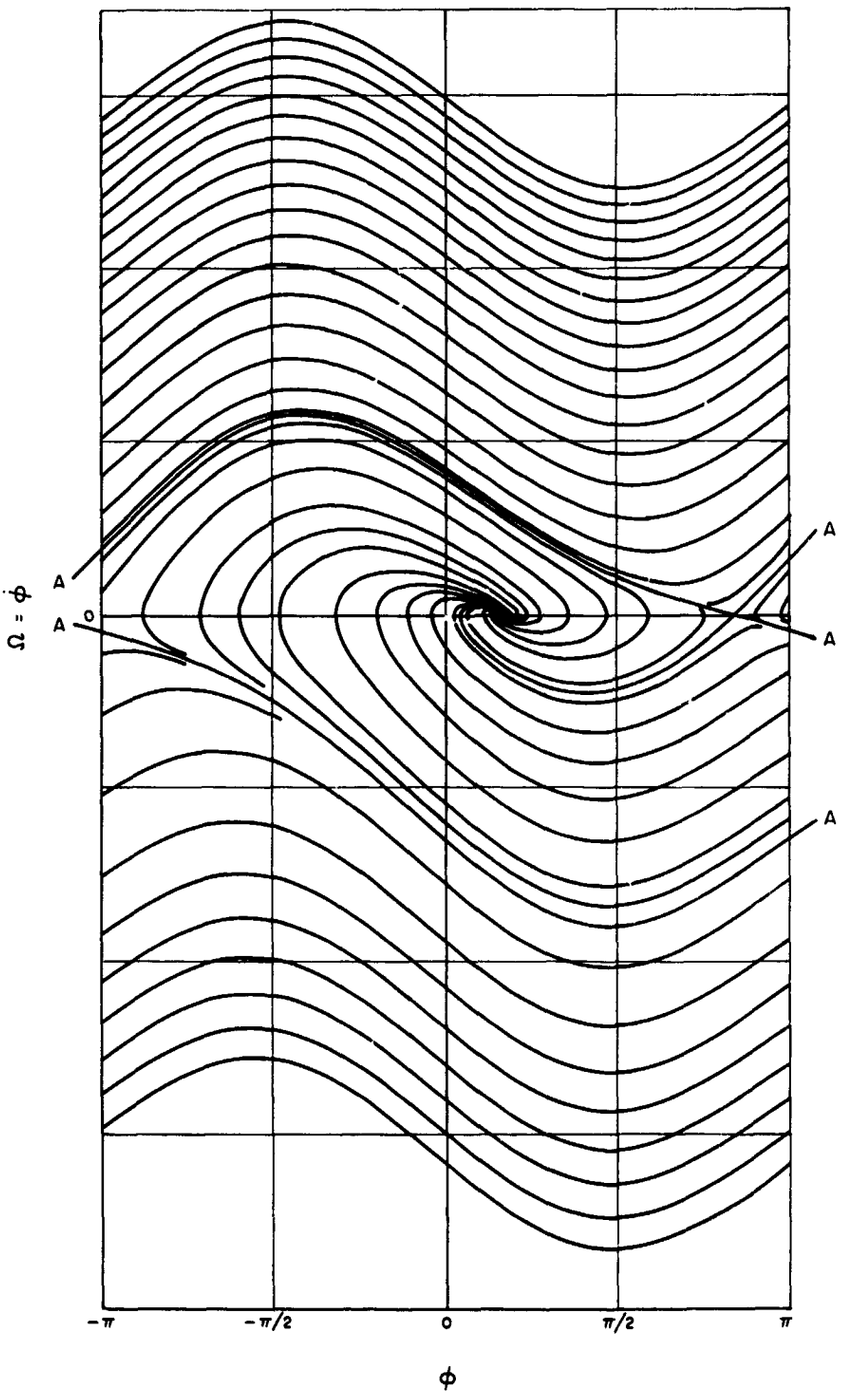


Fig. II-3-5. Phase-plane trajectory of a second-order loop with perfect integrator to a doppler-rate input Λ_0 for $AK\,\tau_2^2/\tau_1 = 2$, and $\Lambda_0 = AK/2\tau_1$

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

With a loop filter of the form

$$F(s) = \frac{1 + \tau_2 s}{\tau_1 s} \tag{3-41}$$

where τ_1 and τ_2 are in seconds, the doppler polynomial $d(t) = \theta_0 + \Omega_0 t + \frac{1}{2}\Lambda_0 t^2$ produces a constant steady-state error given by

$$\phi_{ss} = \sin^{-1}\left(\frac{\Lambda_0}{\beta^2}\right) = \sin^{-1}\left(\frac{\Lambda_0 (r+1)^2}{4r W_L^2}\right) \tag{3-42}$$

where β 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)

$$\begin{aligned} |\Lambda_0| &< k(r)\,\beta^2 \\ \operatorname{sgn}[\Omega(0)] &= -\operatorname{sgn}(\Lambda_0) \end{aligned} \tag{3-43}$$

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 $\Omega_0 \neq 0$ and $\Lambda_0 = 0$, the equations above restrict the VCO input sweep e by

$$\left|K_{VCO}\frac{de}{dt}\right| < k(r)\,\beta^2 \tag{3-44}$$

These relations are approximately true for the passive loop of Section 3-D when $\tau_1 \gg \tau_2$.

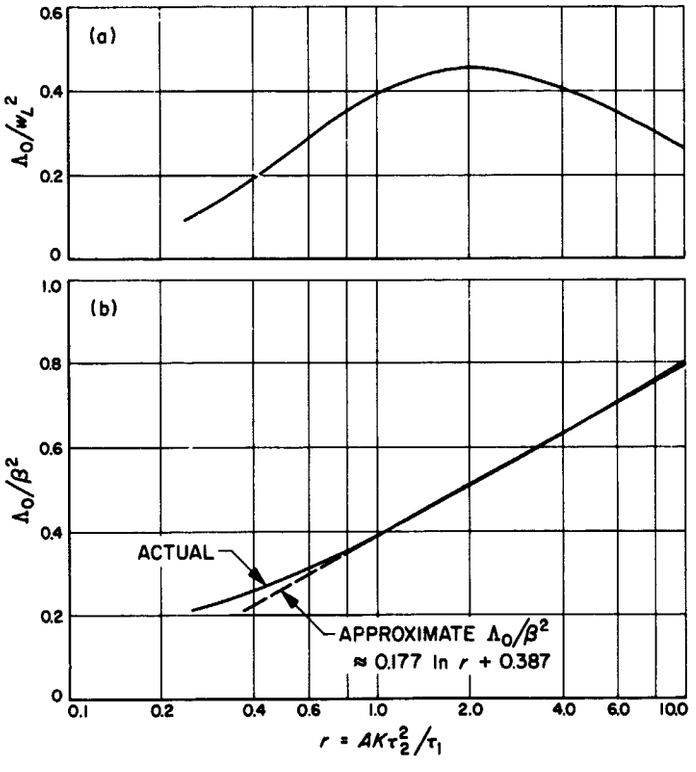


Fig. II-3-6. Normalized doppler rates Λ_0/β^2 and Λ_0/w_L^2 for which lock is guaranteed in the absence of noise, as a function of the loop parameter $v = AK\tau_2^2/\tau_1$. Note that for a fixed bandwidth w_L, the optimum value of r is about two ($\zeta = 0.707$).

CHAPTER 4

BEHAVIOR OF PHASE-LOCKED LOOPS WITH STOCHASTIC INPUTS

With input and VCO noises $n(t)$, $n_v(t)$, the basic equation is

$$\hat{\theta}(t) = \frac{AKF(p)}{p} \sin\left[\theta(t) - \hat{\theta}(t)\right] + \frac{KF(p)}{p} n(t) + \frac{K_{VCO}}{p} n_v(t) \tag{4-4}$$

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 2π, however, the resulting steady-state phase error process is stationary.

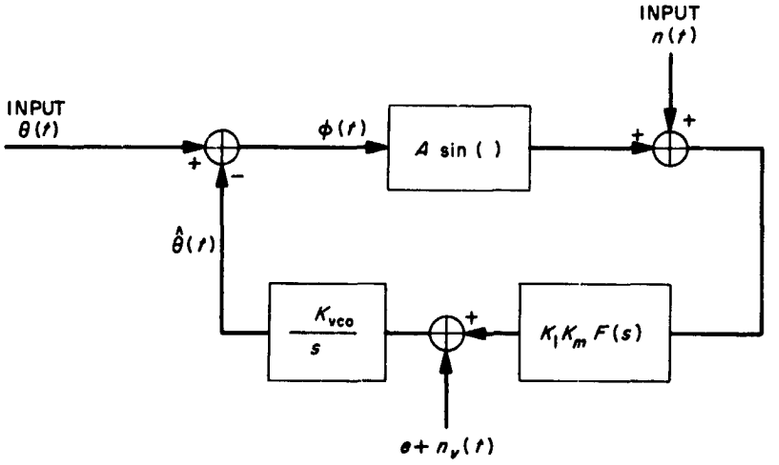


Fig. II-4-1. Exact mathematical equivalent of the phase-locked loop. Sources of external, as well as internal, noises are shown; VCO tuning voltage e is also indicated.

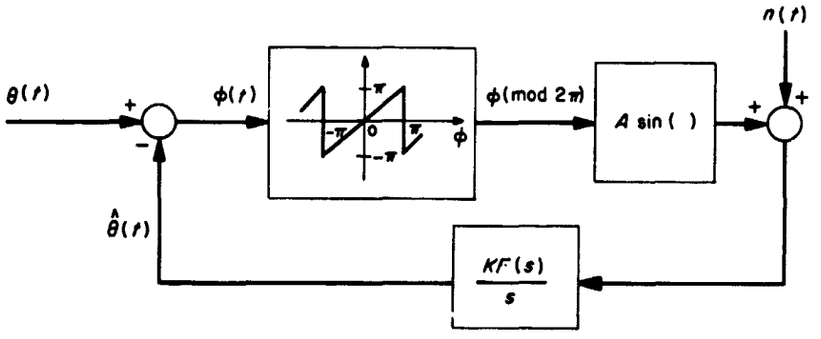


Fig. II-4-2. Equivalent exact mathematical model of phase-locked loop, with explicit reduction of ϕ (mod 2π). The VCO has been replaced by K_{VCO}/s (tuning bias and VCO noise omitted).

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 $\hat{\theta}(t)$,

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

omitting VCO noise. An equivalent circuit appears in Fig. II-5-1. Hence we can define a *loop transfer function*

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

$$F(s) = \frac{sL(s)}{AK[1 - L(s)]}$$

and $L(s)$ has some fiducial bandwidth w_L (see (2-37)).

There are two effects in $\theta(t)$: 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* Σ^2 is of the form

$$\Sigma^2 = \mu^2(t) + \delta^2 + \sigma^2 \quad (\text{rad}^2) \tag{5-6}$$

where $\mu(t)$ is transient (or doppler) distortion due to $d(t)$, δ^2 is modulation distortion, and σ^2 is phase noise:

$$\delta^2 = \frac{1}{2\pi}\int_{-\infty}^{+\infty} |1 - L(j\omega)|^2 S_{\psi\psi}(j\omega)\, d\omega \quad (\text{rad}^2) \tag{5-7}$$

$$\sigma^2 = \frac{N_0 w_L}{A^2} \quad (\text{rad}^2).$$

Total transient distortion ϵ_T^2 is defined by the integral

$$\epsilon_T^2 = \int_0^\infty \mu^2(t)dt \quad (\text{rad-sec})^2 \tag{5-9}$$

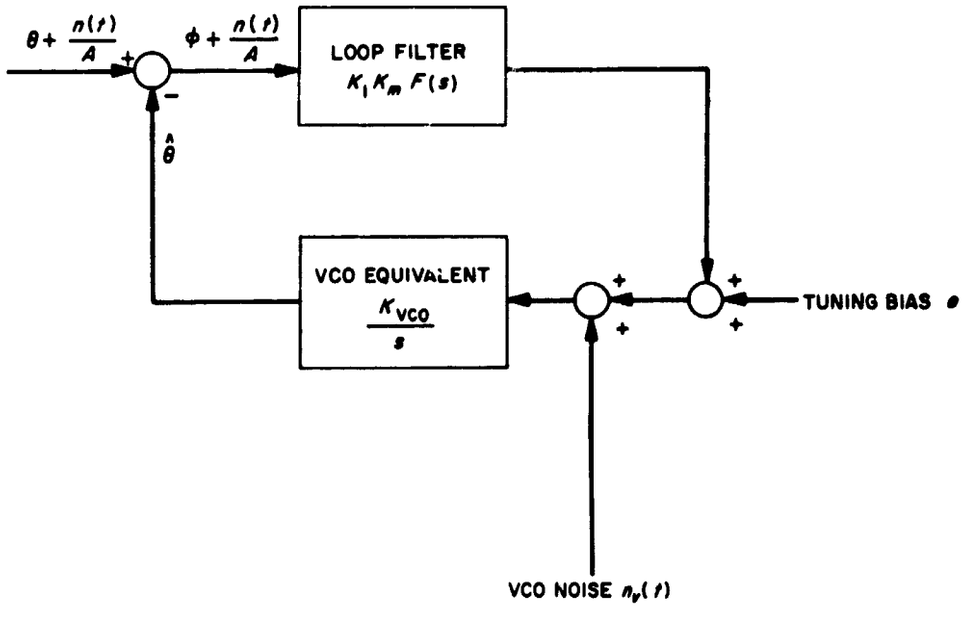


Fig. II-5.1. The linearized model of the phase-locked loop

and the *total phase-error* Σ_T^2 is defined as

$$\Sigma_T^2 = \lambda^2 \epsilon_T^2 + \delta^2 + \sigma^2 \quad (\text{rad}^2). \tag{5-22}$$

Here λ^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 $F(s) = 1$,

$$w_L = W_L = \frac{AK}{2} \quad (\text{cps}) \tag{5-11}$$

$$L(s) = \frac{2w_L}{s + 2w_L}. \tag{5-12}$$

2. Second-Order Loop, Passive Integrator

With a loop filter of the form (Fig. II-5-2),

$$F(s) = \frac{1 + \tau_2 s}{1 + \tau_1 s} \tag{5-13}$$

$$L(s) = \frac{1 + \tau_2 s}{1 + \tau_2\left(1 + \frac{\tau_2}{r\tau_1}\right)s + (\tau_2^2/r)s^2}$$

where r is the loop parameter ratio,

$$r = \frac{AK\tau_2^2}{\tau_1}. \tag{5-14}$$

We shall assume $r\tau_1 \gg \tau_2$ in the formulas that follow. The loop fiducial bandwidth is

$$w_L = \frac{r + 1}{2\tau_2} \quad (\text{cps}). \tag{5-18}$$

The loop has a damping coefficient ζ and natural frequency β given by

$$\zeta = \frac{r^{1/2}}{2} \tag{5-20}$$

$$\beta = \frac{2r^{1/2}}{r + 1} w_L \quad (\text{rad/sec}). \tag{5-20}$$

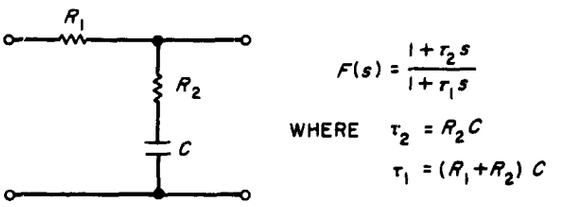


Fig. II-5-2. Passive-integrator loop filter

The maximum gain of the loop occurs at ω_{max}

$$\omega_{max} = \frac{1}{\tau_2}\left[(1 + 2r)^{1/2} - 1\right]^{1/2}$$

$$= \beta\left[\frac{(1 + 2r)^{1/2} - 1}{r}\right]^{1/2} \quad (\text{rad/sec}) \tag{5-15}$$

$$\to 0 \qquad (\text{as } r \to \infty)$$

$$\to \beta \qquad (\text{as } r \to 0).$$

The maximum value of L^2 of loop power gain is

$$L^2 = |L(j\omega)|^2_{max} = \frac{r^2}{2(1 + 2r)^{1/2} + r^2 - 2r - 2}. \tag{5-16}$$

These effects are shown in Fig. II-5-3.

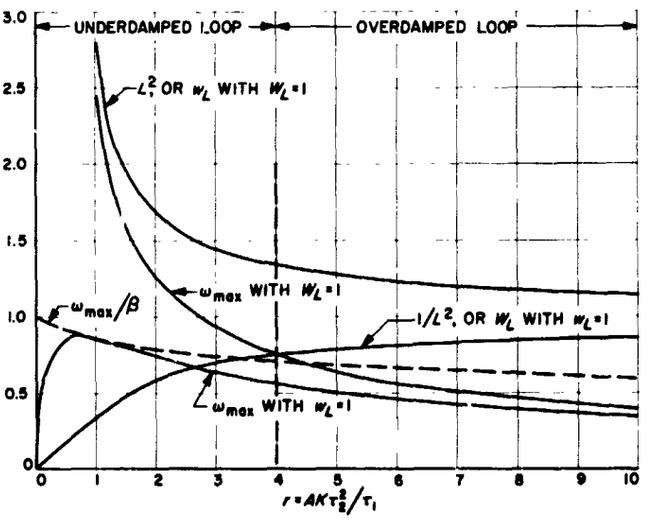


Fig. II-5-3. Variation of maximum loop response L^2, noise bandwidth W_L, fiducial bandwidth w_L, and frequency at maximum loop response, for second-order phase-locked loop, as a function of the parameter $r = AK\tau_2^2/\tau_1$

3. Second-Order Loop, Perfect Integrator

With a loop filter of the form

$$F(s) = \frac{1 + \tau_2 s}{\tau_1 s} \tag{5-21}$$

the values of w_L, ζ, β, ω_{max}, and L^2 are the same as those given in Section 5-B3 when $\tau_2/\tau_1 \to 0$ 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 $\Sigma_T^2 = \lambda^2 \epsilon_T^2 + \delta^2 + \sigma^2$ is achieved by choosing

$$L_{opt}(s) = 1 - \frac{N_0^{1/2}/A}{[S(s)]^+} \tag{5-23}$$

$$S(s) = \lambda^2 E\,[D(s)D(-s)] + S_{\psi\psi}(s) + N_0/A^2\,.$$

5-D. Effects of VCO Noise

The total phase noise in the linear loop due to input, as well as to VCO noise, is

$$\sigma^2 = \frac{N_0 w_L}{A^2} + \left(\frac{r+1}{4r}\right)\frac{N_{0v}}{w_L} + g(r)\frac{N_{1v}}{w_L^2} \quad (\text{rad}^2) \tag{5-27}$$

where N_{0v} is the white-noise-spectrum portion of $K_{VCO}n_v(t)$, and N_{1v} is the coefficient of the $1/f$ noise component of $K_{VCO}n_v(t)$:

$$K_{VCO}^2\, S_{n_v n_v}(j\omega) = N_{0v} + 2\pi N_{1v}/|\omega| \quad (\text{rad}^2/\text{sec}^2\ \text{cps})\,. \tag{5-25}$$

The function $g(r)$ is given in Fig. II-5-4, but to good accuracy we can use $g(r) = 1.55$ for $3 < r < 10$.

Optimum values of r and w_L that minimize σ^2 can be found: σ^2 is fairly insensitive to r, and any value in the range $3 < r < 10$ is acceptable; w_L can be found as the solution to

$$w_L^3 - 0.286\left(\frac{A^2 N_{0v}}{N_0}\right)w_L - 3.1\left(\frac{A^2 N_{1v}}{N_0}\right) = 0\,. \tag{5-29}$$

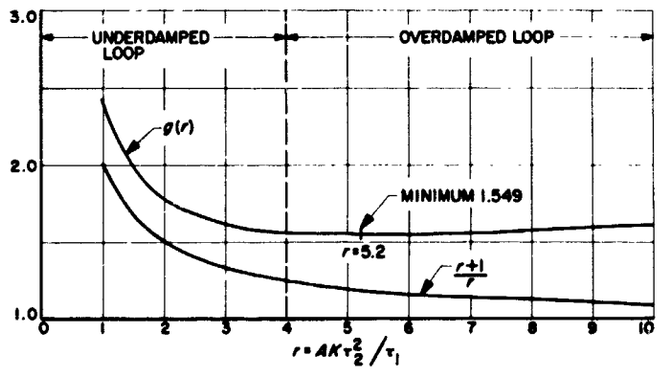


Fig. II-5-4. Factors governing relative contributions of VCO noise to output phase noise

CHAPTER 6

OPTIMIZED DESIGN OF TRACKING FILTERS (LINEAR ANALYSIS)

6-A. Tracking Loop Design

The filter $L_{opt}(s)$ that minimizes Σ_T^2 when the doppler polynomial $d(t)$ has a Laplace transform $D(s)$ given by

$$D(s) = \frac{\theta_0}{s} + \frac{\Omega_0}{s^2} + \frac{\Lambda_0}{s^3} + \cdots = \frac{Q(s)}{s^N} \qquad (6\text{-}1)$$

(the degree of $Q(s)$ is less than N) is given by the Wiener formula

$$L_{opt}(s) = 1 - \frac{s^N}{\left[(-1)^N s^{2N} + \left(\frac{A^2\lambda^2}{N_0}\right) E\,[Q(s)Q(-s)]\right]^+}. \qquad (6\text{-}2)$$

6-B. Optimum Filter for Random Phase Offset

When $d(t) = \theta_0$, where θ_0 is a uniformly distributed random phase offset, we find

$$L_{opt}(s) = \frac{\lambda A \pi}{(3N_0)^{1/3}}\left[\frac{1}{s + \frac{\lambda A \pi}{(3N_0)^{1/3}}}\right] \qquad (6\text{-}6)$$

$$W_L = w_L = \frac{\lambda A \pi}{2(3N_0)^{1/3}} \quad \text{(cps)}. \qquad (6\text{-}7)$$

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

When $d(t) = \theta_0 + \Omega_0 t$ in which θ_0 is uniformly distributed over $(-\pi, \pi)$, it follows that

$$L_{opt}(s) = \frac{s\left(2\beta^2 + \frac{\pi^2\beta^4}{3\Omega_0^2}\right)^{1/3} + \beta^2}{s^2 + \left(2\beta^2 + \frac{\pi\beta^4}{3\Omega_0^2}\right)^{1/3} s + \beta^2} \qquad (6\text{-}13)$$

where the loop natural frequency β is given by

$$\beta = \left(\frac{A\lambda\Omega_0}{N_0^{1/3}}\right). \qquad (6\text{-}12)$$

This $L_{opt}(s)$ can be realized with $F(s) = (1 + \tau_2 s)/\tau_1 s$ by setting

$$\left.\begin{aligned} \tau_2 &= \frac{1}{\beta}\left(2 + \frac{\pi^2\beta^2}{3\Omega_0^2}\right)^{1/3} = \frac{\pi r}{\Omega_0[3r(r-2)]^{1/2}} \\ \frac{AK}{\tau_1} &= \beta^2 = \frac{3\Omega_0^2}{\pi^2}(r-2) \\ r &= 2 + \frac{\pi^2\beta^2}{3\Omega_0^2} \\ w_L &= \frac{r+1}{2\pi r}\,\Omega_0[3r(r-2)]^{1/2} \end{aligned}\right\} \theta_0 \text{ is random.} \qquad (6\text{-}14)$$

When $\theta_0 = 0$ above (not random), the value of β given in (6-12) remains unchanged, but the loop parameters are somewhat different:

$$\left.\begin{aligned} \tau_2 &= \frac{2^{1/3}}{\beta} \\ \frac{AK}{\tau_1} &= \beta^2 \\ r &= 2 \quad (\text{so } \zeta = 0.707) \\ w_L &= \frac{3\beta}{2(2)^{1/3}} \end{aligned}\right\} \text{when } \theta_0 = 0. \qquad (6\text{-}18)$$

1. Choice of Parameters

If we design the loop optimally to lock onto an Ω_0 appearing at the bandwidth edge (i.e., $\Omega_0 = 2\pi b_L$), the proper values of r and ζ for θ_0 random are

$$\begin{aligned} r &= 2.282 \\ \zeta &= 0.755. \end{aligned} \qquad (6\text{-}19)$$

Hence the loop configuration is described by

$$\begin{aligned} \tau_2 &= 1.643/w_L \\ \frac{AK}{\tau_1} &= \frac{w_L^2}{1.180} \end{aligned} \qquad (6\text{-}20)$$

$$L_{opt}(s) = \frac{1 + 1.643/w_L s}{1 + (1.643/w_L)s + (1.18/w_L^2)s^2}. \qquad (6\text{-}21)$$

The resulting transient error is

$$\Sigma^2_{T\,(\text{opt})} = \frac{\pi^3 (r-1)}{6\Omega_0 \left[3r(r-2)^3\right]^{1/2}} \qquad (6\text{-}25)$$

$$= 0.537/w_L \quad \text{for } r = 2.282 \text{ and } \Omega_0 = \pi w_L .$$

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

If there is a small doppler-rate term Λ_0(rad/sec^2) in $d(t)$, the loop with $F(s) = (1 + \tau_2 s)/\tau_1 s$ has a steady-state error

$$\phi_{ss\,(\text{dopp rate})} = \frac{\Lambda_0 \tau_1}{AK} = \frac{\Lambda_0 (r+1)^2}{4r\, w_L^2} \quad \text{(rad)} . \qquad (6\text{-}28)$$

When the imperfect-integrator loop filter $F(s) = (1 + \tau_2 s)/(1 + \tau_1 s)$ is used, there is a steady-state growth in phase error

$$\phi_{ss\,(\text{dopp rate})} \to \frac{\Lambda_0}{AK}\, t = \frac{\Lambda_0 (r+1)^2}{4r\, w_L^2}\left(\frac{t}{\tau_1}\right) \quad \text{(rad)} . \qquad (6\text{-}27)$$

These values are minimized, for a fixed w_L^2 and τ_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 $\vartheta_0 = 0$) 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.

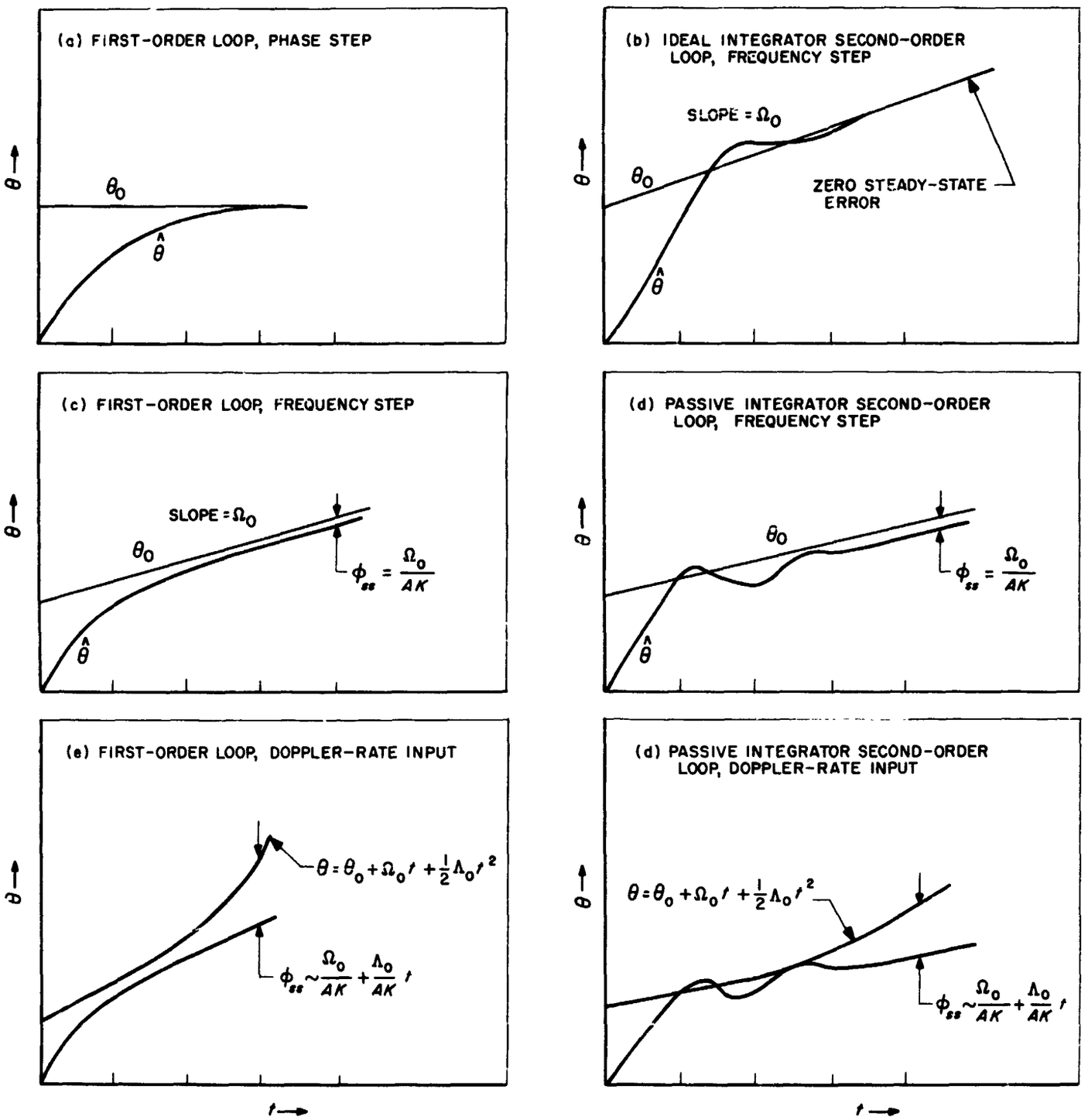


Fig. II-6-1. Response of first- and second-order loops to various inputs

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

$$c(t) = C(p)\left[a(t) + \left(\frac{20 \log e}{e_G}\right) K_D n_i(t) + 20 \log \cos \phi(t) - K_R\right] \tag{7-8}$$

where

$$a(t) = 20 \log A(t) \quad \text{(db-volts}^2\text{)}$$

$$C(p) = \frac{K_{AGC}}{K_1[1 + K_{1GC} Y(s)]}$$

$$K_{AGC} = \frac{K_A e_G K_C}{20 \log e}$$

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

$$a = K_{rec} + 20 \log\left(\frac{K_D}{e_G}\right) + \left(\frac{1}{C(0)}\right) c + (10 \log e)\sigma^2 \quad \text{(db-volts}^2\text{)} \tag{7-10}$$

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

$$\text{var}\,[c] = \left[(20 \log e)^2 \left(\frac{N_0 w_c}{A^2}\right) + 2(10 \log e)^2 \sigma^4 \frac{w_c}{w_L}\right] C^2(0)\,. \tag{7-9}$$

where σ^2 is the phase noise variance (5-7).

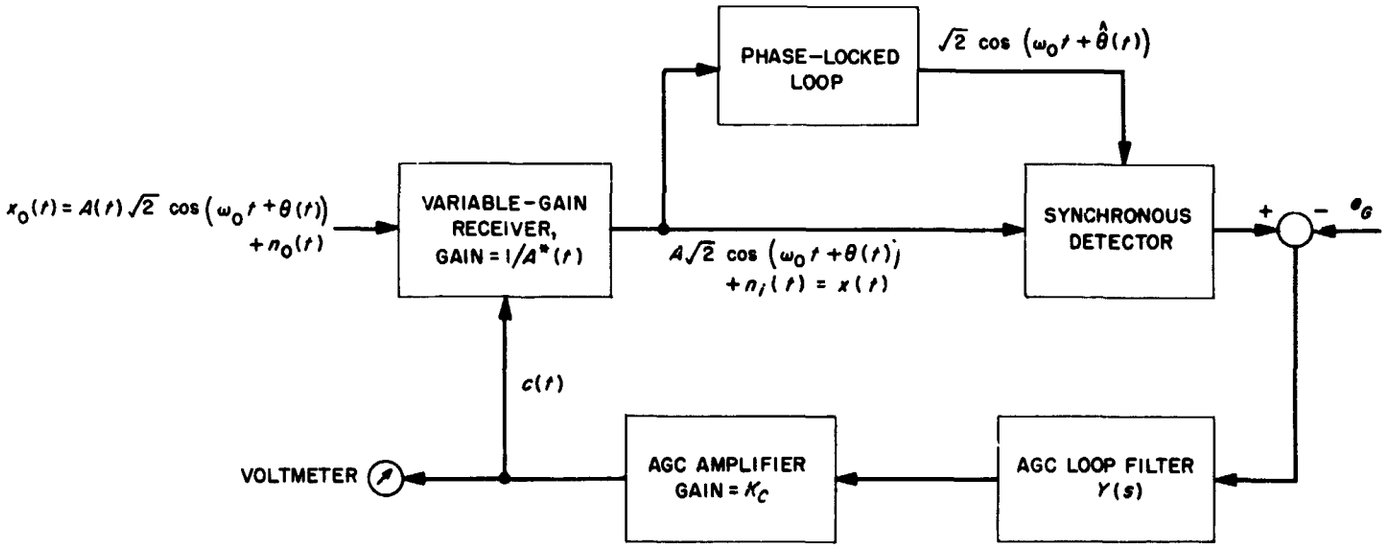


Fig. II-7-1. A synchronous-detector AGC loop, using a phase-locked loop to provide a coherent reference. (The actual phase-locked loop may be part of the gain-controlled receiver.)

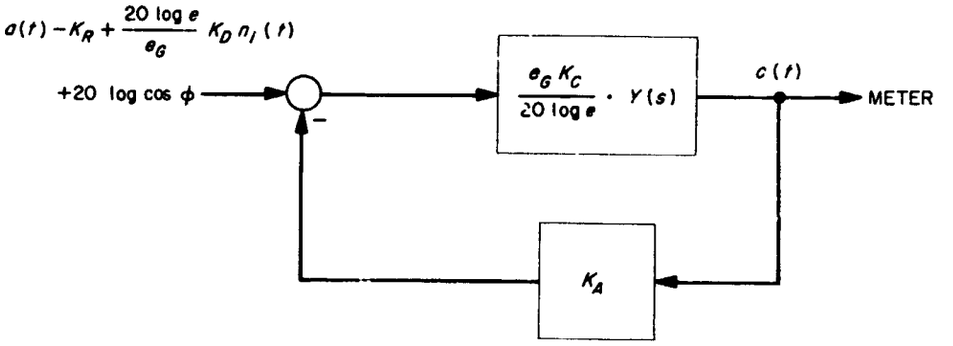


Fig. II-7-2. Equivalent diagram of an AGC loop. Linearized analysis follows the assumptions that $A^*(t)$ is exponentially related to $c(t)$ and that $a(t) \approx a^*(t)$. The input is $a(t) = 20 \log A(t)$. Adjusted loop gain $a^*(t) = 20 \log [A^*(t)\, e_G/K_D]$.

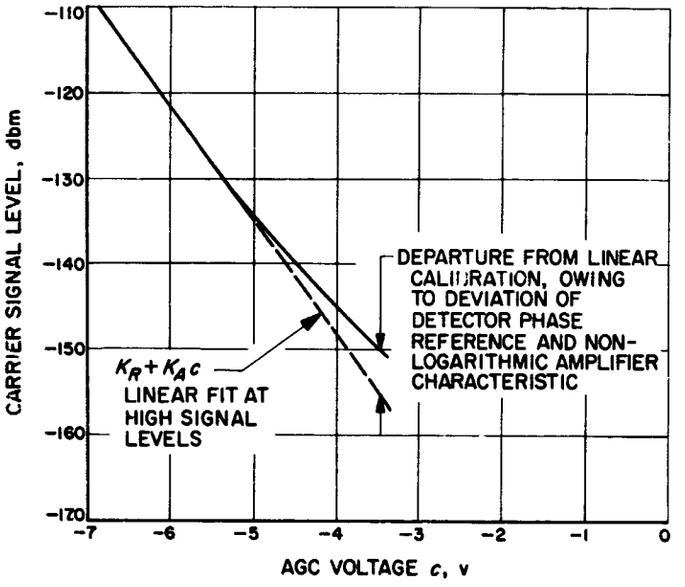


Fig. II-7-3. Measured AGC curve showing departure from linear behavior

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

$$\hat{\theta} = \alpha K_d K_{VCO} M \frac{F(p)}{p} \left\{ \sin\left[\theta - \hat{\theta} + (1 + M_1)\theta_1 + \theta_2\right] + \frac{n(t)}{\alpha} \right\} + \frac{MK_{VCO}}{p} n_v(t). \tag{8-6}$$

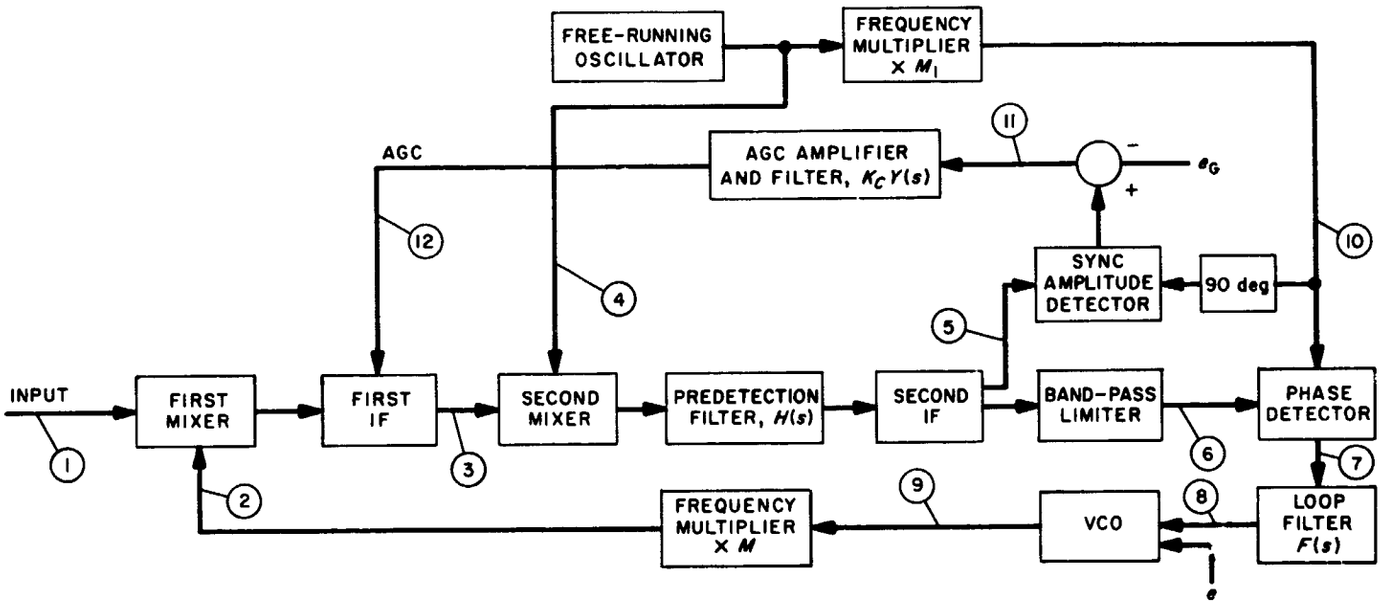
This is the same form that governs the behavior of a simple loop when we set

$$AK = \alpha K_d K_{VCO} M F \quad (\text{sec}^{-1})$$
$$\phi = \theta - \hat{\theta} + (1 + M_1)\theta_1 + \theta_2 \quad (\text{rad}). \tag{8-8}$$

The term F is the dc gain of $F(s)$, viz., $F(0)$.

8-B. Effects of Band-Pass Limiting

The parameter α above is the signal suppression factor derived by Davenport (Fig. II-8-2). In addition to this,



(1) $x_0(t) = \sqrt{2}\, A(t) \cos\left[\omega_0 t + \theta(t)\right] + n_0(t)$

(2) $v(t) \sim \sqrt{2} \cos\left[\omega_{h1} t + \hat{\theta}(t)\right]$

(3) $x_{I1}(t) = K_{I1}\left\{\sqrt{2} \cos\left[(\omega_0 - \omega_{h1})\, t + \theta - \hat{\theta}\right] + n_{I1}(t)\right\}$

(4) $v_2(t) \sim \sqrt{2} \cos(\omega_{h2} t - \theta_1)$

(5) $x(t) = \sqrt{2}\, A \cos\left[(\omega_0 - \omega_{h1} - \omega_{h2})\, t + \theta - \hat{\theta} + \theta_1\right] + n_i(t)$

(6) $x_\ell(t) = \frac{2\ell}{\pi}\left\{\sqrt{2}\, a \cos\left[(\omega_0 - \omega_{h1} - \omega_{h2})\, t + \theta - \hat{\theta} + \theta_1\right] + n_\ell(t)\right\}$

(7) $y(t) = K_d\left\{a \sin\left[(\omega_0 - \omega_{h1} - (1 + M_1)\,\omega_{h2})\, t + \theta - \hat{\theta} + (1 + M_1)\,\theta_1 + \theta_2\right] + n(t)\right\}$

(8) $z(t) = F(\rho)\; y(t)$

(9) $v_1(t) \sim \sqrt{2} \cos\left[\frac{\omega_{h1} t}{M} + \frac{K_{VCO}\, F(\rho)}{\rho}\, y(t)\right]$

(10) $v_3(t) \sim \sqrt{2} \sin(M_1 \omega_{h2} t + M_1 \theta_1 - \theta_2)$

(11) $z_{AGC}(t) = -e_G + K_D\left\{A \cos\left[(\omega_0 - \omega_{h1} - (1 + M_1)\,\omega_{h2})\, t + \theta - \hat{\theta} + (1 + M_1)\,\theta_1 + \theta_2\right] + n(t)\right\}$

(12) $c(t) = K_C\, Y(\rho)\, z_{AGC}(t)$

Fig. II-8-1. The double-heterodyne receiver, with equations for signals at each point

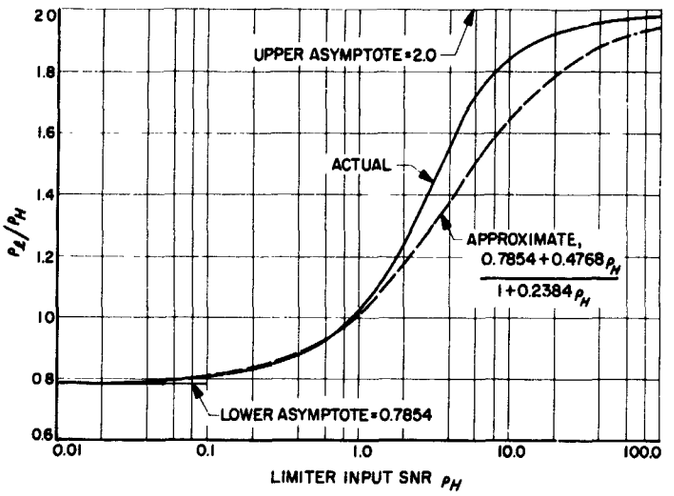


Fig. II-8-2. Davenport's band-pass limiter zonal SNR curve

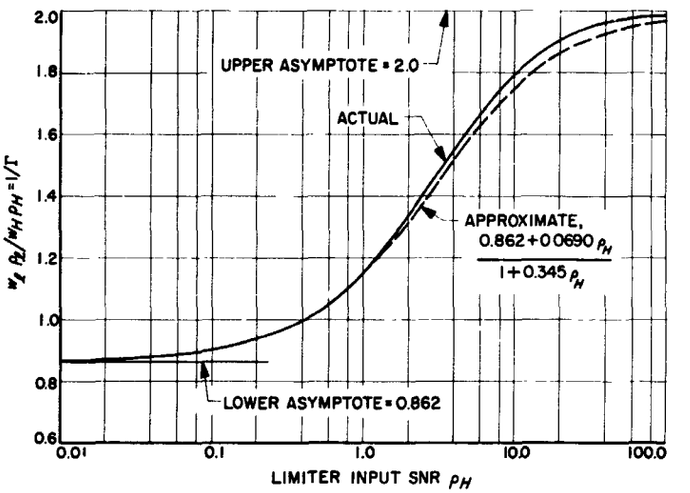


Fig. II-8-3. The ratio of band-pass limiter output signal-to-noise spectral density to that at input. The reciprocal of this curve gives the limiter performance factor Γ

the ratio A^2/N_0 at the limiter input comes out of the limiter as $A^2/N_0\Gamma$, where Γ is a factor shown in Fig. II-8-3. With the loop filter $F(s) = F \cdot (1 + \tau_2 s)/(1 + \tau_1 s)$, we thus have (linear theoretic approximation):

$$\sigma^2 = \frac{N_0 w_L}{A^2}\Gamma = \frac{\Gamma}{m(1+r_0)}\left(1 + \frac{\alpha}{\alpha_0} r_0\right) \quad (\text{rad}^2) \tag{8-19}$$

$$r = \frac{\alpha K_d K_{VCO} M F \tau_2^2}{\tau_1} = \frac{\alpha}{\alpha_0} r_0 \tag{8-14}$$

$$w_L = \frac{1+r}{2\tau_2} = w_{L_0}\left[\frac{1 + \left(\frac{\alpha}{\alpha_0}\right) r_0}{1 + r_0}\right] \quad (\text{cps}) \tag{8-15}$$

$$\zeta = \frac{r^{1/2}}{2} = \left(\frac{\alpha}{\alpha_0}\right)^{1/2} \zeta_0 \tag{5-20}$$

$$N_0 = \frac{kT}{2} \quad (\text{watts/cps}) \tag{2-65}$$

$$\Gamma = \frac{1 + 0.345\rho_H}{0.862 + 0.690\rho_H} \tag{8-18}$$

$$A_0^2 = N_0 w_{L_0} \quad (\text{watts}) \tag{10-1}$$

$$m = \left(\frac{A}{A_0}\right)^2 \tag{10-2}$$

$$\rho_H = m\left(\frac{w_{L_0}}{w_H}\right) \tag{10-3}$$

$$\alpha = \left(\frac{0.7854\rho_H + 0.4768\rho_H^2}{1 + 1.024\rho_H + 0.4768\rho_H^2}\right)^{1/2} \tag{8-13}$$

$$U = \frac{\pi(K_d K_{VCO} M F)^2 \tau_2^3}{8\tau_1^2 w_H} = \frac{r_0^2}{r_0 + 1} \tag{10-5}$$

$$r_0 = \frac{U}{2}\left[1 + \left(1 + \frac{4}{U}\right)^{1/2}\right]. \tag{10-4}$$

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

9-A. The Spectral Equation

The phase-noise spectrum can be approximated by

$$S_{\phi\phi}(s) = \frac{K^2 N_0 F(s) F(-s)}{-s^2 + \eta AK\,[sF(-s) - sF(s)] + (AK\gamma)^2\, F(s)F(-s)} \quad (\text{rad}^2/\text{cps}) \qquad (9\text{-}11)$$

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

$$\gamma = \frac{1 - e^{-\sigma^2}}{\sigma^2} \qquad (9\text{-}19)$$

when $\sigma^2 \leq 1$ (rad²). Then (9-11) can be integrated to give σ^2:

$$\sigma^2 = \frac{N_0 w_{L(\text{eq})}}{A^2 \gamma^2} \quad (\text{rad}^2) \qquad (9\text{-}28)$$

where $w_{L(\text{eq})}$ is the loop equivalent bandwidth,

$$w_{L(\text{eq})} = \frac{1}{2\pi} S_{\phi\phi}(0) \int_{-\infty}^{+\infty} S_{\phi\phi}(j\omega)\, d\omega \quad (\text{cps}). \qquad (9\text{-}27)$$

9-D. Behavior of First-Order Loop

For $F(s) = 1$, a better value of γ than (9-19) is (see Fig. II-9-2)

$$\gamma_{fp} = \left(\frac{\sinh \sigma^2}{\sigma^2\, e^{\sigma^2}}\right)^{1/2} \qquad (9\text{-}18)$$

for $\sigma^2 \leq 1$. The phase-noise variance is thus a transcendental equation in σ^2, whose results are given in Fig. II-9-3,

$$\sigma\, e^{-\sigma^2/2}\, (\sinh \sigma^2)^{1/2} = \frac{N_0 K}{2A} = \frac{N_0 w_L}{A^2}. \qquad (9\text{-}25)$$

At $\sigma^2 = 1$, the value of $N_0\, w_L/A^2$ is 0.357.

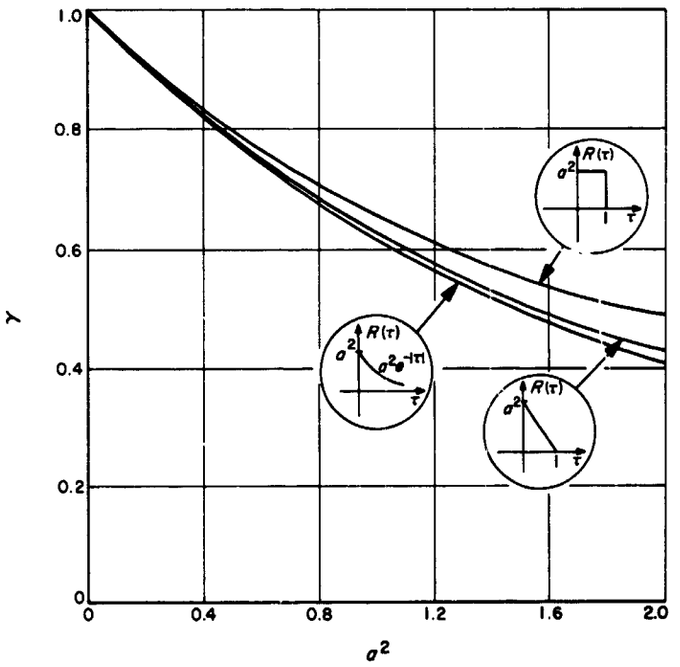


Fig. II-9-2. Variation of the parameter γ as a function of the Gaussian variance σ^2, for various forms of $R_{\phi\phi}(\tau)$. Note that, for $\sigma^2 < 1$, there is not a significant dependence on the form of $R_{\phi\phi}(\tau)$.

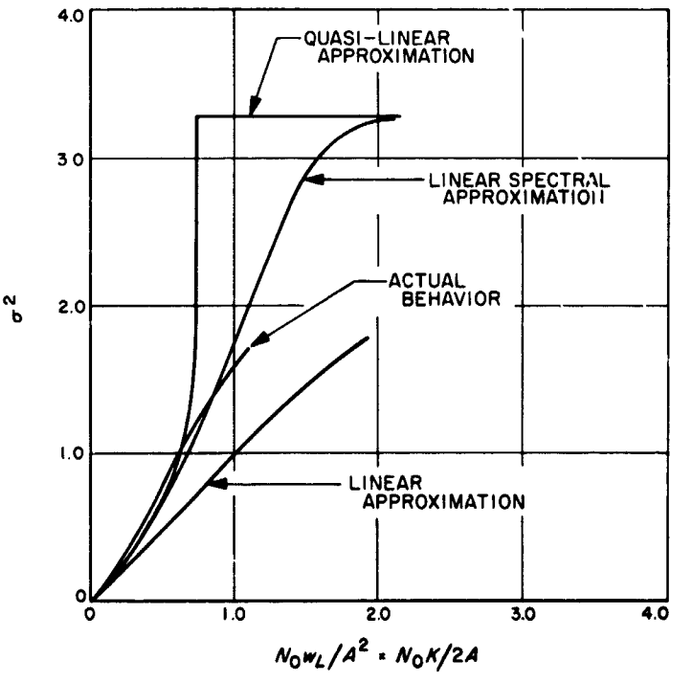


Fig. II-9-3. Comparison of linear, quasi-linear, and linear-spectral approximate methods with the actual behavior of the first-order loop

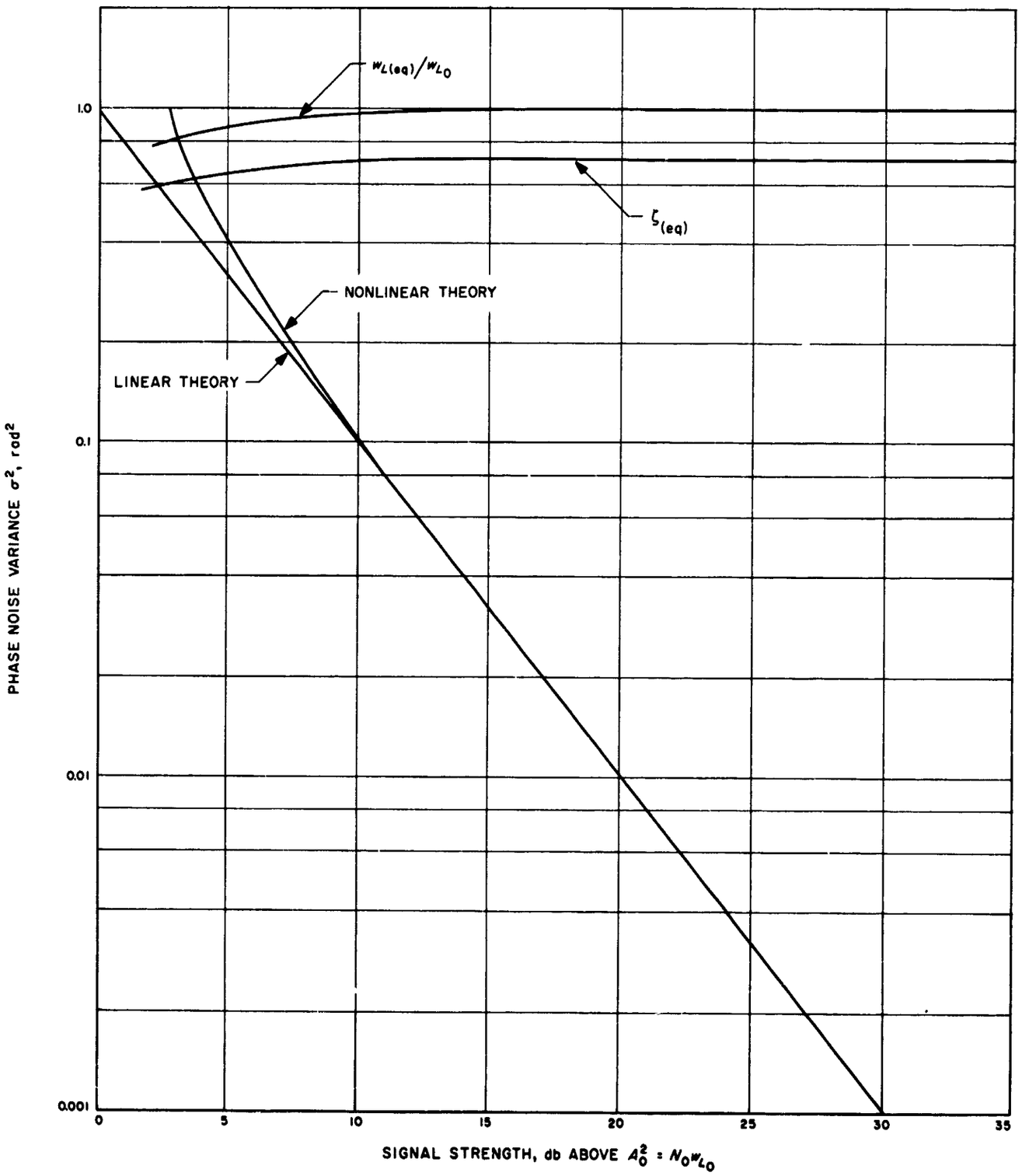


Fig. II-9-4. Comparison of linear and nonlinear theories for second-order, constant linear bandwidth loop, i.e., the value of r is kept constant at $r_0 = 2$.

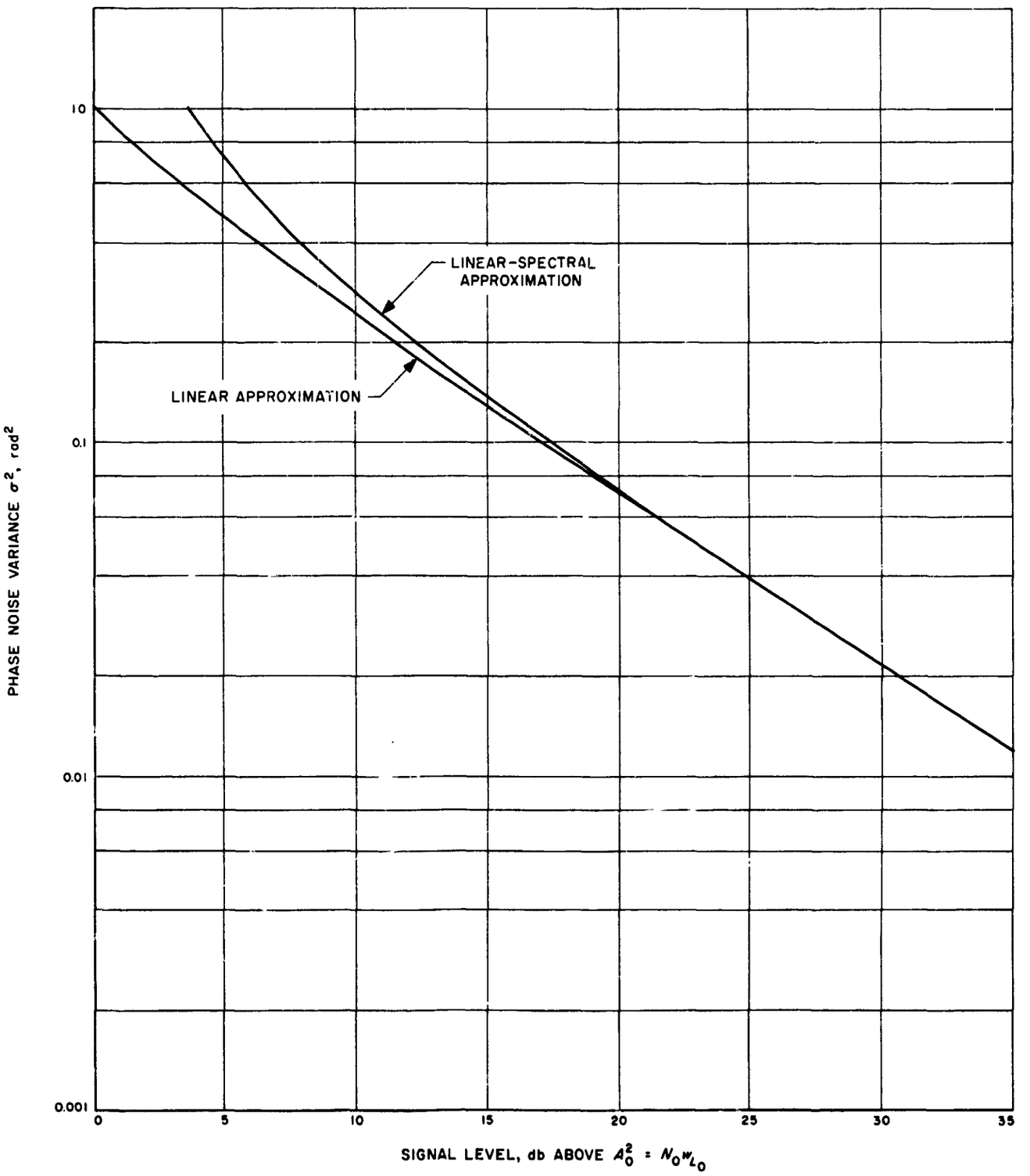


Fig. II-9-5. Comparison of phase-noise variances by linear and linear-spectral approximations. The noise density is fixed, and the signal level is varied. The value of r is taken as $r_0 = 2$ at a reference signal level of $A_0^2 = 3N_0/2\tau_2 = N_0 w_{L_0}$. Note that the ultimate roll-off is 5db/decade, rather than 10db/decade, as in Fig. 9-4. Note also that even the linear approximation produces some curvature of σ^2 near A_0^2.

9-E. Calculation of Behavior of the Second-Order Loop

With the usual passive-integrator filter second-order loop,

$$\frac{w_{L(eq)}}{w_L} = \frac{1 + r\gamma}{(1 + r)\left[1 + \dfrac{2(\gamma - \eta)}{r\gamma^2}\right]^{1/2}} \qquad (9\text{-}29)$$

$$\frac{\zeta_{eq}}{\zeta} = \gamma^{1/2}\left[1 + \frac{2(\gamma - \eta)}{r\gamma^2}\right]^{1/2}. \qquad (9\text{-}33)$$

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.

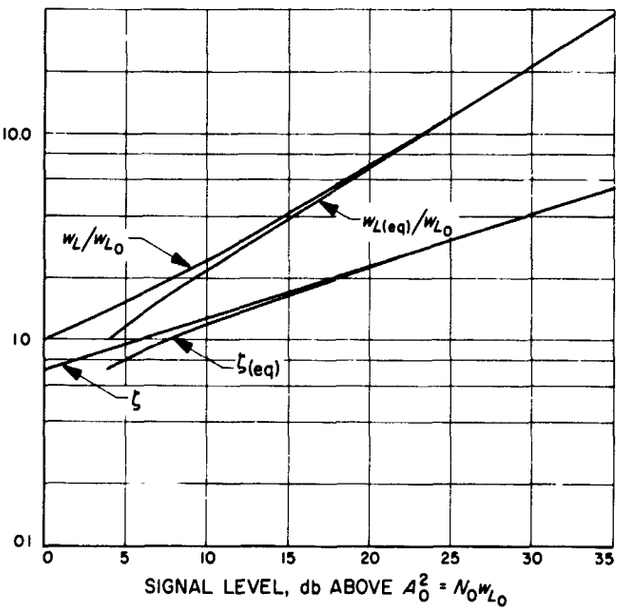


Fig. II-9-6. Variation in bandwidth and damping parameters as a function of signal strength. The value of r at a reference signal level $A_0^2 = 3N_0/2\tau_2 = N_0 w_{L_0}$ was taken as $r_0 = 2$.

CHAPTER 10

DESIGNING A DOUBLE-HETERODYNE TRACKING LOOP

10-A. Definition of Receiver Threshold

Threshold is defined as a signal level A_0^2 such that[18]

$$A_0^2 = N_0 w_{L_0} = \frac{N_0(r_0 + 1)}{2\tau_2} \quad \text{(volts}^2\text{)} \qquad (10\text{-}1)$$

(assuming $r_0\tau_1 \gg \tau_2$). The loop is not acting linearly at this signal level; nevertheless w_{L_0} is the bandwidth a linear loop would have at $A = A_0$.

10-B. Tracking Loop Performance of the Double-Heterodyne Receiver

The ratio $(A/A_0)^2$ is the receiver margin:

$$m = (A/A_0)^2 . \qquad (10\text{-}2)$$

With measured parameters W_H,$K_dK_{VCO}MF$, τ_2, and τ_1, the value of r_0 can be calculated (when $10w_{L_0} < w_H$) as

$$r_0 = \frac{U}{2}\left[1 + \left(1 + \frac{4}{U}\right)^{1/2}\right]$$

$$U = \frac{\pi(K_dK_{VCO}MF)^2\,\tau_2^3}{8\tau_1^2\,w_H} \qquad (10\text{-}4)$$

$$= \frac{r_0^2}{r_0 + 1} .$$

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 $A\gamma$, r by $r\gamma$, and w_L by $w_{L(eq)}$. The resulting equations that specify loop behavior are

$$\rho_H = m\left(\frac{w_{L_0}}{w_H}\right) \qquad (10\text{-}6a)$$

$$\alpha = \left[\frac{0.7854\rho_H + 0.4768\rho_H^2}{1 + 1.024\rho_H + 0.4768\rho_H^2}\right]^{1/2} \qquad (10\text{-}6b)$$

$$\Gamma = \frac{1 + 0.345\rho_H}{0.862 + 0.690\rho_H} \qquad (10\text{-}6c)$$

$$r = \frac{\alpha K_dK_{VCO}MF\tau_2^2}{\tau_1} = \left(\frac{\alpha}{\alpha_0}\right)r_0 = \tau_2^2\left(\frac{\alpha}{\alpha_0}\right)\beta_0^2 \qquad (10\text{-}6d)$$

[18]Since $N_0 = kT'R/2$ (volts²/cps), the threshold signal power is $P_0 = A_0^2/R = kTw_L/2 = kTb_L$ (watts).

$$\gamma = (1 - e^{-\sigma^2})/\sigma^2 \qquad (10\text{-}6f)$$

$$\eta = e^{-\sigma^2}/\sigma^2 \qquad (10\text{-}6g)$$

$$w_L = \frac{1 + r}{2\tau_2} = w_{L_0}\left[\frac{1 + \frac{\alpha}{\alpha_0}r_0}{1 + r_0}\right] \text{(cps)} \qquad (10\text{-}6h)$$

$$w_{L(eq)} = \frac{1 + r\gamma}{2\tau_2\left[1 + \frac{2(\gamma - \eta)}{\gamma^2 r}\right]^{1/2}}$$

$$= w_{L_0}\left[\frac{1 + \left(\frac{\alpha}{\alpha_0}\gamma\right)r_0}{1 + r_0}\right] \text{(cps)} \qquad (10\text{-}6i)$$

$$\zeta = \frac{r^{1/2}}{2} = \left(\frac{\alpha}{\alpha_0}\right)^{1/2}\zeta_0 = \left(\frac{\alpha}{\alpha_0}\right)^{1/2}\frac{r_0^{1/2}}{2} \qquad (10\text{-}6j)$$

$$\zeta_{eq} = \frac{(r\gamma)^{1/2}}{2}\left[1 + \frac{2(\gamma - \eta)}{r\gamma^2}\right]^{1/2} \approx \left(\frac{\alpha}{\alpha_0}\gamma\right)^{1/2}\zeta_0 \qquad (10\text{-}6k)$$

$$\sigma^2 = \frac{N_0 w_{L(eq)}\,\Gamma}{A^2\gamma^2} \approx \frac{\Gamma}{m\gamma^2}\left[\frac{1 + \left(\frac{\alpha}{\alpha_0}\gamma\right)r_0}{1 + r_0}\right] \text{(rad}^2\text{)} \qquad (10\text{-}6m)$$

$$\beta = \frac{2w_L}{r + 1} = \left(\frac{\alpha}{\alpha_0}\right)^{1/2}\frac{2w_{L_0}}{r_0 + 1} \qquad (10\text{-}6n)$$

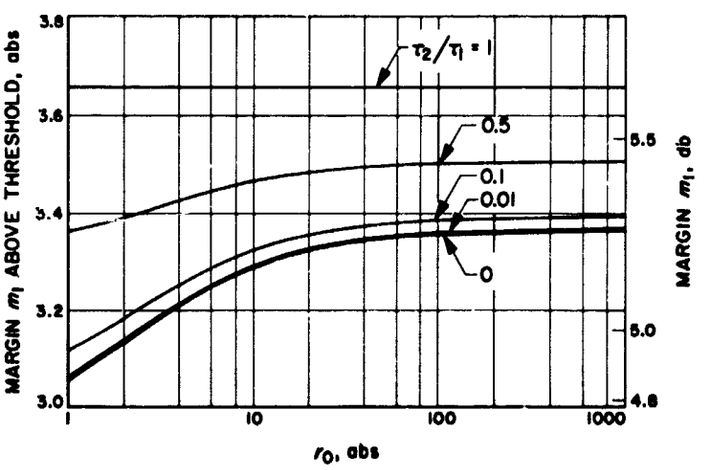


Fig. II-10-1. Variation in margin producing $\sigma^2 = 1$ as a function of threshold design parameter, $r_0 = \alpha_0 K_d K_{VCO} MF\tau_2^2/\tau_1$

10-D. The Signal Level Producing $\sigma^2 = 1$

The signal level A^2, producing $\sigma^2 = 1$, corresponds to a margin m_1 approximately equal to

$$m_1 = \left(\frac{r_0 \Gamma_1}{2\gamma_1(r_0+1)}\right)^2 \left[1 + \left(1 + \frac{4(r_0+1)}{\Gamma_1 r_0^2}\right)^{1/2}\right]^2$$

$$\approx 3.13 \text{ (for } r_0 = 2). \tag{10-8}$$

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, ω_{L_0}, K_d, K_{VCO}, MF, and ω_H specifies a typical receiver design. However, the only parameters required to plot performance are m, γ_n, ρ_{H_0} and perhaps τ_2/τ_1 (τ_2/τ_1 is not needed if $\tau_1 \gg \tau_2$). Such plots are given in Figs. II-10-2 and II-10-3.

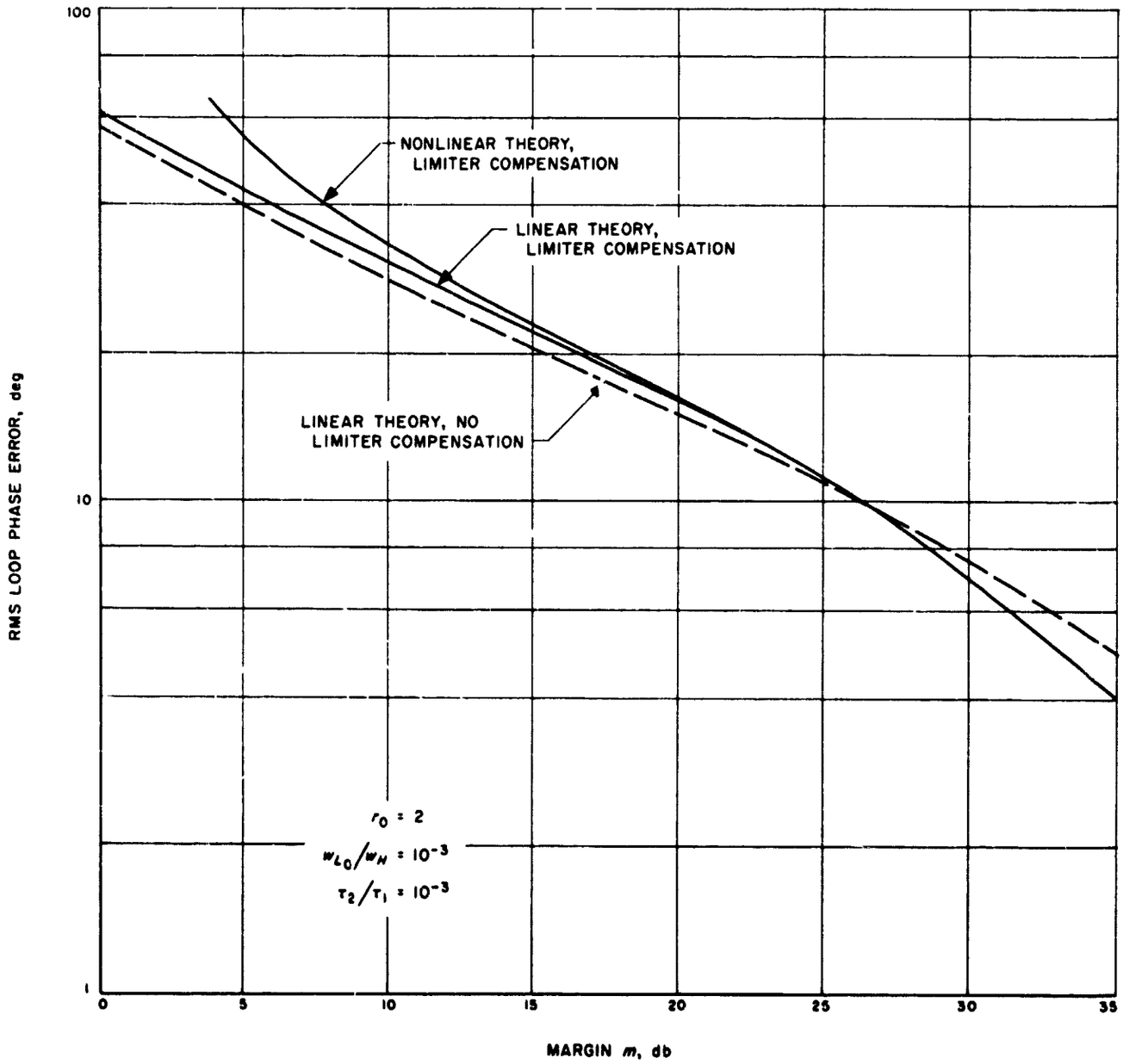


Fig. II-10-2. Comparison of linear and nonlinear approximations to loop rms phase error, as a function of loop margin

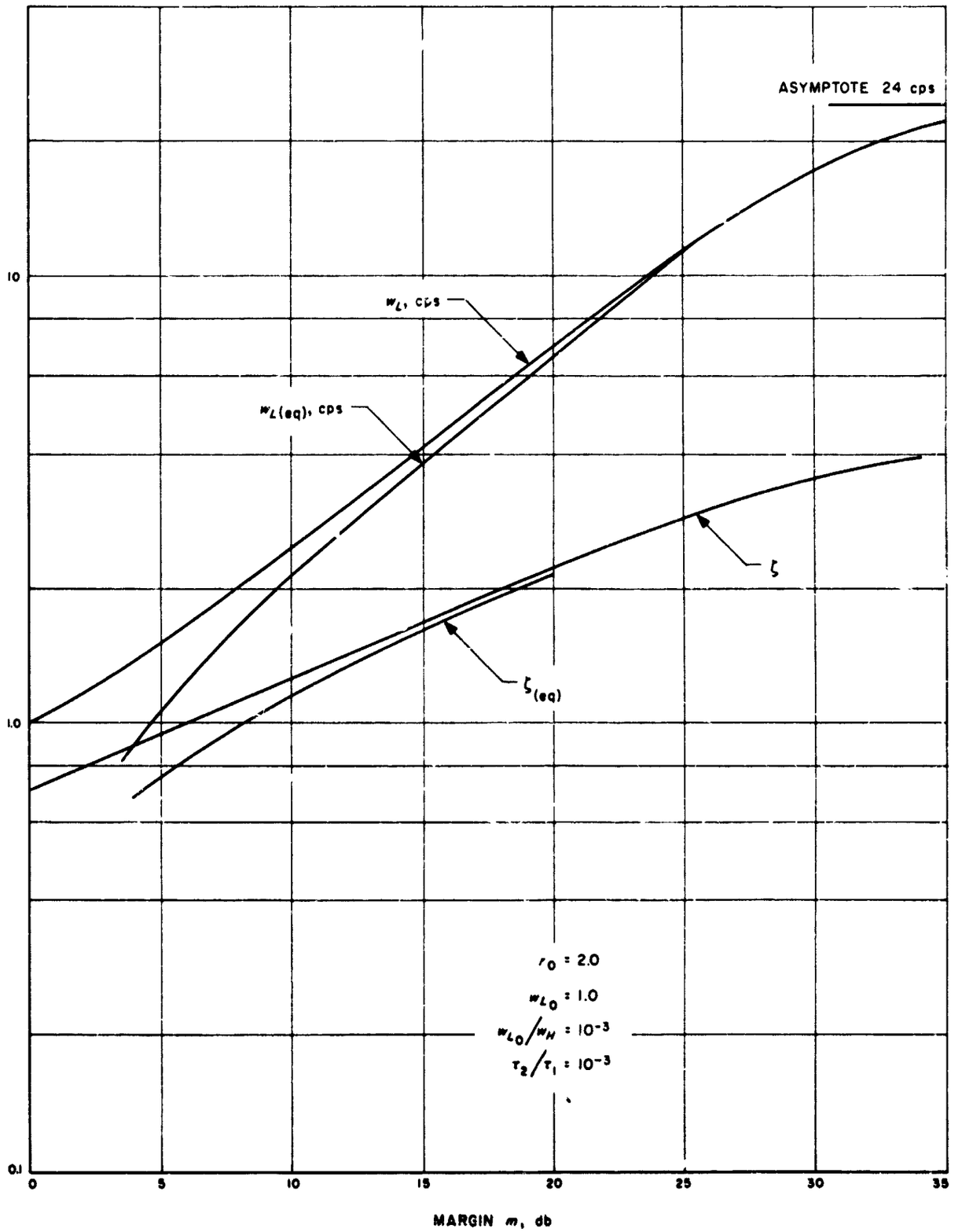


Fig. II-10-3. Variation of loop bandwidths and damping factors, as a function of loop margin

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 nonstationary changes in signal, in db-volts^2, (7-11).

$a_\psi(t)$ that part of $a(t)$ due to stationary random signal fluctuations, in db, (7-11).

$\tilde{a}_d(s)$ 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).

$A_\psi(t)$ stationary gain fluctuation part of $A(t)$, abs, (7-12).

A_1 receiver rms signal level producing $\sigma = 1$ rad, volts, (10-7).

b_H fiducial (one-sided) bandwidth of the linear transfer function $H(s)$, cps, (2-37).

B_H 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, (7-6).

$d(t)$ doppler phase function on input signal, rad, (3-11).

$d_1(t)$ 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_a AGC gain-adjust bias, volts, (7-1).

$E(\ )$ statistical expectation operator (2-1).

f frequency variable, cps.[19]

f_1 first IF frequency, cps, (8-1).[19]

f_2 second IF frequency, cps, (8-1).[19]

f_{h1} first-mixer heterodyne frequency, M times VCO output frequency, cps, (8-1).[19]

f_{h2} second-mixer heterodyne frequency, from internal oscillator, cps, (8-1).[19]

F (finite) dc gain of loop filter, abs, (8-8).

$F(s)$ linear loop-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.

$G_{xx}(j\omega)$ single-sided spectral density of $x(t)$, volts^2/cps, (2-26).

h Planck's constant, 6.625×10^{-34} joule-sec, (2-53).

$h(t)$ unit-impulse response of linear filter H, volts, (2-14).

[19] May take on a negative value.

H a filter operator, Section 2-B.

$H(s)$ linear filter transfer function, $\mathcal{L}[h(t)]$, (2-26).

$i(t)$ current, amperes, (2-54).

j $\sqrt{-1}$, the imaginary unit.

k Boltzmann's constant, 1.38×10^{-23} joule/°K, (2-53).

$k(r)$ maximum normalized lock-on doppler rate, abs, (3-43).

K $K_1 K_m K_{VCO}$, equivalent simple-loop gain, volts^{-1}sec^{-1}, (3-8).

K_1 rms VCO signal output, volts, (3-2).

K_A receiver AGC attenuation, db/volt, (7-3).

K_{AGC} AGC-loop equivalent loop gain, abs, (7-6).

K_C AGC amplifier gain, abs, (7-6).

K_d phase detector gain, volts/rad, (8-2).

K_D AGC detector gain, (volts peak out)/(volts rms in), (7-3).

K_{I1} first IF gain, abs, (8-1).

K_{I2} second IF gain, abs, (8-1).

K_m mixer gain (simple loop), volts^{-1}, (3-4).

K_R 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^2 maximum value of $|L(j\omega)|^2$, 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_1 internal oscillator frequency multiplication ratio, $(f_1/f_{h2})-1$, abs, (8-4). It must be a rational number, and may be negative.

$n_0(t)$ receiver input noise waveform, volts, Section 8-A.

$n_i(t)$ loop input noise waveform, volts, (4-1).

$n(t)$ simple loop baseband noise waveform, volts, (4-3).

$N_v(t)$ equivalent VCO noise referred to VCO input, volts, (5-24).

N noise power out of linear filter, volts^2, (2-23).

N_0 noise (two-sided) spectral density, volts^2/cps, (2-20).

N_{00} receiver input noise (two-sided) spectral density of $n_0(t)$, volts^2/cps, Section 8-A.

N_+ noise (one-sided) spectral density, volts^2/cps, (2-27).

N_{0r} VCO adjusted input white-noise density, (rad/sec)^2/cps, (5-25).

N_{1v} VCO adjusted input $1/f$ noise density, (rad/sec)^3/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_0 resonance quality factor, $\omega_0 L/R$, abs, (2-57).

$Q(s)$ numerator polynomial of $D(s)$, (6-1).

r second-order loop-parameter ratio $AK\tau_2^2/\tau_1$, 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).

$R_{xx}(\tau)$ statistical autocorrelation of stationary process $x(t)$, volts^2, (2-5).

$\mathcal{R}_{xx}(\tau)$ 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).

$S_{xx}(j\omega)$ two-sided spectral density of stationary process $x(t)$; $\mathcal{F}[R_{xx}(\tau)]$, volts^2/cps, (2-8).

$\mathcal{S}_{xx}(j\omega)$ two-sided spectral density of function $x(t)$; $\mathcal{F}[\mathcal{R}_{xx}(\tau)]$, volts^2/cps, (2-9).

$S(s)$ spectral density of total phase-error, rad²/cps, (5-23).

$S'(s)$ spectral density of total AGC error, (db-volts²)²/cps, (7-13).

t time, sec.

t_1, t_2 specific instants of time, sec.

t_{acq} first-order loop phase-acquisition time, sec, (3-27)

$t_{freq\ acq}$ second-order loop frequency-acquisition time, sec, (3-34).

T, T_N equivalent noise temperature of receiver, °K, (2-53).

T_c cathode temperature of diode or triode, °K, (2-66).

U receiver measured-parameter ratio, abs, (10-4).

$v(t)$ VCO output, volts, (3-2).

w_H two-sided fiducial bandwidth of linear filter $H(s)$, cps, (2-37).

w_L two-sided fiducial linear loop bandwidth of phase transfer function $L(s)$, cps, (5-3).

$w_{L(eq)}$ equivalent two-sided fiducial bandwidth of loop output spectrum, cps, (9-29).

W_H two-sided noise bandwidth of a linear filter $H(s)$, cps, (2-24).

W_L two-sided noise bandwidth of a linear loop whose phase transfer function is $L(s)$, cps, (5-4).

$x(t)$ arbitrary time function.

$X(s)$ Laplace transform of $x(t)$, $\mathcal{L}[x(t)]$.

$y(t)$ time function.

$Y(s)$ AGC loop-filter response, (7-5).

$z(t)$ time function.

$Z(j\omega)$ complex impedance function, ohms.

α signal voltage suppression factor, abs, (8-1).

α_0 value of α at threshold, (10-6).

β second-order loop natural frequency $(AK/\tau_1)^{1/2}$, rad/sec, (6-12).

$\gamma(s)$ sinusoidal-error spectral ratio function, abs, (9-7).

ν $\gamma(0)$.

γ_{tp} sinusoidal-error function total-power transfer ratio, abs, (9-15).

δ^2 modulation distortion, rad², (5-6).

δ_{lock} first-order loop "in-lock" constant, rad, (3-27).

$\delta_T(t)$ rectangular unit-pulse function, volts, (2-13).

$\delta(t)$ Dirac delta function, $\lim \delta_T(t)$ as $t \to \infty$, (2-13).

ϵ_T^2 total transient distortion, rad², (5-9).

ϵ_ϕ loop tracking error, rad, (8-7).

ζ linear loop damping factor, abs, (5-20).

$\zeta_{(eq)}$ nonlinear loop equivalent damping factor, abs, (9-33).

η Bussgang coefficient, abs, (9-3).

θ_i phase angle, rad.

θ_0 $\theta(0)$, initial value of loop phase offset, rad, 3-13).

$\theta(t)$ or θ input signal phase function, rad, (3-1).

$\hat{\theta}(t)$ or $\hat{\theta}$ loop estimate of $\theta(t)$, rad, (3-2).

λ^2 Lagrange multiplier, sec^{-1}, (5-22).

λ_n $d^{(n)}(0+)$, the nth doppler moment, rad/sec^n, (3-15).

μ mean of a stationary random variable, volts, (2-12).

$\mu(t)$ mean of a nonstationary random variable, volts, (2-10).

π 3.14169...

ρ_y signal-to-noise ratio of $y(t)$, abs, (2-35).

σ_x^2 variance of stationary random variable x, (x units)², (2-12).

$\sigma_x^2(t)$ variance of nonstationary random process $x(t)$ at time t, (x units)², (2-11).

σ^2 variance of loop phase-noise, rad², (5-7).

σ_{VCO}^2 variance of loop phase-noise due to noise in VCO, rad², (5-26).

τ $t_1 - t_2$, time difference, variable of $R_{xx}(\tau)$, sec, (2-4).

τ_1 second-order loop denominator time constant of $F(s)$, sec, (3-28).

τ_2 second-order loop numerator time constant of $F(s)$, sec, (3-28).

$\phi(t) = \phi$ two-heterodyne loop detector error (same as ϵ_ϕ for simple loop), rad, (5-8), (8-6).

ϕ_{ss} steady-state loop detector phase error, rad, (3-12).

ϕ_0 initial value of loop detection phase error, rad, (3-32)

Φ stationary equivalent phase-error process, rad, Section 4-B.

ω angular frequency variable, rad/sec.

ω_0 VCO short-circuit output frequency, rad/sec, (3-1).

ω_{max} frequency at which $|L(j\omega)|$ is maximum for second-order loops, rad/sec, (5-15).

ω_1 first IF frequency, rad/sec, (8-1).[19]

ω_2 second IF frequency, rad/sec, (8-1).[19]

ω_{h1} first mixer heterodyne frequency, M times VCO output frequency, rad/sec, (8-1).[19]

ω_{h2} second mixer heterodyne frequency, from internal oscillator, rad/sec, (8-1).[19]

Γ limiter performance factor, $w_l\rho_l/w_H\rho_H$, abs, (8-17).

Λ_0 initial value of doppler phase-rate, rad/sec^2, (3-13).

Σ^2 mean-square phase error, rad^2, (5-6).

Σ_T^2 total mean-square phase error, rad^2, (5-22).

$\psi(t)$ modulation phase function, rad, (3-11).

$\Omega = \Omega(t)$ $\dot{\phi}(t)$, loop frequency error, rad/sec, (3-20).

Ω_0 initial value of loop frequency offset rad/sec, (3-12).

Ω_{max} maximum value of Ω for which loop locks in absence of noise, rad/sec, (3-23).

[19]May take on a negative value.