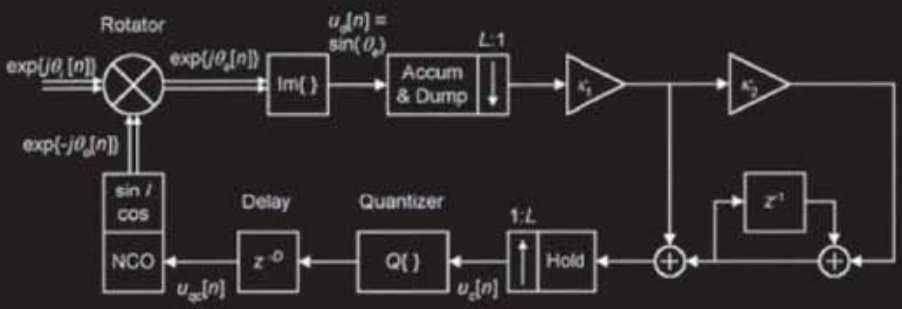
WILEY

Phaselock Techniques

Third Edition

Floyd M. Gardner



PHASELOCK TECHNIQUES

PHASELOCK TECHNIQUES

Third Edition

FLOYD M. GARDNER
Consulting Engineer
Palo Alto, California

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:

Gardner, Floyd Martin, 1929–
Phaselock techniques / Floyd M. Gardner.—3rd ed.
p. cm.
Includes bibliographical references and index.
ISBN-13 978-0-471-43063-6 (cloth)
ISBN-10 0-471-43063-3 (cloth)
1. Phase-locked loops. I. Title.
TK7872.P38G37 2005
621.3815′364—dc22
2004065041

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

To Benjamin

CONTENTS

PREFACE

The first edition of this book was published in 1966 and the second in 1979. Phaselock was an unimaginably exotic subject in 1966, with limited applications and few practitioners. Now phaselock is a mature subject: myriads of phaselock loops are ensconced in the world's electronic devices; numerous applications include phaselock loops; large numbers of practitioners deal with phaselock. No other books on phaselock loops existed when the first edition was published, but more than 20 exist today. Why is a third edition justified at this time?

In 1966, a simple, short introduction to the basics of the subject was needed for an audience for whom phaselock was strange and new. Today, phaselock loops are firmly established in the mainstream of electronics engineering. Much new information on phaselock loops has accumulated over the years, and several topics once thought important have proved to be ephemeral. Experience has taught me that certain explanations would be better presented from revised viewpoints.

There is no need for another introductory text; that function is well served by a number of the books listed in Section 1.3 and probably others as well. Instead, this book reexamines the traditional phaselock topics in greater depth than previously. In addition, much new material has been included, some of it never before published. Examples of additions include revised and expanded material on transfer functions, two chapters related to phase noise, two chapters related to digital phaselock loops, a chapter on charge-pump phaselock loops, expanded material on phase detectors, and a chapter on anomalous phaselocking.

As in the earlier editions, only minimal space has been devoted to circuits. The book is concerned with underlying principles, which remain valid despite technology advances, not with implementations, which change drastically as technology changes. Several parts of the second edition have been omitted: the chapters on

optimization and synchronization, and the mathematical appendix. Formal optimization has not proved to be as important to design as was earlier anticipated; instead, a designer is much more likely to perform a trade-off among the few parameters available in a practical phaselock loop. The mathematical appendix has been omitted on the premise that the level of mathematics presented here should be comfortable for all electrical engineering graduates. Synchronization (recovery of carrier and clock from data signals), a major discipline of its own, was deemed to have grown too large to cover adequately in a book on phaselock loops. See Section 17.1 for a brief guide to synchronization.

Simulation is another absent topic. Information presented in several chapters is based on simulations, certain kinds of new data can be gathered only by simulation, and simulation is crucial for design and verification of integrated circuits. Nonetheless, the book does not tell how to conduct simulations of phaselock loops. That topic deserves a separate book of its own; it is too extensive to include here.

Many thousands of articles and books on phaselock have appeared worldwide over the years, far too many to cite individually. Although many pertinent references have been cited in the individual chapters of the book, it is not possible to discover every valuable publication written on each topic. Nor, after many years of work on the subject, is it possible always to remember who originated every technique that is presented. I apologize in advance to anyone who may have been slighted; the omission is not deliberate.

Several guidelines have been followed in selecting reference citations for each chapter: The reference is to an original work; wherever possible, the reference appeared in a public, archival publication; the reference treats lasting principles rather than transitory details of implementation. A reader will observe a preponderance of citations to IEEE publications and to books published in the United States. This choice reflects the omnipresence of IEEE publications and also the contents of my personal library.

I want to thank my many clients over the years who have afforded me the opportunities to learn so much about such a fascinating subject.

FLOYD M. GARDNER

Palo Alto, California
October 2004

NOTATION

A	Amplitude
B_i	Bandwidth (Hz) of an input bandpass filter
B_L	Noise bandwidth (Hz) of a PLL
b	Number of bits in a digital word
b	Ratio of frequency of a pole to frequency of a zero
f	Frequency (Hz)
f	Transform variable of Fourier transforms
f_c	Comparison frequency (Hz) at a phase detector
f_{ck}	Clock frequency (Hz)
f_m	Frequency (Hz) of modulation
f_s	Sampling frequency (Hz), $= 1/t_s$
Δf	Peak frequency deviation (Hz)
Δf	Frequency offset (Hz) from a carrier
δf	Frequency increment (Hz) in a quantized-tuning oscillator
D	Delay (sample intervals)
$E(f)$	$= E(s)\|_{s=j2\pi f}$
$E(s)$	Closed-loop error transfer function of a PLL
$F(s)$	Transfer function of a loop filter
FP$[x]$	Fractional part of x
$G(s)$	Open-loop transfer function of a PLL
$H(f)$	$= H(s)\|_{s=j2\pi f}$
$H(s)$	Closed-loop system transfer function of a PLL
Im$[x]$	Imaginary part of x
IP$[x]$	Integer part of x
i	Subscript denoting "input"

i	An integer
$J_n(x)$	Bessel function of the first kind, order n, and argument x
j	$\sqrt{-1}$
K	Loop gain (rad/sec) of a PLL
K'	Normalized (dimensionless) loop gain, $= K\tau_2$
K_d	Gain (V/rad or A/rad) of a phase detector
K_{DC}	DC gain (rad/sec) of a PLL
K_i	Gain coefficient in analog PLL, $i = 1, 2, \ldots$
K_m	Gain (V^{-1}) of a multiplier
K_o	Gain (rad/sec·V) of a VCO
K_p	Gain (V/cycle) of a phase detector $= 2\pi K_d$
K_v	Gain (Hz/V) of a VCO, $= K_o/2\pi$
k	An integer
$L\{x\}$	Laplace transform of x
$\mathcal{L}(f)$	Normalized one-sided RF spectrum of a signal
m, M	An integer
$m(t)$	Modulation waveshape
N_0	One-sided spectrum (V^2/Hz) of white noise
n, N	An integer
$n(t)$	Noise voltage (V)
$n_c(t)$, $n_s(t)$	Baseband quadrature components of bandpass noise (V)
o	Subscript denoting "output" or "oscillator"
$P_{\mathrm{RF}}(f)$	Spectrum analyzer representation of the one-sided spectral density of an RF signal
P_s	Signal power (W)
p	Normalized Laplace variable, $= s\tau_2$
Q	Quality factor of a resonator
Q	Number of quantization levels
Q	Division ratio
$\mathrm{Re}[x]$	Real part of x
$r(t)$	Received signal
$s = \sigma + j\omega$	Transform complex variable of a Laplace transform
SNR	Signal-to-noise ratio
SNR_L	Signal-to-noise ratio in PLL noise bandwidth $2B_L$
t	Time (sec)
t_s	Sampling interval (sec), $= 1/f_s$
$u_c[n]$	Sample-n control input (dimensionless) to an NCO
$u_d[n]$	Sample-n output (dimensionless) of a digital phase detector
V_o	Peak output voltage (V) of a VCO
V_s	Peak voltage (V) of an input signal
$v_c(t)$, $V_c(s)$	VCO control voltage (V)
$v_d(t)$, $V_d(s)$	Phase detector output (V)
$W_{n'}(f)$	One-sided spectral density (rad^2/Hz) of the equivalent noise out of a phase detector

$W_{\theta no}(f)$	One-sided spectral density of the VCO phase (rad^2/Hz) due to the noise input to a PLL
$W_{\mathrm{RF}}(f)$	Measured one-sided spectral density (V^2/Hz) of an RF signal
$W_{vo}(f)$	Theoretical one-sided spectral density (V^2/Hz) of an oscillator output
$W_{\phi}(f)$	One-sided baseband spectrum (rad^2/Hz) of phase noise
z	Transform variable of z-transforms

Greek Symbols

α	Signal suppression factor (dimensionless) in a limiter
β	Modulation index (rad) of angle modulation
γ	Crest factor of a signal
$\varepsilon[n]$	Sample n of phase (cycles)
ζ	Damping factor of a second-order PLL
θ	Phase angle (rad)
θ_a	Steady-state phase error (rad) due to frequency-ramp input
θ_e	Phase error (rad) between an input signal and a VCO, $= \theta_i - \theta_o$
θ_i	Phase angle (rad) of an input signal
θ_{no}	Fluctuation of VCO phase (rad) caused by noise
θ_o	VCO phase (rad)
θ_v	Steady-state phase error (static phase error; loop stress) due to frequency offset
$\Delta\theta$	Phase deviation (rad)
$\Delta\theta$	Amplitude (rad) of phase step
κ	Loop gain (dimensionless) in a digital PLL
κ_d	Gain (rad^{-1}) of a digital phase detector
κ_i	Gain coefficient in a digital PLL, $i = 1, 2, \ldots$
κ_o	Gain (rad) of a NCO
κ_p	Gain (cycle^{-1}) of a digital phase detector, $= 2\pi\kappa_d$
κ_v	Gain (cycles) of an NCO, $= \kappa_o/2\pi$
Λ	Rate of change (rad/sec^2) of frequency, $= d\omega/dt$
ρ	Signal-to-noise ratio
σ_x	Standard deviation of x
τ	Timing error (sec)
τ	Delay (sec)
τ_i	Time constant (sec), $i = 1, 2, \ldots$
τ_2	Time constant (sec) of stabilizing zero in a type 2 PLL
$\phi(t)$	Phase noise (rad)
ψ	Angle (rad) around a unit circle
ψ	Normalized frequency (dimensionless), $= \omega t_s$
$\psi(s)$	Phase (rad) of a transfer function
ψ_{gc}	Normalized unity-gain crossover frequency $\omega_{gc}t_s$ of open-loop transfer function of sampled PLL, $\lvert G(e^{j\psi_{gc}})\rvert = 1$

ω	Angular frequency (rad/sec), $= 2\pi f$
ω_c	Comparison frequency (rad/sec) at a phase detector, $= 2\pi f_c$
ω_{gc}	Unity-gain crossover frequency (rad/sec) of open-loop transfer function, $\lvert G(j\omega_{gc})\rvert = 1$
ω_m	Modulating frequency (rad/sec)
ω_n	Natural frequency (rad/sec) of a second-order PLL
ω_π	Phase crossover frequency (rad/sec), $\text{Arg}[G(j\omega_\pi)] = -\pi$
$\Delta\omega$	Frequency offset or frequency step (rad/sec)
$\Delta\omega_H$	Hold-in limit (rad/sec) of a PLL
$\Delta\omega_L$	Lock-in limit (rad/sec) of a PLL
$\Delta\omega_P$	Pull-in limit (rad/sec) of a PLL

CHAPTER 1

INTRODUCTION

A *phaselock loop* (PLL) contains three essential elements (Fig. 1.1): (1) a phase detector (PD), (2) a loop filter (LF), and (3) a voltage-controlled oscillator (VCO). A phase detector compares the phase of a periodic input signal against the phase of the VCO signal; the output of the PD is a measure of the phase error between its two inputs. The error voltage is then filtered by the loop filter, whose control output is applied to the VCO. Control voltage changes the VCO frequency in a direction that reduces the phase error between the input signal and the VCO.

When the loop is *locked*, the control voltage sets the average frequency of the VCO *exactly* equal to the average frequency of the input signal. For each cycle of input there is one and only one cycle of oscillator output. Phaselock does not

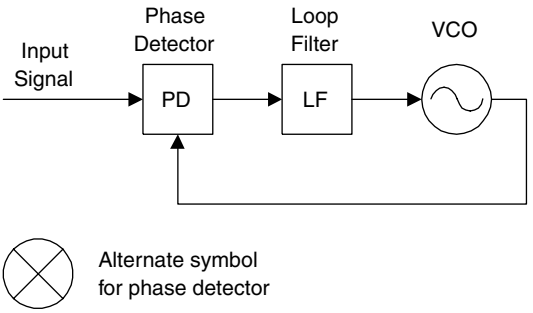


Figure 1.1 Basic phaselock loop.

Phaselock Techniques, Third Edition, by Floyd M. Gardner

imply zero phase error; steady phase errors and fluctuating phase errors can both be present. Excessive phase error causes loss of lock.

1.1 SALIENT PROPERTIES OF PLLs

Certain fundamental properties of phaselock loops are outlined here, properties that arise repeatedly throughout the book.

1.1.1 Bandwidth

Bandwidth is one crucial property; PLLs with a narrow bandwidth are employed quite differently from PLLs with a wide bandwidth.

Narrow Bandwidth Suppose that the input signal carries information in its phase or frequency and that the signal is corrupted by additive noise. The task of a phaselock receiver is to reproduce the original signal adequately while removing as much of the noise as possible. To reproduce the signal, the receiver makes use of a local oscillator whose frequency is very close to that expected in the signal. Waveforms of the local oscillator and incoming signal are compared with one another in the phase detector. Error output from the PD indicates instantaneous phase difference. To suppress noise, the PLL averages the error over some length of time, and the average is used to set the frequency and phase of the oscillator.

If the original signal is well behaved (stable in frequency), the local oscillator will need very little information to be able to track, and that information can be obtained by averaging for a long period of time, thereby eliminating noise that could be very large. The input to the PLL is a noisy signal, whereas the output of the VCO is a cleaned-up version of the input. Therefore, the PLL can be regarded as a kind of filter that passes signals and rejects noise.

Two important characteristics of the PLL as a filter are that (1) its bandwidth can be very small and (2) it tracks the signal frequency automatically. These features, automatic tracking and narrow bandwidth, are the primary reasons for using phaselock in receivers. A narrow bandwidth is capable of rejecting large amounts of noise; it is not at all unusual for a PLL to recover a signal that is deeply embedded in the noise at the input to the PD.

Wide Bandwidth Consider an oscillator with desirable features such as power output or high frequency but with poor stability of frequency. Its frequency can be stabilized by phaselocking that oscillator to a reference oscillator of lesser power, perhaps at a lower frequency but with superior frequency stability. The PLL acts as an electronic servomechanism to suppress unwanted frequency or phase fluctuations in the locked oscillator. The PLL should have fast response—wide bandwidth—to suppress the oscillator fluctuations to the greatest extent possible.

1.1.2 Linearity

Every PLL is nonlinear. Tools for analysis of nonlinear systems are exceedingly cumbersome and provide meager benefits compared to the powerful analytical tools available for linear systems. Fortunately, most (but not all) PLLs of interest can be analyzed by linear techniques when in their locked condition. This book argues throughout that linear methods are sufficient for the bulk of analysis and initial design of most PLLs. Therefore, linear approximations are employed wherever feasible.

Several important instances of inescapably nonlinear PLLs are examined in later chapters. The relative simplicity of linear analysis is vividly emphasized by the obstacles that are encountered when trying to understand nonlinear operations.

1.2 ORGANIZATION OF THE BOOK

The book is divided into several parts. The first part, consisting of Chapters 2 through 8, explains fundamental principles of PLLs. The second part covers the elements within a PLL: oscillators (Chapter 9), phase detectors (Chapter 10), loop filters (Chapter 11), and charge pumps (Chapter 12). Chapters 13 (on digital PLLs) and 14 (on PLL misbehavior) each stand alone. The last part, Chapters 15 through 17, describes various applications of PLLs.

A word on the explanations that follow: The first introduction of a topic is usually simplified, if not oversimplified, with little or no regard for rigor or any warning about complicating factors. Where necessary, complexities are addressed later, after a reader has had a chance to absorb the fundamentals. The essential elements of PLLs are not particularly abstruse even though analysis of many aspects can be formidable. A reader is more likely to be put off by the sheer mass of detail rather than by finding any single topic impenetrable. A system as illustrated in Fig. 1.1 initially appears so simple as to be trivial: How can its treatment fill so many pages? Read the book and find out.

1.3 ANNOTATED BIBLIOGRAPHY

This section lists books, reprint volumes, and journal special issues devoted to PLLs. Items within a heading are entered chronologically. The items listed cover mainly the general topic of PLLs; no claim is made for completeness. More specialized publications are cited in later chapters.

1.3.1 Books

A. J. Viterbi, *Principles of Coherent Communications*, McGraw-Hill, New York, 1966, Chaps. 2–4. (An account of the contributions on PLLs by a noted pioneer in the electronics community.)

W. C. Lindsey, *Synchronization Systems in Communications and Control*, Prentice Hall, Englewood Cliffs, NJ, 1972. (Massive exposition on PLLs in noise. Includes deep theory of stochastic processes and nonlinear analysis.)

W. C. Lindsey and M. K. Simon, *Telecommunication Systems Engineering*, Prentice Hall, Englewood Cliffs NJ, 1973. (High-level presentation of application of PLLs in deep-space receivers.)

A. Blanchard, *Phase-Locked Loops: Application to Coherent Receiver Design*, Wiley, New York, 1976. (Contains data not found elsewhere on the performance of PLL receivers.)

H. Meyr and G. Ascheid, *Synchronization in Digital Systems: Phase-, Frequency-Locked Loops, and Amplitude Control*, Wiley, New York, 1990. (A wealth of material, invaluable to any serious worker on PLLs.)

D. H. Wolaver, *Phase-Locked Loop Circuit Design*, Prentice Hall, Englewood Cliffs, NJ, 1991. (A practical introduction to PLLs. Offers numerous shortcut approximations.)

J. Encinas, *Phase Locked Loops*, Kluwer Academic, Boston, MA, 1993.

P. V. Brennan, *Phase-Locked Loops: Principles and Practice*, McGraw-Hill, New York, 1996.

J. L. Stensby, *Phase-Locked Loops: Theory and Applications*, CRC Press, Cleveland, OH, 1997. (Includes coverage not found elsewhere on nonlinear operations.)

W. Egan, *Phase-Lock Basics*, Wiley, New York, 1998. (Outgrowth of university courses on PLLs. Affords online access to simulations of PLLs.)

D. R. Stephens, *Phase-Locked Loops for Wireless Communications*, Kluwer Academic, Boston, MA, 2001.

R. E. Best, *Phase-Locked Loops*, 5th ed., McGraw-Hill, New York, 2003. (A popular introductory text, profusely illustrated, with accompanying software.)

V. F. Kroupa, *Phase Lock Loops and Frequency Synthesis*, Wiley, Chicester, West Sussex, England, 2003. (Painstaking tour through the fundamentals.)

N. I. Margaris, *Theory of the Non-linear Analog Phase Locked Loop*, Springer-Verlag, Berlin, 2004.

W. H. Tranter, *Phase-Locked Loops and Synchronization Systems: A Matlab-Based Simulation Library*, Prentice Hall, Englewood Cliffs, NJ, 2005.

1.3.2 Reprint Volumes

These volumes are collections of selected papers on the general subject of PLLs. Many of the papers are classic expositions of their subjects. Additional reprint volumes, covering more specialized areas, are cited in later chapters.

W. C. Lindsey and M. K. Simon, eds., *Phase-Locked Loops and Their Applications*, IEEE Press, New York, 1978.

W. C. Lindsey and C. M. Chie, eds., *Phase-Locked Loops*, IEEE Press, New York, 1986.

B. Razavi, *Monolithic Phase-Locked Loops and Clock Recovery Circuits*, IEEE Press, New York, 1996.

B. Razavi, *Phase-Locking in High-Performance Systems*, IEEE Press, New York, and Wiley, Hoboken, NJ, 2003.

1.3.3 Journal Special Issues

Two entire issues of IEEE journals were devoted to phaselock loops.

W. C. Lindsey and C. M. Chie, guest eds., *IEEE Transactions on Communications* ***COM-30***, Oct. 1982.

M. H. Perrott and G.-Y. Wie, guest eds., *IEEE Transactions on Circuits and Systems II* ***50***, Nov. 2003.

CHAPTER 2

TRANSFER FUNCTIONS OF ANALOG PLLs

Although PLLs are inherently and inescapably nonlinear circuits, the main operations of many can be approximated very well by linear models. A linear model typically will be applicable if phase error is small, a condition normally attained when the loop is locked. Most analysis and design of PLLs can be based on the linear approximations; analysis becomes far more challenging when the linear approximations fail.

Among the tools of linear analysis, the Laplace and Fourier transforms—and various concepts derived therefrom—stand out as being particularly valuable. The related concept of a *transfer function*, describing a transform-domain relation between the input and output of a linear circuit, is an extremely powerful tool for dealing with PLLs. Analytical design of PLLs is carried out almost entirely through transfer functions. Take heed that only linear circuits have transfer functions; no such property exists for nonlinear circuits.

Transfer functions of analog PLLs are introduced in this chapter and the next, and transfer functions of digital PLLs are treated in Chapter 4. Results are employed throughout the rest of the book.

2.1 BASIC TRANSFER FUNCTIONS

In ordinary electrical circuits, a transfer function relates voltages or currents of the input and output signals. But in a PLL, the input or output variables of most interest are phases of the signals, not the voltages or currents. The transfer

Phaselock Techniques, Third Edition, by Floyd M. Gardner

functions considered here relate phase modulation of a signal applied in one location of the PLL to phase-modulation response in another location in the PLL.

2.1.1 Transfer Functions of Individual Elements

Consider an elementary loop consisting of a phase detector (PD), a loop filter (LF), and a voltage-controlled oscillator (VCO), as in Fig. 2.1. The phase of the input signal is denoted $\theta_i(t)$ and phase of the VCO output is denoted $\theta_o(t)$, both in radians. Assume that the loop is locked and that the phase detector is linear, so that the PD output voltage is

$$v_d = K_d(\theta_i - \theta_o) \tag{2.1}$$

where K_d is called *the phase-detector gain factor* and is measured in units of volts (or amperes, as applicable) per radian. Define *phase error* as

$$\theta_e = \theta_i - \theta_o \tag{2.2}$$

Error voltage $v_d(t)$ is processed by the loop filter, whose purpose is to establish the dynamic performance of the loop. In addition, noise and high-frequency signal components often are suppressed by the filter, but that is a secondary function, to be ignored for now. The transfer function of the filter is denoted by $F(s)$. Filter output is a control voltage denoted $v_c(t)$ that controls the frequency of the VCO. In the Laplace transform domain, the action of the filter is described by

$$V_c(s) = F(s)V_d(s) \tag{2.3}$$

[**Notation**: $V_c(s) = L\{v_c(t)\}$, where $L\{\cdot\}$ indicates a Laplace transform. Similar relations apply to the other quantities, except that the time- or transform-domain symbol for a phase variable is indicated solely by its argument, t or s in parentheses; that is, $\theta(s) = L\{\theta(t)\}$.]

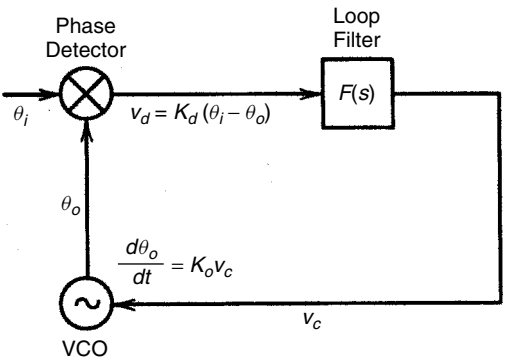


Figure 2.1 Phaselock loop: basic block diagram.

Deviation of the VCO from its center frequency is $\Delta\omega = K_o v_c$ in rad/sec, where K_o is the *VCO gain factor* and has units of rad/sec·V. Since frequency is the derivative of phase, the VCO operation may be described as $d\theta_o/dt = K_o v_c(t)$. Take Laplace transforms to obtain $L\{d\theta_o(t)/dt\} = s\theta_o(s) = K_o V_c(s)$, whereupon

$$\theta_o(s) = \frac{K_o V_c(s)}{s} \tag{2.4}$$

where $s = \sigma + j\omega$ is the Laplace independent variable. Inasmuch as $1/s$ is the Laplace transform of an integration, the phase of the VCO is proportional to the integral of the control voltage.

2.1.2 Combined Transfer Functions

Transfer functions of the individual elements can be combined to obtain the overall loop transfer functions needed for analysis and design purposes. The following combinations are employed in the sequel. The displayed equations generally apply to all PLLs having the configuration of Fig. 2.1, irrespective of the details of the loop filter.

- Open-loop transfer function:

$$G(s) = \frac{\theta_o(s)}{\theta_e(s)} = \frac{K_d K_o F(s)}{s} \tag{2.5}$$

- System transfer function:

$$H(s) = \frac{\theta_o(s)}{\theta_i(s)} = \frac{G(s)}{1 + G(s)} = \frac{K_d K_o F(s)}{s + K_d K_o F(s)} \tag{2.6}$$

- Error transfer function:

$$E(s) = \frac{\theta_e(s)}{\theta_i(s)} = \frac{1}{1 + G(s)} = 1 - H(s) = \frac{s}{s + K_d K_o F(s)} \tag{2.7}$$

[**Comments**: (1) A PLL cannot run properly in an open-loop condition. The open-loop transfer function $G(s)$ is obtained as a formal cascade of the transfer functions of the individual elements. There is no implication that the feedback loop can be broken physically and the indicated response then measured by any straightforward means. Nonetheless, the formal open-loop transfer function is a valuable concept that is used repeatedly in later pages. (2) The *system* transfer function also has been called the *closed-loop* transfer function (e.g., in earlier editions of this book). But the loop error is also described by a closed-loop transfer function, so the term *closed loop* can be ambiguous. The name *system* is proposed as a means to avoid the ambiguity.]

2.1.3 Characteristic Equation

The expression $1 + G(s) = 0$ is known as the *characteristic equation* of the PLL. The roots of the characteristic equation (those values of s that satisfy the equation) are the *poles* of the closed-loop transfer functions. Pole locations are important properties of a PLL. In most PLLs, the open-loop transfer function can be put in the form

$$G(s) = \frac{A(s)}{B(s)} \tag{2.8}$$

where $A(s)$ and $B(s)$ are algebraic polynomials in s. Substitute (2.8) into (2.6) and (2.7) to obtain

$$H(s) = \frac{A(s)}{B(s) + A(s)}, \qquad E(s) = \frac{B(s)}{B(s) + A(s)} \tag{2.9}$$

The polynomial $B(s) + A(s)$ will be called the *characteristic polynomial* because it is derived from the characteristic equation and has the same roots. It turns up repeatedly in the sequel.

2.1.4 Nomenclature, Coefficients, and Units

In the equations above, phase has been indicated by the symbol θ and is measured in radians. The symbol ϕ will also be used. Frequency has been indicated by the symbol ω and is measured in radians per second. But frequency is also commonly indicated by the symbol f and measured in cycles per second (Hz). Correspondingly, it is sometimes convenient to deal with phase measured in cycles instead of radians; the symbol ε is used for this purpose. In the telecommunications industry, units of ε are commonly known as *unit intervals* (UIs).

Rather than phase, it may be convenient to work with a time difference τ, measured in seconds. The relations among θ, ε, τ, and a signal frequency $f_s = 1/T_s$ are

$$\varepsilon = \frac{\tau}{T_s}, \qquad \theta = 2\pi\tau f_s \tag{2.10}$$

The quantities K_d and K_o were introduced as the gain coefficients of the phase detector and the VCO, measured in V/rad and rad/sec·V, respectively. Alternative definitions are equally valid. For a phase detector, a coefficient K_p could be employed, defined by the equation

$$v_d = K_p(\varepsilon_i - \varepsilon_o) \qquad \text{V} \tag{2.11}$$

where K_p has units of V/cycle. Observe that $K_p = 2\pi K_d$. Similarly, a coefficient K_v could be employed for a VCO, defined by the equation

$$\Delta f_o = K_v v_c \qquad \text{Hz} \tag{2.12}$$

where K_v has units of Hz/V. Observe that $K_v = K_o/2\pi$, whereupon $K_d K_o = K_p K_v$. Furthermore, observe that K_d and K_o always appear together as a product in the transfer-function equations (2.5) to (2.7), so that $K_p K_v$ can be substituted freely instead, as may be convenient.

Devices may sometimes be specified in strange units, such as mV/degree for a phase detector. Convert such oddities to one of the standard characterizations before trying to use the data. [**Warning**: Be very careful to use consistent units when working with transfer functions. Mixed units are likely to introduce mistaken factors of 2π, leading to seriously erroneous results.]

2.2 SECOND-ORDER PLLs

The phaselock literature contains innumerable articles on second-order PLLs. The vast majority of practical PLLs either are second order (defined below) or are designed approximately as a second-order loop by neglecting higher-order effects, at least for initial design. Reasons for this predominance are given in later chapters. This section introduces a second-order PLL and some of its properties.

2.2.1 Loop Filters

A particular loop-filter transfer function $F(s)$ has to be specified at this juncture. Figure 2.2 illustrates two loop-filter configurations that lead to a second-order PLL. Figure 2.2*a* shows the circuit of an active filter employing a high-gain DC amplifier. For now, assume that the DC gain is effectively infinite (an easily

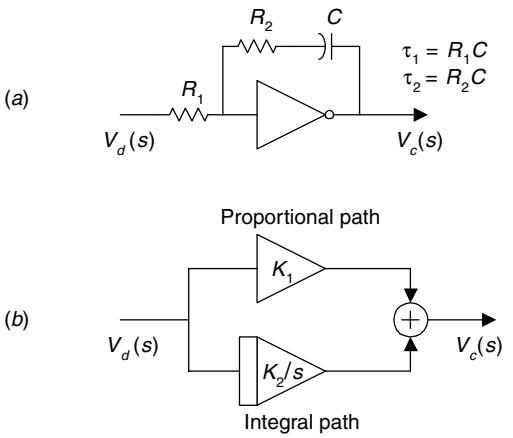


Figure 2.2 Loop filter for a second-order type 2 PLL: (*a*) single-path circuit with operational amplifier; (*b*) two-path proportional-plus-integral configuration.

approximated condition with modern operational amplifiers) so that the equations of that configuration are given by

$$\begin{aligned} \tau_1 &= R_1 C \qquad \text{sec} \\ \tau_2 &= R_2 C \qquad \text{sec} \\ F(s) &\approx -\frac{s\tau_2 + 1}{s\tau_1} = -\left(\frac{\tau_2}{\tau_1} + \frac{1}{s\tau_1}\right) \end{aligned} \tag{2.13}$$

[**Comment**: Observe the minus sign indicating phase inversion in the active filter. The minus sign will be neglected in the sequel; just assume that it is canceled by another minus sign at an unspecified place within the loop to achieve overall negative feedback, as is essential for a stable feedback loop. Later chapters show that some kinds of phase detectors allow the loop to find a negative-feedback operating point *automatically*, irrespective of any phase inversions in the individual elements. But you must make certain that the loop is poled correctly when using other kinds of phase detectors.]

Figure 2.2*b* shows a proportional-plus-integral (P + I) loop filter. There are two parallel paths through this filter, in contrast to the single path of the active filter above. Its transfer-function equation is

$$F(s) = K_1 + \frac{K_2}{s} \tag{2.14}$$

where K_1 is the gain coefficient of the proportional path through the filter and K_2 is the coefficient of the integral path through the filter. The coefficient K_1 is dimensionless but K_2 must have dimensions of (time)$^{-1}$ to make $F(s)$ dimensionless overall. The two configurations are electrically equivalent if $K_1 = \tau_2/\tau_1$ and $K_2 = 1/\tau_1$. Single-path configurations are most widely used in practice, but two-path configurations can offer substantial implementation and analysis advantages; examples are given in later chapters. Additional configurations for loop filters are introduced in later pages; these two are by no means the only ones that might be used.

[**Comment**: In hindsight, the name loop *filter* is unfortunate even though earlier editions of this book disseminated that terminology. Take particular notice that these examples are not *lowpass* filters, despite the mistaken designation as such by some authors. A better name might have been *loop controller*, a term used by our control system colleagues. The main purpose of these circuits is to establish the dynamics of the feedback loop and to deliver a suitable control signal to the VCO. Any filtering of unwanted signals is a secondary task to be accomplished by additional components described later. But the name loop *filter* has become prevalent and would be difficult to overturn, so the remainder of this book retains that terminology.]

2.2.2 Order and Type

Now substitute the loop-filter transfer functions of (2.13) and (2.14) into the basic system transfer function of (2.6) to obtain

$$H(s) = \frac{K_d K_o (s\tau_2 + 1)/\tau_1}{s^2 + sK_d K_o \tau_2/\tau_1 + K_d K_o/\tau_1} = \frac{K_d K_o (K_1 s + K_2)}{s^2 + sK_d K_o K_1 + K_d K_o K_2} \quad (2.15)$$

The denominator polynomial (characteristic polynomial) of this transfer function is of second degree, so that the PLL is said to be *second order*. The two roots of the denominator are the poles of the transfer function, and the root of the numerator is a zero located at $s = -1/\tau_2 = -K_2/K_1$. The zero is a necessary feature of the P + I filter configuration; it is essential for loop stability, a topic addressed in Section 2.3.1.

In past literature (including earlier editions of this book), attention has largely concentrated on the order of the loop (the degree of the characteristic polynomial). However, many of the important properties of a loop relate more correctly to the loop *type*, not the order. *Type*—a term borrowed from the control system community—refers to the number of integrators within the loop. Because the particular loop considered in this section contains two integrators (one in the loop filter and the other in the VCO), it is a type 2 PLL in addition to being second order.

Each integrator contributes one pole to the transfer function, so that the order can never be less than the type. But additional nonintegrator filtering is often present, contributing additional poles and increasing the order, with no effect on the type. Because of the inherent integration in the VCO, a PLL is always at least type 1. For good reasons explained in Section 5.1.1, type 2 PLLs are very common. Occasionally, a type 3 PLL will be employed (for further good reasons), and I once saw a type 4 PLL.

Deliberate additional filtering and unwanted parasitic elements can add multiple nonintegrator poles into the loop, causing the order to be larger than the type. In sufficiently complicated circuits, you might find 10th- to 12th-order loops that are only type 1 or type 2.

2.2.3 Loop Parameters

The second-order transfer function (2.15) contains numerous coefficients (gains and time constants) even though only two poles and one zero are involved. The transfer function is overdetermined as written; a modified notation makes the expressions more compact. The transfer function for a second-order type 2 PLL can be completely specified by just two appropriate *loop parameters*. Initially, loop design and analysis are carried out most conveniently in terms of the loop parameters and, subsequently, are broken down into gains and time constants after satisfactory parameters have been determined. This section defines two different sets of parameters that often appear in the PLL literature.

Natural Frequency and Damping The best known set of parameters for a second-order PLL consists of the *undamped natural frequency* ω_n rad/sec (usually just *natural frequency*) and the dimensionless *damping factor* ζ. In terms of the coefficients and time constants of (2.15), these parameters are defined for a second-order type 2 PLL as

$$\begin{aligned} \omega_n &= \sqrt{\frac{K_d K_o}{\tau_1}} = \sqrt{K_d K_o K_2} \\ \zeta &= \frac{\tau_2}{2}\sqrt{\frac{K_d K_o}{\tau_1}} = \frac{\tau_2 \omega_n}{2} = \frac{K_1}{2}\sqrt{\frac{K_d K_o}{K_2}} \end{aligned} \tag{2.16}$$

and the system transfer function reduces to

$$H(s) = \frac{2\zeta\omega_n s + \omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \tag{2.17}$$

Natural frequency and damping are a convenient description of the properties of a pole pair and so are well suited for second-order loops. If $\zeta < 1$, the poles are a complex-conjugate pair; if $\zeta = 1$, the poles are real and coincident; and if $\zeta > 1$, the poles are real and separate. For $\zeta < 1$, the vector from the origin of the s-plane to a pole location has length ω_n, and the cosine of the angle from the negative real axis to the vector is ζ, as shown in Fig. 2.3. For $\zeta \geq 1$, the geometric mean of the pole locations is equal to ω_n, and the ratio of the two pole locations is given by

$$2\zeta^2 + 2\zeta\sqrt{\zeta^2 - 1} - 1$$

Values of ζ typically lie between 0.5 and 2, with 0.707 often a preferred value, but much larger values—up to 20 or 30—are sometimes needed. Loops with damping smaller than ~0.5 have excessive overshoot in their transient responses and so are dynamically unsatisfactory. Damping factors much larger than ~1

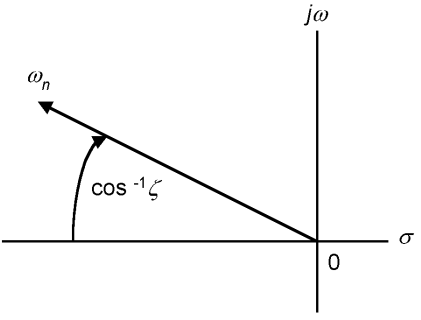


Figure 2.3 Geometry of a complex pole, illustrating ω_n and ζ.

are ordinarily needed only in special circumstances, one example of which is described in Section 2.2.4. Values of ω_n can take on an extremely wide range of values, from $\sim 10^{-5}$ to $\sim 10^{8}$ rad/sec or more, as requirements may dictate.

Natural frequency and damping are an attractive set of parameters because of their intuitive physical description and because of their widespread occurrence in the PLL literature. Strictly speaking, though, they apply only to second-order loops. An extended definition of ω_n is feasible for type 2 loops of higher than second order (as introduced in Chapter 3), but the concept becomes meaningless for a first-order PLL or for a PLL of type 3 or larger. Moreover, ω_n is often used as a measure of loop bandwidth in a second-order loop. It is demonstrated subsequently to be only a mediocre indicator of bandwidth; something better is needed.

Loop Gain K For the second-order type 2 PLL under consideration, define a *loop gain*

$$K = K_d K_o K_1 = \frac{K_d K_o \tau_2}{\tau_1} \qquad \text{rad/sec} \tag{2.18}$$

which will be recognized as the open-loop gain through the proportional path, exclusive of the $1/s$ factor caused by integration in the VCO. A broader definition of K, applicable to PLLs of other orders and types, is provided in Section 2.3.1. A second-order loop needs two parameters: K alone does not suffice. One might choose ζ as the second parameter, or as employed extensively later, τ_2 can be very useful. Any two parameters of a second-order type 2 PLL can be defined in terms of any other two. Some such relations include

$$\begin{aligned} K &= 2\zeta\omega_n, \qquad & \omega_n &= \sqrt{\frac{K}{\tau_2}} \\ K\tau_2 &= 4\zeta^2, & \zeta &= \frac{1}{2}\sqrt{K\tau_2} \end{aligned} \tag{2.19}$$

Using K, the parameterized system transfer function of (2.15) becomes

$$H(s) = \frac{K(s + K/4\zeta^2)}{s^2 + Ks + K^2/4\zeta^2} = \frac{K(s + 1/\tau_2)}{s^2 + Ks + K/\tau_2} \tag{2.20}$$

The corresponding error response is

$$E(s) = \frac{s^2}{s^2 + sK + K^2/4\zeta^2} = \frac{s^2}{s^2 + sK + K/\tau_2} \tag{2.21}$$

DC Gain K_{DC} Define the *DC gain* of the PLL as

$$K_{\mathrm{DC}} = \left| \lim_{s \to 0} sG(s) \right| = K_d K_o |F(0)| \qquad \text{rad/sec} \tag{2.22}$$

This definition applies to all orders and types of PLLs. For a type 1 loop, $F(0)$ is finite, whereas integrators in $F(s)$ make it infinite for loops of type 2 or higher. The significance of K_{DC} is explained in Section 5.1.1.

2.2.4 Frequency Response

The *amplitude responses* $|H(j\omega)|$ and $|E(j\omega)|$ of second-order type 2 PLLs are plotted in Figs. 2.4 to 2.7 (amplitude in dB vs. frequency on a log scale) for several values of damping factor ζ. The frequency scales in Figs. 2.4 and 2.5 are normalized to natural frequency ω_n, whereas the frequency scales in Figs. 2.6 and 2.7 are normalized to loop gain K. Several notable properties of the transfer functions are discernible from these graphs.

Phase-Filtering Properties Inspection of the amplitude responses reveals that the system transfer function $H(s)$ performs a lowpass filtering operation on the phase modulation of the input signal, whereas the error-response transfer function $E(s)$ performs a highpass filtering operation. These broad categories of phase-filtering operations hold true for all PLLs. Only specific details differ among different orders and types. This behavior comes about because the PLL necessarily has restricted bandwidth. The loop tracks input phase modulation that is within the loop bandwidth and fails to track phase modulation that is outside the bandwidth. Thus, input phase modulation within the loop bandwidth is transferred to the VCO's phase output, but input phase modulation outside the loop bandwidth is attenuated. Error response is necessarily complementary: Input phase modulation within the loop bandwidth is tracked with small error, whereas

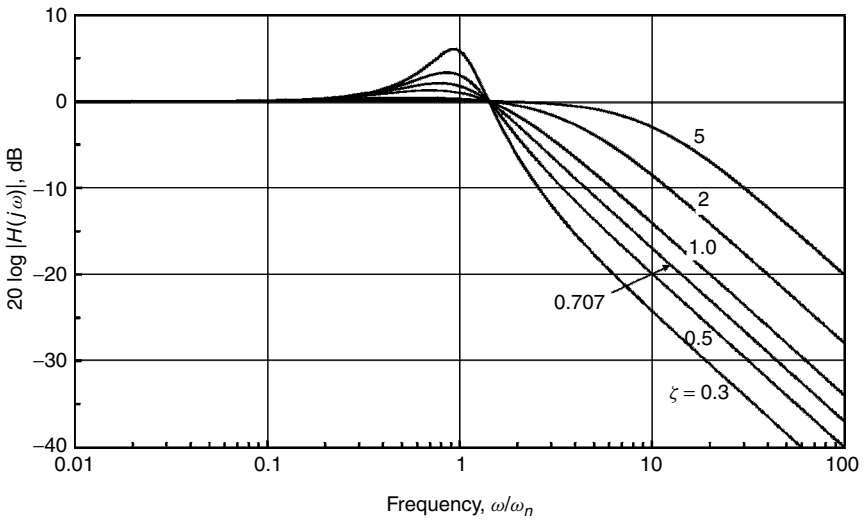


Figure 2.4 Response $|H(j\omega)|$ for a second-order type 2 PLL. Frequency normalized to natural frequency ω_n.

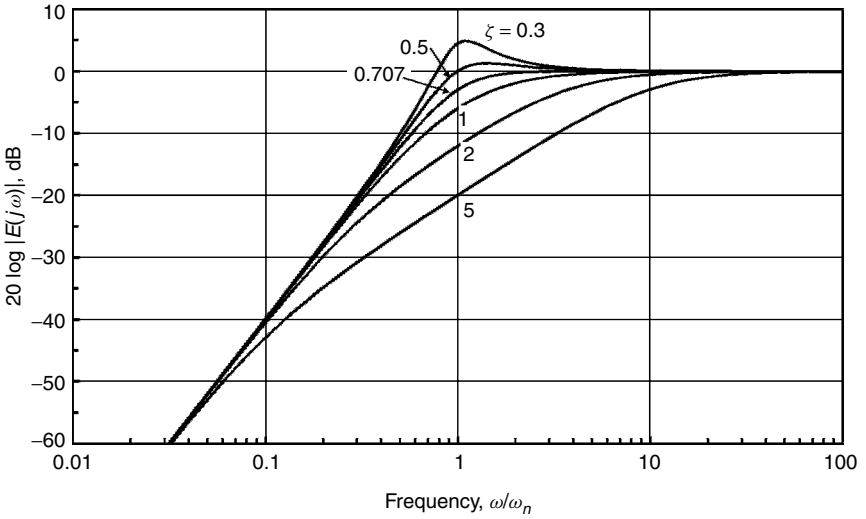


Figure 2.5 Response $|E(j\omega)|$ for a second-order type 2 PLL. Frequency normalized to natural frequency ω_n.

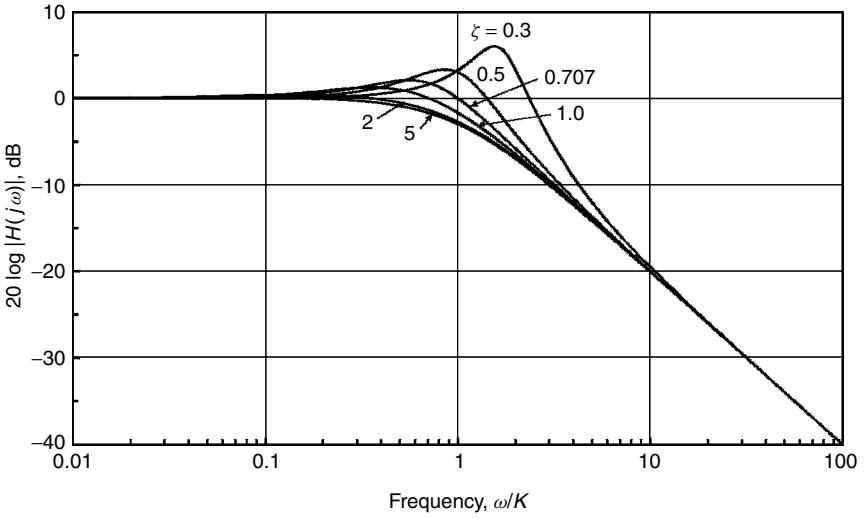


Figure 2.6 Response $|H(j\omega)|$ for a second-order type 2 PLL. Frequency normalized to loop gain K.

input modulation outside the loop bandwidth is hardly tracked at all, resulting in almost 100% tracking error.

Asymptotic Response A look at the asymptotes of the frequency responses $|H(j\omega)|$ and $|E(j\omega)|$ is instructive. From (2.20) and (2.21), the asymptotes of a

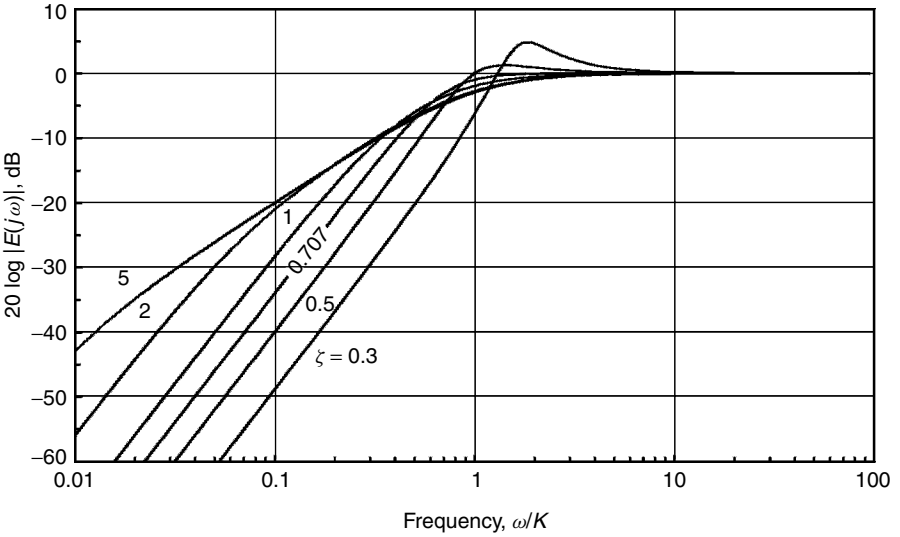


Figure 2.7 Response $|E(j\omega)|$ for a second-order type 2 PLL. Frequency normalized to loop gain K.

second-order type 2 PLL are expressed as

$$|H(j\omega)| \approx \begin{cases} 1, & \omega \ll K \\ \dfrac{K}{\omega}, & \omega \gg K \end{cases}$$
$$|E(j\omega)| \approx \begin{cases} \dfrac{\omega^2}{\omega_n^2}, & \omega \ll \omega_n \\ 1, & \omega \gg \omega_n \end{cases} \tag{2.23}$$

Therefore, the high-frequency asymptote of $|H(j\omega)|$ rolls off at -6 dB/octave, and the low-frequency asymptote of $|E(j\omega)|$ rises at $+12$ dB/octave. These asymptotes are independent of damping ζ. Different asymptotic slopes of $|H(j\omega)|$ are produced in PLLs of different orders, and different slopes of $|E(j\omega)|$ are produced in PLLs of different types.

Bandwidth How should *bandwidth* of a PLL be defined? In fact, there is no single definition that suffices for all purposes. Some candidates are (1) natural frequency ω_n, (2) loop gain K, (3) noise bandwidth B_L, and (4) 3-dB bandwidth $\omega_{3\text{ dB}}$.

Natural frequency is widely used as an indication of bandwidth, but a quick glance at the lowpass curves of Fig. 2.4 shows that it is not a satisfactory measure for $H(s)$ because of strong dependence on damping ζ; a better definition is needed. However, natural frequency does provide a good indication of the

corner frequency of the highpass filtering of $E(s)$, illustrated in Fig. 2.5. This applicability to the highpass response extends to type 2 PLLs of higher order, but the term *natural frequency* is meaningless for any PLL that is neither second order nor type 2.

Figure 2.6 reveals that K is a good indication of the lowpass corner frequency of $H(s)$. Furthermore, K is a good indicator of the lowpass corner for just about any PLL of any order or type. Section 2.3.1 demonstrates additional general properties of K that suit it well as a definition of bandwidth. Henceforth, unless otherwise qualified, use of the term *bandwidth*, will mean K. Notice, though, in Fig. 2.7 that K is a poor indicator of the highpass corner frequency of $E(s)$.

Noise bandwidth B_L, defined in Section 6.1.3, is an appropriate measure of PLL bandwidth if additive white noise is a significant disturbance, although it is less applicable in low-noise situations, where K might be preferable. The relations between B_L and K are detailed in Chapter 6. This book always fully describes B_L as noise bandwidth.

Ordinary filters are commonly specified in terms of their 3-dB bandwidth. Bandwidth of phaselock loops could also be so specified, but that is rarely useful. One could extract 3-dB bandwidths from Figs. 2.4 to 2.7, but there is no apparent significance to such values. The 3-dB bandwidth of the lowpass $|H(j\omega)|$ of a second-order type 2 PLL can be calculated as

$$\omega_{3\text{ dB}} = K\left(\frac{1}{2} + \frac{1}{4\zeta^2} + \frac{1}{2}\sqrt{1 + \frac{1}{\zeta^2} + \frac{1}{2\zeta^4}}\right)^{1/2} \quad \text{rad/sec} \tag{2.24}$$

which approaches K for large ζ, as is evident from Fig. 2.6.

Gain Peaking Figures 2.4 to 2.7 exhibit obvious peaking of the response curves, especially for low values of damping. The highpass responses of $|E(j\omega)|$ have peaking only if $\zeta < \sqrt{0.5} \approx 0.707$ and no peaking for larger ζ. Lowpass response $|H(j\omega)|$ has large peaking for small damping, but the peaking never disappears entirely from a second-order type 2 PLL, no matter how large the damping may be.

Why does peaking always occur for the lowpass curves? For small damping ($\zeta < 0.707$), the complex poles of the transfer function are near the imaginary axis of the s-plane, so incipient resonance effects become prominent. But peaking still occurs for larger damping and even for overdamped loops ($\zeta > 1$): that is, for loops with real poles, so that no resonance effects are present. Peaking in the presence of large damping occurs because of the zero at $s = -1/\tau_2$ in the numerator of $H(s)$. The zero causes $|H(j\omega)|$ to rise with frequency; the rise is terminated by the rolloff of the poles. Spacing between the zero and the closest pole decreases as K is increased (damping increases), but the pole never coincides with the zero for any finite K. Therefore, a second-order type 2 PLL always exhibits some gain peaking in $|H(j\omega)|$.

If the zero causes gain peaking, why not remove the zero to avoid the peaking? Section 2.3.3 shows that a zero is required for a type 2 PLL to be stable. More broadly, $n-1$ zeros are needed for a type n loop to be stable. Some engineers have claimed that third-order loops can suppress the peaking, but firm evidence has been lacking. For the PLL configuration of Fig. 2.1, gain peaking appears to be an inevitable cost of the benefits obtained from integrators in the loop filter.

Moderate peaking is immaterial in many applications, but not all. Consider a situation in which a large number of PLLs are connected in cascade, such as in a chain of repeaters of a telecommunications system. If each repeater has only 1 dB of peaking (corresponding to $\zeta \approx 1$, a damping not ordinarily considered to be small), if the chain contains 100 repeaters (a not unreasonable number), and if no protective measures are taken, the chain will have peaking of 100 dB—a disaster. Common standards for repeaters in the telecommunications plant specify maximum peaking of only 0.1 dB.

From analysis of the transfer function (2.20) for $H(s)$, gain peaking of a second-order type 2 PLL is found to be

$$\text{gain peaking} = 10\log\frac{8\zeta^4}{8\zeta^4 - 4\zeta^2 - 1 + \sqrt{8\zeta^2+1}} \quad \text{dB} \tag{2.25}$$

Peaking vs. damping is plotted in Fig. 2.8. Peaking smaller than 0.1 dB requires that $\zeta > 4.4$.

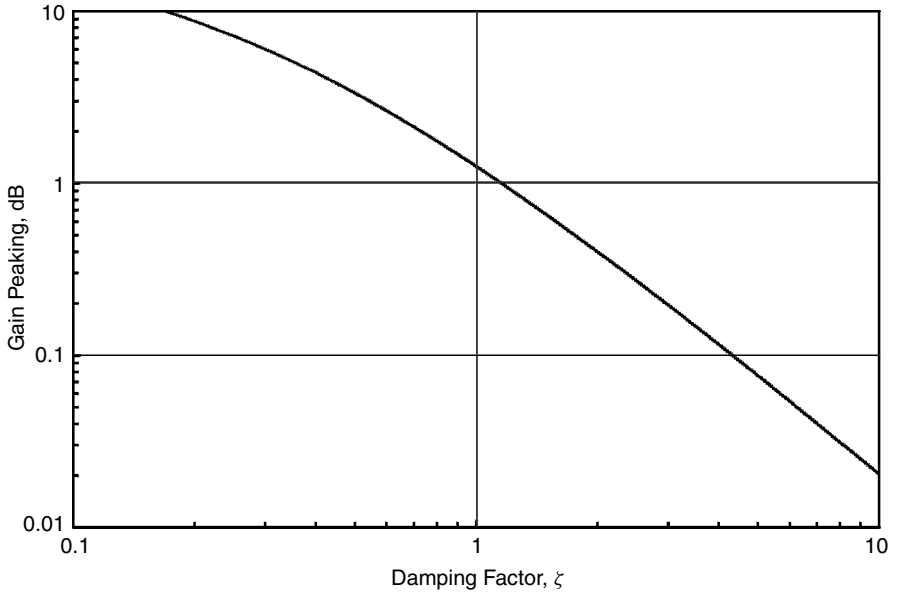


Figure 2.8 Gain peaking in $|H(j\omega)|$ for a second-order type 2 PLL.

Loop Stability Design of feedback loops is incomplete without assurance of stability—phaselock loops are no exception. Literature on a basic second-order type 2 PLL typically says little about loop stability since this kind of loop is unconditionally stable for all values of gain K. Many other loop types and orders are not so robust; they require close attention to stability, as pursued in later pages.

2.3 OTHER LOOP TYPES AND ORDERS

Despite the emphasis in the foregoing sections, a reader should not assume that all phaselock loops are second-order and type 2. Quite the contrary: A great many practical PLLs vary from that base either slightly or substantially. Understanding the variants is essential for skilled engineering of PLLs. Selected examples of other kinds of PLLs are presented in this section, and additional examples arise in later chapters.

2.3.1 General Definition of Loop Gain *K*

As a preliminary matter, the gain K is redefined to make it applicable to almost any PLL that may be encountered. The previous definition (2.18) applies only to a second-order type 2 PLL. To proceed with an extended definition, first divide the transfer function of the loop filter into a cascade of two separate sections in the form

$$F(s) = F_{\mathrm{p+i}}(s)F_{\mathrm{hf}}(s) \tag{2.26}$$

where the subscript "p + i" represents proportional plus integral (P + I) and the subscript "hf," represents high frequency. That is, F_{hf} has its greatest effect at "high" frequencies, typically mostly outside the bandwidth of the PLL. The only constraint on $F_{\mathrm{hf}}(s)$ is that $F_{\mathrm{hf}}(0)$ be finite and nonzero.

In this general formulation, the P + I factor may have an arbitrary number of integrators, not just one as in a type 2 PLL. The expression for the P + I factor can be written as

$$F_{\mathrm{p+i}}(s) = K_1 + \frac{K_2}{s} + \frac{K_3}{s^2} + \cdots \tag{2.27}$$

where K_i has dimensions of $(\text{time})^{-(i-1)}$, thereby assuring that each term in the sum is dimensionless. Most often, no more than one integrator will be employed, but sometimes two integrators will be appropriate (see Section 2.3.4). Only very rarely will more than two integrators be used.

[**Comment**: Frequently, the integrator(s) will be imperfect—a lowpass filter with a low cutoff frequency is used instead of an ideal integrator. In that situation, it is often convenient to substitute a term $1/(s + s_i)$ where needed within $F_{\mathrm{p+i}}(s)$ rather than in $F_{\mathrm{hf}}(s)$, thereby pretending that the lowpass filter serves as an integrator after all. This expedient is especially necessary with analog

circuits since it is almost impossible to realize a perfect analog integrator. As a practical matter, imperfect integrators can be made to be sufficiently close to ideal $1/s$ that the approximation is acceptable. An example is provided in Section 2.3.2.]

Building on (2.5), the open-loop transfer function with the more general loop filter becomes

$$\begin{aligned} G(s) &= \frac{K_d K_o F_{\text{p+i}}(s) F_{\text{hf}}(s)}{s} = \frac{K_d K_o}{s}\left(K_1 + \frac{K_2}{s} + \frac{K_3}{s^2} + \cdots\right) F_{\text{hf}}(s) \\ &= \frac{K_d K_o K_1 F_{\text{hf}}(0)}{s}\left(1 + \frac{K_2}{K_1 s} + \frac{K_3}{K_1 s^2} + \cdots\right)\frac{F_{\text{hf}}(s)}{F_{\text{hf}}(0)} \\ &= \frac{K}{s}\left(1 + \frac{K_2}{K_1 s} + \frac{K_3}{K_1 s^2} + \cdots\right)\frac{F_{\text{hf}}(s)}{F_{\text{hf}}(0)} \end{aligned} \tag{2.28}$$

from which the general definition of K is

$$K = K_d K_o K_1 F_{\text{hf}}(0) \qquad \text{rad/sec} \tag{2.29}$$

Of course, the same definition applies to a second-order type 2 PLL for which $K_i = 0$ for $i > 2$ and $F_{\text{hf}}(s) = 1$. [**Comment**: Although a particular configuration of the P + I section may seem to be implied by (2.28), the transfer function of almost any realistic loop filter can be put into that format, no matter how its circuit is configured. See Appendix 3B for examples.]

Observe that K is determined entirely in the proportional path; integrators and high-frequency effects do not enter into the definition at all. Yet K has a dominant influence on the speed of response and bandwidth of the PLL, an influence that is reiterated in the sequel. This feature points up the following lesson:

Any actual filtering within the loop filter typically has only a secondary effect on loop bandwidth and speed. For the primary influence, look to the loop gain K.

Gain Crossover Frequency Define a *gain crossover frequency* ω_{gc} rad/sec such that $|G(j\omega_{gc})| = 1$ (i.e., 0 dB). An exact expression for ω_{gc} can be extracted from (2.28) if all coefficients are known, but a simple approximation is more informative. If $K_2/K_1\omega_{gc} \ll 1$ ($\omega_{gc} \gg 1/\tau_2$ in a type 2 PLL) and if $|F_{\text{hf}}(j\omega_{gc})|/|F_{\text{hf}}(0)| \approx 1$, then

$$\omega_{gc} \approx K \tag{2.30}$$

Thus, under rather loose conditions, open-loop gain crosses over 0 dB at a frequency close to K rad/sec.

How good is the approximation? Consider a second-order type 2 PLL for which the open-loop gain at $s = jK$ is readily found to be $|G(jK)|^2 = 1 +$

$1/(2\zeta)^4$. Selected values follow:

ζ	20 log $\lvert G(jK)\rvert$ (dB)
0.5	3.0
0.707	0.97
1.0	0.26
2	0.017
5	0.0004

For this example, the approximation is poor for small damping and good to excellent for moderate to large damping. Similar results can be expected for other loop types and orders.

[**Comments**: (1) Loop gain K is employed throughout this book as the most important parameter of any PLL. (2) Emphasis (or recognition) of the importance of K is not widespread in the PLL literature; many other authors rely on ω_n instead, a much less satisfactory choice. A notable exception is in the book by Wolaver [2.1], who uses K in the same way as in this book. (3) Caution is needed in reading other publications; the symbol K is often employed for gain factors that are not the same as the K defined in (2.29). Alternative meanings of K are not wrong, just different; be careful. (4) Some authors mistakenly specify "loop gain" in dB—that is, as a dimensionless quantity. Loop gain of a PLL, no matter how defined, has dimensions of frequency; a dimensionless specification is meaningless.]

2.3.2 Examples of Type 1 PLLs

A type 1 PLL has only one integrator in the loop, that from the VCO. Several varieties are of interest.

First-Order PLL The very simplest PLL contains no loop filter at all. It can be described in the formal notation of Section 2.3.1 by $F_{\text{p+i}}(s) = K_1$ and $F_{\text{hf}}(s) = 1$. Its gains and transfer functions are given by

$$K = K_d K_o K_1 = K_{\text{DC}} \qquad \text{rad/sec} \tag{2.31}$$

and

$$G(s) = \frac{K}{s}, \qquad H(s) = \frac{K}{s+K}, \qquad E(s) = \frac{s}{s+K} \tag{2.32}$$

so loop gain K (which is equal to the 3-dB bandwidth in this case) is the only parameter available to a designer. If it is necessary to have large DC gain (often needed to assure good tracking; see Section 5.1.1), the bandwidth also must be large. Therefore, narrow bandwidth is incompatible with good tracking in the first-order loop; for this reason it is not often used. Nonetheless, first-order loops frequently appear in the literature, mainly because they are the easiest to treat

analytically and also because selected behaviors of first-order loops often can be extended approximately to more complicated PLLs that are more difficult to analyze.

Second-Order PLL with Lag Filter A slightly more complicated PLL contains a loop filter with $F_{\text{p+i}}(s) = K_1$ and $F_{\text{hf}}(s) = 1/(s\tau + 1)$. This arrangement has the same loop gain K and DC gain K_{DC} as the simple first-order PLL, but its system and error transfer functions are second order:

$$\begin{aligned} H(s) &= \frac{K/\tau}{s^2 + s/\tau + K/\tau} = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \\ E(s) &= \frac{s(s + 1/\tau)}{s^2 + s/\tau + K/\tau} = \frac{s(s + 2\zeta\omega_n)}{s^2 + 2\zeta\omega_n s + \omega_n^2} \end{aligned} \tag{2.33}$$

where the parameters are defined by

$$\omega_n = \sqrt{\frac{K}{\tau}}, \qquad \zeta = \frac{1}{2\sqrt{K\tau}} \tag{2.34}$$

With only two coefficients, K and τ, available, it is not possible to specify three parameters, ω_n, ζ, and K_{DC}, independently. Large τ is required if it is necessary to have large DC gain and small ω_n, whereupon the loop will be badly underdamped. [**Caution**: These definitions of ω_n and ζ are substantially different from those of a second-order type 2 PLL.]

Although this is strictly a second-order PLL, it might better be regarded as a first-order PLL with extra filtering to suppress high frequencies in the system response. Indeed, $|H(j\omega)|$ asymptotically rolls off at -12 dB/octave instead of the -6 dB/octave of a second-order type 2 PLL. Moreover, because $H(s)$ has no zero in its numerator, it has no gain peaking whatever for any $\zeta > 0.707$. These features combine to make this form of PLL useful where small bandwidth is not needed. Large numbers of such PLLs are in service in telecommunications repeaters.

Second-Order PLL with Lag-Lead Filter Simultaneous requirements for small bandwidth and large DC gain were recognized from the early days of phaselock loops (i.e., the 1950s), but good DC amplifiers were not available then, so satisfactory integrators were not practicable. Instead, much literature of those earlier times and many PLLs employed a passive *lag-lead filter* with transfer function

$$F(s) = \frac{s\tau_2 + 1}{s\tau_1 + 1} = \frac{\tau_2}{\tau_1}\left(1 + \frac{1/\tau_2 - 1/\tau_1}{s + 1/\tau_1}\right) \tag{2.35}$$

where $\tau_2 < \tau_1$. The expression can also be interpreted as a proportional-plus-imperfect integrator filter with $K_1 = \tau_2/\tau_1$ and $K_2/K_1 = 1/\tau_2 - 1/\tau_1$. Loop gain

is defined as $K = K_d K_o K_1$, DC gain is $K_{\mathrm{DC}} = K_d K_o$, and the loop transfer functions are

$$\begin{aligned} G(s) &= \frac{K}{s}\left(1 + \frac{1/\tau_2 - 1/\tau_1}{s + 1/\tau_1}\right) \\ H(s) &= \frac{K(s + 1/\tau_2)}{s^2 + s(K + 1/\tau_1) + K/\tau_2} \\ E(s) &= \frac{s(s + 1/\tau_1)}{s^2 + s(K + 1/\tau_1) + K/\tau_2} \end{aligned} \tag{2.36}$$

Sufficient degrees of freedom are available to permit independent specification of DC gain, loop bandwidth, and damping, in contrast to the preceding type 1 examples. This freedom accounts for the prevalence of lag-lead filters in early second-order PLLs.

2.3.3 Examples of Type 2 PLLs

Section 2.2 provided extensive descriptions of the important second-order type 2 PLL. This section examines some further variations on type 2 loops. Notice that $K_{\mathrm{DC}} = \infty$ for all PLLs of type 2 or higher, thereby obviating the trade-off between loop bandwidth and DC gain inherent to type 1 loops.

Integrator-Only Loop Filter Suppose that the transfer function of the loop filter is simply $F(s) = K_2/s$; the proportional-path gain $K_1 = 0$ in this formulation. Loop transfer functions are found as

$$G(s) = \frac{K_d K_o K_2}{s^2}, \qquad H(s) = \frac{K_d K_o K_2}{s^2 + K_d K_o K_2}, \qquad E(s) = \frac{s^2}{s^2 + K_d K_o K_2} \tag{2.37}$$

Two features stand out:

1. $H(s)$ has no zero in its numerator. [Compare (2.15), the transfer function applicable for the usual second-order type 2 PLL.]
2. The poles of the denominator lie on the imaginary axis at $s = \pm j\sqrt{K_d K_o K_2}$.

The absence of numerator zeros in $H(s)$ earlier was deemed to imply the absence of gain peaking, but the pole configuration in (2.37) causes infinite gain at the pole frequency. Without the zero, a type 2 PLL is at its stability boundary. Any slight disturbance generates undamped sinusoidal oscillations at the pole frequency. Any slightest additional phase lag within the loop causes full instability evidenced as oscillations of exponentially growing amplitude. The lesson is clear:

The loop filter of a type 2 feedback loop must contain at least one zero in its transfer function to attain stable closed-loop operation. More generally, the filter

transfer function in a type n loop ($n > 1$) must contain at least $n - 1$ zeros or else the loop will be unstable.

Third-Order Type 2 PLL The second-order type 2 PLL is a simplification found widely in the PLL literature but not as often in practice. Most actual PLLs contain additional poles at higher frequencies. Some of these poles might be inserted deliberately to obtain steeper rolloff of the system frequency response or to suppress higher-frequency disturbances emanating from the phase detector. Examples of needs for extra filtering are provided in subsequent chapters. Other poles are parasitic, arising from inescapable frequency-response limitations of practical elements within the feedback loop: stray capacitances, amplifier bandwidth restrictions, or lowpass circuits in the control path of the VCO circuit.

In many instances, the pole frequencies are sufficiently larger than the desired loop bandwidth that these high-frequency poles can be ignored, at least in a first-cut design analysis. In other instances, some or all of the high-frequency poles must be taken into consideration from the outset. This section deals with the simplest case—very important in practice—in which there is only one additional pole of significance.

Transfer Functions The open-loop transfer function with just one additional pole can be represented by

$$\begin{aligned} G(s) &= \frac{K_d K_o}{s} \frac{s\tau_2 + 1}{s\tau_1(s\tau_3 + 1)} = \frac{K}{s}\left(1 + \frac{1}{s\tau_2}\right)\frac{1}{s\tau_3 + 1} \\ &= \frac{K}{s}\left(1 + \frac{1}{s\tau_2}\right)\frac{1}{1 + s\tau_2/b} \end{aligned} \tag{2.38}$$

where the third pole is located at $s = -1/\tau_3$, $K = K_d K_o \tau_2/\tau_1$, and $b = \tau_2/\tau_3$. Since this is a third-order PLL, it has three parameters, chosen here as K, τ_2, and b. After some manipulation, the closed-loop transfer functions become

$$\begin{aligned} H(s) &= \frac{K\tau_2(s\tau_2 + 1)}{s^3\tau_2^3/b + s^2\tau_2^2 + Ks\tau_2^2 + K\tau_2} \\ E(s) &= \frac{s^2\tau_2^2(s\tau_2/b + 1)}{s^3\tau_2^3/b + s^2\tau_2^2 + Ks\tau_2^2 + K\tau_2} \end{aligned} \tag{2.39}$$

Normalization One parameter can be hidden by normalization, thereby simplifying the transfer-function expressions. My preference has been to normalize on τ_2, although other normalization choices are also valid. To that end, define dimensionless normalized gain and normalized frequency as

$$K' = K\tau_2, \qquad p = s\tau_2 \tag{2.40}$$

whereupon the normalized closed-loop transfer functions become

$$H(p) = \frac{K'(p+1)}{p^3/b + p^2 + K'p + K'}$$
$$E(p) = \frac{p^2(p/b+1)}{p^3/b + p^2 + K'p + K'} \tag{2.41}$$

[**Comment**: Observe that in a second-order type 2 PLL, the product $K\tau_2 = 4\zeta^2$. That relationship does not carry over into the third-order loop, where the concept of damping factor in the presence of three poles becomes ambiguous at best. Possible alternative meanings of damping factor are pursued below.]

Frequency Response Let the rectangular components of the complex variable p be defined by $p = u + jv = \sigma\tau_2 + j\omega\tau_2$. For any specified values of K' and b it is a simple matter to evaluate $|H(jv)|$ or $|E(jv)|$ for a range of values of v, as exemplified in Fig. 2.9, where the frequency scale is normalized to $v/K' = \omega/K$. However, specific values of K' and b give only single curves; an entire family of curves would be much more useful. With two independent parameters remaining after normalization, the frequency response cannot be presented as a single two-dimensional family of curves as was done in Figs. 2.4 to 2.7. Instead, the family would have to be drawn either as a three-dimensional chart of multiple surfaces, or as multiple charts of families of two-dimensional curves. Each two-dimensional chart would apply for a single value of one parameter (say, K') and would have multiple curves for the other parameter (say, b). Neither alternative provides a satisfactory display. The problem becomes even more intractable for higher-order loops. Alternative graphical methods are treated in Chapter 3.

Large-K' Approximation Usual values of K' will be in the range 1 to 10, but occasional applications will demand much larger values, 50 to 100 or more. Although such large values are not often encountered, it is instructive to examine the approximate system transfer function for large K'. To that end, consider an approximation of the transfer function for low frequencies (i.e., for small $|p|$), denoted $H_L(p)$, for which the cubic term in the denominator may be neglected, and an approximation for high frequencies (i.e., for large $|p|$), denoted $H_H(p)$, for which the zero-order terms in the numerator and denominator may be neglected. As a further constraint, b should not be small either. The two approximate transfer functions then become

$$H_L(p) = \frac{K'(p+1)}{p^2 + K'p + K'}, \qquad H_H(p) = \frac{K'}{p^2/b + p + K'} \tag{2.42}$$

Poles of the expressions of (2.42) are readily determined. Making further approximations based on the expansion of a binomial, the lower-frequency pole of $H_L(p)$ is found to be near $p_L \approx -(1 + 1/K')$, just a small distance beyond the location of the zero at $p = -1$. This zero–pole pair is responsible for gain peaking as

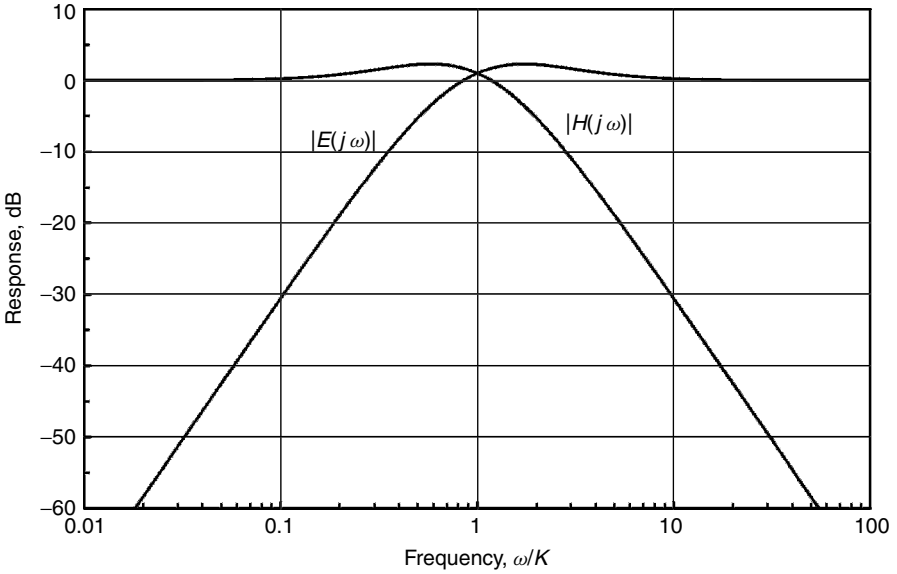


Figure 2.9 Amplitude–frequency responses for a third-order type 2 PLL. Parameters: $b = 9$; $K' = K\tau_2 = 3$.

described in (2.25). As K' increases, the pole moves closer to the zero, thereby reducing the gain peaking.

By similar means, the two poles of $H_H(p)$ are found close to

$$p_H = -\frac{b}{2}\left(1 \pm \sqrt{1 - 4K'/b}\right) \tag{2.43}$$

These poles are real and separate if $K' < b/4$, but they are equal and coincident at $p_H = -b/2$ if $K' = b/4$ (approximately). For $K' > b/4$, the high-frequency poles become complex, with real part $-b/2$. Such behavior would not be discovered if the third pole were simply neglected.

Further inspection of (2.42) reveals that $H_L(p)$ has the same format as that of the second-order type 2 PLL of (2.20), whereas $H_H(p)$ has the same format as that of the second-order type 1 PLL of (2.33). Because the denominators are second order, each of the approximations can be assigned meaningful values of damping (ζ_L and ζ_H) and of normalized natural frequency ($\omega_L\tau_2$ and $\omega_H\tau_2$), defined by

$$\omega_L\tau_2 \approx \sqrt{K'}, \qquad \zeta_L \approx \frac{1}{2}\sqrt{K'} \tag{2.44}$$

$$\omega_H\tau_2 \approx \sqrt{bK'}, \qquad \zeta_H \approx \frac{1}{2}\sqrt{\frac{b}{K'}}$$

If K' is large (as required by the approximation), ζ_L will also be large.

Example If $K' \approx 50$, then $\zeta_L \approx 3.5$. Thus, $H_L(p)$ is always substantially overdamped, within the constraints of the approximation. On the other hand, ζ_H becomes unacceptably small if K' is allowed to be larger than b.

Stability Bounds If $\tau_3 = \tau_2$ (i.e., $b = 1$) is substituted into (2.38), the third open-loop pole cancels the stabilizing zero, and the resulting transfer functions are identical to those of (2.37), in which the loop filter is only a simple integrator. Thus, $b = 1$ is a stability boundary; the loop is unstable if $b < 1$. On the other hand, the approximate high-frequency poles revealed in (2.43) indicate that a third-order type 2 PLL will be stable for all K' no matter how large, provided only that $b > 1$. Of course, damping will be unacceptably small if K' is too large.

2.3.4 Higher-Type PLLs

Section 5.1.1 explains why a type 3 PLL is sometimes needed. Transfer functions of a third-order type 3 PLL can be represented as

$$\begin{aligned} G(s) &= \frac{K}{s}\left(1 + \frac{K_2}{K_1 s} + \frac{K_3}{K_1 s^2}\right) \\ H(s) &= \frac{K\left(s^2 + \frac{K_2}{K_1}s + \frac{K_3}{K_1}\right)}{s^3 + K\left(s^2 + \frac{K_2}{K_1}s + \frac{K_3}{K_1}\right)} \\ E(s) &= \frac{s^3}{s^3 + K\left(s^2 + \frac{K_2}{K_1}s + \frac{K_3}{K_1}\right)} \end{aligned} \tag{2.45}$$

Three independent parameters are needed to characterize this PLL, so its properties are not grasped as readily as those of a second-order loop. Some of the properties are explored further in Chapter 3.

Phaselock loops of type 4 or higher are extremely rare; they are not considered in this book.

REFERENCE

2.1 D. H. Wolaver, *Phase-Locked Loop Circuit Design*, Prentice Hall, Englewood Cliffs, NJ, 1991.

CHAPTER 3

GRAPHICAL AIDS

Families of frequency-response curves based on transfer functions are just one form of graphical aid for understanding PLL properties. Various additional graphical representations of transfer functions have been devised over the years, including root-locus plots [3.1], Bode plots [3.2], Nyquist diagrams [3.3], and Nichols charts [3.4]. These are well-established methods that are explained in great detail in numerous books on control systems; the references cited are those of the early originators of each method.

Two of these methods have been used extensively for analysis of PLLs: root-locus plots and Bode plots. Nyquist diagrams and Nichols charts employ the same data as Bode plots, but they graph the data differently. Although root-locus and Bode plots have predominated for analysis of PLLs, the PLL community might benefit from Nichols charts as well. The following pages give accounts of the application of root-locus plots and Bode plots to several PLL transfer functions of interest, followed by a few words on Nyquist diagrams and then a description of Nichols charts. A concluding section tells how closed-loop frequency response graphs are readily produced from the same open-loop data used for Bode plots or Nichols charts.

All material in this chapter applies to analog (i.e., time-continuous, amplitude-continuous) PLLs. Furthermore, since all four techniques are intended for graphical display of transfer functions, and since transfer functions apply only to linear circuits, none of these methods can be applied to nonlinear PLLs.

Phaselock Techniques, Third Edition, by Floyd M. Gardner

3.1 ROOT-LOCUS PLOTS

Considerable insight into the behavior of a phaselock loop can be acquired from the locations of the poles of the closed-loop responses [i.e., the roots of $1 + G(s)$]. Poles change their locations as the loop gain (or some other parameter) is changed. The paths that the poles trace out in their migrations in the complex s-plane are known as the *root loci*. Salient features of the plot (e.g., number of paths, intersections with the axes) can be determined from the locations of the known open-loop poles and zeros, utilizing a few simple rules (see Appendix 3A).

3.1.1 Description of Root-Locus Plots

A locus commonly is drawn for a range of gain, from near zero to very large. The plot starts (zero gain) on the open-loop poles and terminates (infinite gain) on the open-loop zeros, some of which may be located at infinity. The open-loop transfer function for any PLL is given by $G(s) = K_o K_d F(s)/s$. Thus, at least one pole is always located at $s = 0$, in addition to the poles of any integrators within $F(s)$. The open-loop zeros are the zeros of $F(s)$.

Several methods can be used for constructing the plots:

1. The root loci for simple enough transfer functions (e.g., first- and second-order PLLs) can be laid out by inspection. The loci are simple geometric forms, as explained in the sequel. Examples are shown in Figs. 3.1 and 3.2.
2. Intersections of complex branches of the loci with the real or imaginary axis of the s-plane can be determined by methods described in Appendix 3A.
3. Root locations of the characteristic equation are calculated for a range of values of gain K and specified fixed values of other parameters, and the locations then plotted. The locus is the set of all such locations for one fixed set of parameters as K (typically) is varied. Figures 3.3 and 3.4 are

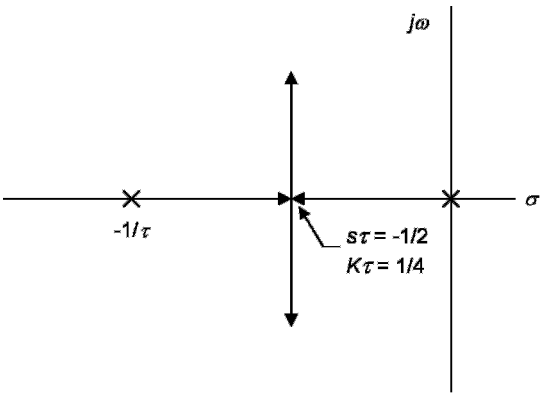


Figure 3.1 Root locus of a second-order type 1 PLL with lag filter $F(s) = 1/(s\tau + 1)$.

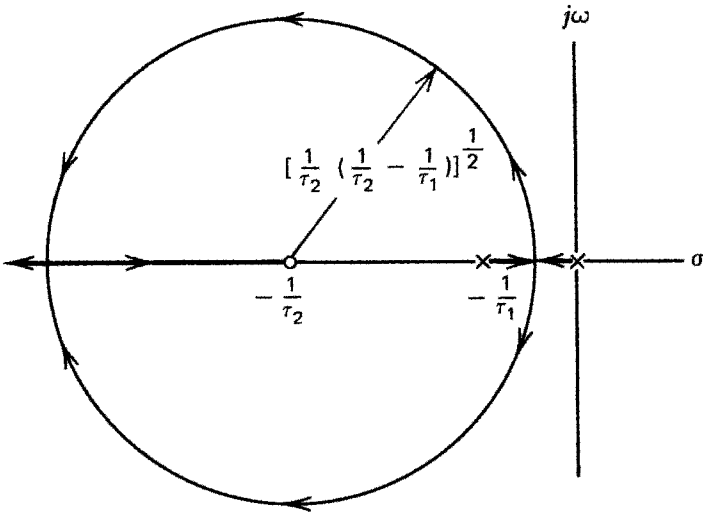


Figure 3.2 Root locus of a second-order type 1 PLL with lag-lead filter $F(s) = (s\tau_2 + 1)/(s\tau_1 + 1)$.

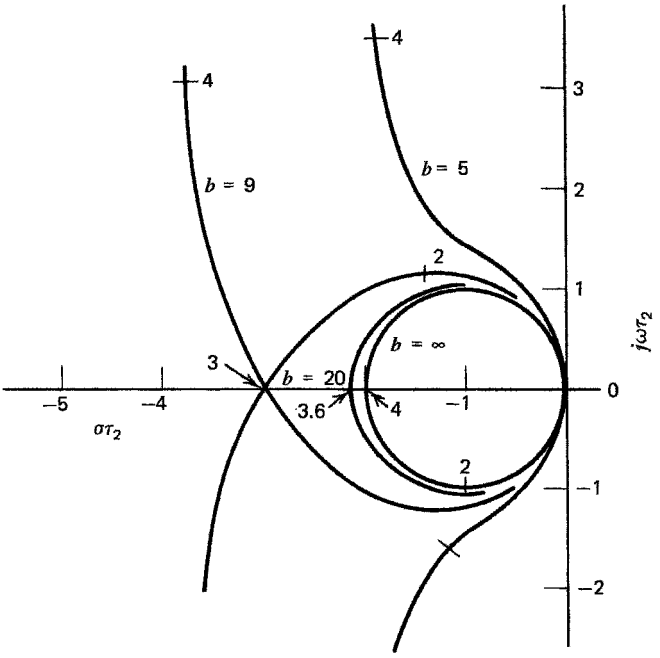


Figure 3.3 Root loci of a third-order type 2 PLL with a zero at $s\tau_2 = -1$ and a third pole at $s\tau_2 = -b$.

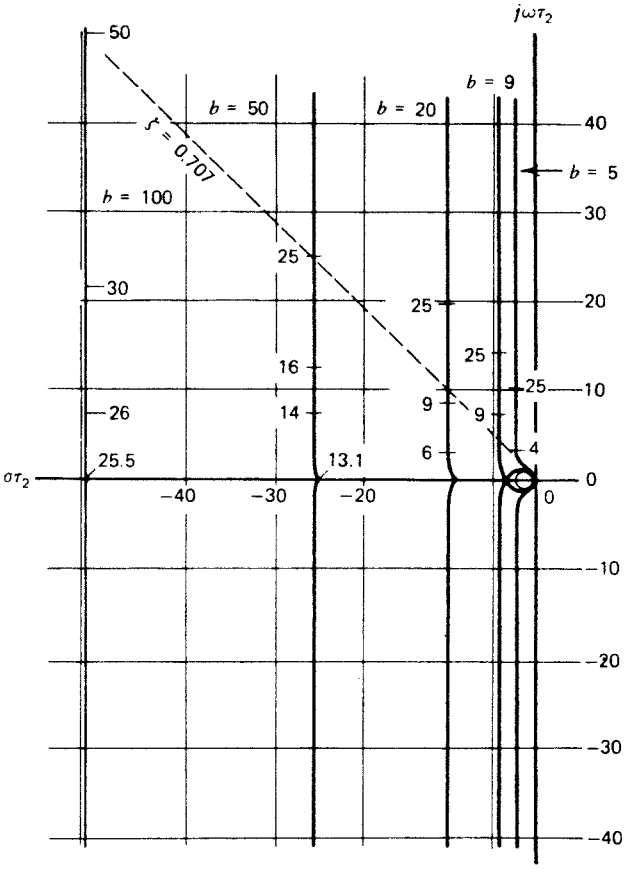


Figure 3.4 Root loci of a third-order type 2 PLL with a third pole at $s\tau_2 = -b$.

examples for which the roots were found numerically with the aid of a calculator and the graphs were plotted manually. Figures 3.6 and 3.7 are newer examples for which the roots were found numerically with a root-solver program on a computer (much quicker than manual iterations on a calculator) and the plots were generated by computer after transferring root locations into a spreadsheet. Calculations available in an ordinary spreadsheet are not adequate for the root-finding task.

4. The processes grow excessively tedious as the number of poles increases, especially in the near vicinity of coincident roots, where tiny changes in K cause comparatively large changes in the root locations.
5. The MATLAB Control-System Toolbox contains a routine `rlocus` that generates a root-locus plot automatically from a specified open-loop transfer function. Similar capabilities may exist in other high-end computer-mathematics programs.

3.1.2 Stability Criterion

A feedback loop is *stable* if all poles lie within the left half of the s-plane and is *unstable* if any pole lies within the right-half plane. The imaginary axis of the s-plane is the dividing line between stability and instability; no closed-loop pole should be on the imaginary axis. Moreover, as a matter of good engineering, no pole should even be close to the imaginary axis because of consequent poor stability margin, inadequate damping, and excessive gain peaking.

3.1.3 Root Loci of Type 1 PLLs

Example root loci of type 1 PLLs are shown to illustrate the simplest cases and earlier design practice.

First-Order Loop As might be expected, the first-order loop [$F(s) = 1$] has the simplest root locus. There is a single open-loop pole at the origin and a single zero at infinity. The closed-loop pole migrates along the negative real axis from zero to infinity as the gain increases.

Loop with Lag Filter A loop that uses only a lag filter $F(s) = 1/(s\tau + 1)$ has two open-loop poles, one at zero and one at $s = -1/\tau$, and two zeros at infinity. The root locus is sketched in Fig. 3.1. As gain increases from zero, the poles migrate toward each other on the negative real axis. After they meet at $K\tau = 1/4$, they become a complex-conjugate pair and move toward infinity along a vertical line at $\sigma = -1/(2\tau)$ as gain increases further. Damping becomes very poor for large values of gain.

Loop with Lag-Lead Filter A lag-lead filter has a transfer function $F(s) = (s\tau_2 + 1)/(s\tau_1 + 1)$. The benefit obtainable from the lead term may be seen in Fig. 3.2. The poles initially migrate toward one another along the negative real axis and become complex where they meet. Because of the finite zero, the complex portion of the locus is now a circle centered at $-1/\tau_2$, not the straight vertical line of Fig. 3.1. Damping is small for moderately small gains, but beyond a minimum gain the damping increases with increasing gain. With sufficiently high gain, the locus eventually returns to the real axis and the loop becomes overdamped. One branch of the locus terminates at the finite zero; the other terminates at infinity on the negative real axis.

3.1.4 Root Loci of Type 2 PLLs

Type 2 PLLs predominate over all others in practice.

Second-Order Loops The plot of Fig. 3.2 is for a second-order type 1 loop. If the loop filter contained a perfect integrator, both open-loop poles would be

located at $s = 0$, the circle portion of the locus would originate at $s = 0$, the center of the circle would lie at $s = -1/\tau_2$, and the radius of the circle would be $1/\tau_2$. Other than that, the plot would be little altered from Fig. 3.2.

Third-Order Loops Root loci of a third-order type 2, PLL are plotted on two different scales in Figs. 3.3 and 3.4. For large values of b (third pole far beyond the τ_2-normalized location of the zero at $p = s\tau_2 = -1$), Fig. 3.3 shows that the close-in root locus (i.e., for relatively small values of $K' = K\tau_2$) is almost the same as that of a second-order type 2 PLL. But Fig. 3.4 shows that the third pole makes its influence strongly felt if K' becomes large enough; the outbound pole (one of those originating from $p = 0$) meets the inbound third pole and the pair become complex, approaching a vertical asymptote at $p = -(b - 1)/2$. This behavior validates the "high gain" analysis in Section 2.3.3, which found the vertical asymptote to lie at $p = -b/2$, approximately. Actually, the vertical asymptote is found at $p = -(b - 1)/2$ for all values of b, not only large values.

Moreover, if b is small, the third pole has a major influence even for small gain values. In fact, if $b < 9$, the two poles originating at $p = 0$ never return to the real axis but remain complex for all values of gain. Small values of b mandate that the third pole be taken into explicit account and not be ignored. The condition $b \approx 9$ may have practical design interest. Situations arise in which you want as large a bandwidth as possible (large K') and as much extra filtering as possible from the third pole (large τ_3), consistent with good damping. The condition $b = 9$ and $K' = 3$ puts all three closed-loop poles coincident at $p = -3$. A smaller value of b impairs the damping, while a larger value of b does not provide as much extra filtering. In a sense, the parameter choice of $b = 9$ and $K' = 3$ (or its close vicinity) is nearly optimum under some requirements. [**Caution**: Practical tolerances on the parameter values are often too loose to permit this quasioptimum choice of parameters to be employed.] If $b < 1$, the two poles originating at $p = 0$ migrate directly into the right-half plane and the loop is unstable for all values of $K' > 0$.

3.1.5 Root Loci of Type 3 PLLs

A third-order type 3 PLL has all three of its open-loop poles at $s = 0$. For illustrative purposes, assume that the two zeros are coincident at $s = -1/\tau_2$. Figure 3.5 shows the root locus for this choice of zeros. The general characteristics of the plot are fairly typical of any type 3 loop that would be considered useful. One feature of the plot is striking: The locus enters the right-half plane for low values of gain; the loop is unstable for low gain. Larger gain brings the poles into the left-half plane and provides stable operation. The loop is said to be *conditionally stable*. This behavior is in direct contrast to the preceding type 1 and type 2 loops, which were unconditionally stable for all values of gain. When a type 3 loop is used, the gain must be prevented from falling into the unstable region.

[**Comments**: (1) Coincident zeros have been found useful by several designers of practical type 3 PLLs. That choice is recommended as optimal by Tausworthe [3.5] and Tausworthe and Crow [3.6]. (2) Countless publications cover

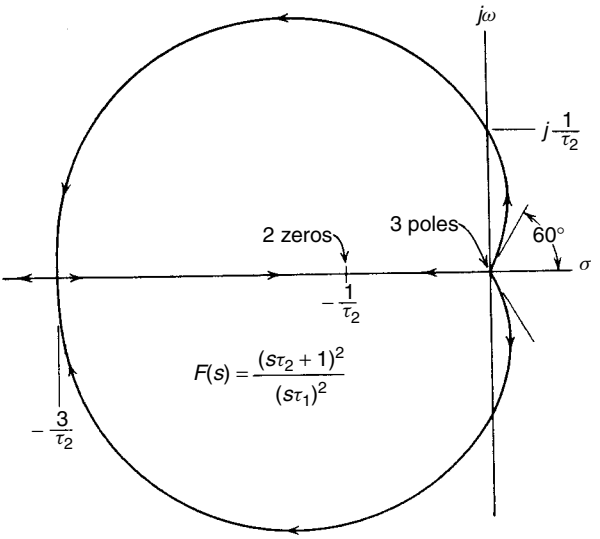


Figure 3.5 Root locus of a third-order type 3 PLL with two zeros coincident at $s\tau_2 = -1$. Intersection with the imaginary axis occurs for $K\tau_2 = \frac{1}{2}$, and intersection with the real axis occurs for $K\tau_2 = \frac{27}{4}$.

the properties of type 1 and type 2 PLLs in exhaustive detail; the references in Chapter 1 list many sources. By contrast, the literature on type 3 PLLs is exceedingly sparse. Besides [3.5] and [3.6], the book by Meyr and Ascheid [3.7] provides analysis methods, design guidance, and performance curves. Several other references are cited in Chapter 5, where situations that warrant a type 3 PLL are described, and in Chapter 8.]

3.1.6 Root Loci of Higher-Order PLLs

Suppose that the PLL has order higher than third: What is the impact on the root-locus plot? If the open-loop poles and zeros are all specified, a single root locus can be calculated and plotted without extreme effort. But the labor increases substantially if a family of root loci is needed for different values of parameters. Consider one of the simpler, but fairly common higher-order cases: a fourth-order type 2 PLL, with the extra poles at $s = -1/\tau_3$ and $-1/\tau_4$. Normalizing on τ_2 as before and taking $K' = K\tau_2$ as the independent variable, there are still two parameters, $b_3 = \tau_2/\tau_3$ and $b_4 = \tau_2/\tau_4$, to take into account. The complete family of root loci can no longer be plotted on one or two drawings, as in Figs. 3.3 and 3.4, but needs a whole sheaf of drawings to cover the $b_3 b_4$ space reasonably. The situation worsens further as more poles are included.

If the pole notation of a fourth-order type 2 PLL is defined such that $\tau_3 \geq \tau_4$, the fourth pole has its largest time constant and greatest effect if $\tau_4 = \tau_3$; example plots for this extreme condition, $\tau_3 = \tau_4 = \tau_2/b$, are shown in Figs. 3.6 and 3.7

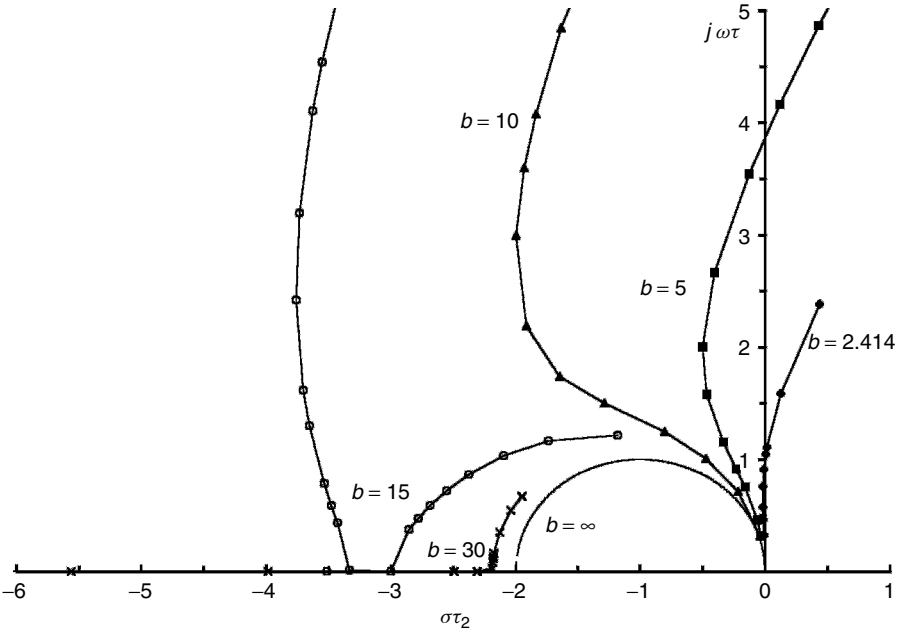


Figure 3.6 Root loci of a fourth-order type 2 PLL with a zero at $s\tau_2 = -1$ and two high-frequency poles coincident at $s\tau_2 = -b$.

for several values of b. To minimize clutter, the plots do not show the lower complex pole of a conjugate pair nor more than a few of the real-pole locations.

Two open-loop poles originate at $p = 0$ and two originate at $p = -b$. Of the two latter open-loop poles, one migrates leftward along the negative real axis toward infinity and the other migrates rightward toward the open-loop zero at $p = -1$. The two open-loop poles originating at $p = 0$ first migrate as a complex-conjugate pair as K' is increased from zero. They always remain complex for all $K' > 0$ if $(8 - 4\sqrt{3}) \approx 1.0718 < b < (8 + 4\sqrt{3}) \approx 14.928$. If $b = (8 + 4\sqrt{3})$, three poles meet at $p = -2(1 + 1/\sqrt{3}) \approx -3.15470$ for $K' = 4(1 + 2/\sqrt{3})/3 \approx 2.8729$, but two poles for that choice of b will be complex for any other value of K'. For larger values of b, there is a range of K' such that all four poles are real and distinct, but for any b there exist large enough values of K' such that two poles will be a complex-conjugate pair.

For any value of $b > 0$, the complex poles eventually migrate into the right-half plane for sufficiently large $K' > 0$, making the loop unstable. This behavior is contrary to that of all the other examples examined so far in which the loop remained stable (although perhaps severely underdamped) no matter how large K' might be. As an extreme case, the fourth-order type 2 PLL of this example is unstable for all $K' > 0$ if $b < 2$.

Figure 3.6 depicts close-in loci for relatively small values of b. Inspection reveals that satisfactory damping ($\zeta > 0.707$, corresponding to pole angles of

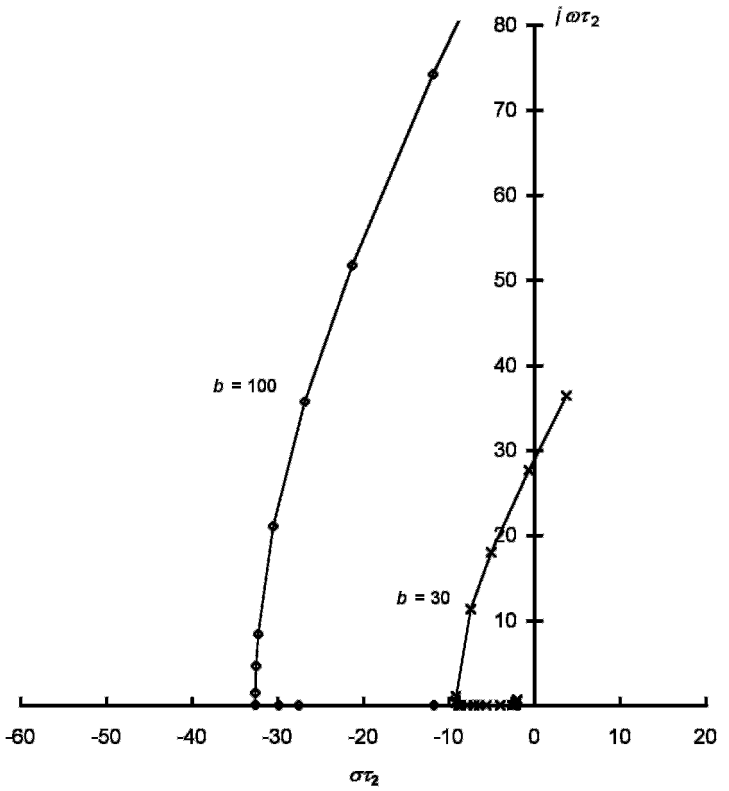


Figure 3.7 Root loci of a fourth-order type 2 PLL with a zero at $s\tau_2 = -1$ and two high-frequency poles coincident at $s\tau_2 = -b$.

$\pm 45^\circ$ from the negative real axis) is unattainable for $b < 10$ and that achieving dominant pole locations close to those of a second-order type 2 PLL (half-circle labeled "$b = \infty$") requires $b = \sim 30$ or more. Examples of complex-pole migration into the right-half plane appear for two of the loci displayed, but the scale is too small to exhibit the instability for all of the loci in this figure.

Take note of the locus for $b = 15$ in Fig. 3.6, particularly its two intersections with the negative real axis at $p = -3.0$ and $-3.33\underline{3}$, occurring at $K' = 2.88$ (exactly) and 2.88066049..., respectively. A gain increase of 0.023% causes a pole shift of 11% in this region of the locus. Regard such extreme sensitivity as a warning to check PLL design parameters for effects of tolerances, especially in the vicinity of multiple poles. It is also a warning to beware of small values of b.

Figure 3.7 plots root loci on a larger scale for larger values of b. This scale is too large to show the close-in behavior in the near vicinity of the zero at $p = -1$, but a good approximation to the close-in behavior of a second-order type 2 PLL can be expected, as has been demonstrated in Fig. 3.6. More pertinent is the large-gain behavior of the locus, which becomes complex and eventually

crosses into the right-half plane for sufficiently large K'. Asymptotically, complex branches of the loci of a fourth-order type 2 PLL approach a straight line that is inclined at $\pm 60^\circ$ to the real axis. Refer to Appendix 3A for methods to calculate the locations and gain values of the locus intersections with the axes, as well as other salient features.

3.1.7 Effect of Loop Delay on Root Locus

A delay circuit with delay τ_d has a transfer function of $\exp(-s\tau_d)$. Consider the simplest first-order PLL with delay whose characteristic equation is

$$s + Ke^{-s\tau_d} = 0 \tag{3.1}$$

The roots of (3.1) are those values of s that satisfy the equation. But unlike the delay-free first-order PLL, which has only one pole, (3.1) has an infinite number of roots because of the exponential term. Root-locus methods cannot be employed for a PLL containing delay within the feedback loop.

3.2 BODE PLOTS

Another useful tool in the study of PLLs is the Bode plot: a pair of curves that displays the polar components of the open-loop transfer function $G(j\omega)$ vs. the radian frequency ω. Customarily, frequency is displayed on a logarithmic abscissa, amplitude $|G(j\omega)|$ is displayed on a dB ordinate, and phase $\text{Arg}[G(j\omega)]$ is displayed in degrees on a linear ordinate. Also, since amplitude typically has straight-line asymptotes on log-log scales, the amplitude plot is often drawn with only the asymptotes as an adequate and convenient approximation to the actual $|G(j\omega)|$.

Bode plots are valuable for several reasons: (1) They provide visual insight to PLL properties that are not apparent from the algebraic transfer-function equations; (2) several loop parameters appear as distinctive points on the graphs; (3) they are well suited for experimental analysis of loop stability; and (4) they can be generated with a tolerable expenditure of labor.

3.2.1 Presentation Options

Bode [3.2] plotted both phase and exact amplitude (not asymptotic approximation) on the same chart; this form is produced readily with a spreadsheet program. Subsequent authors typically have displayed the information in two separated charts with their frequency axes aligned with one another. Separation may be conducive to clearer explanation. Moreover, asymptotic approximation of the amplitude plot allowed quick hand sketching of the amplitude and clearly reveals certain important parameters of the PLL. Pedagogic considerations aside, though, the prevalence of spreadsheets today makes the single chart with phase and exact amplitude the most convenient option for engineering purposes.

Bode plots offer an engineer relief from the display complexity of higher-order loops. Changes in the loop-gain parameter K trivially show up as vertical displacement of the amplitude plot relative to the zero-dB level, without alteration in the shapes of either of the plots or any change in the phase plot. It is not necessary to construct multiple plots for different values of K. If the plot is produced in a spreadsheet, changes in parameters can be entered and observed very quickly without a need for families of curves.

3.2.2 Stability

The degree of stability is a crucial feature of any PLL; a Bode plot is a useful tool for evaluating stability.

Stability Criterion The Bode criterion of stability is simple; a PLL will be stable if its phase lag at the gain crossover frequency ω_{gc} is less than 180°. This criterion is valid under the following restrictions: (1) The amplitude plot crosses 0 dB at only one frequency, and (2) the open-loop transfer function $G(s)$ is stable (no poles in the right-half plane). Since the vast majority of PLLs meet these conditions, the restrictions have imposed little practical constraint on the use of Bode plots. Gain-crossover frequency ω_{gc} was defined in Section 2.3.1 by $|G(j\omega_{gc})| = 1$ (i.e., 0 dB). The term *phase lag* implies a negative value for $\text{Arg}[G(j\omega)]$, so the PLL stability criterion is stated more correctly, in a strict algebraic sense, as $\text{Arg}[G(j\omega_{gc})] > -\pi$ radians.

Stability Margins *Phase margin* in radians is defined as $\text{Arg}[G(j\omega_{gc})] + \pi$. A PLL is stable if its phase margin is positive and unstable if its phase margin is negative. Phase margin not only tells whether a loop is stable but also gives a qualitative indication of the loop damping. The following table shows the phase margin for several values of damping of a second-order type 2 PLL.

Damping, ζ	ω_{gc}/K	Phase Margin (deg)
0.5	1.27	51.8
0.707	1.10	65.6
1.0	1.03	76.4
2.0	1.002	86.4
5.0	1.00005	89.4

Some pertinent observations:

- Since $\zeta = 0.5$ is a rough lower bound on reasonable damping factor, a PLL should have a phase margin of at least 45°, and preferably 60° or more.
- With large damping, the phase margin approaches 90°, a limit imposed by the presence of the pole at $s = 0$ inherent in the VCO.

- The table also shows $\omega_{gc} \approx K$ to be a good to excellent approximation for any damping $\zeta \geq 0.707$.

Gain margin is defined as $-20 \log |G(j\omega_\pi)|$ dB, where phase-crossover frequency ω_π is defined by $\text{Arg}[G(j\omega_\pi)] = -\pi$ (excluding the frequency $\omega = 0$). The phase-crossover frequency does not exist for a second-order type 2 PLL since the phase of $G(j\omega)$ does not cross $-\pi$ for any $\omega > 0$. Moreover, this definition of gain margin is applicable only to feedback loops with *absolute stability* (a loop is stable for all $K > 0$ or for all sufficiently small $K > 0$) and does not apply to *conditionally stable* loops as exemplified in Section 3.2.5. For these reasons, phase margin tends to be a more useful tool than gain margin in many PLLs. Margin concepts are depicted in Figs. 3.8 and 3.9, which show Bode plots for a stable and an unstable loop, respectively. These two charts are representative of feedback amplifiers rather than PLLs, but they illustrate gain and phase margins, nonetheless.

3.2.3 Bode Plots of Type 1 PLLs

The next few sections provide Bode plots for examples of PLLs of various types and orders, beginning with type 1 loops.

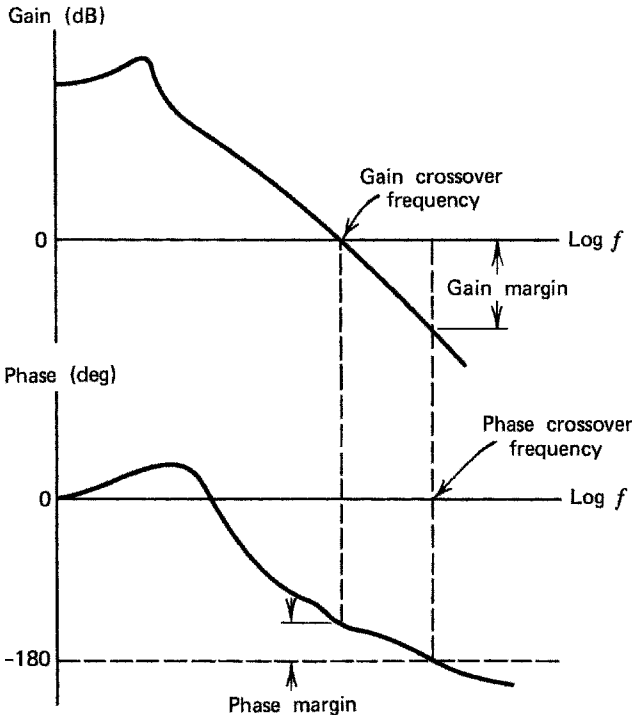


Figure 3.8 Example Bode plot of a stable loop, showing the phase and gain margins.

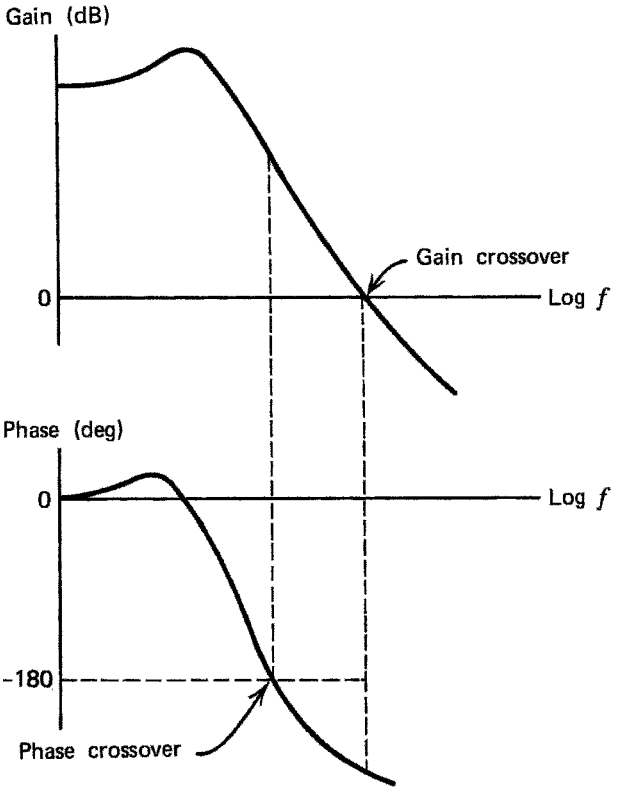


Figure 3.9 Example Bode plot of an unstable loop.

First-Order Loop The Bode plot for a first-order loop is shown in Fig. 3.10. The only frequency-selective term arises from the integration action of the VCO; the magnitude plot is a straight line on the log-log scale, with a slope of -6 dB/octave, and the phase is constant at -90°. A straight-line magnitude curve is exact in this example, not approximate. Since a VCO is present in every PLL, the Bode plot of the VCO is embedded in the plot of every higher-order loop. Gain crossover of the first-order loop occurs at $\omega = K$. The straight line and its crossover completely define the linear dynamics of the first-order loop.

Loop with Lag Filter Insertion of a simple lag filter $F(s) = 1/(s\tau + 1)$ into the loop causes a break in the magnitude curve to an asymptotic -12 dB/octave for frequencies above $\omega = 1/\tau$, as in Fig. 3.11. The break is usually placed at a frequency well beyond gain crossover, so as to obtain a satisfactory value of damping. If the break is at crossover, damping is $\zeta = 0.5$. If the break is at a frequency below crossover, damping will be less than 0.5—a condition to be avoided. Gain crossover of the asymptotic -6 dB/octave straight-line segment (or its extension) occurs at $\omega = K$. Phase is -90° at low frequencies but

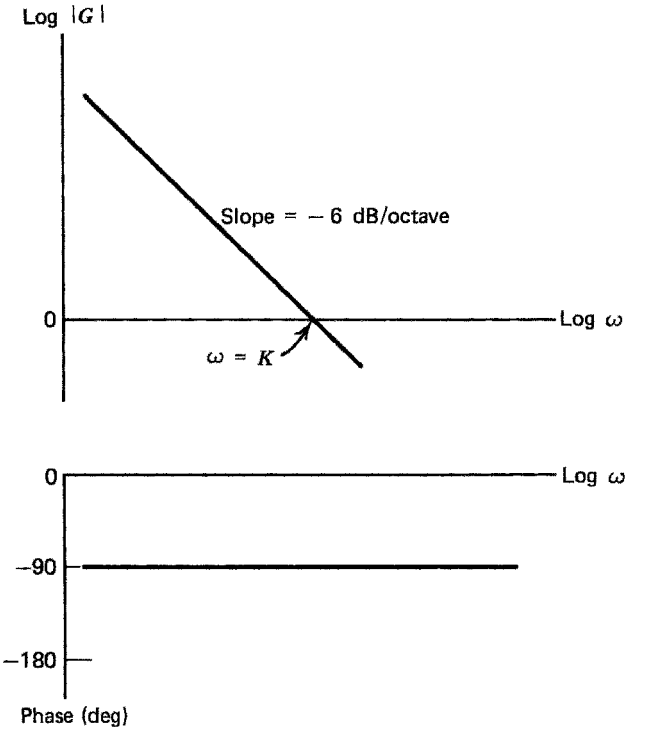


Figure 3.10 Bode plot of a first-order PLL.

approaches -180° at high frequencies. The additional phase lag is 45° at the break frequency.

Loop with Lag-Lead Filter A Bode plot for type 1 loops with a lag-lead filter $F(s) = (s\tau_2 + 1)/(s\tau_1 + 1)$ is shown in Fig. 3.12. At very low frequencies, the VCO integration is dominant, so the asymptotic amplitude slope is -6 dB/octave and the phase is -90°. The pole of the loop filter introduces another corner at $\omega = 1/\tau_1$. Asymptotic slope becomes -12 dB/octave and phase approaches a value of -180° at middle frequencies. The stabilizing zero at $s = -1/\tau_2$ introduces a lead that causes the asymptotic slope to revert to -6 dB/octave and the phase to approach -90° for high frequencies. The break in slope occurs at a frequency $\omega = 1/\tau_2$. Placing the lead break at the unity-gain point yields damping $\zeta = 0.5$. Since smaller damping is rarely wanted, the gain-crossover frequency is almost invariably placed above the break induced by the zero. Crossover of the final -6 dB/octave straight-line segment (or its extension if damping is less than 0.5) occurs at a frequency $\omega = K$. Natural frequency ω_n is the frequency at which the extension of the -12-dB/octave straight-line segment crosses the unity-gain ordinate.

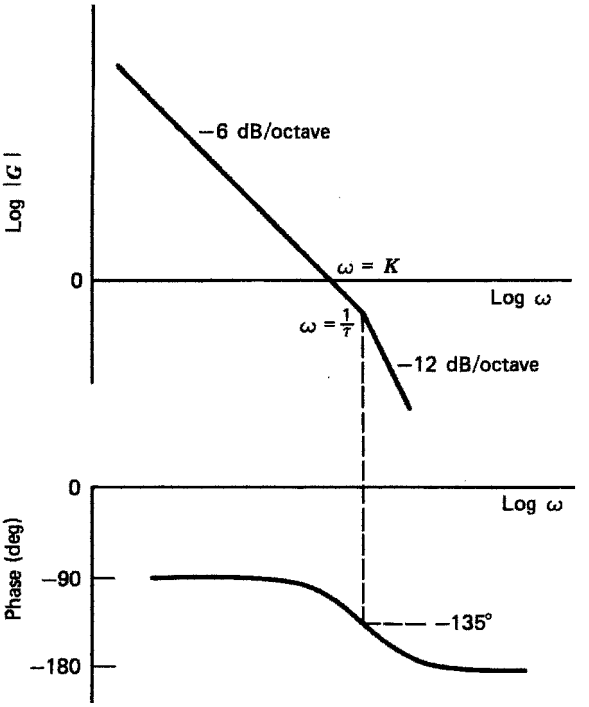


Figure 3.11 Bode plot of a second-order type 1 PLL with lag filter $F(s) = 1/(s\tau + 1)$.

3.2.4 Bode Plots of Type 2 PLLs

A large majority of existing PLLs are type 2 or good approximations thereto. Bode plots for type 2 PLLs build on those shown above for type 1 PLLs.

Second-Order Loops The Bode plot of a second-order type 2 PLL differs from the plot in Fig. 3.12 only at the lowest frequencies: the low-frequency asymptotic magnitude slope remains at -12 dB/octave all the way down to zero frequency—there is no corner at $\omega = 1/\tau_1$—and the low-frequency phase approaches -180° toward zero frequency. Otherwise, mid- and high-frequency behavior is the same as for a type 1 PLL with a lag-lead filter. The graphical depiction of K and ω_n remains unchanged from Fig. 3.12.

Figure 3.13 offers a different presentation for a Bode plot of a second-order type 2 PLL. Instead of separate charts for magnitude and phase, Fig. 3.13 plots both components on the same chart. Moreover, instead of the straight-line asymptotes of the magnitude, Fig. 3.13 plots the true dB-magnitude. The straight-line approximate magnitude is easier to hand-plot, but since Fig. 3.13 was prepared with a spreadsheet program, it was actually easier to plot the true magnitude.

The following comparisons of Figs. 3.12 and 3.13 are germane: (1) The gain-crossover frequency is clearly evident in either figure; (2) as expected, the salient

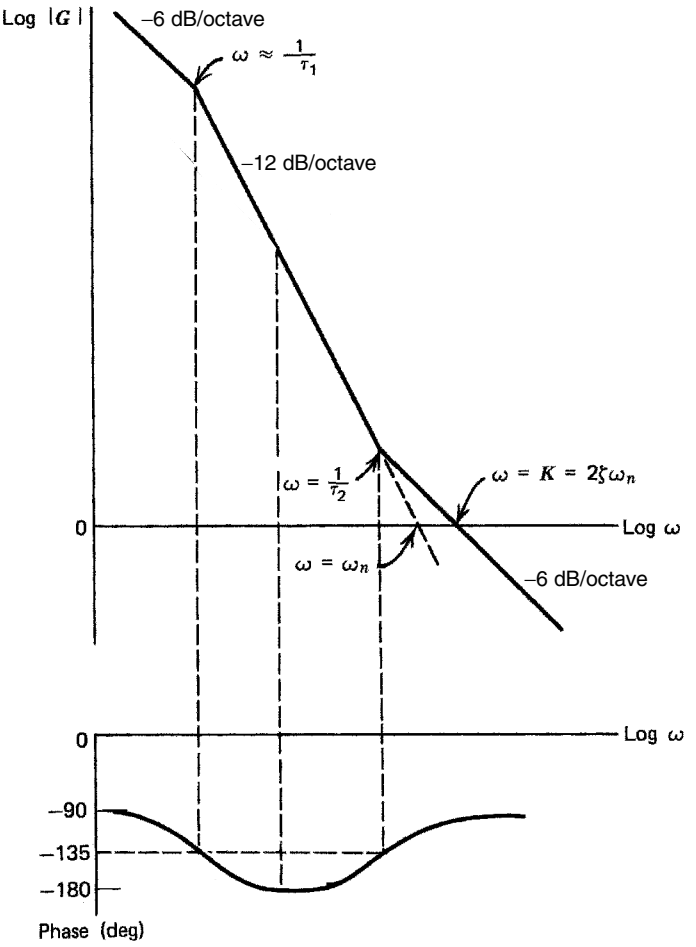


Figure 3.12 Bode plot of a second-order type 1 PLL with lag-lead filter $F(s) = (s\tau_2 + 1)/(s\tau_1 + 1)$.

frequencies at $\omega = 1/\tau_2$ or $\omega = \omega_n$ are easier to make out from the straight-line asymptotes of the magnitude; and (3) a spreadsheet has voluminous numerical data behind the chart. You can refer to the tabular data to obtain any accessible numerical datum, a facility not readily available from hand plotting or burdensome to retrieve from a chart at a reasonable scale. For example, the spreadsheet readily divulges that the normalized gain depicted in Fig. 3.13 is $K\tau_2 = 3$ and that the phase margin is 72°.

Third-Order Loop Figure 3.14 is a Bode plot of a third-order type 2 PLL with $K\tau_2 = 3$ and $b = 9$ (the same normalized parameters as the third-order type 2 PLL whose frequency response is shown in Fig. 2.9 and the same normalized gain as the second-order type 2 PLL whose Bode plot is shown in Fig. 3.13).

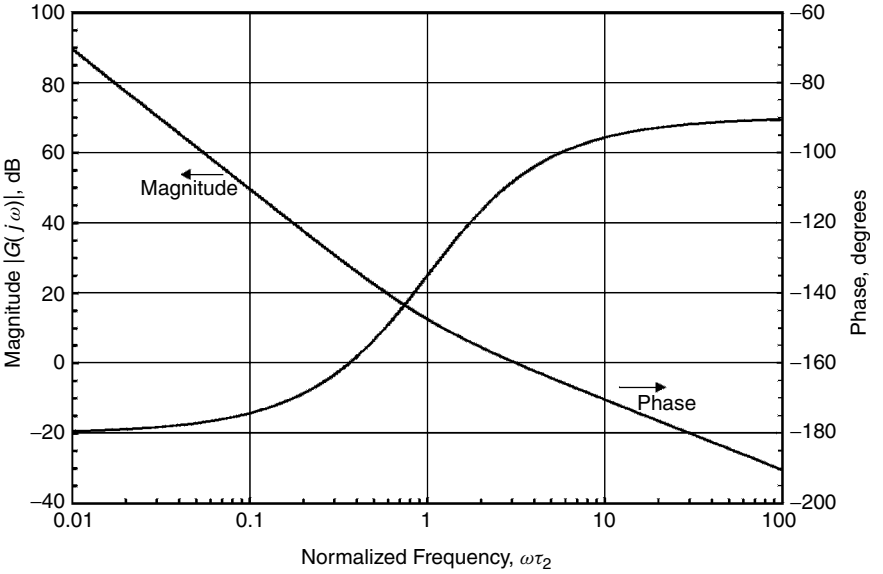


Figure 3.13 Bode plot of a second-order type 2 PLL with a zero at $s\tau_2 = -1$, gain $K\tau_2 = 3$ ($\zeta = 0.866$), and a phase margin of 72°; the same parameters as in the Nichols chart of Fig. 3.18 and the frequency-response graphs of Figs. 3C.1 and 3C.2.

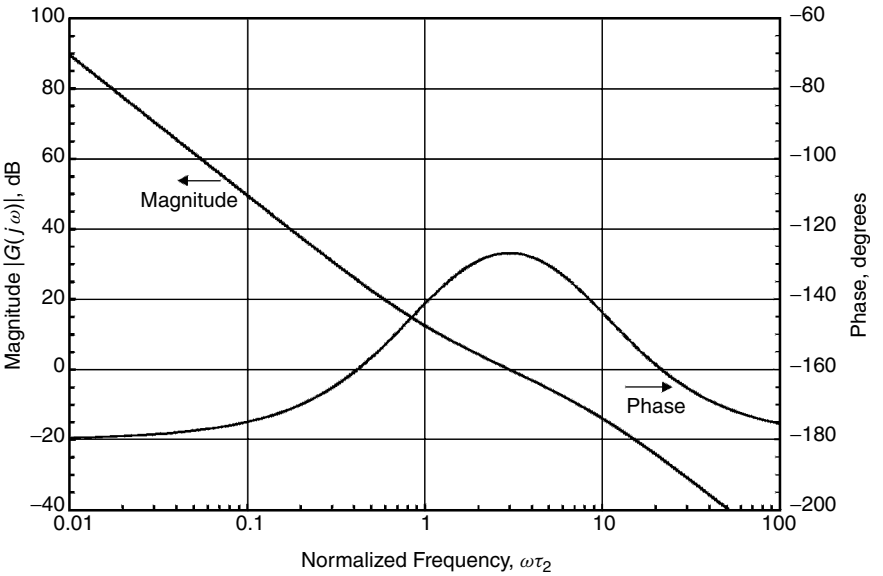


Figure 3.14 Bode plot of a third-order type 2 PLL with $b = 9$, $K\tau_2 = 3$, and a phase margin of 53°; the same parameters as in the frequency-response graph of Fig. 2.9 and the root-locus plots labeled "$b = 9$" in Figs. 3.3 and 3.4.

Comparing Fig. 3.14 to Fig. 3.13, it can be seen that the magnitude curves are but little changed for frequencies $\omega\tau_2 < 9$ (the corner frequency of the third pole), but the phase is affected for frequencies greater than $\omega\tau_2 \approx 1$.

In particular, the gain-crossover frequency is not altered noticeably by the presence of the extra pole. This behavior is typical of feedback loops with comfortable stability margins; additional high-frequency elements impair the phase margin but tend to have little effect on the location of the gain crossover. Under these circumstances, you may find it convenient to approximate the effect of multiple high-frequency elements as a single delay term that has the correct phase at the gain-crossover frequency. In Fig. 3.14 the phase margin is reduced to 53° from a more ample 72° of Fig. 3.13. Since phase lag never reaches −180° for any finite $\omega > 0$, gain margin is undefined for a third-order type 2 PLL (and also for a second-order type 2, PLL).

Also observe that the additional pole has no influence on the 0-dB crossover frequency of the extension of the −12 dB/octave portion of the amplitude curve—the location of $\omega = \omega_n$ in the second-order type 2 PLL. Thus, although the third-order PLL has three poles, so that the standard definition of "natural frequency" for a pole pair (introduced in Section 2.2.3) cannot apply here, the expanded definition remains valid.

Fourth-Order Loop Figure 3.15 is the Bode plot for a fourth-order type 2 PLL with $K\tau_2 = 3$ and the two additional pole corners coincident at $\omega\tau_2 = 30$.

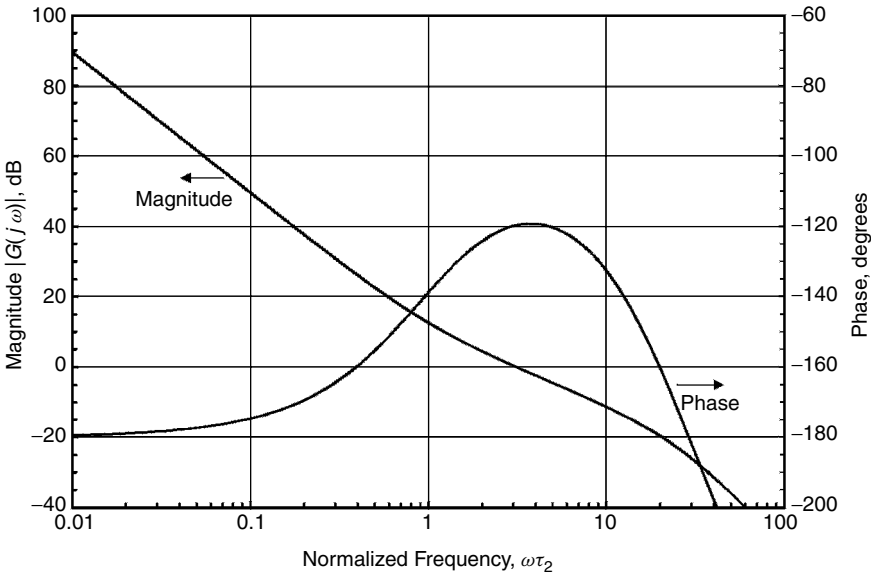


Figure 3.15 Bode plot of a fourth-order type 2 PLL with two high-frequency poles coincident at $s\tau_2 = -b$, gain $K\tau_2 = 3$, $b = 30$, phase margin = 60°, and gain margin ≈ 25 dB; the same parameters as in the root-locus plot labeled "$b = 30$" in Figs. 3.6 and 3.7.

The root-locus plot for this PLL is the one labeled "$b = 30$" in Figs. 3.6 and 3.7. Once again, the filtering has little effect on magnitude at frequencies much below the pole corners, but has an appreciable effect on phase, starting from below the gain crossover and increasing beyond. The fourth pole causes the phase to approach -270° at high frequencies, so phase crossover frequency is now meaningful. Phase crossover occurs near $\omega\tau_2 = 30$ and the gain margin is found to be about 25 dB. The phase margin for this set of parameters is $\sim 60^\circ$. These are still respectable margins but not as ample as for the corresponding second-order type 2 PLL illustrated in Fig. 3.13.

Effect of Delay on Bode Plots Figure 3.16 is the Bode plot for a second-order type 2 PLL with $K\tau_2 = 3$ and a transport delay $\tau_d = \tau_2/10$. Earlier in this chapter it was concluded that root-locus plots cannot cope with transport delay in the loop. However, Bode plots (or Nyquist diagrams or Nichols charts) have no such trouble; just add a phase of $-\omega\tau_d$ radians to the phase of each plot for all ω. Figure 3.16 demonstrates that delay has no influence whatever on the magnitude [since $|\exp(j\omega)| \equiv 1$ for all real ω] but can have a drastic effect on phase. Phase margin is reduced to 54° and gain margin is about 14 dB at a phase-crossover frequency $\omega\tau_2 \approx 15$. The presence of delay in the loop always assures the existence of a phase-crossover frequency because the phase lag of a delay term increases strictly monotonically with frequency.

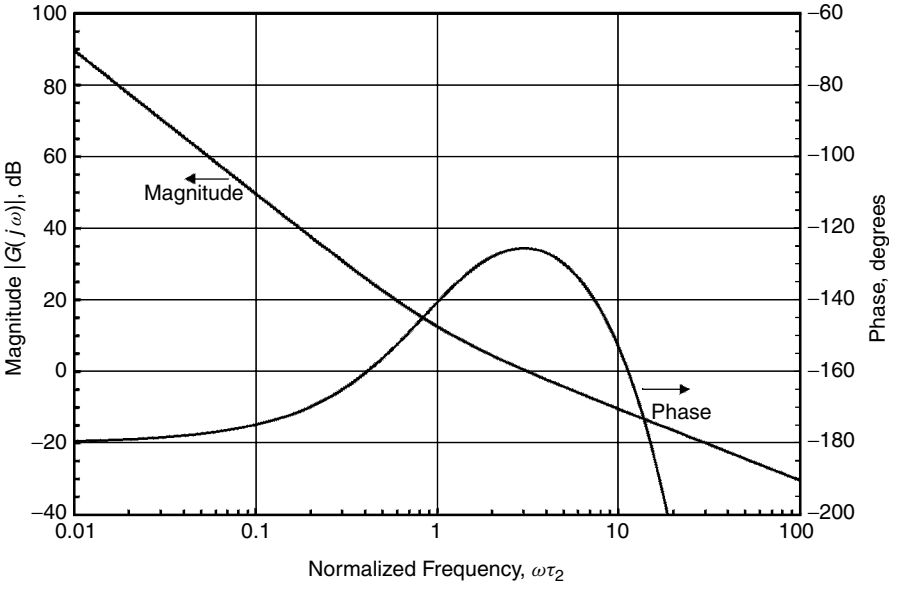


Figure 3.16 Bode plot of a second-order type 2 PLL with an in-loop delay of $\tau_d = \tau_2/10$, gain $K\tau_2 = 3$, phase margin $\approx 54^\circ$, gain crossover at $\omega\tau_2 \approx 15$, and gain margin ≈ 14 dB; the same parameters as in the Nichols chart of Fig. 3.20 and the frequency-response graphs of Figs. 3C.3 and 3C.4.

3.2.5 Bode Plots of Type 3 PLLs

As a last Bode example, Fig. 3.17 shows a Bode plot for the third-order type 3 loop whose root locus was shown in Fig. 3.5. Because the loop filter now contains two ideal integrators, the low-frequency asymptotic slope is -18 dB/octave and the zero-frequency phase is -270°. Two lead zeros are needed to break the asymptotic slope to -6 dB/octave around the gain-crossover frequency; the zeros are arbitrarily shown as coincident at $\omega = 1/\tau_2$. Gain crossover occurs again at $\omega = K$ on the asymptotic straight-line amplitude curve. Observe that this plot offers no definition of ω_n, contrary to the definition as the extended intercept of a -12-dB/octave slope in plots of type 2 PLLs. Absence of a definition reflects the fact that natural frequency, strictly speaking, is a property of a second-order system; its absence is inherent in the type 3 character of the example PLL and is not a deficiency of the Bode plot.

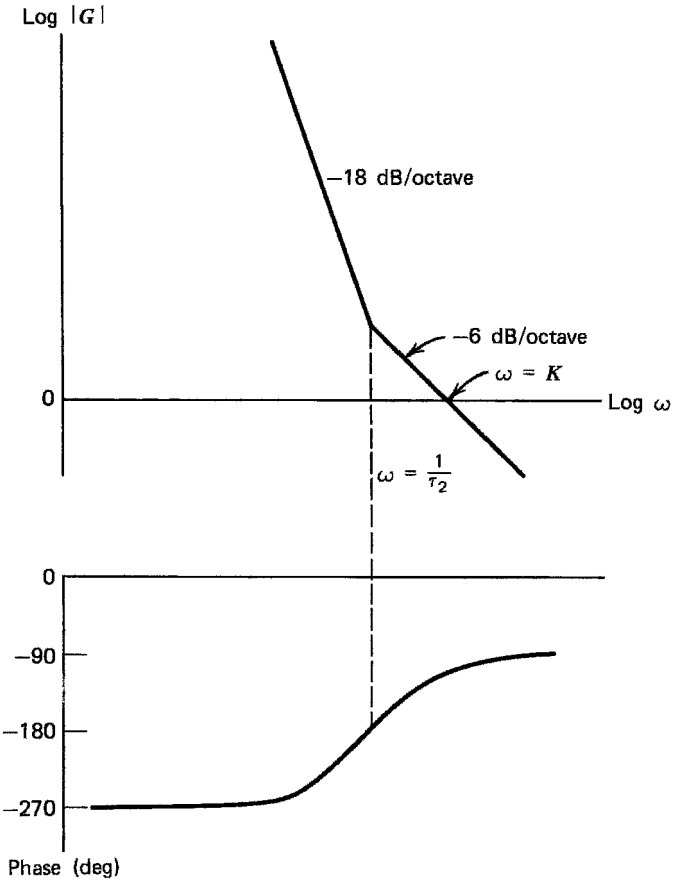


Figure 3.17 Bode plot of a third-order type 3 PLL with two zeros coincident at $\omega = 1/\tau_2$; the same parameters as in the root-locus plot of Fig. 3.5.

The loop becomes unstable if gain is *reduced* so that gain crossover occurs at a frequency of less than ω_π, the frequency at which phase $= -180^\circ$. (For the example plot, ω_π happens to coincide with $\omega = 1/\tau_2$ since the two zeros each contribute exactly 45° of lead at that frequency.) Instability from reduced gain is characteristic of the conditional stability encountered in all type 3 (or higher-type) PLLs.

3.3 NYQUIST DIAGRAMS

A Nyquist diagram is a plot of $G(j\omega)$ in the complex $G(j\omega)$-plane, with frequency ω as the parametric variable. A Nyquist diagram is not limited by the same constraints that apply to Bode plots (and to Nichols charts as well). Nyquist plots can cope with multiple feedback loops and with open-loop poles in the right-half plane. But since almost no PLLs possess these characteristics, the broader applicability of Nyquist diagrams has not mattered. Loop stability is indicated in a Nyquist diagram by encirclements of the $-1 + j0$ point in the complex $G(j\omega)$-plane. The Nyquist stability criterion is less convenient to evaluate than those for Bode plots or Nichols charts, even though it is more broadly applicable. Because $G(j\omega)$ is plotted on linear axes (not dB for magnitude as in Bode plots or Nichols charts), the highest- and lowest-amplitude portions of the diagram are either off-scale or too small to see. Thus, although Nyquist diagrams are powerful tools, they have not seen much use for analysis and design of PLLs. They are not considered further in this book.

3.4 NICHOLS CHARTS

A Nichols chart is a rectangular-coordinate plot of the polar components of $G(j\omega)$: that is, $|G(j\omega)|$ in dB as ordinate vs. Arg$[G(j\omega)]$ in degrees as abscissa. Frequency ω is a parametric variable and does not appear explicitly in the plot. The plot encompasses only a single curve, a feature that might facilitate quicker visual interpretation than is achievable from the two curves of a corresponding Bode plot, even though extracted from exactly the same data. Because of the dB scale for $|G|$, the Nichols chart clearly displays the large and small $|G|$ regions that are obscured in a Nyquist diagram. Nichols charts can be enhanced by *M-contours* (described in Section 3.4.2), which are used to evaluate the peak gain of the closed-loop system response. A Nichols chart does not display frequency information; a Bode plot is superior when visible frequency information is required. Nichols charts have been widely employed by our control system colleagues for many years to design feedback systems; you might do well to look into their use for PLLs.

3.4.1 Stability Criterion

The stability criterion of Nichols charts is subject to the same restrictions as those applicable to Bode plots: The open-loop transfer function must be stable

and the Nichols curve must cross 0 dB at only a single point. A PLL is stable if its gain crossing in the Nichols chart is at a phase that is more positive than -180°. Phase margin is the sum of 180° plus the phase at the gain crossing. Gain margin is the negative of the dB gain at the -180° phase crossing (if any). Margins are more readily visible on a Nichols chart than on Bode plots or Nyquist diagrams.

3.4.2 *M*-Contours

Closed-loop system response $H(j\omega) = G(j\omega)/[1 + G(j\omega)]$ can be calculated from knowledge of open-loop response $G(j\omega)$. Every point in a Nichols chart corresponds to a distinct value of G and thus to H. Represent the closed-loop response in polar format as $H = Me^{j\alpha}$. Then each point in the Nichols chart corresponds to a distinct pair of values of M and α. Curves connecting equal values of M are known as *M-contours*, and curves connecting equal values of α are known as *α-contours*. Methods for calculating these contours are explained in [3.4] and many later texts on control systems.

The contours are properties of the underlying Nichols chart itself and do not depend on particular $G(j\omega)$ transfer functions. Thus, the contours can be produced as part of an otherwise blank chart and used for evaluating any admissible transfer function. The M-contours are particularly useful since they identify the peak gain of $|H|$ for any G that might be plotted. Gain peaking can be an important issue in design of PLLs.

3.4.3 Examples of Nichols Charts

Several examples of Nichols charts for type 2 PLLs are shown in Figs. 3.18 to 3.20. These charts were produced with the same spreadsheet as that used for the Bode plots of Figs. 3.13 to 3.16. Figure 3.18 depicts the chart for a second-order type 2 PLL with $K\tau_2 = 3$ ($\zeta = 0.866$). Gain crossover occurs at a phase of -107° to give a phase margin of 73°. Gain margin is infinite since the curve does not cross -180° at any nonzero frequency.

Gain peaking (of $|H|$) can be estimated from the M-contours; the plotted curve passes about midway between the M-contours for $+1$ and $+2$ dB, so the peak of $|H|$ is approximately 1.5 dB. Exact calculation with (2.25) yields 1.55 dB. Exact peaking formulas have not been derived for more complicated PLLs, but the M-contour evaluation of peaking does not require a peaking formula.

A second example in Fig. 3.19 is the same as the first example except that the gain has been increased to $K\tau_2 = 81$ ($\zeta = 4.5$). The shape of the Nichols curve has not been altered; it has merely been shifted upward by $20\log(81/3) = 28.6$ dB. Inspection of the curve reveals that it barely grazes the $+0.1$-dB M-contour [so peaking of $|H(j\omega)|$ does not exceed 0.1 dB] and the 0-dB gain crossing is at a phase of approximately -91° (so the phase margin is about 89°).

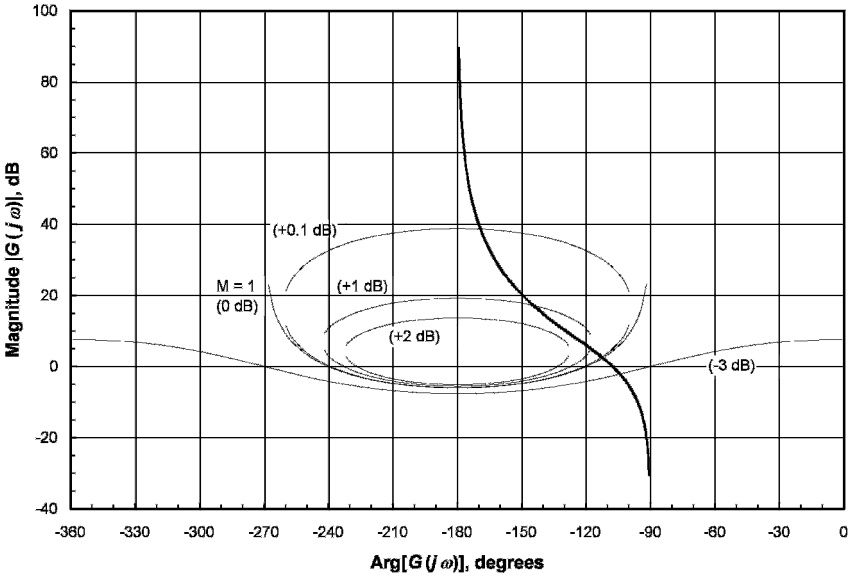


Figure 3.18 Nichols chart of a second-order type 2 PLL with $K\tau_2 = 3$ ($\zeta = 0.866$); the same parameters as in the Bode plot of Fig. 3.13 and the frequency-response graphs of Figs. 3C.1 and 3C.2. The heavy curve is $|G(j\omega)|$ in dB plotted against $\text{Arg}[G(j\omega)]$ in degrees. Light oval curves are M-contours. Gaps at the ends of the ovals are due to computational quantization of the abscissa.

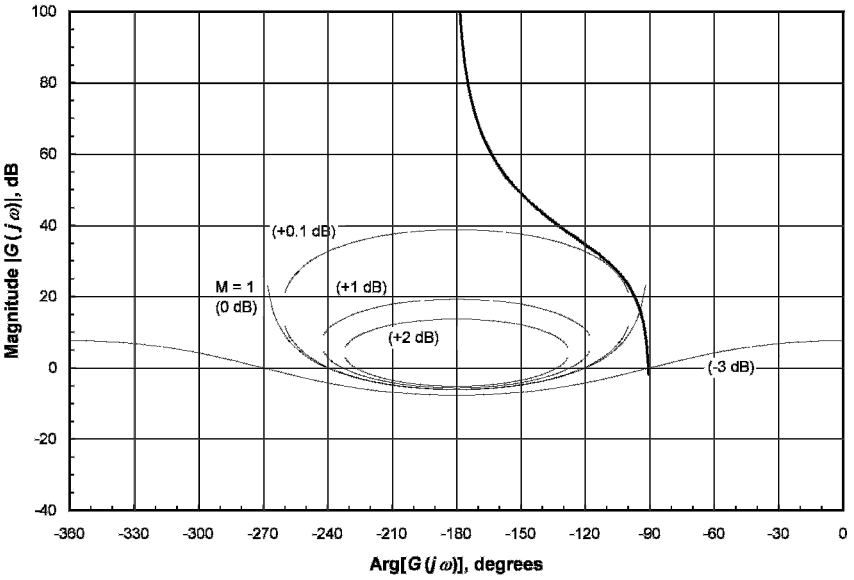


Figure 3.19 Nichols chart of a second-order type 2 PLL with $K\tau_2 = 81$ ($\zeta = 4.5$) and phase margin $\approx 89^\circ$.

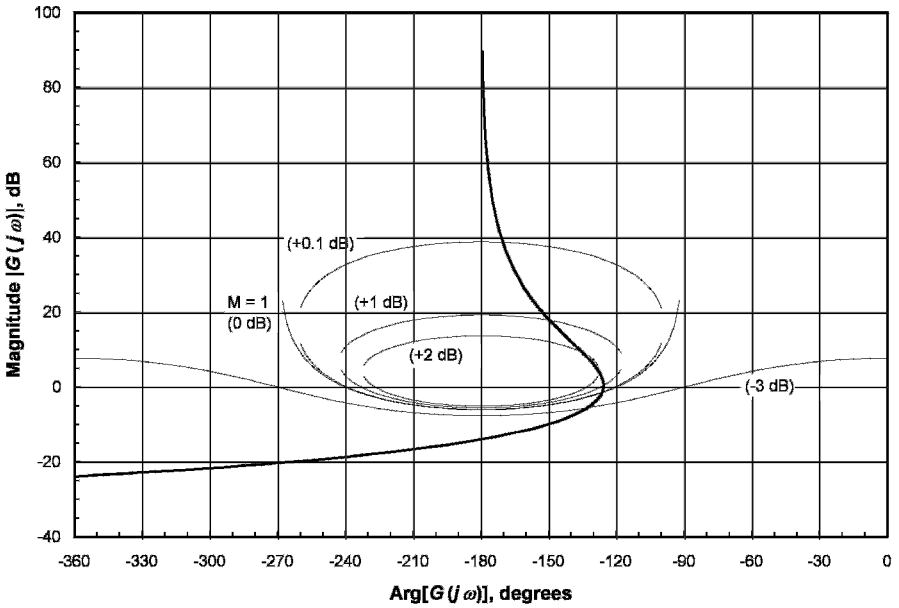


Figure 3.20 Nichols chart of a second-order type 2 PLL with an in-loop delay of $\tau_d = \tau_2/10$, gain $K\tau_2 = 3$, phase margin $\approx 54^\circ$, and gain margin ≈ 14 dB; the same parameters as in the Bode plot of Fig. 3.16 and the frequency-response graphs of Figs. 3C.3 and 3C.4.

The third example, shown in Fig. 3.20, is the same as the example in Fig. 3.18, with the addition of a delay of $\tau_d = \tau_2/10$. It has the same conditions as in the example for the Bode plot of Fig. 3.16. Inspection of the Nichols curve quickly reveals gain peaking of slightly more than 2 dB, a phase margin of 54°, and a gain margin of about 14 dB.

3.5 CLOSED-LOOP FREQUENCY-RESPONSE CURVES

It turns out that the closed-loop frequency responses $|H(j\omega)|$, $\text{Arg}[H(j\omega)]$, $|E(j\omega)|$, and $\text{Arg}[E(j\omega)]$ vs. frequency ω are easily calculated in a spreadsheet that generates Bode plots or Nichols charts. All necessary data are present already; only a few additional formulas and data columns are needed. Principles and examples are provided in Appendix 3C.

APPENDIX 3A: SALIENT FEATURES OF ROOT LOCI

A root locus consists of all of the points in the s-plane that satisfy the characteristic equation $1 + G(s) = 0$, as the value of a coefficient in $G(s)$ (most typically, the gain K) is varied over an appropriate range (e.g., $K = 0$ to ∞). The zeros of

$1 + G(s)$ define the poles of the closed-loop transfer functions $H(s)$ and $E(s)$. If, as is usual, the coefficients of $G(s)$ are all real, then any complex roots of $1 + G(s)$ occur in conjugate pairs. This appendix tells how to calculate several prominent features of a root locus.

3A.1 Branches of Root Loci

The number of distinct branches of a root locus is equal to the number of finite roots of the characteristic equation: that is, the number of poles in the closed-loop transfer functions. Some of the branches terminate on the finite zeros, while others extend to $|s| = \infty$. If the magnitude of s is sufficiently large, the characteristic equation is approximated by

$$G(s) = -1 \approx \frac{aKs^{N_z}}{s^{N_p}} = \frac{aK}{s^{N_p - N_z}}$$

where N_p and N_z are the number of finite poles and zeros of $1 + G(s)$ and aK is the ratio of coefficients of the highest powers of s in $G(s)$. [**Comment**: The *zeros* of $1 + G(s)$ are the *poles* of $H(s)$ or $E(s)$.]

For any point s on the locus, the angle of $G(s)$ is an odd multiple of π inasmuch as $-1 = e^{j(2k-1)\pi}$. That is,

$$\text{Arg}[G(s)] = \text{Arg}\left[\frac{aK}{s^{N_p - N_z}}\right] = (2k - 1)\pi$$

for an integer k, where Arg$[x]$ means the phase angle of the complex quantity x. The product aK is real and positive, so the asymptotic angles of $G(s)$ are established by those values of k that satisfy

$$\text{Arg}\left[\frac{1}{s^{N_p - N_z}}\right] = (2k - 1)\pi$$

which reduces to

$$\text{Arg}[s] = \frac{(2k - 1)\pi}{N_p - N_z}$$

For example, if $N_p - N_z = 3$, then the asymptotic angles for $k = 0$, 1, and 2 are -60°, $+60^\circ$, and 180°, respectively. Any other value of k for $N_p - N_z = 3$ yields an angle such as $300^\circ \equiv -60^\circ$ that reduces to one of the three angles found for $k = 0$, 1, or 2. The asymptotes do not generally meet at $s = 0$ but at some point on the negative real axis.

3A.2 Locus on the Real Axis

Any portion of the locus on the real axis is defined by $G(\sigma + j0) = -1$, where σ is the real part of s. Equivalently, Arg$[G(\sigma)] = (2k - 1)\pi$. Complex-conjugate

pairs of poles or zeros of $G(s)$, irrespective of location, contribute 0° to the angle of $G(s)$ at $s = \sigma + j0$. Real poles or zeros to the left of σ also contribute 0° to the angle of $G(\sigma)$, but each real pole or zero to the right of σ contributes $-180°$ or $+180°$, respectively, to the angle of $G(\sigma)$. Therefore, any portion of the root locus that lies on the real axis has an odd number of real open-loop poles and zeros lying to its right. If, as is true in just about every PLL arrangement, no open-loop poles or zeros are on the positive real axis, any real portions of the root locus can appear only on the negative real axis or at $s = 0$.

3A.3 Locus Intersections with Axes

The locations of the intersections of a root locus with the imaginary or real axis is valuable information for a PLL engineer. An example from a particular PLL configuration is shown here rather than a derivation of the general conditions for intersections. The case selected is the fourth-order type 2 PLL of Section 3.1.6. This PLL has two open-loop poles at the origin, one zero at $s = -1/\tau_2$, and two coincident open-loop, high-frequency poles located at $s = -b/\tau_2$. Normalizing on τ_2 gives $p = s\tau_2$, so that the open-loop zero is at $p = -1$ and the two open-loop, high-frequency poles are at $p = -b$. The characteristic polynomial of this case is

$$\frac{p^4}{b^2} + \frac{2p^3}{b} + p^2 + K'p + K' = 0 \tag{3A.1}$$

where $K' = K\tau_2$ is the normalized gain.

Intersections with the Imaginary Axis The imaginary axis is the stability boundary of the PLL; the loop is unstable if the locus crosses into the right half of the s-plane. At a crossing, the imaginary poles will be a conjugate pair at $p = \pm jp_0$, where p_0 (defined as a real value) is to be determined. In addition, there are two more poles at unknown locations $-p_1$ and $-p_2$. Now divide (3A.1) by $p^2 + p_0^2$ to obtain a remainder of $p(K' - 2p_0^2/b) + K' - p_0^2(1 - p_0^2/b^2)$. The remainder is zero for all p if

$$K' - \frac{2p_0^2}{b} = 0 \quad \text{and} \quad K' - p_0^2 + \frac{p_0^4}{b^2} = 0 \tag{3A.2}$$

Subtract the first condition from the second to obtain

$$p_0^2\left(\frac{p_0^2}{b^2} + \frac{2}{b} - 1\right) = 0$$

which identifies pairs of poles on the imaginary axis at $p_0 = 0$ (the open-loop integrator poles of the type 2 PLL) and at

$$p_0 = \pm b\sqrt{1 - \frac{2}{b}} \tag{3A.3}$$

Equation (3A.3) is the result that has been sought for p_0.

A real value exists for p_0 only for $b > 2$; no nonzero intersection with the imaginary axis exists for smaller values of b. That finding implies that the complex branch of the locus is entirely in the right-half plane for $b < 2$ and any $K' > 0$; the loop is unstable for $b < 2$. To find the value of K' that corresponds to p_0 (i.e., the stability-boundary value of K'), solve the first equation of (3A.2) for p_0 in terms of b and K', substitute the result into the second equation of (3A.2), and thereby find the roots $K' = 0$ (the two open-loop poles at $p = 0$) and

$$K' = 2(b - 2) \tag{3A.4}$$

which is applicable only for $b > 2$.

Intersections with the Real Axis Refer to Fig. 3A.1, which shows the preceding example's open-loop poles and zeros in the p-plane along with a point on the root locus located at $p = -d + j\Delta$. Consider Δ to be infinitesimally small, so that $p = -d$ is an intersection of a complex portion of the locus with the real axis. Angles of the vectors from the open-loop poles and zeros to the point $p = -d + j\Delta$ must add up to an odd multiple of 180°. The magnitudes of contributions from those poles and zeros to the right of $-d$ are each slightly less than 180°, while the contributions from the two poles to the left of $-d$ are each slightly more than 0°. Zeros contribute a lead (positive) phase to $G(s)$ and poles contribute a lag (negative) phase. That is,

$$\begin{aligned}(2k-1)\pi &= -2\left(\pi - \tan^{-1}\frac{\Delta}{d}\right) + \left(\pi - \tan^{-1}\frac{\Delta}{d-1}\right) - 2\tan^{-1}\frac{\Delta}{b-d} \\ &= -\pi + 2\tan^{-1}\frac{\Delta}{d} - \tan^{-1}\frac{\Delta}{d-1} - 2\tan^{-1}\frac{\Delta}{b-d}\end{aligned} \tag{3A.5}$$

Set $k = 0$ so that the π terms cancel, which leaves

$$0 = +2\tan^{-1}\frac{\Delta}{d} - \tan^{-1}\frac{\Delta}{d-1} - 2\tan^{-1}\frac{\Delta}{b-d}$$

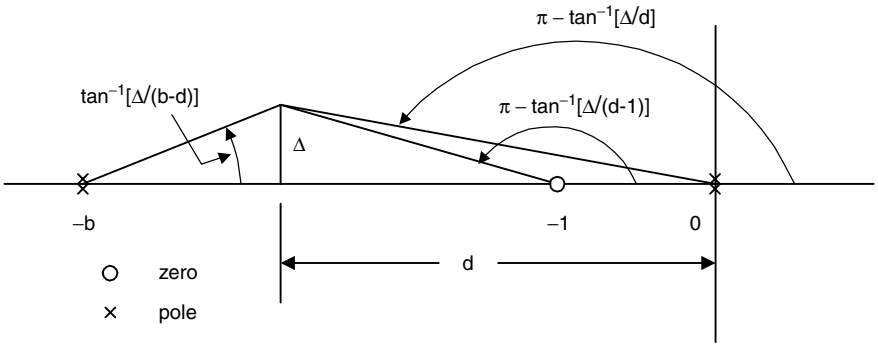


Figure 3A.1 Geometric construction for determining intersection with the negative real axis of the root-locus plot of a fourth-order type 2 PLL.

Since Δ is infinitesimal, these three angles are all small and can be approximated by their tangents to obtain

$$-\frac{2\Delta}{d} + \frac{\Delta}{d-1} + \frac{2\Delta}{b-d} = 0$$

After combining all terms over a common denominator and factoring out Δ, the resulting numerator is

$$3d^2 - (4+b)d + 2b = 0$$

Roots of this equation lie at

$$d = \frac{4 + b \pm \sqrt{b^2 - 16b + 16}}{6} \tag{3A.6}$$

The discriminant is nonnegative for $b \leq 4(2-\sqrt{3}) \approx 1.072$ or $b \geq 4(2+\sqrt{3}) \approx 14.928$. Since $b = 2$ is a stability boundary, the smaller limit on b represents an unstable condition and is of no practical use in a PLL design. The larger limit at $b = 4(2+\sqrt{3})$ indicates the existence of three coincident real poles on the negative real axis at $p = -(4+b)/6 = -2(1+1/\sqrt{3}) \approx -3.15470$.

APPENDIX 3B: FORMATS OF THE OPEN-LOOP TRANSFER FUNCTION $G(s)$

Section 2.3.1 presented a broadly general formulation of the open-loop transfer function $G(s)$ in the form

$$G(s) = \frac{K_d K_o F_{\mathrm{p+i}}(s) F_{\mathrm{hf}}(s)}{s}$$

wherein the loop filter is partitioned into two cascaded sections: $F_{\mathrm{p+i}}(s)$, a proportional-plus-integral (P + I) section, and $F_{\mathrm{hf}}(s)$, a high-frequency filtering section. This appendix examines several alternative configurations for $F_{\mathrm{p+i}}(s)$. It also shows how each of these two transfer functions should be arranged for calculations of Bode plots or Nichols charts.

3B.1 Proportional-Plus-Integral Section

Equation (2.27) defined the P + I section in the form

$$F_{\mathrm{p+i}}(s) = K_1 + \frac{K_2}{s} + \frac{K_3}{s^2} + \frac{K_4}{s^3} + \cdots \tag{3B.1}$$

Although nearly any P + I arrangement can be expressed in this manner by appropriate algebraic manipulations, the literal format of (3B.1) implies a configuration

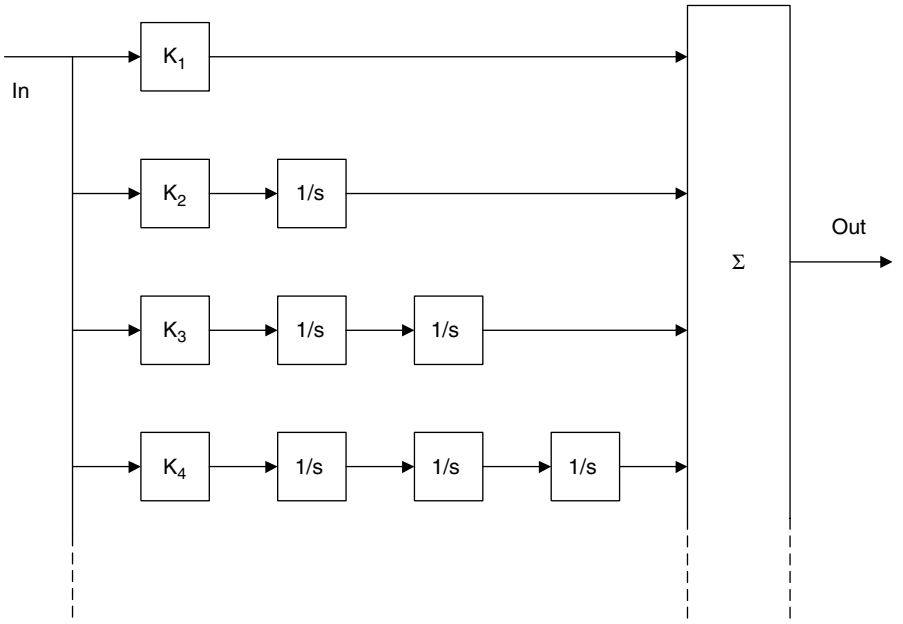


Figure 3B.1 Parallel configuration of the P + I portion of a loop filter for a high-type PLL.

with multiple parallel arms, as depicted in Fig. 3B.1. This fully parallel configuration is workable, but it clearly has more integrators than needed if the loop type is higher than 2. If a PLL is type n, it can be implemented with exactly n integrators. One integrator is always furnished by the VCO, so the loop filter need supply only $n-1$. But the fully parallel configuration of Fig. 3B.1 requires $\sum_{i=1}^{n}(i-1) = n(n-1)/2$ integrators, so a more economical configuration would be better.

Various configurations can be devised in which the $n-1$ integrators required are connected in cascade and the $n-1$ stabilizing zeros are produced by suitable weighted combinations of the proportional input and the $n-1$ outputs of the integrators. One such configuration (many others are possible) is illustrated in Fig. 3B.2. This configuration has a transfer function of

$$F_{\text{p+i}}(s) = K_1 + \frac{K_1 a_2}{s} + \frac{K_1 a_2 a_3}{s^2} + \frac{K_1 a_2 a_3 a_4}{s^3} + \cdots$$

Its coefficients can be related to those of (3B.1) by

$$\begin{aligned} K_2 &= K_1 a_2 \\ K_3 &= K_1 a_2 a_3 \\ K_4 &= K_1 a_2 a_3 a_4 \\ &\vdots \end{aligned}$$

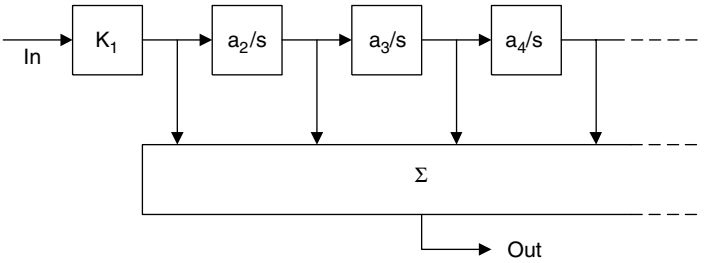


Figure 3B.2 Cascade integrators, parallel adder configuration for the P + I portion of a loop filter.

yet only $n-1$ integrators are used. Observe that each a_i has dimensions of $(\text{time})^{-1}$.

For a type 2 PLL, the transfer function of the P + I filter is simply

$$\frac{F_{\text{p+i}}}{K_1} = 1 + \frac{a_2}{s} = \frac{s + a_2}{s}$$

so $1/a_2$ is seen to be the same as τ_2, the time constant of the stabilizing zero of a type 2 PLL, as in Fig. 2.2.

For a type 3 PLL, the transfer function of the P + I filter is expressed as

$$\frac{F_{\text{p+i}}(s)}{K_1} = 1 + \frac{a_2}{s} + \frac{a_2 a_3}{s^2} = \frac{s^2 + s a_2 + a_2 a_3}{s^2}$$

Applying the quadratic formula, the two zeros of this expression are located at

$$s = -\frac{a_2}{2}\left(1 \pm \sqrt{1 - \frac{4a_3}{a_2}}\right)$$

The zeros are coincident (a desirable design goal) at $s = -a_2/2$ if $a_3 = a_2/4$ but will be conjugate complex (an undesirable condition) if a_3 is any larger. You could design for coincident zeros if the tolerance on a_3/a_2 were suitably tight, but the risk of complex zeros might be too high if the tolerances were very loose (as is common in practical analog PLLs).

Reasons for choosing coincident zeros and avoiding complex zeros are discussed in Section 8.3.1. Complex zeros are possible with the configurations of Figs. 3B.1 and 3B.2—or any of the variations on them—for all PLLs of type 3 or higher. It is the multiple parallel connections that have the potential for generating complex zeros, even if the integrators themselves are strictly in cascade. A configuration that avoids complex zeros because of its inherent structure would be safer, especially if component tolerances are loose.

Figures 3B.3 and 3B.4 show two configurations that entirely prevent complex zeros for any number of cascaded integrators. Prevention is accomplished by

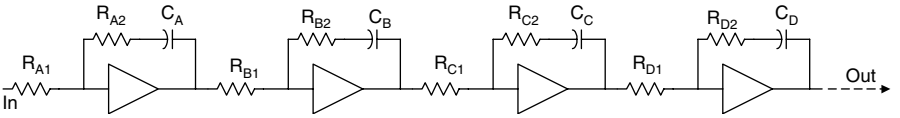


Figure 3B.3 Cascaded active lag-lead configuration for the P + I portion of a loop filter.

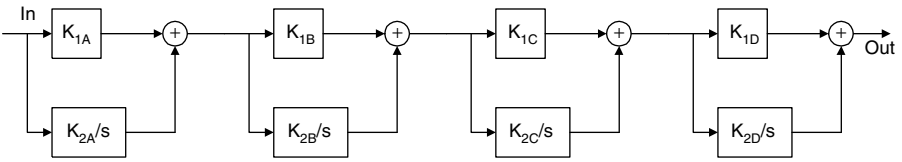


Figure 3B.4 Cascaded individual first-order P + I cells for the P + I portion of a loop filter.

associating each real zero with a single integrator rather than generating zeros by parallel connections from two or more integrators. These two configurations are simply cascades of the first-order P + I filters introduced in Fig. 2.2. [**Caution**: An odd number of inverting operational amplifiers in the cascade will introduce a net polarity inversion in the feedback loop. Be sure to take that into account in the overall design.]

Neglecting polarity inversions, the transfer function of the P + I loop filter of Fig. 3B.3 is

$$\begin{aligned} F_{\mathrm{p+i}}(s) &= \frac{(sC_AR_{A2}+1)(sC_BR_{B2}+1)\cdots}{s^{n-1}C_AR_{A1}C_BR_{B1}\cdots} \\ &= \frac{R_{A2}R_{B2}\cdots}{R_{A1}R_{B1}\cdots}\cdot\frac{(s+1/\tau_A)(s+1/\tau_B)\cdots}{s^{n-1}} \\ &= K_1\frac{(s+1/\tau_A)(s+1/\tau_B)\cdots}{s^{n-1}} \end{aligned} \tag{3B.2}$$

where $\tau_i = C_iR_{i2}$, $K_1 = (R_{A2}R_{B2}\cdots)/(R_{A1}R_{B1}\cdots)$, and n is the loop type. Similarly, the transfer function of the configuration of Fig. 3B.4 is the same as the last line of (3B.2) with the definitions $K_1 = K_{1A}K_{1B}\cdots$ and $\tau_i = K_{1i}/K_{2i}$. For purposes of calculation (e.g., in a spreadsheet), the P + I transfer function would be split into magnitude and phase according to

$$20\log\left|\frac{F_{\mathrm{p+i}}(j\omega)}{K_1}\right| = -20(n-1)\log(\omega) + 10\sum_{i=1}^{n-1}\log\left(\omega^2+\frac{1}{\tau_i^2}\right) \quad \text{dB} \tag{3B.3}$$

$$\mathrm{Arg}[F_{\mathrm{p+i}}(j\omega)] = \frac{180}{\pi}\left[-\left(\frac{\pi}{2}\right)^{n-1} + \sum_{i=1}^{n-1}\tan^{-1}\omega\tau_i\right] \qquad \text{deg} \tag{3B.4}$$

3B.2 High-Frequency Section

The transfer function of almost any high-frequency filter that might be found in a PLL can be written as products of first- and second-order zeros and poles and a delay factor, in the form

$$F_{\mathrm{hf}}(s) = F_{\mathrm{hf}}(0)\frac{\prod_m (s\tau_m + 1)\prod_q (\alpha_q s^2 + \beta_q s + 1)}{\prod_k (s\tau_k + 1)\prod_r (\alpha_r s^2 + \beta_r s + 1)} e^{-s\tau_d} \tag{3B.5}$$

where τ_d is a transport delay. The associated magnitude and phase representations are

$$\begin{aligned} 20\log\left|\frac{F_{\mathrm{hf}}(j\omega)}{F_{\mathrm{hf}}(0)}\right| &= 10\sum_m \log(1+\omega^2\tau_m^2) + 10\sum_q \log[(1-\alpha_q\omega^2)^2 + \omega^2\beta_q^2] \\ &\quad - 10\sum_k \log(1+\omega^2\tau_k^2) \\ &\quad - 10\sum_r \log[(1-\alpha_r\omega^2)^2 + \omega^2\beta_r^2] \qquad \mathrm{dB} \end{aligned} \tag{3B.6}$$

$$\begin{aligned} \mathrm{Arg}[F_{\mathrm{hf}}(j\omega)] &= \frac{180}{\pi}\left(-\omega\tau_d + \sum_m \tan^{-1}\omega\tau_m + \sum_q \tan^{-1}\frac{\omega\beta_q}{1-\omega^2\alpha_q}\right. \\ &\quad \left. - \sum_k \tan^{-1}\omega\tau_k - \sum_r \tan^{-1}\frac{\omega\beta_r}{1-\omega^2\alpha_r}\right) \qquad \mathrm{deg} \end{aligned} \tag{3B.7}$$

Complex-conjugate zero or pole pairs in F_{hf} are unusual, but they do occur; they require second-order factors if they are to be represented with real coefficients. Significant delay within an analog PLL is not usual either, but it too does occur; the provisions in (3B.5) and (3B.7) accommodate delay when it must be treated.

3B.3 Calculations

- A spreadsheet can be set up with more poles and zeros in F_{hf} than needed in any particular instance, simply by setting excess coefficients τ, α, and β to zero, thereby reducing any affected factor to unity.
- That same expedient is not available for $F_{\mathrm{p+i}}$ because any time constant set to zero in (3B.2) will cause division by zero.
- The rule for formatting $F_{\mathrm{p+i}}$ is $\lim_{s\to\infty} F_{\mathrm{p+i}}(s)/K_1 = 1$.
- The rule for formatting F_{hf} is $\lim_{s\to 0} F_{\mathrm{hf}}(s)/F_{\mathrm{hf}}(0) = 1$.
- Both K_1 and $F_{\mathrm{hf}}(0)$ are to be incorporated as factors of K.

APPENDIX 3C: CLOSED-LOOP FREQUENCY RESPONSES

Once a spreadsheet has been set up for Bode plots or Nichols charts, it is a simple matter to add a few formulas to generate frequency responses of the closed-loop transfer functions $E(j\omega)$ and $H(j\omega)$. Formulas are developed and examples given in this appendix.

3C.1 Frequency-Response Formulas

Start with the closed-loop error transfer function $E(j\omega)$, which is related to open-loop transfer function $G(j\omega)$ in polar components by

$$|E|e^{j\text{Arg}[E]} = \frac{1}{1+|G|e^{j\phi}} = \frac{1}{1+|G|\cos\phi + j|G|\sin\phi} \tag{3C.1}$$

where the $j\omega$ argument has been dropped for compactness and $\phi = \text{Arg}[G]$. The polar components of E are readily separated from (3C.1). First, the phase is

$$\text{Arg}[E] = -\tan^{-1}\frac{\sin\phi}{\cos\phi + 1/|G|} \qquad \text{rad} \tag{3C.2}$$

which typically would be plotted in degrees after multiplication by $180/\pi$. Then the magnitude is

$$|E|^2 = \frac{1}{1+2|G|\cos\phi + |G|^2}$$

and the magnitude in dB is

$$10\log|E|^2 = -10\log[1+2|G|\cos\phi + |G|^2] \qquad \text{dB} \tag{3C.3}$$

The closed-loop system transfer function written in polar components is

$$|H|e^{j\text{Arg}[H]} = \frac{|G|e^{j\phi}}{1+|G|e^{j\phi}}$$

which leads to the formulas

$$\text{Arg}[H] = \phi + \text{Arg}[E] \qquad \text{rad} \tag{3C.4}$$

$$|H|^2 = |G|^2|E|^2$$

$$10\log|H|^2 = 10\log|G|^2 + 10\log|E|^2 \qquad \text{dB} \tag{3C.5}$$

3C.2 Example Frequency-Response Graphs

Figures 3C.1 to 3C.4 show example plots of closed-loop frequency responses of E and H. Figures 3C.1 and 3C.2 are for a second-order type 2 PLL with

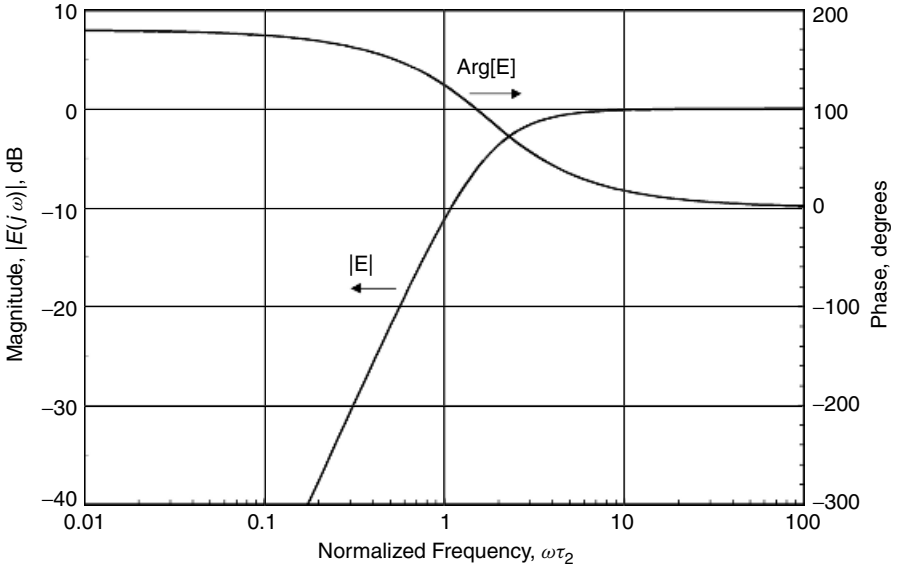


Figure 3C.1 Frequency response of $E(j\omega)$ for a type 2 PLL with $K\tau_2 = 3$; the same parameters as for the Bode plot of Fig. 3.13 and the Nichols chart of Fig. 3.18.

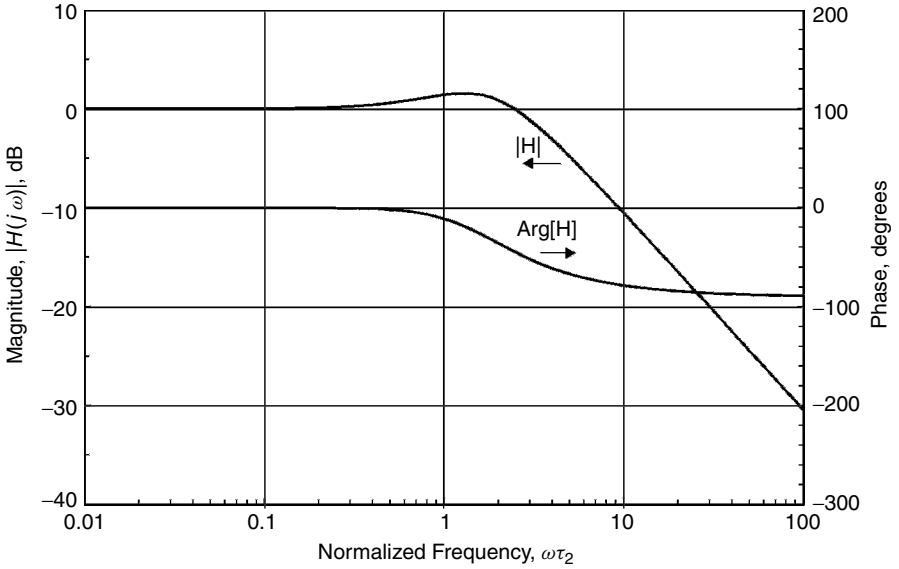


Figure 3C.2 Frequency response of $H(j\omega)$ for a type 2 PLL with $K\tau_2 = 3$; the same parameters as for the Bode plot of Fig. 3.13 and the Nichols chart of Fig. 3.18.

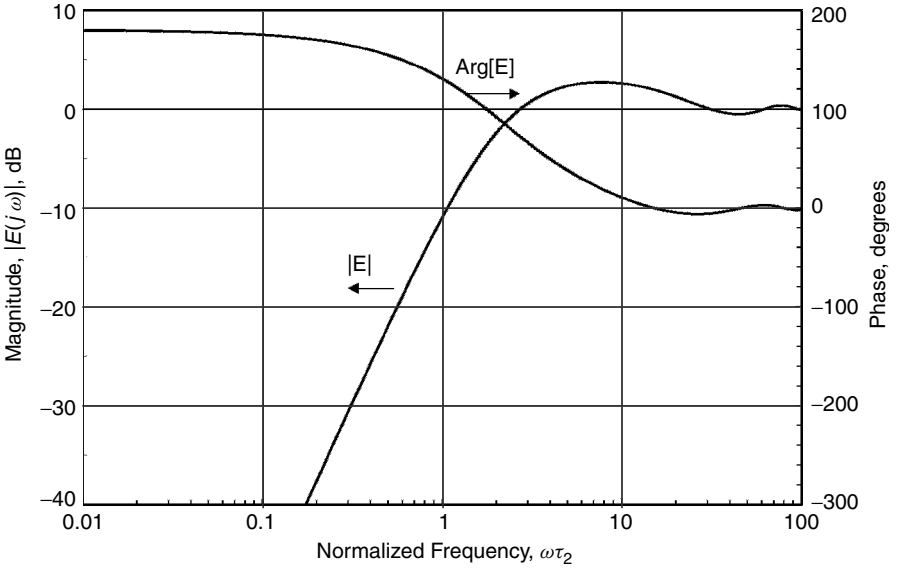


Figure 3C.3 Frequency response of $E(j\omega)$ for a type 2 PLL with $K\tau_2 = 3$ and an in-loop delay of $\tau_d = \tau_2/10$; the same parameters as for the Bode plot of Fig. 3.16 and the Nichols chart of Fig. 3.20.

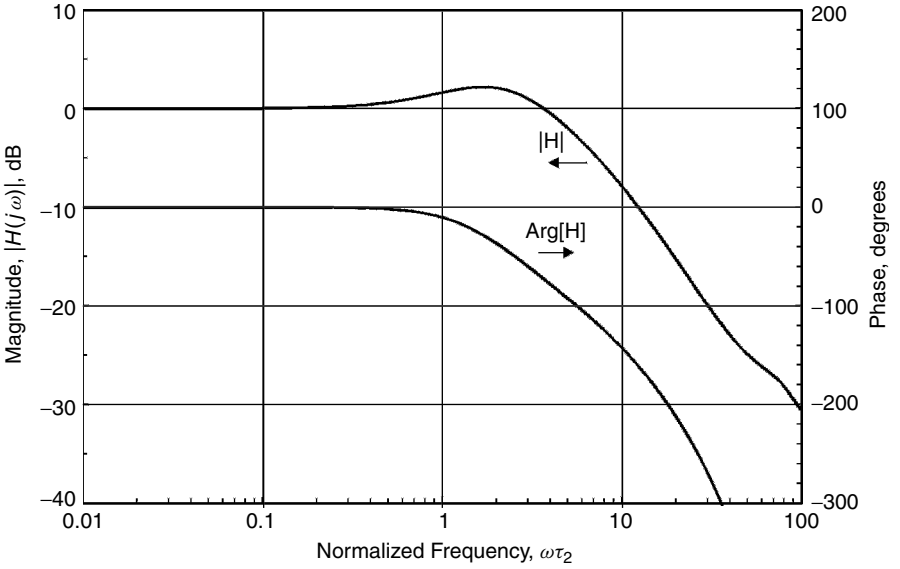


Figure 3C.4 Frequency response of $H(j\omega)$ for a type 2 PLL with $K\tau_2 = 3$ and an in-loop delay of $\tau_d = \tau_2/10$; the same parameters as for the Bode plot of Fig. 3.16 and the Nichols chart of Fig. 3.20.

$K\tau_2 = 3$ ($\zeta = 0.866$)—the same conditions as in the Bode plot of Fig. 3.13 and the Nichols chart of Fig. 3.18. Figures 3C.3 and 3C.4 show the effect of an in-loop delay of $\tau_d = \tau_2/10$ on the frequency responses. The effect is minimal at frequencies $\omega\tau_2 < 1$, but an increasing effect is observed at higher frequencies. Particularly affected is Arg[H], a not surprising result since $H(j\omega) \to G(j\omega)$ for high frequencies.

REFERENCES

3.1 W. R. Evans, *Control-System Dynamics*, McGraw-Hill, New York, 1954.

3.2 H. W. Bode, *Network Analysis and Feedback Amplifier Design*, Van Nostrand, New York, 1945.

3.3 H. Nyquist, "Regeneration Theory," *Bell Syst. Tech. J.* ***11***, 126, 1932.

3.4 H. M. James, N. B. Nichols, and R. S. Phillips, *Theory of Servomechanisms* (Rad. Lab. Ser. 25), McGraw-Hill, New York, 1947, Sec. 4–11.

3.5 R. C. Tausworthe, "Improvements in Deep-Space Tracking by Use of Third-Order Loops," *JPL Q. Tech. Rev.* ***1***, 96–106, July 1971.

3.6 R. C. Tausworthe and R. B. Crow, "Improvements in Deep-Space Tracking by Use of Third-Order Loops," *IEEE Int. Conf. Commun.*, 1972, pp. 577–583.

3.7 H. Meyr and G. Ascheid, *Synchronization in Digital Communications*, Wiley, New York, 1990, Sec. 2.5.

CHAPTER 4

DIGITAL PLLs: TRANSFER FUNCTIONS AND RELATED TOOLS

Like many other electronic devices, more and more phaselock loops are being implemented in digital versions. The usual reasons for digital implementation also apply to PLLs: lower cost, easier fabrication, drift-free components, and absence of tolerance problems. It is easy to store digital signals but almost prohibitive to store analog signals. A digital integrator has no offset or volatility problems. Ordinary digital operations are feasible with a complexity that is unimaginable through analog methods. All of these advantages notwithstanding, the need for compatibility with other digital operations in a system is by far the strongest motivation for digital implementation of PLLs.

4.1 DISTINCTIVE PROPERTIES OF DIGITAL PLLs

Digital PLLs (DPLLs) operate under several conditions that are generic to digital signal processing:

- Signals exist as sequences of discrete *samples*.
- The information in each sample is a *dimensionless digital number*.
- Digital numbers necessarily have finite precision; they are *quantized*.
- Operations within the digital PLL are *computed*.

A computed PLL is sometimes called a *software* PLL, but the underlying algorithms can be used equally well in either software or hardware, depending

Phaselock Techniques, Third Edition, by Floyd M. Gardner

on the speed required and the hardware available. The term *computed* is used throughout the book to refer to either hardware or software implementation.

All PLLs exhibit nonlinear behavior. In Chapters 2 and 3 it is assumed that operations can be approximated by a linear model if the phase error is small enough. This assumption is very good for many analog PLLs and yields the enormous benefit of analysis and design through transfer functions.

More so than with analog PLLs, numerous useful digital PLLs have gross nonlinearities that cannot be approximated away, even for small phase errors. These inherently nonlinear PLLs cannot be analyzed with transfer functions; examples are considered in Chapter 13. But even in the absence of gross nonlinearities, every digital PLL suffers from quantization effects; quantization is a nonlinear operation whose consequences are most significant at small phase errors. To avoid the severe complications of nonlinear analysis, common practice assumes that quantization is fine enough to be ignored to first order and that the DPLL can be analyzed by a linear approximation. Effects of quantization are treated as a separate problem in Chapter 13. The present chapter develops transfer functions of several digital PLLs by ignoring quantization and confining itself to digital PLLs that do not include other significant nonlinearities for small phase errors.

4.2 DIGITAL TRANSFER FUNCTION

Just as an analog circuit is described in the time domain by a differential equation, a digital circuit is described in the shift domain by a difference equation. (Shift can be related to discrete time, but need not be.) Just as a linear, time-invariant differential equation is converted to the transform domain by means of Laplace transforms, a linear, shift-invariant difference equation is converted to the transform domain by means of z-transforms. This section develops difference equations and z-transforms for a representative configuration of a digital PLL. Transfer functions are developed from the z-transforms of the digital PLL elements.

4.2.1 Configuration of a Digital PLL

The generic block diagram of a PLL in Fig. 1.1 also applies to a DPLL, with some minor changes. A digital phase detector and a digital loop filter are present, but a *number*-controlled oscillator (NCO) replaces the *voltage*-controlled oscillator (VCO). Also, a delay of D sample intervals (D is a positive integer) is a crucial element within the loop.

In Chapters 2 and 3, input and output signal phases were measured in radians and given the symbol θ. Although the same nomenclature could be employed for digital PLLs, the phases presented here will be measured in cycles [equivalently, unit intervals (UIs)] instead and given the symbol ε. Resulting transfer functions are exactly the same for either convention; the differing approach is taken solely to provide an example of how to apply the alternative nomenclature.

The input signal is a sequence of dimensionless digital numbers that includes a periodic component characterized, in part, by its phase $\varepsilon_i[n]$, where n is the sample index. The phase of the NCO output sequence is designated $\varepsilon_o[n]$. [**Notation**: Brackets $[\cdot]$ enclose discrete-index arguments and parentheses $(\cdot)$ enclose continuous arguments.]

4.2.2 Difference Equations

For small-enough phase errors, and neglecting quantization, the nth sample output of the phase detector is the dimensionless number

$$u_d[n] = \kappa_p\{\varepsilon_i[n] - \varepsilon_o[n]\} \tag{4.1}$$

where κ_p is the phase detector gain; it determines the PD output $u_d[n]$ in response to a phase error of $\varepsilon_e[n] = \{\varepsilon_i[n] - \varepsilon_o[n]\}$ cycles. Although κ_p is dimensionless, it should be associated with a notation—a pseudodimension—of $(\text{cycles})^{-1}$ to distinguish it from its sibling κ_d, which expresses the PD output in response to a phase error of θ_e radians. A pseudodimension of $(\text{radians})^{-1}$ should be associated with κ_d. [**Note**: All coefficients, such as κ_p, are always considered to be positive unless explicitly designated otherwise.]

Loop filters considered in this section comprise proportional elements, integral elements, and delay elements. A generic proportional element has a gain κ_m, input $x_{mi}[n]$, output $x_{mo}[n]$, and its difference equation is represented as

$$x_{mo}[n] = \kappa_m x_{mi}[n] \tag{4.2}$$

The subscript m connotes *multiplier*; modified notation will be substituted shortly. The proportional element is modeled as delay-free and memoryless; that is, the nth input and the scaling coefficient uniquely determine the nth output.

A digital integrator has a difference equation of

$$y_{Io}[n] = \kappa_I x_{Ii}[n-1] + y_{Io}[n-1] \tag{4.3}$$

where the subscript I connotes an integrator, κ_I is a scaling coefficient, $y_{Io}[n]$ is the nth sample output of the integrator, and $x_{Ii}[n]$ is the nth sample input. The integrator register for y_{Io} in a loop filter must saturate at its two extreme values, never recycle. Good design of the DPLL would avoid saturation under normal operating conditions.

An NCO is a special kind of integrator with a difference equation

$$\varepsilon_o[n] = \{\kappa_v u_c[n-1] + \varepsilon_o[n-1]\} \quad \text{mod-1} \qquad \text{cycles} \tag{4.4}$$

where κ_v is the NCO scaling coefficient, $\varepsilon_o[n]$ is the nth sample output, and $u_c[n]$ is the nth sample of the control input. The notation mod-1 means that $\varepsilon_o \in [0, 1)$; the NCO integrator discards any integer part of ε_o. That is, the register in the NCO

recycles, contrary to that of an integrator in the loop filter. This phase wrapping is a nonlinearity that typically is counteracted by some means not addressed at this juncture. For purposes of developing transfer functions, the nonlinearity is treated as nonexistent. (Think of the NCO as a circular up-down counter in which the apparent discontinuity is an artifact of trying to label position around a circle with numbers taken from a straight line. No genuine discontinuity occurs in the operation of the NCO.)

The product $\kappa_v u_c$ is a phase increment in fractional cycles, so the dimensionless NCO gain coefficient κ_v has a pseudodimension of cycles. Common NCOs have a gain coefficient $\kappa_v = 1$. If the NCO is clocked at a frequency f_s, the frequency of its output (the average rate of recycling of the phase register) is $\kappa_v u_c f_s$ Hz. Take note that negative frequencies can be physically meaningful if $\kappa_v u_c$ is negative. The output frequency of an NCO can pass through zero, unlike an ordinary analog VCO. The magnitude $|\kappa_v u_c|$ of the phase increment must be less than 0.5 to conform to the Nyquist sampling condition, to avoid frequency aliasing.

The difference equations for the integrator and NCO include delays of one sample interval; the $(n-1)$th input does not appear in the output until the nth sample of output. Also, an integrator has memory; the nth output is the sum of all preceding scaled inputs up to the $(n-1)$th.

The difference equation of a delay of D sample intervals is simply

$$x_{do}[n] = x_{di}[n - D] \tag{4.5}$$

The choice of the delay models in the preceding difference equations has been a compromise. On the one hand, delay-free operation (the nth input contributes to the nth output) may be feasible in a simulation or in post-time processing of stored signals. On the other hand, if the frequency of the system clock is comparable to the sampling rate of the PLL (typical of high-speed hardware systems), the loop will include pipelining delays, so that all elements, even the proportional scaling elements, might require more than one sample interval to fully accomplish their tasks. Delay D has been incorporated into the model to account for necessary integer delays, but be aware that not all possible configurations allow delay to be lumped in one location as in this model.

If the frequency of the clock is fast enough compared to the sampling rate of the PLL (typical of software systems), the multiple operations in the PLL might be performed within one sampling interval, but a real-time loop still includes some processing delay that is less than one sample interval. An integer delay D does not account for fractional delay. Only integer delays are considered in this book. A feedback loop must have a delay of at least $D = 1$. Without delay, the loop would not be computable; it would have to produce an output before it had a chance to generate the phase error (difference between input and output phases) that is needed for computing the output.

These difference equations have been written with the tacit assumption that all elements run at the same sampling rate. All following material in this chapter

is based on that single-rate assumption. Multirate operations are examined in Chapter 13.

4.2.3 *z*-Transforms of the Loop Elements

The z-transform of the phase-detector difference equation is simply

$$U_d(z) = \kappa_p\{\varepsilon_i(z) - \varepsilon_o(z)\} \tag{4.6}$$

and similarly, the z-transform of a delay-free proportional element is

$$X_{mo}(z) = \kappa_m X_{mi}(z) \tag{4.7}$$

where $\varepsilon(z)$ and $X(z)$ are the z-transforms of their respective sequences. The z-transform for an integrator with unit delay is

$$Y_{Io}(z) = \frac{\kappa_I z^{-1} X_{Ii}(z)}{1 - z^{-1}} \tag{4.8}$$

and ignoring the mod-1 nonlinearity, the NCO z-transform is

$$\varepsilon_o(z) = \frac{\kappa_v z^{-1} U_c(z)}{1 - z^{-1}} \tag{4.9}$$

Finally, the z-transform of an integer delay is

$$X_{do}(z) = z^{-D} X_{di}(z) \tag{4.10}$$

Notice that the z-transform representation of integer delay for a sampled feedback loop is algebraic and thus easier to manipulate than the transcendental Laplace transform representation of delay for a time-continuous feedback loop.

Figure 4.1 depicts a linear model of a digital PLL with z-transform transfer functions indicated for each element. This model is the basis for most of the material that follows in this chapter.

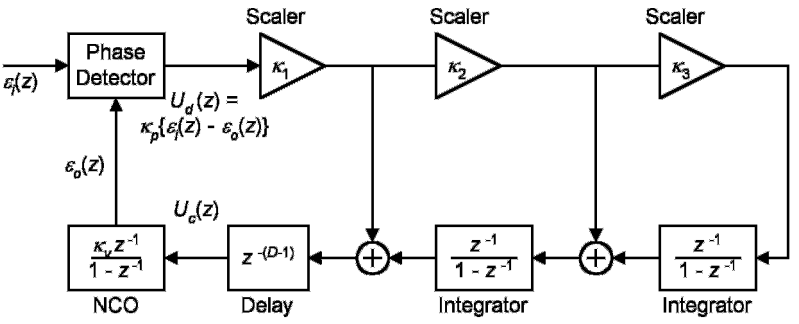


Figure 4.1 Block diagram of a type 3 DPLL.

4.2.4 Loop Filter

Denote the z-transform transfer function of the loop filter as $F(z)$. The loop filter consists of a combination of proportional elements, integrators, and delays. Additional high-frequency filter elements might also be included, but those are not considered until later. A transfer-function expression $F(z)$ is constructed by combining the z-transforms of its constituents. Only the one loop filter shown in Fig. 4.1 will be treated for now, a filter that produces a type 3 digital PLL. No implication is intended that a type 3 PLL is commonplace (it is not) or that digital PLLs are necessarily devoid of high-frequency filtering. A transfer function containing the example filter can be reduced to type 2 or type 1 by setting appropriate coefficients to zero.

Several variations on loop filter configurations were shown in Appendix 3B for analog PLLs. The same configurations could be used for digital PLLs. Similarly, the configuration of Fig. 4.1 could have been included in Appendix 3B as one more possible variation for analog PLLs. An astute reader will observe that the configuration of Fig. 4.1 employs the minimum number of integrators (because the integrators are in cascade) but has the potential for undesirable complex zeros in the transfer function (because of the parallel paths for combining signals). In actuality, since digital implementations need have no tolerance problems on coefficients, there is no risk of complex zeros; the zeros can be set exactly where desired, as demonstrated further below.

Notice that not only the integrators are cascaded but so are the coefficients κ_1, κ_2, and κ_3. This arrangement is a tidy way of drawing the filter block diagram, and it also has implementation benefits in that all coefficients almost always will be less than 1, κ_2 is almost always less than κ_1, and κ_3 is almost always less than κ_2. Thus, the attenuation needed for the input to the rightmost integrator is shared among κ_1, κ_2, and κ_3 and is not all placed in a single scaling element. Common practice selects values for the scaling coefficients κ_1, κ_2, and κ_3 as integer powers of 0.5 so that scaling can be performed as shifts instead of multiplications, thereby simplifying the computational effort.

For simplicity, excess delay $D-1$ has been included within the loop filter rather than separated out into the individual elements where it actually resides. Configurations of realistic digital PLLs may not permit element delays to be extracted to fit this model accurately; modified transfer functions must be derived for those configurations.

Combining all of the scalers, integrators, and excess delay of Fig. 4.1 gives a transfer function for the example loop filter of

$$F(z) = \frac{U_c(z)}{U_d(z)} = z^{-(D-1)}\kappa_1\left[1 + \frac{\kappa_2 z^{-1}}{1-z^{-1}}\left(1 + \frac{\kappa_3 z^{-1}}{1-z^{-1}}\right)\right]$$
$$= \frac{z^{-(D-1)}\kappa_1}{(1-z^{-1})^2}[(1-z^{-1})^2 + \kappa_2 z^{-1}(1-z^{-1}) + \kappa_2\kappa_3 z^{-2}] \qquad (4.11)$$

which has two finite poles at $z = 1$ (equivalent to $s = 0$ for time-continuous systems), plus $D-1$ poles at $z = 0$ (introduced by the excess delay) and two

finite zeros at

$$z = 1 - \frac{\kappa_2}{2} \pm \frac{\kappa_2}{2}\sqrt{1 - \frac{4\kappa_3}{\kappa_2}} \tag{4.12}$$

The two zeros are coincident at $z = 1 - \kappa_2/2$ if $\kappa_3 = \kappa_2/4$ (an easy condition to assure exactly, with no tolerance problems whatever in a digital implementation) and will be complex only if κ_3 exceeds $\kappa_2/4$.

4.2.5 Loop Transfer Functions

The loop transfer functions can now be written simply by combining $F(z)$ with the z-transforms for the phase detector and NCO and defining the dimensionless loop gain

$$\kappa = \kappa_p \kappa_v \kappa_1 \tag{4.13}$$

leading to the transfer functions listed below. Subscripts on G, H, and E indicate the loop type (the total number of integrators within the loop).

- Open-loop transfer function:

$$G_3(z) = \frac{\varepsilon_o(z)}{\varepsilon_e(z)} = \frac{\kappa z^{-D}[(1 - z^{-1})^2 + \kappa_2 z^{-1}(1 - z^{-1}) + z^{-2}\kappa_2\kappa_3]}{(1 - z^{-1})^3} \tag{4.14}$$

- System transfer function:

$$\begin{aligned} H_3(z) &= \frac{\varepsilon_o(z)}{\varepsilon_i(z)} = \frac{G_3(z)}{1 + G_3(z)} \\ &= \frac{\kappa z^{-D}[(1 - z^{-1})^2 + \kappa_2 z^{-1}(1 - z^{-1}) + \kappa_2\kappa_3 z^{-2}]}{(1 - z^{-1})^3 + \kappa z^{-D}[(1 - z^{-1})^2 + \kappa_2 z^{-1}(1 - z^{-1}) + \kappa_2\kappa_3 z^{-2}]} \end{aligned} \tag{4.15}$$

- Error transfer function:

$$\begin{aligned} E_3(z) &= \frac{\varepsilon_e(z)}{\varepsilon_i(z)} = \frac{1}{1 + G_3(z)} = 1 - H_3(z) \\ &= \frac{(1 - z^{-1})^3}{(1 - z^{-1})^3 + \kappa z^{-D}[(1 - z^{-1})^2 + \kappa_2 z^{-1}(1 - z^{-1}) + \kappa_2\kappa_3 z^{-2}]} \end{aligned} \tag{4.16}$$

4.2.6 Poles and Zeros

The example transfer functions of (4.14) to (4.16) are type 3, as evidenced by the $(1 - z^{-1})^3$ term in $G(z)$, which indicates three digital integrators in the loop. But the denominators of the closed-loop transfer functions are of higher degree than the type unless $D = 1$, its minimum possible value. The presence of excess delay

increases the order of a digital PLL by an amount $D - 1$; the additional open-loop poles are located at $z = 0$. This increased order has adverse implications for stability, as explored later.

Setting $\kappa_3 = 0$ and dividing through numerators and denominators of all transfer functions by a common factor $(1 - z^{-1})$ gives the transfer functions for a type 2 digital PLL of order $D + 1$. The system transfer function reduces to

$$H_2(z) = \frac{\kappa z^{-D}(1 - z^{-1} + \kappa_2 z^{-1})}{(1 - z^{-1})^2 + \kappa z^{-D}(1 - z^{-1} + \kappa_2 z^{-1})} \tag{4.17}$$

which has a zero at $z = 1 - \kappa_2$. If $D = 1$, $H_2(z)$ has a pair of poles at

$$z = 1 - \frac{\kappa}{2} \pm \frac{\kappa}{2}\sqrt{1 - \frac{4\kappa_2}{\kappa}} \tag{4.18}$$

The two poles are real and separate if the discriminant in (4.18) is positive, real, and coincident at $z = 1 - \kappa/2$ if the discriminant is zero, and complex if the discriminant is negative.

[**Comments**: (1) The expression (4.18) for the poles of $H_2(z)$ has the same formal structure as the expression (4.12) for the zeros of $F_3(z)$ and thus the zeros of $H_3(z)$. But the easy ability to set the zeros of H_3 accurately does not necessarily carry over to the poles of H_2 because κ in (4.18) includes a factor of κ_p, the phase-detector gain. In many phase detectors (not all), κ_p depends on the amplitude of the input signal or the input signal-to-noise ratio, properties that are rarely established with great accuracy. Influences on phase-detector gain are examined further in Chapter 10. (2) In Chapters 2 and 3, the transfer-function expressions for the type 2 PLL were normalized to the location of the zero at $s = -1/\tau_2$. That normalization is not necessary nor as useful for a digital PLL since z is already a dimensionless (i.e., normalized) quantity.]

Now set $\kappa_2 = 0$ in (4.17) and divide out the common factor $1 - z^{-1}$ from the numerator and denominator to obtain the closed-loop system transfer function of a type 1 digital PLL:

$$H_1(z) = \frac{\kappa z^{-D}}{1 - z^{-1} + \kappa z^{-D}} \tag{4.19}$$

This transfer function has no zero. If $D = 1$, it has a single pole at $z = 1 - \kappa$. If, additionally, $\kappa = 1$, the system transfer function reduces to $H_1(z) = z^{-1}$, a pure delay of 1 unit interval. Several authors over the years have pointed out that this choice of parameters gives a PLL that responds fully to its input with just a one-sample-time delay—as nearly instantaneous as possible in a sampled feedback system. The delayed output is exactly the same as the input, with no filtering distortion. The equivalent PLL in continuous time would have infinite bandwidth, a physical impossibility. Choosing $\kappa = 1$ and $D = 1$ defeats one important purpose of a PLL: filtering of its input. Any noise or other disturbance at the input appears without reduction at the output.

Now suppose that $D = 2$; the poles of a second-order type 1 digital PLL will be located at

$$z = \tfrac{1}{2} \pm \tfrac{1}{2}\sqrt{1 - 4\kappa} \tag{4.20}$$

The two poles are real and separate if $\kappa < 0.25$, real and coincident at $z = 0.5$ if $\kappa = 0.25$, and a complex-conjugate pair if $\kappa > 0.25$.

4.3 LOOP STABILITY

A digital PLL is stable if all of its poles (the roots of its characteristic polynomial) are inside the unit circle and unstable if any pole lies outside the unit circle. Stability conditions for several examples of DPLLs are summarized in this section; analytical details are relegated to Appendix 4A. Further examples related to the effects of delay may be found in [4.1].

Always be aware that pole locations and resulting stability boundaries depend on the particular arrangement of delays within the PLL. Different delay arrangements give different results. The examples shown in this section should be regarded as typical but do not necessarily apply to altered configurations except, perhaps, approximately.

4.3.1 Type 1 DPLLs

Consider the first-order type 1 ($D = 1$) DPLL of (4.19). Its one pole lies on the real axis, within the unit circle if $\kappa < 2$ and outside (therefore unstable) if $\kappa > 2$. Contrast this behavior with that of a time-continuous first-order PLL, which is stable for any positive value of loop gain, no matter how large. The essential delay within a digital PLL introduces unavoidable instability at a finite gain.

Next, increase the delay to $D = 2$, whereupon the two poles are located as in (4.20). To find the value of κ that places the conjugate poles on the unit circle, set $|z|^2 = 1$ and solve to get $\kappa = 1$ for the stability bound with the boundary poles at $z = (1 \pm j\sqrt{3})/2$. In Appendix 4A it is shown that the stability limit on loop gain for a type 1 DPLL with arbitrary integer delay $D > 0$ and without any other filtering in the loop is

$$\kappa = 2 \sin \frac{\pi}{2(2D - 1)} \tag{4.21}$$

Excess delay impairs the stability of a DPLL quite drastically.

4.3.2 Type 2 DPLLs

Next, consider the pole locations of $H_2(z)$ for $D = 1$, as shown in (4.18). Provided that $\kappa_2 < 1$, the instability boundary is defined by

$$\kappa = \frac{4}{2 - \kappa_2} \tag{4.22}$$

which reduces to $\kappa = 2$ for $\kappa_2 = 0$, as in a first-order DPLL. Equation (4.22) is obtained simply by substituting $z = -1$ into (4.18) and rearranging the algebra. The loop is unstable for all $\kappa > 0$ if $\kappa_2 > 1$. Reasons for these results are clarified in the subsequent discussion of root loci of digital PLLs in Section 4.4.

4.3.3 Type 3 DPLLs

For $D = 1$, a type 3 DPLL is unstable for all $\kappa > 0$ if

$$\kappa_2 \geq \frac{4}{4 - 3\kappa_3} \tag{4.23}$$

which reduces to $\kappa_2 > 1$ if $\kappa_3 = 0$ (thereby agreeing with the constraint on κ_2 for a type 2 DPLL with $D = 1$) and reduces to $\kappa_2 > 4/3$ if $\kappa_3 = \kappa_2/4$ (for the desirable coincident zeros of a type 3 PLL).

Within the constraints of (4.23), a type 3 DPLL with $D = 1$ is stable if

$$\frac{\kappa_3}{(1 - \kappa_3)(1 - \kappa_2 + \kappa_2\kappa_3)} < \kappa < \frac{8}{4 - 2\kappa_2 + \kappa_2\kappa_3} \tag{4.24}$$

and is unstable outside these boundaries. The stability condition on κ reduces to that of (4.22) for a type 2 DPLL if $\kappa_3 = 0$ and to

$$\frac{\kappa_2}{(4 - \kappa_2)(1 - \kappa_2/2)^2} < \kappa < \frac{8}{(2 - \kappa_2/2)^2} \tag{4.25}$$

if $\kappa_3 = \kappa_2/4$. The nonzero lower bounds in (4.24) and (4.25) demonstrate that a type 3 digital PLL is conditionally stable (i.e., unstable for small-enough gain) in the same manner as a type 3 analog PLL.

4.4 ROOT-LOCUS PLOTS

An observant reader will have noticed that stability boundaries were not pursued above for $D > 1$ except in type 1 DPLLs. As evidenced in Appendix 4A, the mathematical labor needed for analytical determination of stability becomes increasingly burdensome as the transfer function under consideration becomes more complicated. Root-locus patterns can be helpful in reducing the burden.

Underlying principles of root-locus plots for z-domain transfer functions are the same as those for s-domain transfer functions that were explored in Section 3.1 and Appendix 3A. Significant differences are:

- Open-loop integrator poles originate at $z = 1$ instead of $s = 0$.
- An additional $D - 1$ open-loop poles appear at $z = 0$.
- The stability boundary is the unit circle ($|z| = 1$) in the z-plane instead of the imaginary axis of the s-plane.

- Contours of constant damping are spirals in the z-plane rather than simple straight-line rays in the s-plane. In consequence, a z-plane root-locus plot does not reveal damping of complex poles in nearly so ready a fashion as offered by an s-plane root-locus plot.
- Since integer delays have an algebraic representation in a z-domain transfer function, as opposed to a transcendental representation in the s-domain, a z-domain root-locus plot readily accommodates integer delays.

4.4.1 Root Loci of Type 1 DPLLs

Type 1 DPLLs are frequently encountered in the literature and sometimes are feasible in practice if the frequency of the input signal is known with sufficient accuracy.

Type 1, D = 1 The root locus of a type 1 DPLL with $D = 1$ is a straight line originating at $z = 1$ (the location of the NCO integrator pole) and migrating toward the left along the real axis toward $z = -\infty$ as κ increases. For any κ, the single closed-loop pole lies at $z = 1 - \kappa$. This behavior closely resembles that of a first-order type 1 analog PLL. However, the DPLL locus crosses the unit circle at $z = -1$, so the DPLL becomes unstable for a gain $\kappa = 2$. By contrast, the corresponding analog PLL is stable for all $K > 0$.

Type 1, D = 2 A type 1 DPLL with $D = 2$ closely resembles the type 1 analog PLL with a simple lag filter described in Sections 2.3.2 and 3.1.3. The DPLL poles are at $z = 0.5(1 \pm \sqrt{1 - 4\kappa})$. The poles are real and separate for $\kappa < 0.25$ and conjugate complex for $\kappa > 0.25$. The complex portion of the locus lies on a vertical line at $\mathrm{Re}[z] = 0.5$. [**Notation**: $\mathrm{Re}[z]$ means the real part of z.] Intersection with the unit circle occurs at $z = 0.5(1 \pm j\sqrt{3}) = e^{\pm j\pi/3}$ for $\kappa = 1$.

Type 1, D = 3 Figure 4.2 illustrates the root locus for a type 1 DPLL with $D = 3$. One open-loop pole (due to the NCO integrator) lies at $z = 1$, and two open-loop poles (due to the excess delay) lie at $z = 0$. As κ increases, one pole originating at zero migrates to the left on the real axis and the other migrates to the right. As the integrator pole migrates leftward, it meets the right-moving delay pole at $z = \frac{2}{3}$ for $\kappa = \frac{4}{27}$, whereupon the pair become complex for larger gain values. The complex branch crosses the unit circle at $\mathrm{Arg}[z] = \pm\pi/5$ for $\kappa = 0.618$, thereby establishing the stability boundary. The left-moving real pole crosses $z = -1$ for $\kappa = 2$, which is larger than $\kappa = 0.618$ and thus is just an additional crossing of the unit circle by another pole, and is not the stability boundary.

4.4.2 Root Loci of Type 2 DPLLs

For the same reasons that apply to analog PLLs, most digital PLLs also will be type 2.

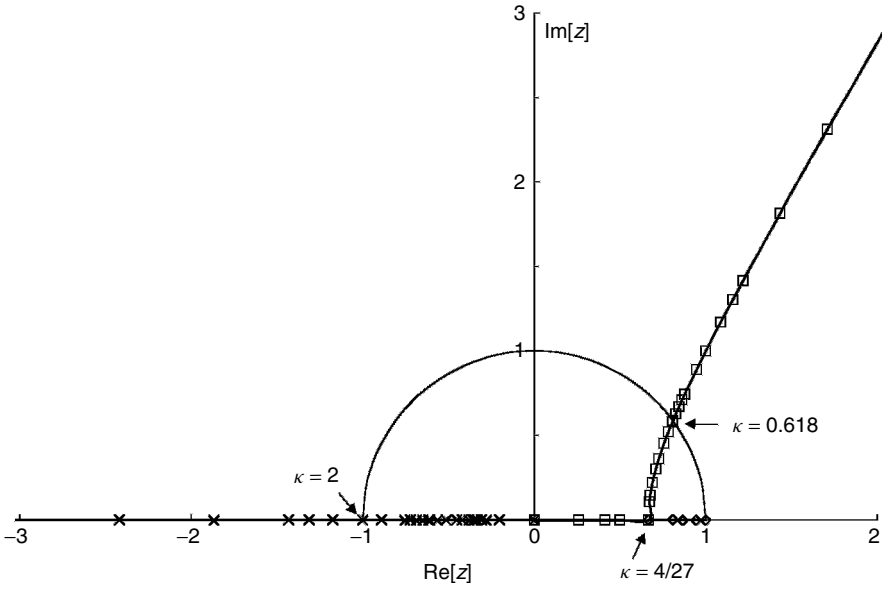


Figure 4.2 Root-locus plot of a type 1 DPLL with $D = 3$. The lower half plane is omitted.

Type 2, D = 1 A root locus for a type 2 DPLL with $D = 1$ has two open-loop poles at $z = 1$ and a zero at $z = 1 - \kappa_2$. As gain increases from zero, the two poles initially migrate along a circle with its center at the location of the zero and with a radius of κ_2. The two poles rejoin the real axis for $\kappa = 4\kappa_2$ at $z = 1 - 2\kappa_2$. From that location, one pole migrates to the right, eventually terminating on the zero (for $\kappa = \infty$), and the other migrates to the left toward $z = -\infty$. This behavior is closely similar to that for a second-order type 2 analog PLL, as described in Chapter 3.

Parameters κ and κ_2 both enter into the conditions of stability. Unless the circle portion of the locus lies inside the unit circle, the PLL will be unstable for all κ. A value of $\kappa_2 = 1$ causes the locus circle to coincide with the unit circle, so one stability criterion requires that $\kappa_2 < 1$. If that criterion is met, the left-migrating real pole crosses the unit circle at $z = -1$, for a gain value of

$$\kappa = \frac{4}{2 - \kappa_2} \tag{4.26}$$

Type 2, D = 2 Figure 4.3 shows a root-locus family for a type 2 DPLL with $D = 2$. Excess delay generates an extra open-loop pole at $z = 0$ but has no effect on the location of the zero. The pole originating at $z = 0$ migrates to the right on the real axis while the two integrator poles are complex for small gain. (The figure shows only the complex pole with positive imaginary part.) The loop is unstable (complex poles are outside the unit circle) for all $\kappa > 0$ if $\kappa_2 > \frac{1}{2}$.

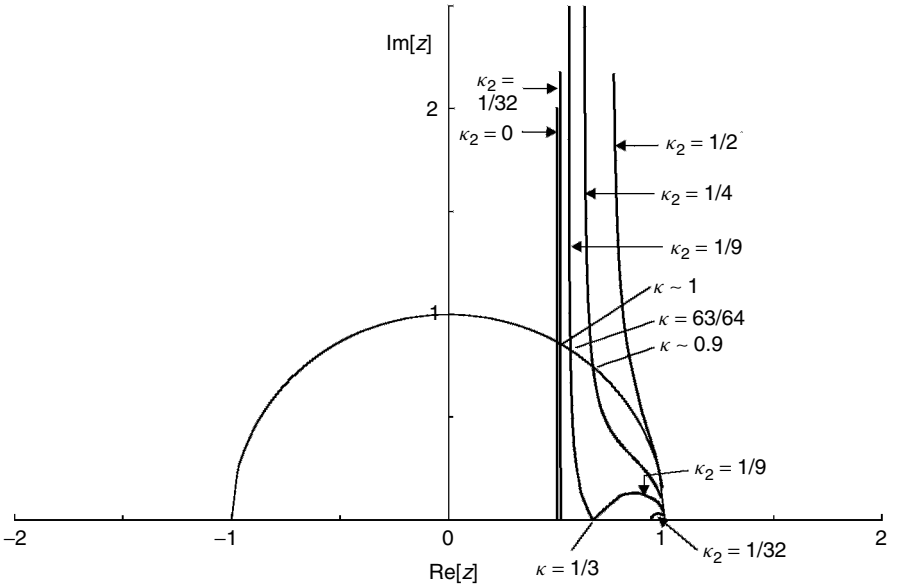


Figure 4.3 Root-locus plots of a type 2 DPLL with $D = 2$ and various κ_2. Real poles and the lower half plane are omitted.

If $\kappa_2 = \frac{1}{9}$, three poles are coincident at $z = \frac{2}{3}$ for $\kappa = \frac{1}{3}$, but two of them form a complex pair for all other values of κ. If $\kappa_2 > \frac{1}{9}$, the two poles originating at $z = 1$ remain complex for all $\kappa > 0$ and the right-migrating real pole terminates on the zero for $\kappa = \infty$. If $\kappa_2 < \frac{1}{9}$, the complex poles return to the real axis for some range of κ. One of these poles migrates rightward along the real axis toward the zero while the other migrates leftward, eventually meeting the third pole, whereupon the two become a complex-conjugate pair on vertical loci.

The root loci of a type 2 DPLL with $D = 2$ are similar to those of a third-order type 2 analog PLL as described in Section 3.1.4, except for the finite stability boundary of the DPLL. For another similarity, as κ_2 is reduced toward zero, the root locus for large κ approaches that of a type 1 DPLL with $D = 2$ with nearly the same stability boundary.

Type 2, D = 3 The open-loop transfer function has a zero at $z = 1 - \kappa_2$, a pair of poles due to the integrators at $z = 1$, and a pair of poles due to the excess delay at $z = 0$. The two poles originating at $z = 1$ initially migrate along complex loci, whereas the poles originating at $z = 0$ initially migrate left and right along the real axis (provided that $\kappa_2 < 1$).

Appendix 4A shows that the loop will be unstable for all $\kappa > 0$ if $\kappa_2 > \frac{1}{3}$. It is noteworthy that the stability bound on κ_2 for a type 2 DPLL has been found to be 1 for $D = 1$, $\frac{1}{2}$ for $D = 2$, and $\frac{1}{3}$ for $D = 3$. Following up on this suggestive result,

further analysis (not included here) revealed that the apparent rule continues for all integer $D > 0$; the stability bound on κ_2 for a type 2 DPLL is

$$\kappa_2 < \frac{1}{D} \tag{4.27}$$

The root loci for a type 2 DPLL with $D = 3$ will be similar to those shown in Fig. 3.6 for a fourth-order type 2 analog PLL: one pole locus terminates at $z = -\infty$, another terminates on the zero at $z = 1 - \kappa_2$, and the poles of a complex-conjugate pair asymptotically approach straight lines at angles of $\pm 60^\circ$ to the real axis. If $\kappa_2 = \frac{1}{2} - \sqrt{3}/4 \approx 0.067$, three poles will coincide at $z = \frac{1}{2} + \sqrt{3}/6 \approx 0.789$ for $\kappa = \sqrt{3}/9 \approx 0.1925$. If κ_2 is larger than $\sim$0.067, the two poles originating at $z = 1$ never return to the real axis. If κ_2 is smaller, those two poles return to the real axis for some value of κ; one then migrates rightward toward the zero while the other migrates leftward. The left-migrating pole eventually meets the right-migrating pole that originated at $z = 0$ and the two become complex.

The stability bound on κ is given by

$$\kappa = \frac{2(1 - \cos\psi)}{2(1 - \kappa_2)\cos\psi - 1} \tag{4.28}$$

with the further proviso that $\kappa_2 < \frac{1}{3}$. Stability condition (4.28) corresponds to complex poles migrating across $z = \exp(\pm j\psi)$. Another crossing, by the real pole at $z = -1$, occurs for larger values of κ, for which the loop is already unstable. See Section 4A.2 for further details.

4.4.3 Root Loci of Type 3 DPLLs

Figure 4.4 illustrates example root loci of a type 3 DPLL with $D = 1$ and $\kappa_3 = \kappa_2/4$. The latter condition causes the two zeros of the system transfer function to be coincident at $z = 1 - \kappa_2/2$. The plot for each value of κ_2 shows only the upper pole loci of complex pairs; all real loci and the lower complex loci have been omitted. The open-loop transfer function has three poles at $z = 1$. One of those poles migrates toward the left on the real axis, eventually terminating on a zero. The other two poles depart $z = 1$ at angles of $\pm 60^\circ$ to the positive real axis, thereby immediately passing out of the unit circle. Therefore, just like a type 3 analog PLL, a type 3 DPLL is conditionally stable, at best (i.e., unstable for low-enough gain).

Moreover, if $\kappa_2 \geq \frac{4}{3}$, the loop is unstable for all $\kappa > 0$, as demonstrated by the outermost locus in Fig. 4.4. The loci of the two complex poles return to the real axis at exactly $z = -1$ for $\kappa_2 = \frac{4}{3}$ and $\kappa = 4.5$; the complex poles are outside the unit circle for all other values of κ. If $\kappa_2 < \frac{4}{3}$, the type 3 DPLL is stable over the range of κ given in (4.25). The lower limit is determined by the intersection of the complex locus with the unit circle and the upper limit is reached where a real pole crosses $z = -1$.

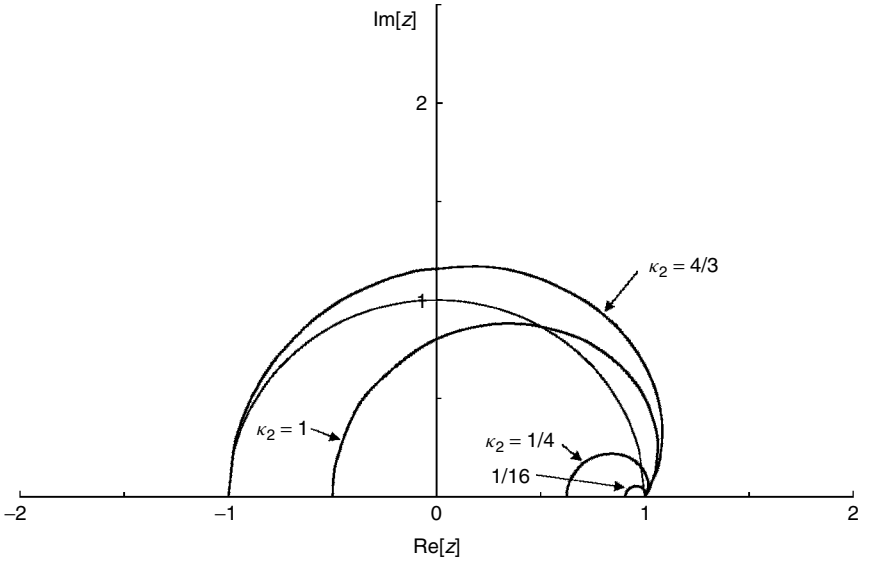


Figure 4.4 Root-locus plots of a type 3 DPLL with $D = 1$, various κ_2, and $\kappa_3 = \kappa_2/4$. Zeros are coincident by the choice of κ_3. Real poles and the lower half plane are omitted.

4.5 DPLL FREQUENCY RESPONSES: FORMULATION

For a transfer function $Y(s)$ of a time-continuous PLL, the frequency response is defined as $Y(s)|_{s=j\omega} = Y(j\omega)$. Equivalently, the frequency response for a transfer function of a sampled PLL is defined as $Y(z)|_{z=e^{j\omega t_s}} = Y(e^{j\omega t_s})$, where $Y = E$, F, G, or H as applicable and t_s is the sampling interval. In the time-continuous transfer function, ω has a range from $-\infty$ to $+\infty$ along the imaginary axis of the s-plane, whereas the product ωt_s in the transfer function of a discrete-time PLL has a range of $-\pi$ to $+\pi$ along the unit circle in the z-plane. In either representation, the frequency ω has dimensions of rad/sec and the sample interval t_s has dimensions of seconds; the product ωt_s is dimensionless.

It is customary in the digital signal processing literature to represent the dimensionless angle around the unit circle by the symbol ω and to suppress mention of t_s (equivalently, pretend that $t_s = 1$). Inasmuch as most PLLs operate in real time, however, and the sample interval often is an item of concern, this book uses the dimensionless product ωt_s to represent the angle. For compactness of notation, the symbol $\psi = \omega t_s$ is used to represent the angle when sampling interval is irrelevant.

Frequency responses are typically displayed graphically in polar components: magnitude and phase, or often just magnitude alone. Frequency response of a sampled system is periodic with normalized period 2π. A plot of normalized frequency from $-\pi$ to $+\pi$ contains the totality of frequency-response information, so the plots do not ordinarily extend beyond these bounds. Moreover, if the coefficients of the DPLL transfer functions are real (as is usual), the frequency

response in $(-\pi, \pi]$ is conjugate symmetric about $\psi = 0$, so it is sufficient to plot only for positive values of ψ. Finally, if frequency is plotted on a logarithmic scale (to be able to display response at very low frequencies), the minimum frequency in the plot has to be slightly larger than zero.

4.6 BODE PLOTS AND NICHOLS CHARTS

Preceding sections of this chapter and Appendix 4A show how mathematical analysis or root-locus methods can be excessively tedious if a transfer function has more than a very few poles. By contrast, Bode plots and Nichols charts of DPLLs are easily generated with the help of spreadsheets, even for complicated transfer functions. This section is devoted mainly to Bode plots of significant DPLLs, but keep in mind that Bode plots and Nichols charts are interchangeable. They employ the same data and obey the same stability rules. The choice of one or the other is a matter of individual taste in the appearance of the displays. Although the following text concentrates on Bode plots, the same discussions apply equally to Nichols charts.

4.6.1 Basis of Bode Plots

A Bode plot for a digital PLL is a graph of the polar components—magnitude and phase—of the open-loop transfer function $G(z)$ for z on the unit circle: that is, for $z = e^{j\psi}$, where $\psi = \omega t_s$ is the radian frequency normalized to the sampling rate $1/t_s$. Magnitude is plotted in decibels and phase is plotted on a linear scale, typically in degrees. Except for (1) taking frequency along the unit circle instead of the imaginary axis, and (2) truncation of the abscissa at $\psi = \pi$, the Bode plot of a digital PLL seems very much like that of an analog PLL as laid out in Chapter 3. Indeed, that likeness is borne out by the examples that follow; only minor differences will appear. A review of the Bode and Nichols materials in Chapter 3 is a good foundation for Bode plots and Nichols charts of digital PLLs as well.

But first it is necessary to recognize divergent approaches in control system textbooks. Some authors simply apply Bode plots directly to the z-plane transfer functions as if there were no question as to the validity of the application. However, other authors point out—correctly—that Bode analysis originated with continuous-time systems whose transfer functions are described in the s-plane. Furthermore, mapping a z-plane transfer function into the s-plane via the relation $z = \exp(st_s)$ yields something very different from the usual s-plane transfer function for which Bode relations were originally derived. From these valid observations, some authors seem to imply that direct computation of Bode plots of discrete-time systems in the z-plane is improper. Instead, they perform yet a different mapping that transforms the z-plane transfer function into something more closely resembling the transfer functions of a continuous-time system for which the original Bode rules should apply.

A nonexpert (including this author) is placed in a dilemma: Who is to be followed? Is it wrong to directly apply Bode methods in the z-domain? Or have unnecessary complications been introduced by those who employ an intermediate transformation? To ease my own misgivings, I explored several example DPLLs by means of the direct approach. These examples were analyzed by other methods: in Sections 4.3 and 4.4 by algebraic analysis and by root-locus plots in Appendix 4A. For each example, interpretation of the z-domain Bode plot agreed exactly with the alternative analyses. Concordance among examples does not constitute a general proof, but it does prove that the z-domain Bode criteria are correct in at least some conditions of practical importance and raises confidence that the Bode criteria are valid for most DPLLs. The rest of this chapter applies the Bode method directly and does not introduce an intermediate transformation.

4.6.2 Bode Stability Criteria

Bode stability criteria for sampled systems do not differ in any important way from those for time-continuous systems as described in Chapter 3. Central to the Bode analysis are the gain-crossover frequency ψ_{gc}, where $|G(e^{j\psi_{gc}})| = 1$ (0 dB) and the phase-crossover frequency ψ_π, where $\text{Arg}[G(e^{j\psi_\pi})] = -\pi$. It is possible that phase crossovers could occur at more than one frequency: for example, in a conditionally stable DPLL. It is also possible that no phase-crossover frequency exists at all, either because phase is more positive than -180° for all frequencies or phase is more negative than -180° for all frequencies. Whatever the number of phase crossovers, strict Bode analysis is confined to transfer functions in which there is one and only one gain crossover.

Phase margin is defined as $\text{Arg}[G(e^{j\psi_{gc}})] + \pi$; stability requires a positive phase margin. A stability boundary has the property that the unique gain-crossover frequency coincides with a phase-crossover frequency. A loop can have more than one stability boundary, each corresponding to a specific phase crossover and each with its own critical value of κ. Although gain margin is often a useful concept, it does not necessarily have a clear definition in all instances. Examples of vagaries or breakdowns of the concept of gain margin are pointed out in the sequel.

4.6.3 Bode Plots of Example DPLLs

This section provides Bode plots of several DPLLs of major practical interest.

Type 1 DPLL Figure 4.5 shows a Bode plot of the very simplest DPLL: type 1 with delay $D = 1$ and no other filter elements within the loop. Phase vs. frequency has two constituents: a constant -90° caused by the integrator inherent in the NCO and a linear $-90\psi/\pi$ degrees caused by the delay. (The logarithmic scale for frequency distorts the linear phase to induce curvature in the display.) Phase crossover of -180° is reached at $\omega t_s = \psi_\pi = \pi$.

The magnitude curve is drawn for $\kappa = 1$; other values of κ are accommodated by shifting the gain curve up or down by the requisite number of decibels.

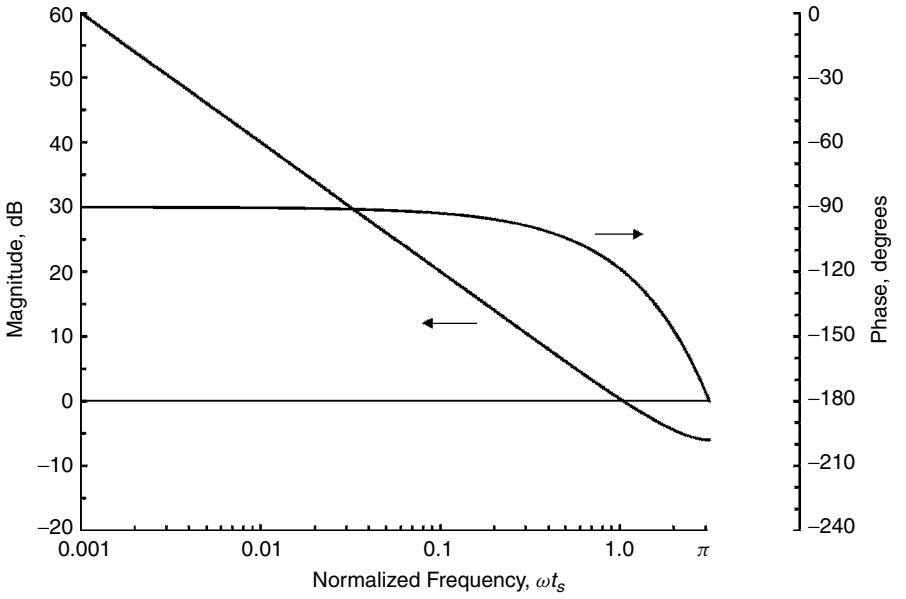


Figure 4.5 Bode plot of a type 1 DPLL with $D = 1$, $\kappa = 1$.

Gain is -6 dB at the phase-crossover frequency, so the gain margin is 6 dB. The loop would be at its stability boundary if the gain were increased to $\kappa = 2$ and would be unstable for larger values of κ. The slope of the curve is approximately -6 dB/octave at low frequencies but gradually changes to flat at $\psi = \pi$. The flattening is a feature of the periodicity of the transfer function and is encountered in most DPLLs. A choice of $\kappa = 1$ is very large; typical choices are appreciably smaller. In consequence, gain flattening often will be invisible, off the bottom of the chart. Therefore, the Bode plot of a type 1 DPLL with $D = 1$ usually will be nearly the same as that of an analog type 1 PLL with comparable delay in the loop.

Observe that if $\kappa > 2$, the Bode plot for this DPLL has no gain crossover; magnitude would exceed 0 dB for all ψ. Strictly speaking, the Bode criterion fails in the absence of a gain crossover. Rigorous confirmation of instability in this particular case comes from analysis of the characteristic equation as in Section 4.3.1 or Appendix 4A or from constructing root loci as in Section 4.4.1.

[**Comment**: A couple of graphic features are worthy of note. (1) The two ordinates have been arranged so that 0 dB on the magnitude axis aligns with -180° on the phase axis; a single horizontal line then defines the key levels of the gain and the phase crossovers. (2) A gap is evident between the end of the frequency axis and the bottom of the phase axis. The gap comes about because the plotting routine in the spreadsheet program will only display full decades of a logarithmic scale, whereas the abscissa of interest in a sampled system ends at $\psi = \pi$. An invisible rectangle overlay hides the unwanted portion of the rightmost decade of the abscissa.]

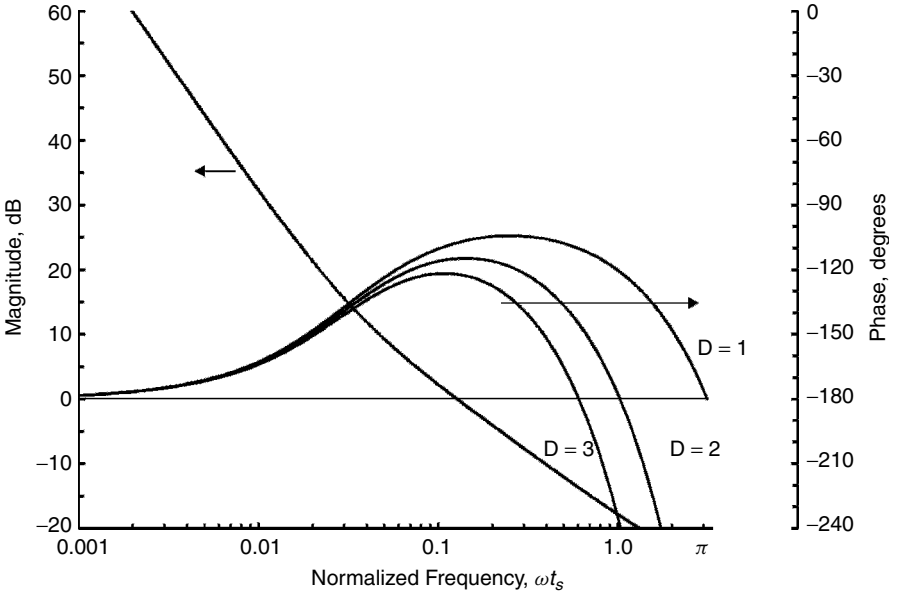


Figure 4.6 Bode plot of a type 2 DPLL with delay D. $\kappa = \frac{1}{8}$, $\kappa_2 = \frac{1}{32}$.

Type 2 DPLL Figure 4.6 shows Bode plots for a type 2 DPLL for several values of delay D. Loop gain is $\kappa = \frac{1}{8}$ (rather large) and the gain of the integrator path is $\kappa_2 = \frac{1}{32}$ (a choice that produces coincident poles if $D = 1$). Slope of the magnitude curve is close to -12 dB/octave at low frequencies and approximately -6 dB/octave at high frequencies. Slope flattening is present at high frequencies but is not apparent to the eye. A corner frequency, where the low-frequency slope gives way to the high-frequency slope, can be seen at $\psi \approx \kappa_2 = 0.03125$. (The approximation is good within 12% for $\kappa_2 = 0.2$—a very large value—and is progressively better as κ_2 decreases.)

Taking numbers from the curves: Gain crossover occurs at $\psi_{gc} \approx 0.12$ and the phase margins are approximately 73°, 62°, and 58° for $D = 1$, 2, and 3, respectively. Gain margin, as taken from the curves, is approximately 18 dB for $D = 2$ and 13 dB for $D = 3$. The gain curve is off the bottom of the chart at the phase-crossover frequency for $D = 1$; the corresponding gain margin was found to be 24 dB from the spreadsheet that generated Fig. 4.6.

4.6.4 Nichols Chart Example

The open-loop transfer function of a type 3 DPLL, evaluated at $z = e^{j\psi}$, can be written as

$$G_3(e^{j\psi}) = \frac{\kappa e^{-j(D-1/2)\psi}[(2 - \kappa_2 + \kappa_2\kappa_3)\cos\psi - 2 + \kappa_2 + j\kappa_2(1 - \kappa_3)\sin\psi]}{8j^3\sin^3(\psi/2)} \tag{4.29}$$

whereupon the polar components are

$$20\log|G_3(e^{j\psi})| = 20\log\frac{\kappa}{8} - 60\log\left(\sin\frac{\psi}{2}\right) + 10\log\{[(2-\kappa_2+\kappa_2\kappa_3)\cos\psi - 2 + \kappa_2]^2 + [\kappa_2(1-\kappa_3)\sin\psi]^2\} \quad \text{dB} \tag{4.30}$$

$$\text{Arg}[G_3(e^{j\psi})] = \frac{180}{\pi}\left[-\frac{3\pi}{2} - \left(D - \frac{1}{2}\right)\psi + \tan^{-1}\frac{\kappa_2(1-\kappa_3)\sin\psi}{(2-\kappa_2+\kappa_2\kappa_3)\cos\psi - 2 + \kappa_2}\right] \quad \text{deg} \tag{4.31}$$

These equations are plotted in a Nichols chart in Fig. 4.7 for three different values of D. Gain coefficients for the chart are $\kappa = \frac{1}{8}$, $\kappa_2 = \frac{1}{32}$, and $\kappa_3 = \kappa_2/4 = \frac{1}{128}$. The choice for κ_3 causes the two zeros to be coincident at $z = 1 - \kappa_2/2$. The choices for κ_2 and κ are the same as those for the preceding type 2 DPLL, thereby permitting some interesting comparisons.

Gain and phase crossings, along with the associated margins, are readily seen in a Nichols chart. Margin values can be extracted even more easily than from a Bode plot (compare Fig. 4.6), especially from a Nichols chart drawn in the format of Fig. 4.7. Despite the multitude of crossings, stability is evident immediately from a glance at Fig. 4.7. As a penalty, a Nichols display loses the frequency information contained in a Bode plot.

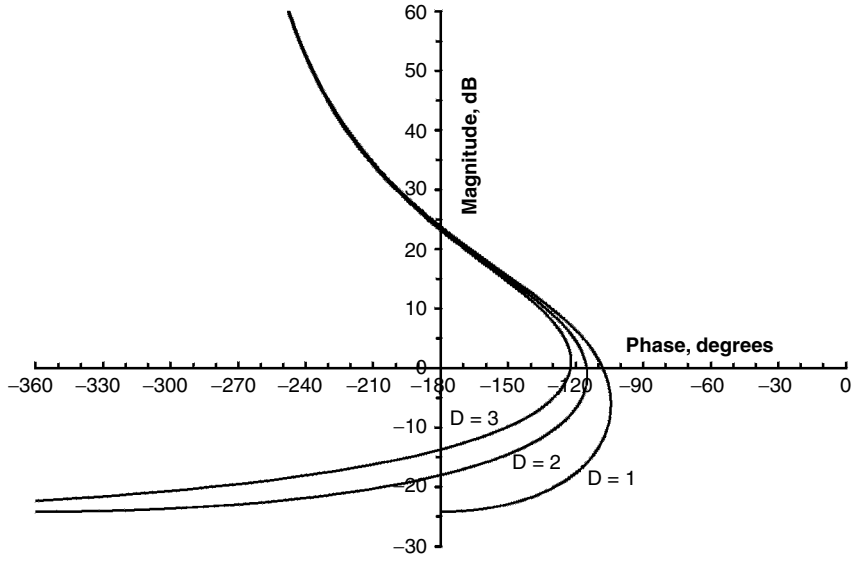


Figure 4.7 Nichols chart for a type 3 DPLL. $\kappa = \frac{1}{8}$, $\kappa_2 = \frac{1}{32}$, $\kappa_3 = \frac{1}{128}$.

Phase margins are approximately 71°, 64°, and 58° for $D = 1$, 2, and 3, respectively. Each curve has two phase crossings (this example, like all type 3 PLLs, is conditionally stable): one at a large gain and comparatively low frequency and another at a small gain and high frequency. The high-frequency phase crossings of the example have gain margins of approximately 24, 18, and 13 dB for $D = 1$, 2, and 3, respectively. These gain and phase margins are negligibly different from those found for the type 2 DPLL of the preceding example, which had the same values of κ and κ_2. Close similarity is to be expected since the high-frequency phase crossings and the gain crossings are all band-edge features in a PLL, features that are little influenced by such a low-frequency parameter as κ_3 or even κ_2.

Each curve shows a low-frequency phase crossing with a gain margin of about 23 dB, nearly the same for all three curves. Since these curves differ only in their values of D, which has its strongest influence at high frequencies, it is to be expected that the low-frequency phase crossovers should nearly coincide. The gain margin at high frequencies is the amount of gain *increase* that will cause the loop to become unstable. But on the contrary, observe that the low-frequency gain margin is the amount of gain *reduction* that will cause the loop to become unstable. A definition of gain margin would have to include both cases to be generally applicable; the most common definition includes only the high-frequency case.

4.7 TIME-CONTINUOUS APPROXIMATION FOR A DPLL

Bode plots of example DPLLs have been found to be similar to those of analog PLLs, at least for small-enough gain coefficients. Engineering folklore holds that behavior of a time-discrete system is close to that of a time-continuous system if the bandwidth (however "bandwidth" may be defined) of the time-discrete system is small compared to the sampling rate. Intuition would suggest that the properties of a DPLL should be close to those of an analog PLL if the poles and zeros of a stable digital PLL are close to $z = 1$. This section provides a first-order quantification of these vague qualitative concepts. Meanings of "small-enough" and "close" have to be defined in the context of each particular project.

Consider the time-discrete system transfer function $H_2(z)$ of (4.17). The complex variable z used in z-transforms is defined in terms of the complex variable s of Laplace transforms as $z = \exp(st_s)$. The following first-order approximations are valid if $|Dst_s| \ll 1$ is applicable:

$$z^{-1} \approx 1, \qquad z^{-D} \approx 1, \qquad 1 - z^{-1} \approx st_s \tag{4.32}$$

Applying these approximations to (4.17) yields

$$H_2(e^{st_s}) \approx \frac{s\kappa/t_s + \kappa\kappa_2/t_s^2}{s^2 + \kappa s/t_s + \kappa\kappa_2/t_s^2} \tag{4.33}$$

Comparison of (4.33) to (2.16) yields

$$\omega_n \leftrightarrow \frac{1}{t_s}\sqrt{\kappa\kappa_2}, \qquad \zeta \leftrightarrow \frac{1}{2}\sqrt{\frac{\kappa}{\kappa_2}} \tag{4.34}$$

and comparison to (2.19) yields

$$K \leftrightarrow \frac{\kappa}{t_s}, \qquad \tau_2 \leftrightarrow \frac{t_s}{\kappa_2} \tag{4.35}$$

where the arrows $\leftrightarrow$ indicate equivalence when the approximations (4.32) are valid.

Thus, the choice of $\kappa/\kappa_2 = 4$ as in Fig. 4.6 is equivalent to a damping $\zeta = 1$ (the two poles are coincident) in an analog PLL. In fact, (4.18) reveals that the two poles will be coincident in a type 2 DPLL with $D = 1$ if $\kappa/\kappa_2 = 4$, irrespective of κ. Such coincidence at $\kappa/\kappa_2 = 4$ is not to be expected with larger D.

4.8 FREQUENCY-RESPONSE EXAMPLES

Appendix 3C for analog PLLs showed how easy it is to extend a Bode–Nichols spreadsheet to plot closed-loop frequency responses as well. Those techniques also apply, without modification, to digital PLLs. This section depicts several examples of magnitude responses of type 2 DPLLs with assorted properties.

4.8.1 Effect of Delay

Figure 4.8 depicts the magnitude response $|H_2(e^{j\psi})|$ of a type 2 DPLL with $\kappa = 4\kappa_2$ for several values of delay D. From (4.18), a type 2 DPLL with $D = 1$ and these gain coefficients has coincident poles, as does an second-order type

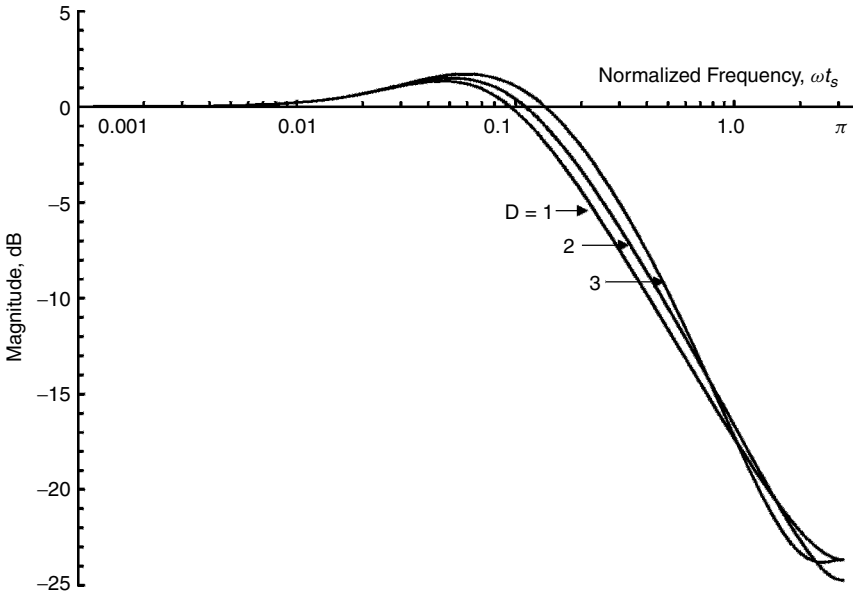


Figure 4.8 Magnitude response $|H_2[\exp(j\omega t_s)]|$ of a type 2 DPLL. $\kappa = \frac{1}{8}$, $\kappa_2 = \frac{1}{32}$.

2 analog PLL with $\zeta = 1$. Delay affects gain peaking, but not greatly in this example. Gain peaks are 1.34, 1.50, and 1.72 dB for $D = 1$, 2, and 3, respectively. Compare these values to gain peaking of 1.25 dB from (2.25) for a second-order type 2 analog PLL with $\zeta = 1$. Delay also affects high-frequency rolloff, but not by much if the delay is moderate. Since a gain $\kappa = \frac{1}{8}$ is usually regarded as fairly large, these results bear out the approximations of Section 4.7 rather well.

4.8.2 Effect of Bandwidth

The frequency scale for a DPLL runs from 0 to π/t_s; in consequence, bandwidth is wide or narrow relative to π/t_s. Alteration of relative bandwidth of a DPLL causes changes in the shapes of frequency responses, unlike the simple dilation that accompanies bandwidth changes in an analog PLL (provided that the bandwidth of the analog PLL is small compared to the PD comparison frequency f_c). These shape changes mean that relative bandwidth is a design parameter of a DPLL, a parameter that typically is ignored in the usual concept of a narrowband analog PLL (but see Chapter 12 for a counterexample).

Figure 4.9 illustrates the effect of relative bandwidth on the magnitude of the system closed-loop frequency response $|H(e^{j\omega t_s})|$. The DPLL is type 2 and has delay $D = 1$. Loop gain κ is taken as a measure of relative bandwidth. In the figure, κ is assigned the value $4\kappa_2$ and the curves are labeled by κ_2. The curve with the smallest κ_2 is the same as one of the curves in Fig. 4.8. This curve is not drastically different from that of an analog PLL with the same nominal damping per (4.34), at least up to around $\omega t_s \approx 1$.

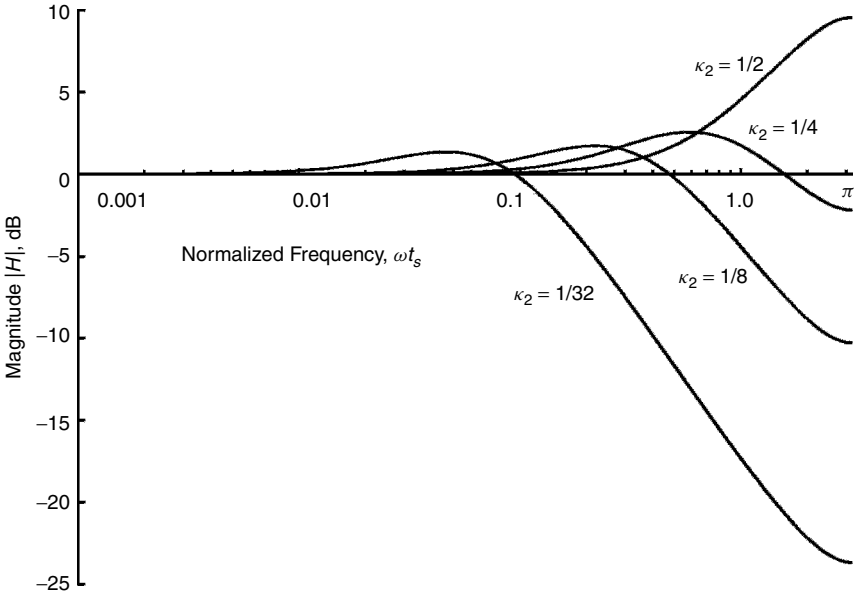


Figure 4.9 Magnitude response $|H_2[\exp(j\omega t_s)]|$ of a type 2 DPLL. $\kappa = 4\kappa_2$, $D = 1$.

For larger values of κ_2, and therefore κ as well, the gain peaking increases and the attenuation at the highest frequencies decreases. The magnitude response loses its lowpass character completely (i.e., $|H| \geq 1$ for all frequencies) for $\kappa_2 > 1 - \sqrt{0.5} \approx 0.292$. All of the example DPLLs in Fig. 4.9 are stable, even though stability margins are meager for the largest relative bandwidths.

4.9 LOWPASS FILTERS IN THE LOOP

Chapters 2 and 3 indicated that lowpass filters might be included within the feedback loop of an analog PLL, either deliberately to filter unwanted high-frequency signals or noise, or unavoidably because of the limited frequency responses of loop components. No lowpass filters have yet been introduced in this chapter on digital PLLs. The only explicit high-frequency effects considered for DPLLs are those due to unavoidable excess delay, and excess delay does not constitute a lowpass filter. Situations arise in which high-frequency response needs to be rolled off more steeply than is possible with the DPLLs examined so far. Steeper rolloff is attained by incorporating lowpass filters into the feedback loop. This section describes two simple lowpass filters for DPLLs.

4.9.1 Infinite Impulse Response Lowpass Filter

The simplest lowpass filter for use within analog PLLs has a single pole. Multipole lowpass filters within a loop often are cascades of single-pole filters. A filter with poles has infinite impulse response (IIR). Equivalently, the simplest IIR lowpass filter for use within a digital PLL also has a single pole. Its difference equation is

$$y[n] = ax[n-1] + (1-a)y[n-1] \tag{4.36}$$

where x and y are input and output, respectively, and a delay of one sample interval has been included to allow for high-speed clocking. Need for multiplication is avoided by choosing the gain parameter a as an integer power of 0.5. The z-transform of the difference equation yields the filter transfer function

$$F_{\text{iir}}(z) = \frac{az^{-1}}{1-(1-a)z^{-1}} = \frac{a}{z-1+a} \tag{4.37}$$

Salient Properties of a One-Pole IIR Lowpass Filter

- Pole location: $z = 1 - a$
- Stability bound: $a = 2$
- DC gain: $F_{\text{iir}}(+1) = 1$
- Maximum attenuation: $F_{\text{iir}}(-1) = \dfrac{a}{a-2}$

- Frequency response:

$$F_{\text{iir}}(e^{j\psi}) = \frac{a}{e^{j\psi} - 1 + a} = \frac{a}{\cos\psi - 1 + a + j\sin\psi}$$

$$\text{Arg}[F_{\text{iir}}(e^{j\psi})] = -\tan^{-1}\frac{\sin\psi}{\cos\psi - 1 + a} \tag{4.38}$$

$$20\log|F_{\text{iir}}(e^{j\psi})| = 20\log(a) - 10\log[(\cos\psi - 1 + a)^2 + \sin^2\psi]$$

- 3-dB frequency:

$$\psi_{3\text{ dB}} = \cos^{-1}\left[1 - \frac{a^2}{2(1-a)}\right] \tag{4.39}$$

If $a \ll 1$, then $\psi_{3\text{ dB}} \approx a/\sqrt{2}$. Magnitude response is monotonic in ψ. Response decreases with ψ if $a < 1$ but increases if $a > 1$. In fact, the lowpass character of the filter is lost entirely if $a > 1$.

Example It is evident from the maximum-attenuation formula above that a small value of a must be used to achieve substantial attenuation. In Chapters 2 and 3 it was shown that the corner frequency of a lowpass pole had to exceed the gain-crossover frequency to achieve acceptable phase margin and acceptable damping. Similar constraints apply in a DPLL; the 3-dB frequency $\psi_{3\text{ dB}}$ of the lowpass filter should exceed κ by a substantial margin.

Figure 4.10 shows Bode plots of a type 2 DPLL with loop parameters of $D = 1$, $\kappa = \frac{1}{8}$, and $\kappa_2 = \frac{1}{32}$. Curves are drawn for the filter in the loop and for

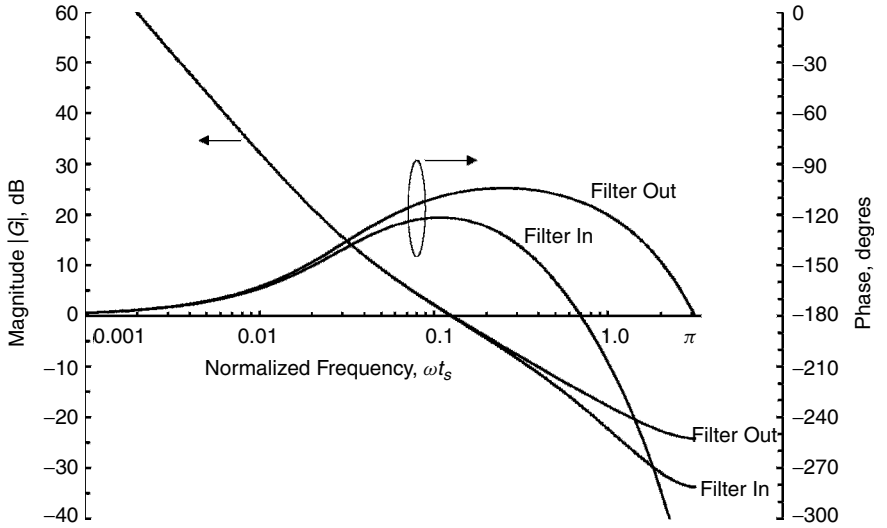


Figure 4.10 Bode plot of a type 2 DPLL with an IIR lowpass filter in the loop. The filter pole is at $z = 1 - a$. $D = 1$, $\kappa = \frac{1}{8}$, $\kappa_2 = \frac{1}{32}$, $a = \frac{1}{2}$.

the filter out. The filter parameter in the example is $a = 4\kappa = \frac{1}{2}$, which provides 9.5 dB of attenuation at $\psi = \pi$: not very much. Phase margin in the example is $\sim 60^\circ$, so a somewhat smaller a could be used, but not much smaller without unduly damaging the phase margin. To achieve appreciably greater attenuation at high frequencies while preserving phase margin requires not only a smaller value of a but smaller values of κ and κ_2 as well.

4.9.2 Finite Impulse Response Lowpass Filter

All lowpass filters that are practical for use within an analog PLL have infinite impulse response. Such filters are also applicable to digital PLLs, but DPLLs can also employ lowpass filters with finite impulse response (FIR). This option is one minor way in which digital implementation is more versatile than analog. The simplest FIR lowpass filter has a two-term difference equation

$$y[n] = 0.5(x[n] + x[n-1]) \tag{4.40}$$

and a transfer function

$$F_{\text{fir}}(z) = \frac{Y(z)}{X(z)} = \frac{1}{2}(1 + z^{-1}) \tag{4.41}$$

[**Comment**: The factor of 0.5 is inserted to make the DC gain equal to 1. In practice, the filter would be implemented without the factor of 0.5 and the resulting DC gain of 2 would be incorporated into the value for κ.]

Salient Properties of a Two-Tap FIR Lowpass Filter

- Zero location: $z = -1$
- Stability bound: unconditionally stable
- DC gain: $F_{\text{fir}}(+1) = 1$
- Maximum attenuation: $F_{\text{fir}}(-1) = 0$ ($-\infty$ dB)
- Frequency response:

$$\begin{aligned} F(e^{j\psi}) &= e^{-j\psi/2} \cos\frac{\psi}{2} \\ \text{Arg}[F(e^{j\psi})] &= -\frac{\psi}{2} \\ |F(e^{j\psi})| &= \cos\frac{\psi}{2} \end{aligned} \tag{4.42}$$

- 3-dB frequency: $\psi_{3\text{ dB}} = \pi/2$

Example Figure 4.11 shows the effect of the FIR lowpass filter on Bode plots of a type 2 DPLL. All parameters of the DPLL are the same as that of Fig. 4.10, except for the FIR filter and delay. Because the difference equation (4.40) is

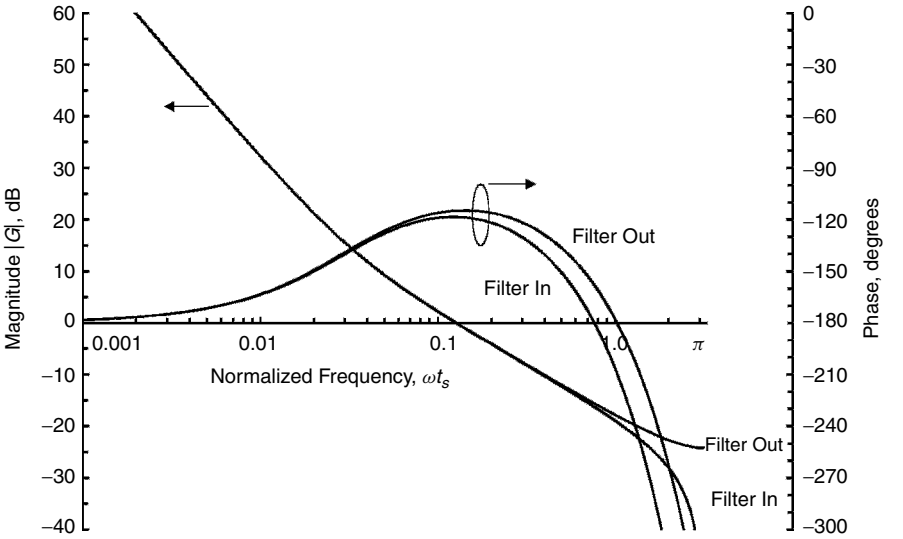


Figure 4.11 Bode plot of a type 2 DPLL with a two-tap FIR lowpass filter in the loop. The filter zero is at $z = -1$. $D = 2$, $\kappa = \frac{1}{8}$, $\kappa_2 = \frac{1}{32}$.

delay-free, and since delay-free operation would not be feasible if the signal-sampling frequency were equal to the fastest clock frequency, an extra delay has been inserted to make the model more nearly realistic. Therefore, the example uses $D = 2$.

Filtering action clearly is confined to the highest frequencies. That is a useful property if a disturbance is also confined to the highest frequencies but not if filtering is needed at lower frequencies as well. This simplest FIR filter cannot be adjusted to provide attenuation at lower frequencies. Other FIR filters can be designed to have almost any desired attenuation characteristic; the literature on digital signal processing has abundant examples. Be careful, though, with large FIR filters; they introduce extra delay into the loop, thereby impairing stability.

APPENDIX 4A: STABILITY OF DIGITAL PHASELOCK LOOPS

Since a DPLL is stable if all poles lie inside the unit circle, a stability boundary is defined by the intersection of a root locus with the unit circle at $z = \exp(j\psi)$. If ψ can be determined, the gain κ associated with the crossing of the circle can be calculated from the characteristic equation. Determination of ψ involves straightforward algebra and trigonometry for simple DPLLs but the labor grows more burdensome as the complexity increases. Calculations are worked out for several examples in this appendix.

Be aware that every locus crossing of the unit circle does not necessarily identify a stability boundary. A crossing might be from outside to inside, with other poles remaining outside. Or other poles of the DPLL might cross out of the unit circle for lower values of gain, so that the loop is already unstable at another crossing. Each crossing in a multiple-pole system has to be checked to verify whether it is a true stability boundary.

4A.1 Type 1 DPLL

The characteristic equation of a type 1 DPLL of (4.19) with delay D is

$$1 - z^{-1} + \kappa z^{-D} = 0 \tag{4A.1}$$

Multiply through by z^D and substitute $z = \exp(j\psi)$ to obtain the characteristic equation for a crossing of the unit circle as

$$e^{jD\psi} - e^{j(D-1)\psi} + \kappa = 0 \tag{4A.2}$$

whose real and imaginary parts are

$$\begin{aligned} \cos D\psi - \cos[(D-1)\psi] + \kappa &= 0 \\ \sin D\psi - \sin[(D-1)\psi] &= 0 \end{aligned} \tag{4A.3}$$

Using standard trigonometric identities, the imaginary part can be rewritten as

$$\sin D\psi - \sin[(D-1)\psi] = 2\sin\frac{\psi}{2}\cos\frac{\psi(2D-1)}{2} = 0 \tag{4A.4}$$

Either $\sin(\psi/2) = 0$ (a trivial case) or $\cos[\psi(2D-1)/2] = 0$, which is the useful solution. The cosine vanishes for angles of $(2k-1)\pi/2$, $k =$ integer. The angle sought is

$$\psi = \frac{\pi(2k-1)}{2D-1} \tag{4A.5}$$

Substituting this result into the real part of (4A.3) and applying trigonometric identities yields values of κ at the unit circle as

$$\kappa = 2\sin\frac{\psi}{2} \tag{4A.6}$$

This type 1 DPLL has D poles and no zeros in its open-loop transfer function. Each pole locus crosses the unit circle for some value of κ; the smallest such κ has to be determined. Confine attention to the upper half circle (because complex poles occur in conjugate pairs) and recognize that $\sin(\psi/2)$ is monotonically increasing over that range of ψ so that the smallest value of ψ obtained from (4A.5) is the one corresponding to the smallest gain at a stability boundary.

The smallest angle on the upper half circle corresponds to $k = 1$, so the stability boundary of this type 1 DPLL is

$$\kappa = 2 \sin \frac{\pi}{2(2D - 1)} \tag{4A.7}$$

displayed previously as (4.21).

4A.2 Type 2 DPLL

The denominator of (4.17) is the characteristic polynomial of a type 2 DPLL. Substituting $z = \exp(j\psi)$ and solving for κ yields

$$\kappa = \frac{-(1 - e^{-j\psi})^2}{e^{-jD\psi}(1 - e^{-j\psi} + \kappa_2 e^{-j\psi})} \tag{4A.8}$$

After substituting $\exp(-j\psi) = \cos\psi - j \sin\psi$ and applying several trigonometric identities, the expression for κ becomes

$$\kappa = \frac{4 \sin^2 \psi/2}{\text{Re[denom]} + j\text{Im[denom]}} \tag{4A.9}$$

where the real and imaginary parts of the denominator are given by

$$\text{Re[denom]} = \kappa_2 \cos\left[\left(D - \frac{1}{2}\right)\psi\right] \cos\frac{\psi}{2} + (2 - \kappa_2) \sin\left[\left(D - \frac{1}{2}\right)\psi\right] \sin\frac{\psi}{2} \tag{4A.10}$$

$$\text{Im[denom]} = (2 - \kappa_2) \cos\left[\left(D - \frac{1}{2}\right)\psi\right] \sin\frac{\psi}{2} - \kappa_2 \sin\left[\left(D - \frac{1}{2}\right)\psi\right] \cos\frac{\psi}{2} \tag{4A.11}$$

The gain κ has to be real and positive. The numerator of (4A.9) is real and positive, so (4A.11), the imaginary part of the denominator, has to be zero at $z = \exp(j\psi)$. The condition for vanishing of the imaginary part is found to be

$$(1 - \kappa_2) \sin D\psi - \sin(D - 1)\psi = 0 \tag{4A.12}$$

Provided that the imaginary part of the denominator is zero, the real part reduces to

$$\text{Re[denom]} = \cos(D - 1)\psi - (1 - \kappa_2) \cos D\psi \tag{4A.13}$$

Although expressions (4A.9), (4A.12), and (4A.13) appear simple enough, they were sufficiently complicated to block attempts at a general solution for the stability boundaries for arbitrary D. Instead, solutions for specific $D = 1$, 2, and 3 are outlined below.

Type 2, D = 1 For $D = 1$, (4A.12) yields $\sin\psi = 0$, so $\psi = 0$ or π. The zero solution identifies the open-loop poles and occurs for $\kappa = 0$, so $\psi = \pi$ is the critical angle that is sought. For $\psi = \pi$, (4A.13) becomes $2 - \kappa_2$, so the stability boundary is

$$\kappa = \frac{4}{2 - \kappa_2} \tag{4A.14}$$

the same as (4.22). An additional constraint is imposed by the condition $\kappa_2 < 1$. If $\kappa_2 > 1$, the circle centered on $1 - \kappa_2$ described by the complex portion of the root locus is everywhere outside the unit circle, and the crossing found from (4A.12) is directed outside to inside.

Type 2, D = 2 If $D = 2$, (4A.12) becomes $2(1 - \kappa_2)\sin\psi\cos\psi - \sin\psi = 0$, so that either $\sin\psi = 0$ or $2(1 - \kappa_2)\cos\psi - 1 = 0$. From the root-locus plots in Fig. 4.3, it is evident that $\sin\psi = 0$ is not a feasible solution, leaving only $2(1 - \kappa_2)\cos\psi - 1 = 0$, from which the cosine of the angle of the crossing is determined to be

$$\cos\psi = \frac{1}{2(1 - \kappa_2)} \tag{4A.15}$$

Inserting this result into (4A.13) and simplifying yields the stability boundary as

$$\kappa = \frac{1 - 2\kappa_2}{(1 - \kappa_2)^2} \tag{4A.16}$$

Any $\kappa_2 > 0.5$ gives an impermissible negative value for κ, so $\kappa_2 = 0.5$ is also a stability boundary irrespective of κ. For small-enough κ_2, the stability boundary approaches $\kappa = 1$ and $\psi = \pm 60°$.

Type 2, D = 3 For $D = 3$, (4A.12) becomes

$$\begin{aligned}&(1 - \kappa_2)\sin 3\psi - \sin 2\psi \\ &\quad = (1 - \kappa_2)\sin\psi(3 - 4\sin^2\psi) - 2\sin\psi\cos\psi = 0\end{aligned} \tag{4A.17}$$

from which one concludes that either $\sin\psi = 0$ (i.e., $\psi = \pi$) or

$$\begin{aligned}&(1 - \kappa_2)(3 - 4\sin^2\psi) - 2\cos\psi \\ &\quad = 4(1 - \kappa_2)\cos^2\psi - 2\cos\psi - (1 - \kappa_2) = 0\end{aligned} \tag{4A.18}$$

Both alternatives have to be pursued since both identify crossings of the unit circle. If $\psi = \pi$ is the correct crossing, (4A.13) reduces to $2 - \kappa_2$ and the gain at the circle crossing is identified as

$$\kappa = \frac{4}{2 - \kappa_2} \tag{4A.19}$$

which is formally the same as (4A.14) for $D = 1$.

Before accepting (4A.19), the constraints imposed by (4A.18) have to be investigated. To that end, (4A.18) was solved for κ_2, which was then plotted vs. ψ as shown in Fig. 4A.1. Regions of negative κ_2 are inadmissible and can immediately be excluded from further consideration. If (4A.18) is true, (4A.9) becomes

$$\kappa = \frac{2(1 - \cos\psi)}{2(1 - \kappa_2)\cos\psi - 1} \tag{4A.20}$$

That expression is negative if $\cos\psi$ is negative and $\kappa_2 < 1$, a condition that occurs for the region of ψ between 0.5π and 0.6π. Similarly, (4A.20) is negative in the region $\psi = \pi/3$ to $\pi/2$. Angles in these regions are inadmissible.

In the region from $\psi = 0$ to $\pi/5$ [i.e., $\cos\psi = 1$ to $(1 + \sqrt{5})/4$], the resulting κ_2 varies from $\frac{1}{3}$ to 0. From (4A.20), the stability boundary is $\kappa = 0$ if $\psi = 0$; this angle of crossing is attained for $\kappa_2 = \frac{1}{3}$. That is, the two poles originating at $z = 1$ will be outside the unit circle for all $\kappa > 0$ if $\kappa_2 > \frac{1}{3}$. Thus, $\kappa_2 < \frac{1}{3}$ is a stability constraint for a type 2 DPLL with $D = 3$. Admissible values of κ_2 are $\frac{1}{3} > \kappa_2 > 0$. This constraint applies in (4A.19) as well as (4A.20).

At the other end of this range, where $\psi = \pi/5$ and $\kappa_2 = 0$, (4A.20) yields $\kappa = 0.618$, the same boundary that was found for a type 1 DPLL with $D = 3$, which is to be expected when $\kappa_2 = 0$. Since the result from (4A.20) is smaller than the gain yielded by (4A.19), one concludes that (4A.20) discloses the true stability boundary of a type 2 DPLL with $D = 3$, whereas (4A.19) simply discloses a crossing of the unit circle at a gain that already exceeds the stability limit.

The region from $\psi = 2\pi/3$ to π remains to be considered. It is a region of instability since κ_2 is too large, but what does it reveal about the poles? The key

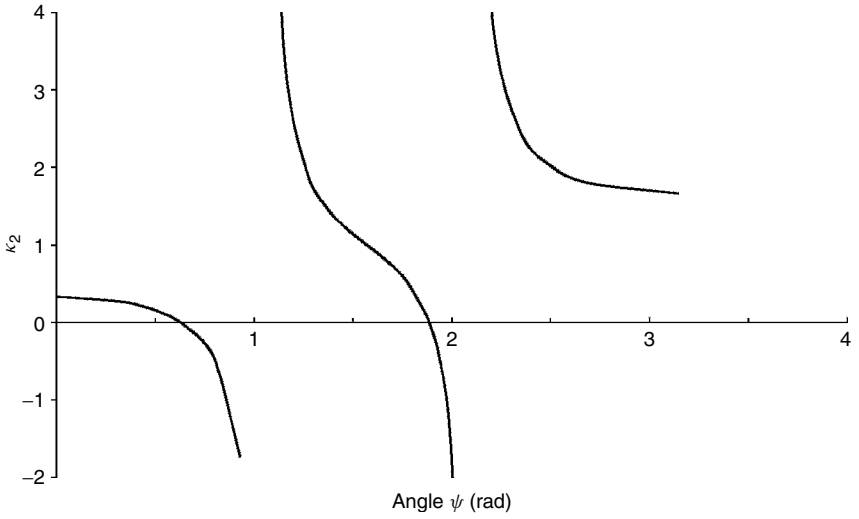


Figure 4A.1 Integral path gain κ_2 vs. angle ψ at the intersection of a root locus with a unit circle. Type 2 DPLL with $D = 3$.

feature is that $\kappa_2 > 1$ in this region, so that the zero of the transfer function at $1 - \kappa_2$ is on the negative real axis. Therefore, the two poles due to excess delay cannot depart $z = 0$ along the real axis; their loci must be complex initially, circling around the location of the zero, and eventually returning to the real axis for large enough κ. Then one pole migrates toward $z = -\infty$ and the other toward the zero. Equation (4A.20) discloses the κ value at which the complex portion of these loci crosses out of the unit circle, and (4A.19) discloses the larger κ value at which the right-moving real pole reenters the unit circle. But since the excessive κ_2 makes the loop unstable in this region for all $\kappa > 0$, this behavior is only of academic interest.

REFERENCE

4.1 J. W. M. Bergmans, "Effect of Loop Delay on Stability of Discrete-Time PLL," *IEEE Trans. Circuits & Syst. I*, ***42***, 229–231, Apr. 1995.

CHAPTER 5

TRACKING

A locked PLL is said to *track* its input signal. Tracking is studied through the phase errors θ_e that result from various input phases θ_i. A small phase error is usually desired and is considered to be the criterion of good tracking performance. If the error should become so large that the VCO slips cycles, tracking is considered to have failed (the loop has lost lock), even if only momentarily.

This chapter deals first with phase error small enough that a linear approximation is valid. Linearity allows the powerful tools of transfer-function analysis to be applied for the determination of PLL responses to inputs of engineering importance. One pivotal analysis explains why a type 2 PLL is found so overwhelmingly in practice. Next, nonlinear behavior is explored, especially the limits on phaselocking. That is, what input conditions will cause the PLL to slip cycles or otherwise lose lock? Cycle slipping is a crucial topic reappearing in later chapters.

5.1 LINEAR TRACKING

Transfer functions are helpful for defining phase errors in the steady state, in response to transients, and in response to sinusoidal angle modulation of the input. The phase error transfer function $E(s)$ for a time-continuous PLL is given by (2.7) as

$$E(s) = \frac{\theta_e(s)}{\theta_i(s)} = \frac{1}{1 + G(s)} = \frac{s}{s + K_d K_o F(s)} \tag{5.1}$$

Phaselock Techniques, Third Edition, by Floyd M. Gardner

Suitable modifications yield $E(z)$ for a time-discrete PLL. Although this chapter concentrates on time-continuous PLLs, similar results can be expected for many time-discrete PLLs.

5.1.1 Steady-State Phase Errors

The simplest phase errors to analyze are the steady-state errors remaining after any transients have died away. These errors are readily evaluated by means of the final-value theorem of Laplace transforms, which states that

$$\lim_{t\to\infty} y(t) = \lim_{s\to 0} sY(s) \tag{5.2}$$

or for z-transforms,

$$\lim_{n\to\infty} y[n] = \lim_{z\to 1}(1 - z^{-1})Y(z) \tag{5.3}$$

That is, the steady-state value of a function in the time domain is readily determined from inspection of its transform in the transform domain. Application of the final-value theorem to the phase-error equation (5.1) yields

$$\lim_{t\to\infty} \theta_e(t) = \lim_{s\to 0} \frac{s^2\theta_i(s)}{s + K_d K_o F(s)} \tag{5.4}$$

Phase Offset As a first example, consider the steady-state error resulting from a step change $\Delta\theta$ of input phase. The Laplace transform of the input is $\theta_i(s) = \Delta\theta/s$, which may be substituted into (5.4) to give

$$\lim_{t\to\infty} \theta_e(t) = \lim_{s\to 0} \frac{s\,\Delta\theta}{s + K_d K_o F(s)} = 0 \tag{5.5}$$

[provided that $F(0) > 0$]. In other words, the loop eventually will track out any change of input phase; there is no steady-state phase error resulting from a step change of input phase in any PLL.

Frequency Offset For another example, consider the steady-state error resulting from a step change (or initial offset) $\Delta\omega$ of input frequency. The input phase is a ramp $\theta_i(t) = \Delta\omega t$, so $\theta_i(s) = \Delta\omega/s^2$. Substitution of this value of θ_i into (5.4) results in

$$\theta_v = \lim_{t\to\infty} \theta_e(t) = \lim_{s\to 0} \frac{\Delta\omega}{s + K_o K_d F(s)} = \frac{\Delta\omega}{K_o K_d F(0)} \tag{5.6}$$

The product $K_o K_d F(0)$, introduced in Section 2.2.3 as the *DC gain*, is also called the *velocity constant* and is denoted by the symbol K_{DC}. Those familiar with servo terminology will recognize it as the *velocity-error coefficient*. Note that K_{DC} has the dimensions of frequency. A similarly defined but dimensionless DC gain also exists in digital PLLs.

The frequency of an incoming signal almost never agrees exactly with the zero-control-voltage frequency of the VCO. As a rule, there is a frequency difference $\Delta\omega$ between the two. The difference may be due to an actual difference between the transmitter and receiver frequencies or it may be due to a Doppler shift. In either case, the resulting phase error is often called the *velocity error, loop stress*, or *static phase error* and is given by

$$\theta_v = \frac{\Delta\omega}{K_{\text{DC}}} \qquad \text{rad} \tag{5.7}$$

A heuristic derivation of (5.7) provides better physical insight. The control voltage increment v_c needed to retune the VCO by an amount $\Delta\omega$ is $\Delta\omega/K_o$. In the steady state, the control voltage $v_c = v_d F(0)$, where v_d is the DC output of the phase detector. But phase-detector output is produced by a phase error $\theta_e = v_d/K_d$. Therefore, to produce the necessary control voltage requires the phase error $\theta_e = \Delta\omega/K_o K_d F(0)$, as in (5.6).

Now the reason for the popularity of type 2 PLLs becomes evident. In a type 1 PLL, the DC gain is finite, so static phase error is unavoidable. Static phase error impairs the performance of the PLL. In contrast, in a type 2 PLL the DC gain is infinite because of the integrator in the loop filter [whereby $F(0) = \infty$], so static phase error is zero. You might object, rightly, that no physical analog integrator has infinite DC gain, but the DC gain in most practical PLLs can easily be made large enough to reduce the static phase error to insignificance.

Frequency Ramp Next, suppose that the input frequency is changing linearly with time at a rate of Λ rad/sec^2; that is, $\theta_i(t) = \Lambda t^2/2$. Such input behavior might arise from accelerated motion between transmitter and receiver, from changing Doppler frequency during an overhead pass of a satellite, or from sweep-frequency modulation. Laplace-transformed phase is $\theta_i(s) = \Lambda/s^3$, and it can be shown that phase error will grow without bound if K_{DC} is finite.

However, suppose that the PLL is type 2 and second order. Then from (2.7), (2.14), and (2.16), the phase error in the Laplace transform domain may be written as

$$\theta_e(s) = \frac{s^2\theta_i(s)}{s^2 + 2\zeta\omega_n s + \omega_n^2} \tag{5.8}$$

Applying the final-value theorem and the Laplace transform for a frequency ramp leads to the *acceleration error* (sometimes called *dynamic tracking error* or *dynamic lag*)

$$\theta_a = \lim_{t\to\infty} \theta_e(t) = \lim_{s\to 0} \frac{\Lambda}{s^2 + 2\zeta\omega_n s + \omega_n^2} = \frac{\Lambda}{\omega_n^2} \tag{5.9}$$

Equation (5.9) can be deduced from physical considerations. Apply a DC voltage v_d to the integrator of the loop filter. Integrator output is $v_c(t) = v_c(0) + v_d t/\tau_1$, so the rate of change of VCO frequency is $\Lambda = K_o v_d/\tau_1$. The DC voltage v_d

must be generated by a phase error $\theta_e = v_d/K_d$, which when substituted into the expression for frequency rate gives $\Lambda = K_o K_d \theta_e/\tau_1$. From (2.16), $K_o K_d/\tau_1 = \omega_n^2$, whereupon (5.9) follows.

Sometimes it is necessary to track a frequency ramp without incurring steady-state tracking error. What form of $F(s)$ is needed to reduce θ_a to zero? The expression for the final-value acceleration error is

$$\theta_a = \lim_{s\to 0} \frac{\Lambda}{s[s + K_o K_d F(s)]} \tag{5.10}$$

For θ_a to be zero it is necessary that $F(s)$ have the form $Y(s)/s^2$, where $Y(0) \neq 0$. The factor $1/s^2$ implies that the loop filter must contain two cascaded integrators. Along with the integrator of the VCO, the loop contains three integrators and so is type 3. Because of this property of eliminating the steady-state acceleration error, a type 3 PLL can be useful in tracking signals from satellites or missiles [5.1–5.3]. A type 2 PLL requires a large natural frequency and therefore a large bandwidth to handle a rapidly changing input frequency. By using a type 3 PLL instead, the frequency rate can be accommodated in a loop with small bandwidth.

DC Offset Ever present in analog PLLs is another steady-state error, one caused by unwanted DC offsets in the active filter and in the phase detector. The loop acts to produce a DC balance that includes the effect of offset. The resulting phase error needed to counteract the offset is simply the offset voltage divided by K_d, the PD gain factor. Offsets are discussed further in Chapters 10 and 11. Drift and DC offset are shortcomings of analog circuits that are absent from all-digital PLLs.

5.1.2 Transient Response

Besides steady-state behavior, it is often necessary to determine the transient phase error caused by particular inputs. The signal phases considered in Section 5.1.1 are:

- A step of phase, $\Delta\theta$ rad.
- A step of frequency (phase ramp), $\Delta\omega$ rad/sec.
- A step of acceleration (frequency ramp), Λ rad/sec^2.

For these inputs, the L-transformed input phases are $\Delta\theta/s$, $\Delta\omega/s^2$, and Λ/s^3 respectively. To compute transient phase errors, each input is substituted into (5.8) and inverse L-transforms are then computed or looked up in tables to determine time response. The analyses in this section are all predicated on a linear approximation and all fail if the loop is driven into a nonlinear region.

Type 1 PLL In a first-order loop, the resulting transient phase errors are simple exponentials:

$$\begin{aligned} &\Delta\theta e^{-Kt} && \text{phase step} \\ &\frac{\Delta\omega}{K}(1 - e^{-Kt}) && \text{frequency step} \\ &\frac{\Lambda}{K^2}(Kt + e^{-Kt} - 1) && \text{frequency ramp} \end{aligned}$$

A couple of features are worthy of note. For one, observe that the initial slope of the phase-error response to a frequency step is $\Delta\omega$ rad/sec, independent of K. Further analysis (not shown here) reveals that an initial slope of $\Delta\omega$ rad/sec occurs similarly for any PLL, irrespective of type, order, or any parameters of the loop. This phenomenon arises because the input phase begins to change abruptly at a rate of $\Delta\omega$ rad/sec at the instant of the frequency step, but corrective feedback is necessarily delayed within the loop filter and VCO. Observe also that the error response of a first-order PLL to a frequency ramp increases with time and that linear bounds are eventually passed; a first-order PLL is not suitable for tracking a long-lasting frequency ramp.

Type 2 PLL Analytic expressions for transient phase error of the important second-order type 2 loop are given in Table 5.1, normalized to natural frequency ω_n, and are plotted in Figs. 5.1 to 5.3. These curves have appeared widely in the PLL literature and have contributed to the popularity of ω_n as a descriptive parameter of PLLs. Additional plots of the same transient response are shown in Figs. 5.4 to 5.6 normalized to loop gain K and in Figs. 5.7 to 5.9 normalized to noise bandwidth $2B_L$ (see Chapter 6 for a definition of noise bandwidth). These six charts are obtained from the same equations listed in Table 5.1 after substitutions of

$$K = 2\zeta\omega_n, \qquad B_L = \frac{\omega_n}{2}\left(\zeta + \frac{1}{4\zeta}\right) \tag{5.11}$$

Why bother with multiple plots of the same underlying equations, plots that differ only in the normalization parameter? For one reason, natural frequency ω_n, although traditionally popular in the literature, is much less advantageous to a design engineer; loop gain K is a much more useful parameter for wideband PLLs, and noise bandwidth B_L is often much more useful for narrowband PLLs. Refer to Section 2.2.3 for further discussion of parameters. For another reason, unexpected properties emerge from a study of plots with differing normalizations, as brought out in the following examples. But before inspecting the examples, notice that all the charts exhibit large over- or undershoots for a small damping factor ζ. Large oscillatory transients are ordinarily unacceptable, so small damping values are to be avoided except in highly unusual circumstances.

TABLE 5.1 Transient Phase Error $\theta_e(t)$ (rad) of Second-Order Type 2 PLL

	Phase Step, $\Delta\theta$ (rad)	Frequency Step, $\Delta\omega$ (rad/sec)	Frequency Ramp, Λ (rad/sec^2)
$\zeta < 1$	$\Delta\theta\left(\cos\sqrt{1-\zeta^2}\ \omega_n t - \frac{\zeta}{\sqrt{1-\zeta^2}}\sin\sqrt{1-\zeta^2}\ \omega_n t\right)e^{-\zeta\omega_n t}$	$\frac{\Delta\omega}{\omega_n}\left(\frac{1}{\sqrt{1-\zeta^2}}\sin\sqrt{1-\zeta^2}\ \omega_n t\right)e^{-\zeta\omega_n t}$	$\frac{\Lambda}{\omega_n^2} - \frac{\Lambda}{\omega_n^2}\left(\cos\sqrt{1-\zeta^2}\ \omega_n t + \frac{\zeta}{\sqrt{1-\zeta^2}}\sin\sqrt{1-\zeta^2}\ \omega_n t\right)e^{-\zeta\omega_n t}$
$\zeta = 1$	$\Delta\theta(1-\omega_n t)e^{-\omega_n t}$	$\frac{\Delta\omega}{\omega_n}(\omega_n t)e^{-\omega_n t}$	$\frac{\Lambda}{\omega_n^2} - \frac{\Lambda}{\omega_n^2}(1+\omega_n t)e^{-\omega_n t}$
$\zeta > 1$	$\Delta\theta\left(\cosh\sqrt{\zeta^2-1}\ \omega_n t - \frac{\zeta}{\sqrt{\zeta^2-1}}\sinh\sqrt{\zeta^2-1}\ \omega_n t\right)e^{-\zeta\omega_n t}$	$\frac{\Delta\omega}{\omega_n}\left(\frac{1}{\sqrt{\zeta^2-1}}\sinh\sqrt{\zeta^2-1}\ \omega_n t\right)e^{-\zeta\omega_n t}$	$\frac{\Lambda}{\omega_n^2} - \frac{\Lambda}{\omega_n^2}\left(\cosh\sqrt{\zeta^2-1}\ \omega_n t + \frac{\zeta}{\sqrt{\zeta^2-1}}\sinh\sqrt{\zeta^2-1}\ \omega_n t\right)e^{-\zeta\omega_n t}$

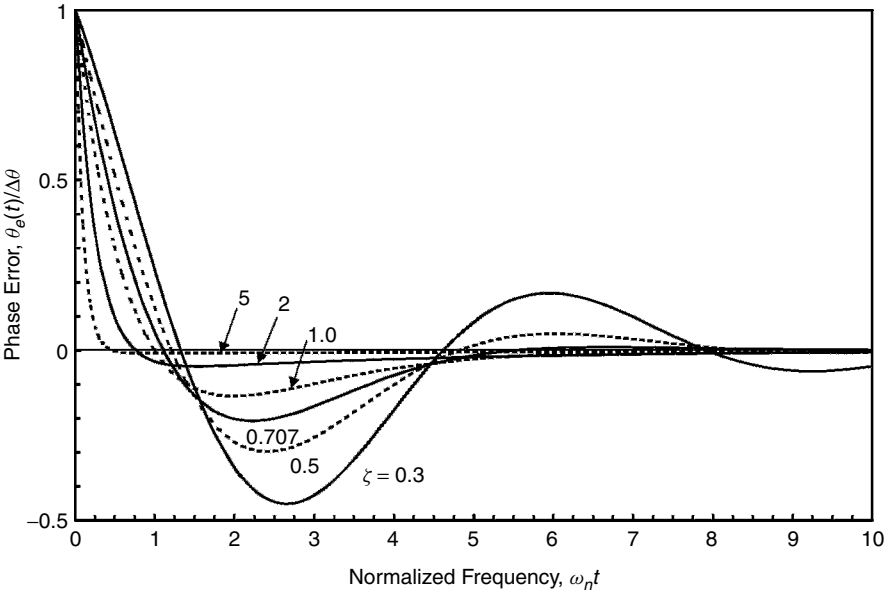


Figure 5.1 Transient response to a phase step of a second-order type 2 PLL.

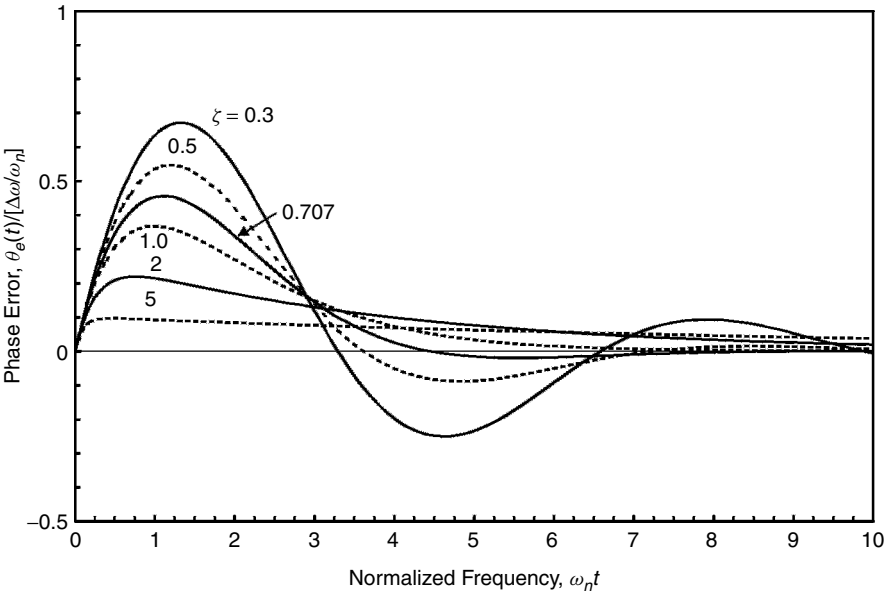


Figure 5.2 Transient response to a frequency step of a second-order type 2 PLL.

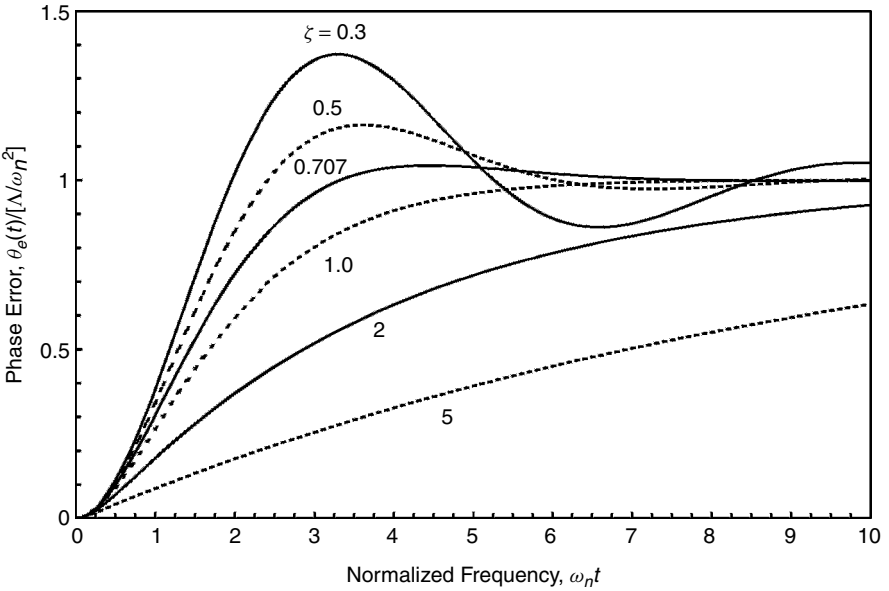


Figure 5.3 Transient response to a frequency ramp of a second-order type 2 PLL.

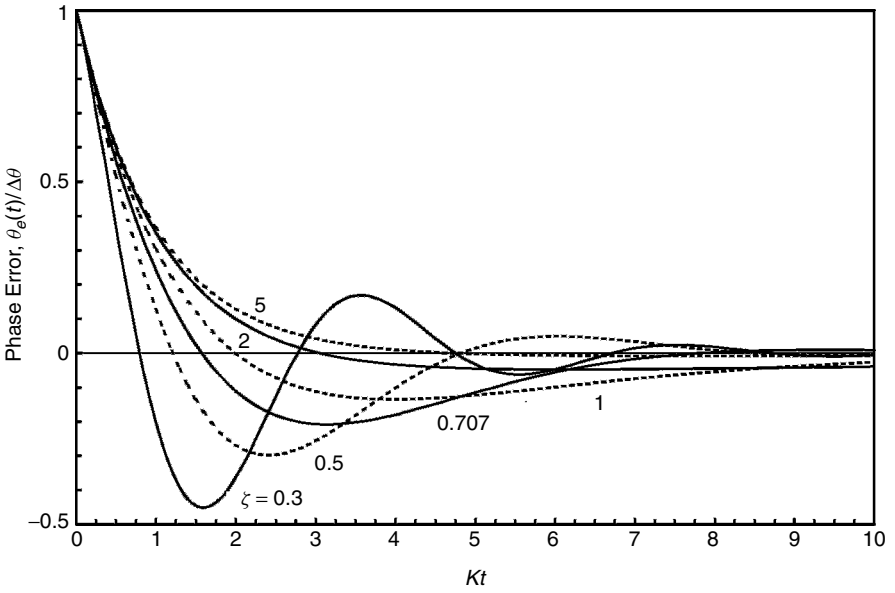


Figure 5.4 Transient response to a phase step of a second-order type 2 PLL.

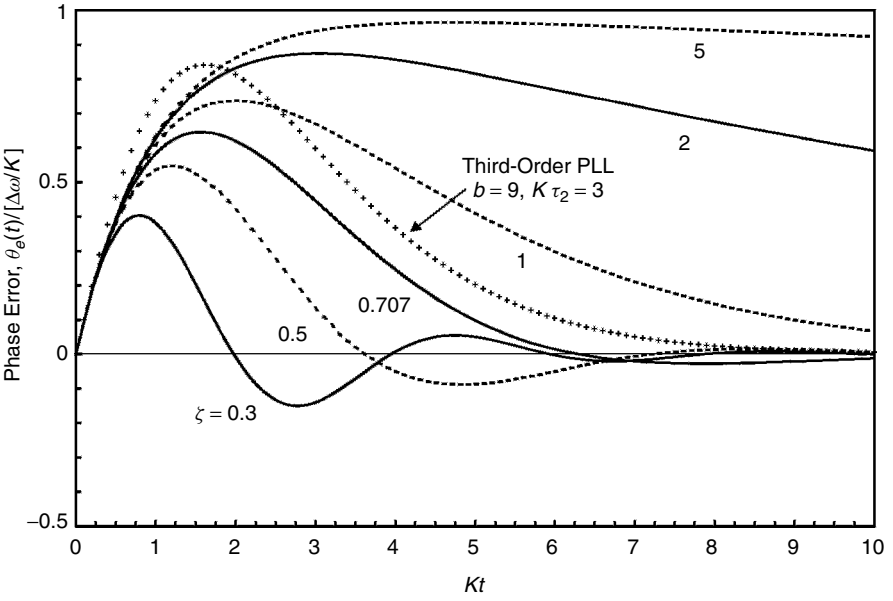


Figure 5.5 Transient response to a frequency step of a second-order type 2 PLL.

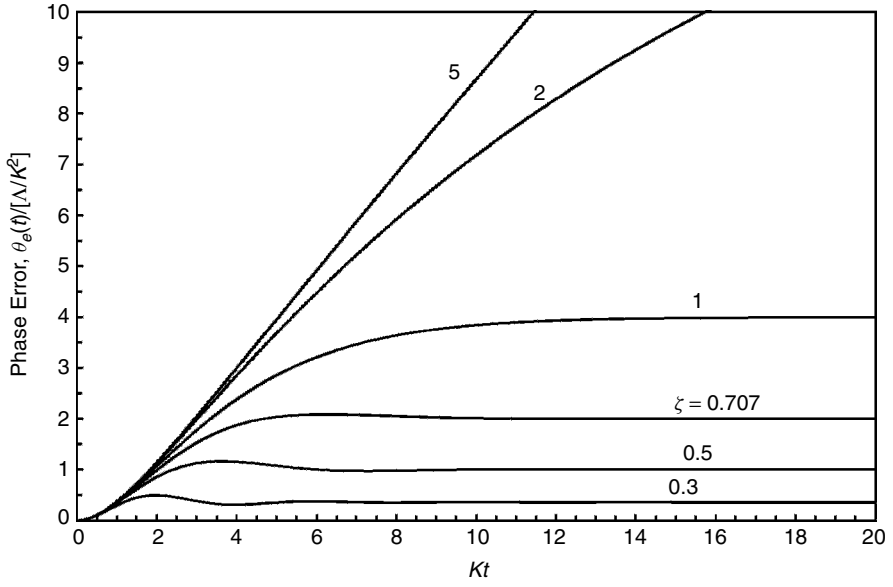


Figure 5.6 Transient response to a frequency ramp of a second-order type 2 PLL.

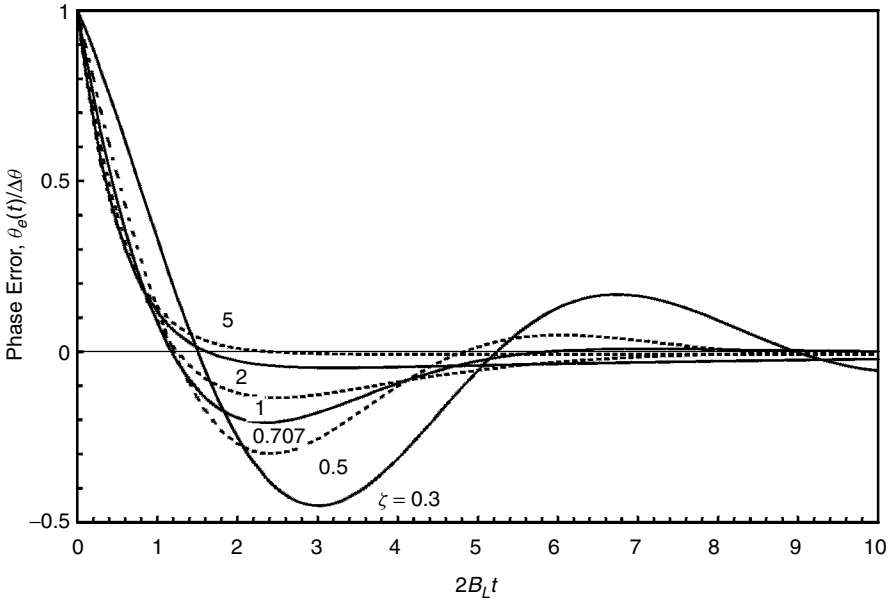


Figure 5.7 Transient response to a phase step of a second-order type 2 PLL.

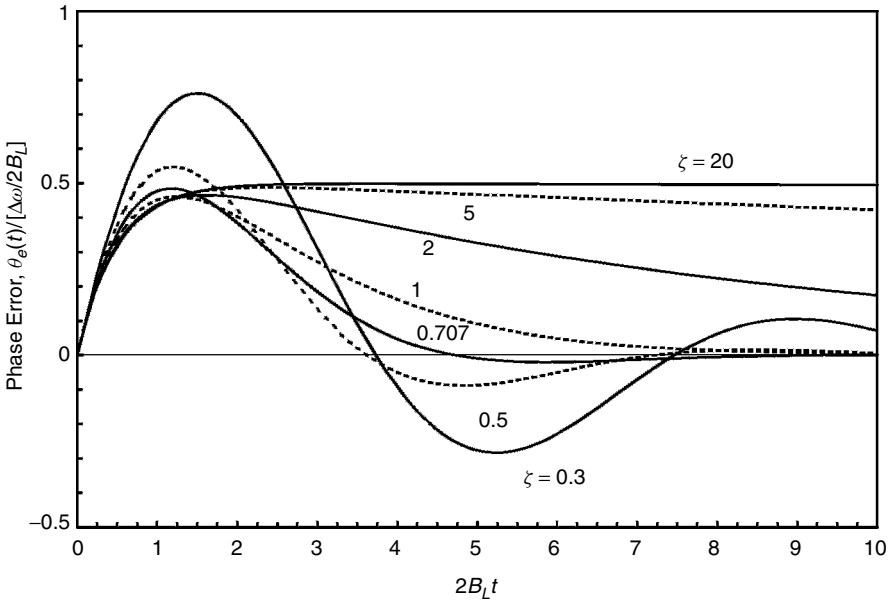


Figure 5.8 Transient response to a frequency step of a second-order type 2 PLL.

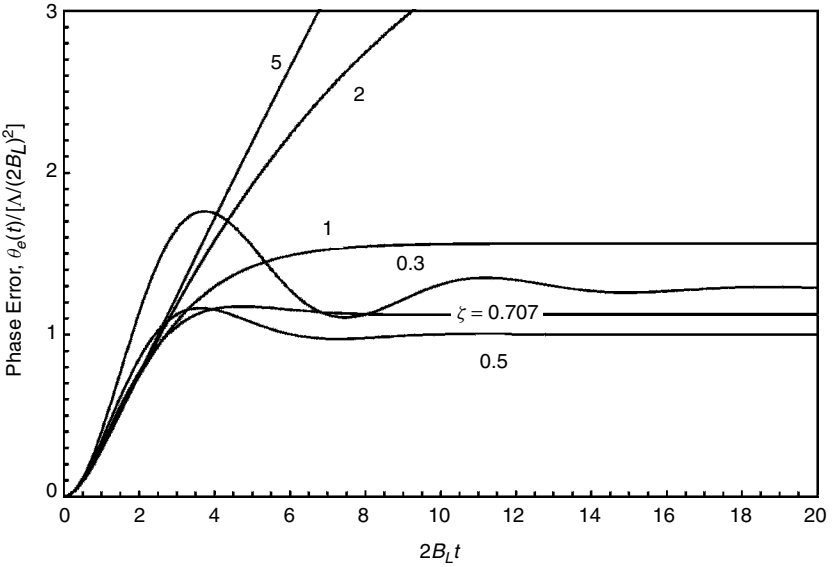


Figure 5.9 Transient response to a frequency ramp of a second-order type 2 PLL.

Phase-Step Transients Figure 5.1 (normalized to ω_n) appears to show the fastest initial rate of descent of the error response occurring for the largest damping factor, but Fig. 5.4 (normalized to K) shows the initial rate to be slowest for the largest damping factor, and Fig. 5.7 (normalized to $2B_L$) shows the initial rates to be about equal for all $\zeta > 0.5$. Figure 5.1 is misleading because a larger damping factor with fixed ω_n implies a larger bandwidth through the relation $K = 2\zeta\omega_n$ and a more lively response is to be expected with larger bandwidth.

Frequency-Step Transients The same reversal of effects can be seen more clearly when comparing Fig. 5.2 (normalized to ω_n) to Fig. 5.5 (normalized to K); the apparent advantage of high damping in Fig. 5.2 arises solely because of the consequent larger bandwidth noted above. Consider the following argument for an advantage of small damping for quick recovery from frequency steps. (But small damping has other disadvantages; you should rarely choose damping—or any other parameter—based solely on a single criterion.) In each of Figs. 5.2, 5.5, and 5.8, initially the phase error rises approximately as a straight line with slope independent of damping factor. After some time interval (depending on damping), the slope levels off and reverses. Think of the initial portion of the transient as phase error accumulating gradually at the rate $\Delta\omega$ rad/sec, due to the changed frequency, while the leveling off and reversal of slope is due to feedback through the combined proportional and integral paths of the loop. In Fig. 5.5, the gain in the proportional path is the same for all curves, whereas the gain in the integral path is an inverse function of the damping. Small damping (large integral-path gain) leads to quick reversal of the transient, whereas large damping (small integral-path gain) leads to very slow reversal. For very large damping,

the peak error approaches $\Delta\omega/K$, which is the steady-state error of a type 1 PLL. Figure 5.2 is misleading because proportional path gain K increases with ζ for fixed ω_n. Figure 5.8 (normalized to $2B_L$) shows an unexpected phenomenon: The peak phase error is $\Delta\omega/B_L$ radians $\pm$ 10% for all $\zeta > 0.5$, independent of ζ, a property to be remembered when designing narrowband PLLs.

Frequency-Ramp Transients Figure 5.3 (normalized to ω_n) shows all error curves converging to the same steady-state value Λ/ω_n^2, in agreement with (5.9). Normalization to ω_n is most informative in this one instance, where the steady-state phase error depends exclusively on ω_n and not so directly on one of the other bandwidth parameters.

Higher-Order Type 2 PLLs The transient responses in Figs. 5.1 to 5.9 each take place over a time span that is long compared to either $1/\omega_n$, $1/K$, or $1/B_L$. Accordingly, one would not expect additional high-frequency poles to have a strong effect on the transient response, especially if the high-frequency poles are well above the gain-crossover frequency.

An extreme example, the response of a particular third-order type 2 PLL, has been included in Fig. 5.5 along with the responses of second-order type 2 PLLs with the same loop gain K. The third-order example has $b = 9$ and $K\tau_2 = 3$. (See Section 2.3.3 for further details on this kind of PLL.) Analysis yields its Laplace transform response to a frequency step $\Delta\omega$ as

$$\theta_e(s) = \frac{\Delta\omega(s + 3K)}{(s + K)^3} \tag{5.12}$$

from which the time-domain transient

$$\theta_e(t) = \frac{\Delta\omega}{K}[Kte^{-Kt}(Kt + 1)] \tag{5.13}$$

is plotted as the curve labeled "third-order PLL" in Fig. 5.5. Compare the curve for the third-order loop to that of the second-order loop with $\zeta = 1$. The two PLLs are similar in that each has all its poles coincident on the negative real axis: at $s = -K$ for the third-order PLL and at $s = -K/2$ for the second-order PLL. Because the poles are real, both transient responses are unipolar; there is no backswing over zero error as the transient settles.

A condition $K\tau_2 = 3$ would cause a damping factor $\zeta = 0.866$ in a second-order type 2 PLL but clearly not in the example third-order PLL. Nonetheless, the settling curve for the third-order PLL lies between $\zeta = 0.707$ and $\zeta = 1$, as would that for a second-order PLL with $\zeta = 0.866$. Thus, the settling behavior for the third-order PLL is roughly the same as that of a second-order PLL with the same value of $K\tau_2$, despite the third pole being extremely close in. The major discrepancy in transient behavior occurs in the rise to the peak error, which is higher and steeper for the third-order PLL than it would be for a second-order PLL with the same $K\tau_2$. Why is the peak higher? The extra lowpass filtering causes the filtered phase error indication from the phase detector to reach the

VCO later than it would in the absence of the third pole, so more phase error accumulates before significant corrective feedback can take effect.

Type 3 PLL A type 3 PLL can be treated in the same manner as a type 2 PLL, but published results [5.1–5.7] are few and widely scattered. The reason lies partly in the far greater popularity of the type 2 loop, but also in the extra complexity of type 3. There are (at least) three loop parameters in a type 3 loop, so many pages of figures would be needed to present the same kind of data, as in, for example, Fig. 5.1. As a rule of thumb, one can assume that the transient error of a type 3 loop in response to a phase step or frequency step would be roughly the same as that of a type 2 loop with the same loop gain K and similar positions of the dominant poles. Examples are provided in [5.4]. The main difference between the two loop types arises in the response to a frequency ramp. Whereas the response of a type 2 PLL has a steady-state error of Λ/ω_n^2 radians, a type 3 PLL with the same input has zero steady-state error. Nonetheless, type 3 and type 2 PLLs of comparable bandwidths will exhibit approximately the same peak phase error in response to the sudden onset of a frequency ramp, a crucial fact to remember when dealing with dynamically varying signals.

More Complicated Inputs A phase step, a frequency step, or a frequency ramp are useful simplifications of signal properties encountered in practice. Knowledge of the response to a simplified input is a valuable guide to the behavior of a PLL under more general circumstances. Nonetheless, if properties of the input signal depart significantly from the simplifications, the methods and results presented here may not be adequate; an engineer may have to resort to numerical calculations of the signals and responses. Some examples appear in [5.3]. Computer programs for calculations of time-domain responses of linear circuits are well suited to this task.

Digital PLLs The foregoing material on transient response was developed for analog PLLs. From Section 4.7 one can anticipate that similar behavior should be expected from digital PLLs whose bandwidths are small compared to the sampling rate. What is meant by "small"? Perhaps loop gain $\kappa \approx 0.1$ is a plausible dividing line. Time-discrete analysis will be required for determination of the actual transient behavior if larger values of loop gain are employed.

5.1.3 Response to Sinusoidal Angle Modulation

Next, let's investigate loop behavior in the presence of an angle-modulated input signal. For sinusoidal phase modulation

$$\theta_i(t) = \Delta\theta \sin \omega_m t \tag{5.14}$$

and for sinusoidal frequency modulation

$$\theta_i(t) = \frac{\Delta\omega}{\omega_m} \cos \omega_m t \tag{5.15}$$

where $\Delta\theta$ is the peak phase deviation, $\Delta\omega$ is the peak frequency deviation, and ω_m is the modulating frequency. Phase error is sinusoidal (in the linear approximation) and may be calculated simply as the steady-state frequency response of the closed-loop error response $E(s)$. Examples are shown in Figs. 5.10 to 5.12. Error response to phase modulation is a highpass function of modulating frequency, as shown in Figs. 2.5 and 5.10. At low frequencies, the response amplitude rises at $6n$ dB/octave for a type n loop. At high modulating frequencies, the loop is unable to follow the modulation, so the full modulation phase appears as error at the phase detector. Accordingly, the high-frequency asymptote in Fig. 5.10 is constant at 0 dB.

The curves of Fig. 5.10 are sketches; the examples are loops of different types that have the same corner frequencies in the error response $E(s)$. It is apparent that for any frequency within the loop bandwidth, a higher-type loop tracks the modulation better than does a lower-type loop. Expect that any additional high-frequency filtering within the loop would have its main effect in the vicinity of the corner frequency of Fig. 5.10 and little effect on the asymptotes.

Error response to sinusoidal FM is shown in Fig. 5.11 for three different kinds of loops. Note that the high-frequency asymptote is the same for all loops; the response differences lie at low frequencies within the loop bandwidth. The 6-dB/octave rolloff at high frequencies occurs solely because input phase deviation $\Delta\theta = \Delta\omega/\omega_m$ is inversely proportional to the modulating frequency. The curves of Fig. 5.11 have all been drawn for loops with equal loop gain K. The first-order loop has a lowpass response in accordance with its one-pole transfer function, whereas the type 2 loop is more effective at tracking out the lower frequencies.

Figure 5.12 shows phase error in response to FM of a second-order type 2 PLL, plotted against natural frequency with damping as a plot parameter. Phase

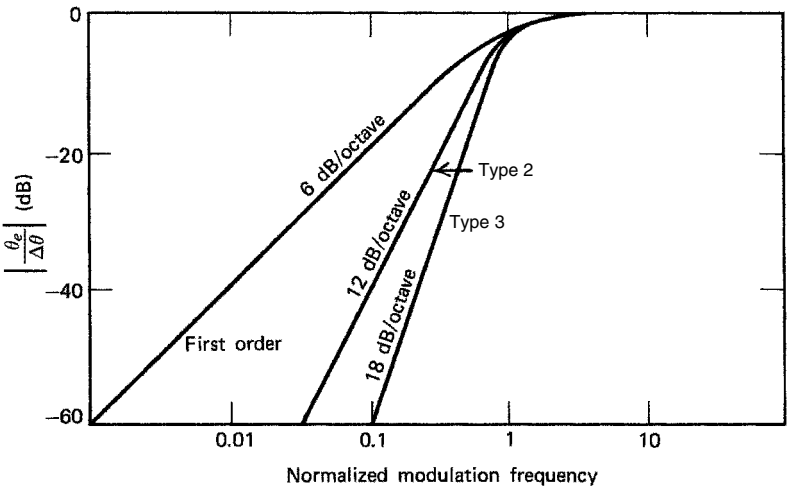


Figure 5.10 Peak steady-state phase error due to sinusoidal PM with peak deviation $\Delta\theta$ and modulation frequency ω_m.

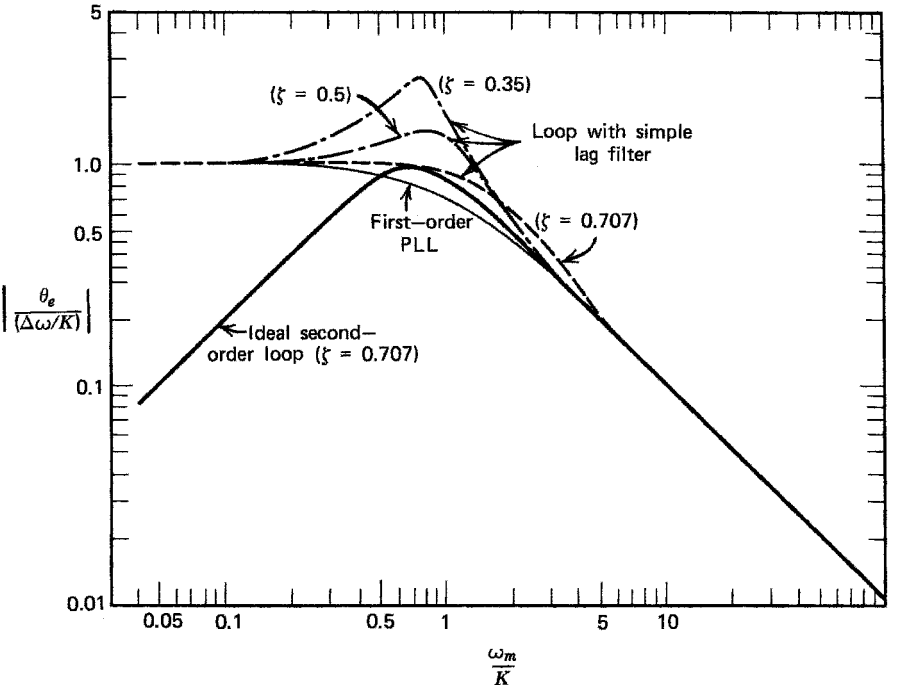


Figure 5.11 Peak steady-state phase error due to sinusoidal FM with peak deviation $\Delta\omega$ and modulation frequency ω_m.

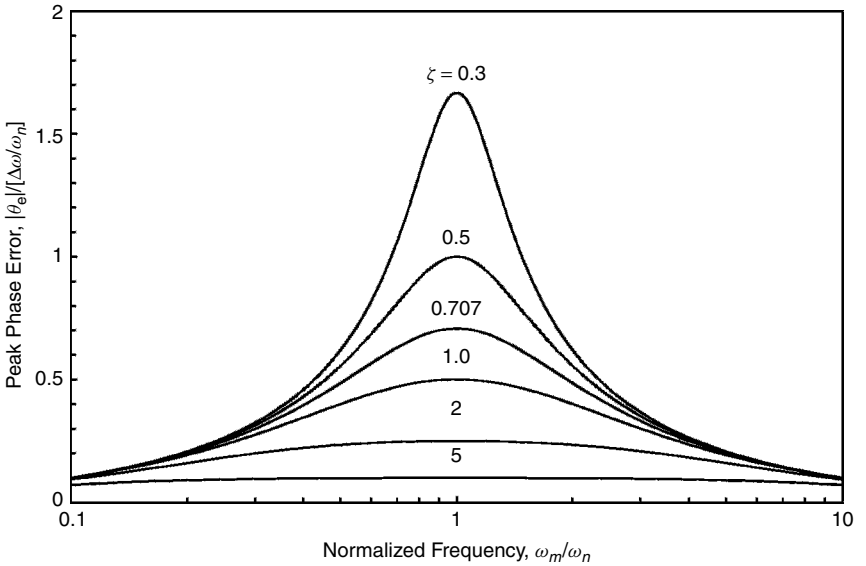


Figure 5.12 Peak steady-state phase error of a second-order type 2 PLL due to sinusoidal FM with peak deviation $\Delta\omega$ and modulation frequency ω_m.

error is maximum at a modulating frequency equal to natural frequency ω_n, irrespective of damping. Peak amplitude is $\Delta\omega/K$ and the phase shift between the phase error and the input frequency modulation passes through zero at the amplitude peak at $\omega_m = \omega_n$. These properties are sometimes used as the basis for experimental measurement of ω_n.

5.2 NONLINEAR TRACKING: LOCK LIMITS

All the preceding material on tracking and phase error is based on the assumption that the phase error is small enough for the loop to be considered linear in its operation. This assumption becomes progressively less accurate as error increases, until finally the loop drops out of lock and the assumption becomes totally worthless. In this section the linear assumption is discarded and the limiting conditions for loop lock are investigated.

5.2.1 Phase-Detector Nonlinearity

As explained further in Chapter 10, phase detector *s-curves* (output vs. phase error) are periodic and necessarily nonlinear. An s-curve is a function of the phase error, customarily denoted $g(\theta_e)$; the slope $dg/d\theta_e$ evaluated at $\theta_e = 0$ is the phase detector gain K_d. An s-curve is a bounded function; a phase detector can deliver no more than some maximum amount of output. One very common and important s-curve is sinusoidal:

$$g(\theta_e) = K_d \sin\theta_e \tag{5.16}$$

Lock limits imposed by sinusoidal phase detectors have received prominent attention in the literature and also in the next sections of this chapter. Other s-curves will be introduced presently. Despite the emphasis on phase-detector nonlinearity, be aware that tracking limits are likely to be imposed instead by other elements within the loop, as described in the sequel.

5.2.2 Steady-State Limits

The first topic considered is the input frequency range over which the loop will hold lock. In (5.7) the linear approximation of steady-state phase error due to a frequency offset is shown to be $\theta_v = \Delta\omega/K_{\mathrm{DC}}$. However, for a phase detector with a sinusoidal s-curve, the true expression should be $\sin\theta_v = \Delta\omega/K_{\mathrm{DC}}$. The sine function cannot exceed unit magnitude; therefore, there is no solution to this equation if $\Delta\omega > K_{\mathrm{DC}}$. Instead, the loop falls out of lock and the phase-detector voltage becomes a beat note rather than a DC level. The *hold-in range* of a PLL with sinusoidal phase detector may therefore be defined as

$$\Delta\omega_H = \pm K_{\mathrm{DC}} \qquad \text{rad/sec} \tag{5.17}$$

Equation (5.17) states that the hold-in range can be made arbitrarily large simply by using very high DC gain. Of course, unlimited increase of gain is not a panacea since another component in the loop will then overload before the phase detector does. Consider this reasoning: Some definite control voltage is needed to achieve any given frequency deviation of the VCO. However, the loop amplifier (if one is used) has some maximum voltage excursion that it can deliver and the VCO has some maximum control voltage excursion that it can accept. The loop will unlock if either of these limits is exceeded. In practice, it is common to find PLLs with such high DC gain that the amplifier or VCO saturates when static phase error is only a very few degrees. If the PLL is truly type 2 (as is usual in digital PLLs), static phase error due to frequency offset is exactly zero and the hold-in range is determined entirely by bounds on some element other than the phase detector.

Dynamic error in a type 2 PLL was approximated previously (5.9) as $\theta_a = \Lambda/\omega_n^2$. The correct expression for a phase detector with a sinusoidal s-curve should be $\sin\theta_a = \Lambda/\omega_n^2$, from which it may be deduced that the maximum permissible rate of change of input frequency is

$$\Lambda = \omega_n^2 \tag{5.18}$$

The loop falls out of lock if the input rate exceeds this amount.

Many phase detectors have greater linear spans and larger maximum outputs than afforded by the sinusoidal s-curve of (5.16). Several examples are shown in Fig. 5.13. All curves of Fig. 5.13 are shown with the same slope at $\theta_e = 0$,

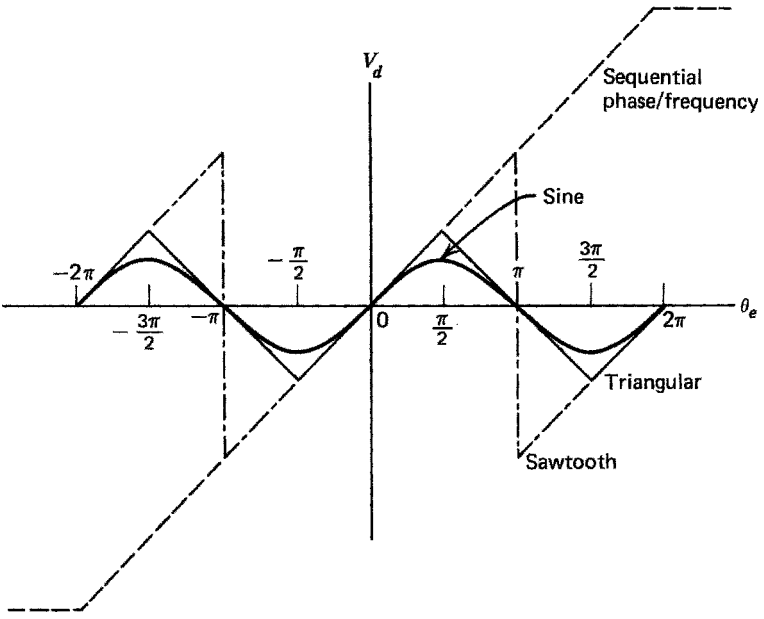


Figure 5.13 Phase detector s-curves.

which means that the various PDs all have the same gain factor K_d. Circuits that provide these and other extended s-curves are described in Chapter 10.

Increased PD output capability provides a larger tracking range (i.e., a larger lock limit) than is obtainable from a sinusoidal PD. (Of course, the extended range of the PD is helpful only if the limit is set by the PD and not by some other nonlinearity, such as clipping in the operational amplifier.) Lock-limit extension of each of the kinds of PDs shown in Fig. 5.13 is given in the following table. Hold-in (5.17) and rate limits (5.18) are both extended by the same factor.

PD Type	Extension Factor
Sinusoidal	1
Triangular	$\pi/2$
Sawtooth	π
Sequential phase/frequency	2π

5.2.3 Transient Limits

Figures 5.1 to 5.9 demonstrate that transient phase error can be much larger than steady-state phase error, implying that a loop can be pulled out of lock on a transient basis by an input change that could be tracked easily in the steady state. This section examines a number of such excessive transient conditions.

Most phase detectors are periodic and so cannot distinguish a phase step of $\Delta\theta + 2\pi n$ from one of $\Delta\theta$. Therefore, in the absence of other stress, an ordinary PLL should never lose lock when subjected to a phase step, irrespective of magnitude or loop order. (This ability applies only if the phase detector's s-curve has a nonzero output of the correct polarity for all phase errors within one period; see Fig. 14.3 for a counterexample.) A frequency step can break the lock. A first-order loop loses lock if and only if the frequency step exceeds the hold-in limit as given by (5.17) for a sinusoidal PD and extended as above for other shapes of s-curves. Limits for a type 2 PLL are treated after introducing a tool for nonlinear analysis.

Phase-Plane Principles *Phase-plane portraits* are useful for the study of transient nonlinear behavior of second-order PLLs. Descriptions of phase planes are found in numerous control system texts. Viterbi [5.8, 5.9] has adapted phase-plane analysis to PLLs with sinusoidal s-curves. The dynamics of a second-order loop may be described by a pair of first-order nonlinear differential equations using time as the independent variable and using phase error θ_e and frequency error $d\theta_e/dt = \omega_e$ as the dependent variables. Eliminating the time variable between the equations produces a single second-order nonlinear differential equation that relates phase and frequency errors.

Solutions of the second-order equation are in terms of $d\theta_e/dt = \omega_e$ vs. θ_e; these can be plotted in the phase plane, which has ω_e and θ_e as its coordinates. Solutions cannot be obtained analytically; computer assistance is needed. A plot

of a single solution in the phase plane is known as a *phase-plane trajectory*. A family of trajectories is known as a *phase-plane portrait*. A trajectory shows the dynamic behavior of a loop as it settles (or fails to settle) toward equilibrium.

Figure 5.14 is a sketch of one particular phase-plane portrait for a type 2 PLL with sinusoidal phase detector and critical damping $\zeta = 1$. Different portraits are obtained for different choices of loop damping, phase-detector s-curve, loop stress, or signal modulation. The best source of portraits may be found in Viterbi's original report [5.8] if accessible. His book [5.9] contains the same portraits but at an inconveniently reduced scale. Blanchard's book [5.10] has a few of the portraits on a larger scale. Many of the results that follow in this section and in Chapter 8 were obtained by use of the portraits in [5.8]. Phase-plane analysis is central to an understanding of nonlinear dynamics of second-order loops.

The phase-plane portrait of a PLL with a periodic phase detector is itself periodic with period 2π in the variable θ_e and aperiodic in ω_e. The pattern repeats indefinitely along the phase axis; two complete periods are shown in Fig. 5.14. Trajectories proceed clockwise only, as shown by the flow arrows. Intersection of trajectories can occur only at singular points, which can be either stable or unstable. Equilibrium occurs (the trajectory reaches a rest point) at a stable singularity, which is called a *stable node* if the loop is overdamped or a *stable focus* if it is underdamped. Equilibrium is a steady-state tracking condition that may be reached after an infinite time. (Conditions can be imposed such that no equilibrium exists; see Viterbi [5.8] or [5.9] for examples.)

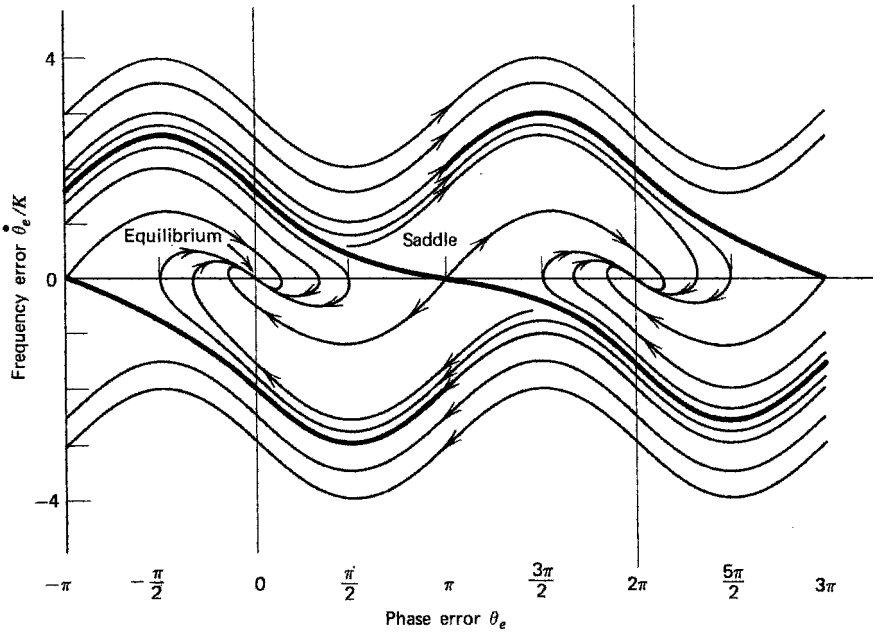


Figure 5.14 Phase-plane portrait for a second-order type 2 PLL with $\zeta = 1$ and a sinusoidal phase detector.

An unstable singularity is called a *saddle point*; the loop state cannot remain at a saddle point indefinitely because any slight disturbance sets it on an active trajectory. A trajectory that terminates on a saddle point is called a *separatrix*. The separatrices of Fig. 5.14 are indicated by heavy curves. (The designation "separatrix" applies only in the 2π interval in which the trajectory terminates on a saddle point and not all the way back into the infinite past.)

If a trajectory lies between two separatrices, it will terminate at the equilibrium point of that particular 2π interval. If a trajectory lies outside the separatrices, the loop slips one or more complete cycles before arriving at equilibrium (if, indeed, it ever reaches equilibrium; unending slippage is also possible). A *cycle slip* is an excursion of phase error by 2π radians.

Application of a Phase Plane Now consider transients in a second-order type 2 PLL, a loop with infinite DC gain. In principle, this kind of loop can never lose lock permanently. If a large frequency step is applied, the loop unlocks, slips cycles for awhile and then locks up once again. The phase error is a ringing oscillation for a number of cycles corresponding to the number of cycles slipped. There is some frequency-step limit below which the loop does not slip cycles but remains in lock; denote this limit as the pullout frequency and give it the label $\Delta\omega_{\mathrm{PO}}$. If the loop is at equilibrium of $\theta_e = 0$ and $\omega_e = 0$ at the instant the frequency step is applied, the pullout limit is simply the intercept of the separatrix with the $\theta_e = 0$ axis. Using the portraits of [5.8], the pullout limit for a sinusoidal PD was found to have the values indicated in Fig. 5.15. These data

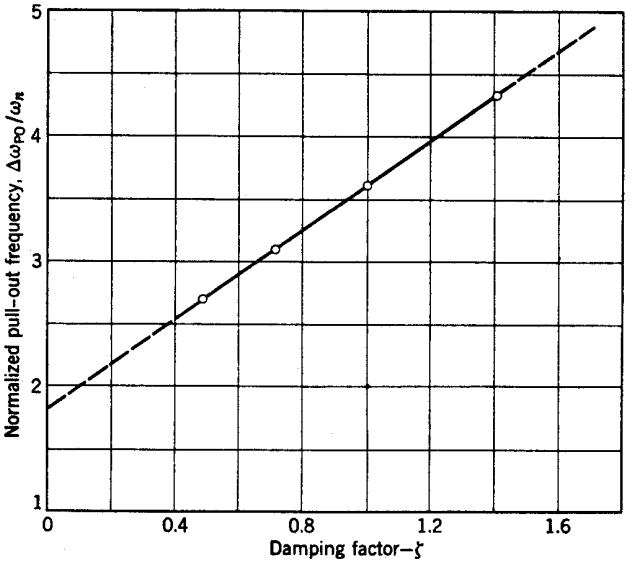


Figure 5.15 Pullout frequency of a second-order type 2 PLL with a sinusoidal phase detector.

points fit the empirical relation

$$\Delta\omega_{\mathrm{PO}} = 1.8\omega_n(\zeta + 1) \tag{5.19}$$

for ζ between 0.5 and 1.4.

The phase portrait can also be used to determine peak phase error for large steps of frequency. Peak phase error is 180° for $\Delta\omega = \Delta\omega_{\mathrm{PO}}$. However, the error increases rapidly as soon as it exceeds 90°, and therefore the frequency step causing 90° peak error is only slightly less than $\Delta\omega_{\mathrm{PO}}$. Figure 5.16 shows the situation for the special case of $\zeta = 0.707$.

A phase plane is applicable only to a second-order loop (or a degenerate phase plane to a first-order loop). A third-order loop has three state variables—phase, frequency, and frequency rate—so it must have a three-dimensional phase space to represent it completely. Presentation of such a space is very difficult to achieve in two dimensions. As a result, much less is known about the nonlinear transient response of higher-order loops than of second-order loops.

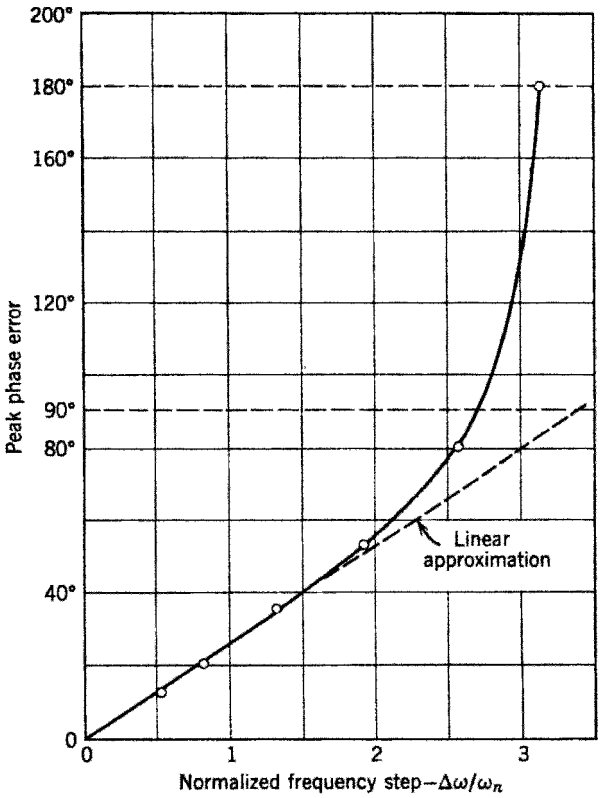


Figure 5.16 Peak transient phase error in response to a frequency step in a second-order type 2 PLL with $\zeta = 0.707$ and a sinusoidal phase detector.

5.2.4 Modulation Limits

An engineer must also be concerned with unlock problems when the input signal is angle modulated; a PLL is unable to remain locked to the signal if the modulation index is excessive. Distinguish between *carrier tracking* loops in which the modulation spectrum is entirely outside the loop bandwidth and *modulation tracking* loops in which the modulation spectrum is inside the loop bandwidth. The first kind is used primarily for demodulation of small-index PM signals, whereas the second accommodates large-index FM or PM signals.

Carrier-Tracking PLLs Modulation applied to a carrier-tracking loop must be restricted so that a trackable carrier actually exists. If sinusoidal phase modulation of peak deviation θ is applied, the carrier strength is proportional to the zero-order Bessel function $J_0(\theta)$. This function passes through its first zero for $\theta = 2.4$ rad (137°). Experiments have demonstrated that lock is lost for modulation index very close to that first null. As deviation is increased somewhat beyond 2.4 rad, lock is regained and is held until deviation reaches the next null of J_0 at $\theta = 5.5$ rad. In principle, a carrier-tracking loop loses lock on a sinusoidally modulated signal only in the immediate vicinity of the carrier nulls and holds lock for all other modulation indices.

Modulation-Tracking PLLs: Sinusoidal Modulation The behavior of a modulation-tracking loop cannot be explained nearly as easily. To introduce the problem, imagine a laboratory experiment in which a sinusoidally modulated signal (PM or FM) is applied to a PLL with sinusoidal phase detector. The loop gain K is required to be very much larger than the modulation frequency ω_m so that the PLL can track the modulation. Otherwise, the ensuing explanation no longer applies.

Phase-detector output voltage is observed on an oscilloscope and the modulation index is adjusted appropriately. At small deviations the observed waveshape is sinusoidal, as would be expected. Amplitude of the PD output increases with increasing deviation. If the deviation is made too large, the loop begins slipping cycles and severe distortion appears on the scope face (slip details are given later).

Absence of Distortion However, PD output appears to remain nearly sinusoidal (i.e., nearly undistorted) from small index all the way up to the break-lock condition. This behavior is rather surprising since a PD operates well into its nonlinear region before break lock is reached. How can there be low-distortion operation in a nonlinear device? The answer, of course, is that negative feedback cancels out most of the distortion at the PD output, provided that feedback gain is large at the modulation frequency. Reduction of distortion is a familiar property of feedback loops in general, a property that is shared by the PLL in particular.

If the PD output is almost undistorted, the peak phase error must increase as the inverse sine of the input deviation, to a good approximation. In other words, the distortion expected because of the PD nonlinearity appears in the phase error θ_e but not in the PD output $v_d = K_d \sin\theta_e$. This relation is used in the next

paragraphs to determine the waveshape of the phase error as a function of the peak PD output.

If the PD output is sinusoidal, it must be of the form $v_d(t) = aK_d \sin\theta_m t$, where a is a factor between 0 and 1. Maximum possible output voltage from a sinusoidal phase detector is K_d volts, so a is the ratio of peak output to maximum possible output. Furthermore, $v_d(t) = K_d \sin\theta_e(t)$, which leads to (valid in the first quadrant)

$$\theta_e(t) = \sin^{-1}(a \sin\omega_m t) \tag{5.20}$$

Examples for several values of a are plotted in Fig. 5.17. Considerable distortion of θ_e is evident for large values of a, but the curves for v_d are all sinusoidal.

Modulation Limits Once it is recognized that v_d is substantially undistorted, it becomes a simple matter to establish the modulation limits. Knowing the input modulation, the amplitude of v_d is found from the frequency response

$$V_d(j\omega_m) = K_d E(j\omega_m)\theta_i(j\omega_m) \tag{5.21}$$

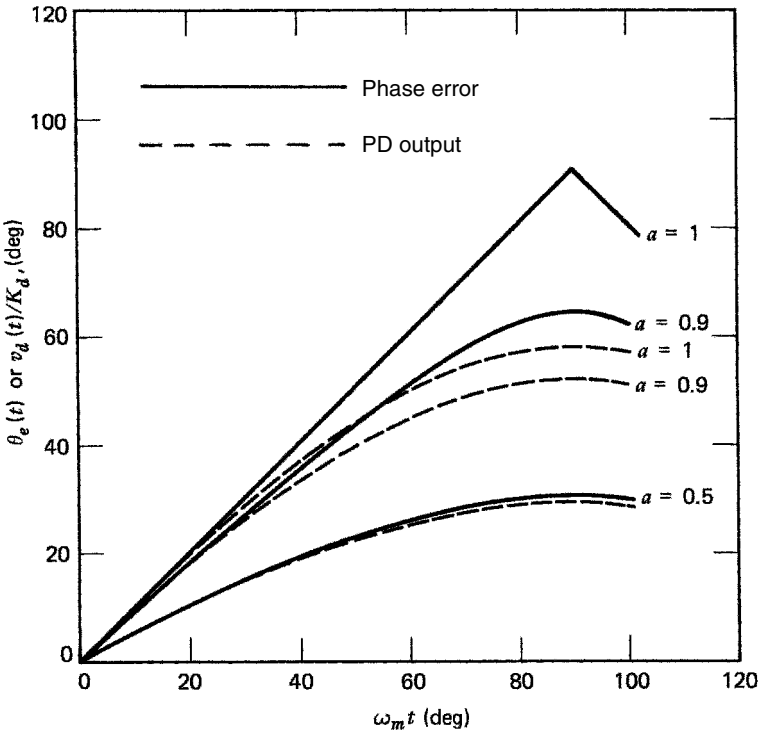


Figure 5.17 Phase error $\theta_e(\omega_m t)$ (solid curves) and PD output $v_d(\omega_m t)/K_d$ (dashed curves) in a PLL with a sinusoidal phase detector in response to sinusoidal angle modulation with frequency ω_m, based on $v_d(t)/K_d = a \sin\omega_m t = \sin\theta_e(t)$.

where θ_i represents the phase modulation on the input signal. The magnitude of V_d in (5.21) yields the peak value v_{dp} of the sinusoidal PD output at radian frequency ω_m. The loop remains in lock if the calculated value of $v_{dp} < K_d$ and slips cycles if the calculated value of $v_{dp} > K_d$. (This criterion is applicable to a sinusoidal PD and must be modified for other PD characteristics.) If modulation is sinusoidal with modulating frequency ω_m and peak frequency deviation $\Delta\omega$, the deviation limit is found to be [5.11]

$$\Delta\omega = \begin{cases} K & \text{type 1 PLL; } \omega_m \ll K \\ \dfrac{\omega_n^2}{\omega_m} & \text{type 2 PLL; } \omega_m \ll \omega_n \end{cases} \tag{5.22}$$

Unlock Behavior Detailed behavior [5.12] of the loop at the unlock threshold is rather curious. If modulation is sinusoidal, the PD output remains nearly sinusoidal for any deviation up to the lock limit. An infinitesimal increase beyond the limit causes a drastic change in the PD output. For a first-order loop (Fig. 5.18) large spikes suddenly appear upon unlock. Each spike represents the slip of one cycle. Slip spikes occur only while the instantaneous deviation is beyond the lock limit, and the first-order loop relocks immediately when the instantaneous input deviation returns within the lock bounds. A single spike appears for each modulation peak if the overmodulation is slight, and additional spikes appear in bursts as the overmodulation is increased.

The unlock behavior of a type 2 PLL is quite different. Phase-detector output is sinusoidal right up to the lock limit, but an infinitesimal increase of deviation

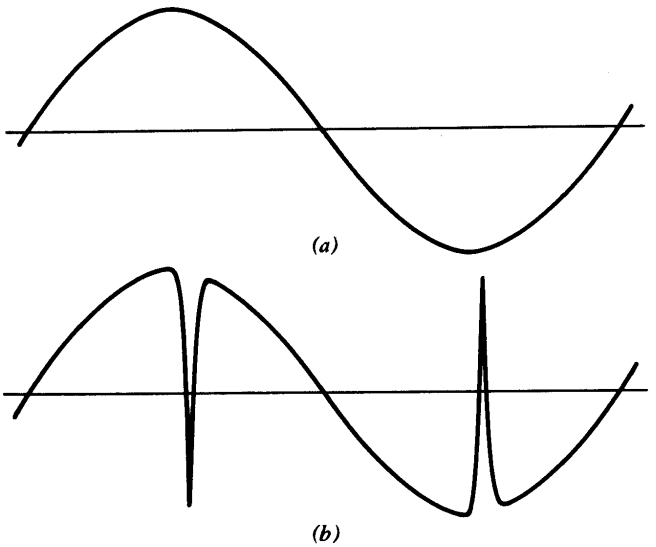


Figure 5.18 Waveforms at the phase detector output of a first-order PLL subjected to sinusoidal angle modulation: (*a*) modulation peak deviation within the lock limits; (*b*) modulation peak deviation slightly beyond the lock limits.

causes the loop to go completely out of lock and stay out; only a beat note appears at the PD output. The loop does not relock until peak deviation is reduced below the lock limit. It is impossible to adjust the deviation to obtain the single spikes of Fig. 5.18.

Why should a type 2 loop behave so differently from a first-order loop? Insight is gained upon recognizing that the peak phase error does not occur at the peak of the frequency-modulation cycle (as in a first-order loop). Instead, peak phase error coincides with maximum rate of change of frequency, which corresponds to zero instantaneous-frequency deviation for sinusoidal modulation. In fact, the sinusoidal unlock criterion (5.22) can be shown to be the same as $\Lambda = \omega_n^2$, which is the sweep-rate limit (5.18) for a type 2 PLL.

Why does the loop not relock as soon as the modulation cycle has passed through the region of excessive frequency rate? After all, the first-order loop relocks as soon as the region of excessive frequency deviation has been passed. Consider that the signal and the VCO are in frequency agreement only at the instant of maximum rate of change of frequency modulation (occurring at zero deviation of the frequency modulation) and the loop cannot lock there if the rate is excessive. At other points in the cycle, the rate is lockable, but the frequency difference between signal and VCO prevents quick relocking.

Modulation-Tracking PLLs: Gaussian Modulation Modulation limits have also been worked out for frequency modulation by a Gaussian message [5.11] with a baseband spectrum flat from DC to a cutoff frequency B_m Hz and rms frequency deviation σ_f Hz. Gaussian signals have unbounded peaks so there is some small probability that the loop will occasionally slip a cycle no matter how small the rms deviation may be. As an engineering tool, invoke the concept of *crest factor* and give it the symbol γ. Choose γ such that the magnitude of the instantaneous deviation is less than $\gamma\sigma_f$ almost all the time. A value of $\gamma = 3.5$ has been found as a good empirical fit to laboratory observations of modulation unlock. Using these concepts, the lock limits for Gaussian modulation are found to be [5.11]

$$\sigma_f = \begin{cases} \dfrac{K}{2\pi\gamma} & \text{type 1 PLL; } B_m \ll K/2\pi \\ \dfrac{\sqrt{3}\,\omega_n^2}{4\pi^2\gamma B_m} & \text{type 2 PLL; } B_m \ll \omega_n/2\pi \end{cases} \tag{5.23}$$

REFERENCES

5.1 R. C. Tausworthe, "Improvements in Deep-Space Tracking by Use of Third-Order Loops," *JPL Q. Tech. Rev.* ***2***, 96–106, July 1971.

5.2 R. C. Tausworthe and R. B. Crow, "Improvements in Deep-Space Tracking by Use of Third-Order Loops," *IEEE Int. Conf. Commun.*, 1972, pp. 577–583.

5.3 P. H. Lewis and W. E. Weingarten, "A Comparison of Second, Third, and Fourth Order Phase-Locked Loops," *IEEE Trans. Aerosp. & Electron. Syst.* ***AES-3***, 720–727, July 1967.

5.4 H. Meyr and G. Ascheid, *Synchronization in Digital Communications*, Wiley, New York, 1990, Sec. 2.5.

5.5 S. L. Goldman, "Jerk Response of a Third-Order Phase-Lock Loop," *IEEE Trans. Aerosp. & Electron. Syst.* ***AES-12***, 293–295, Mar. 1976.

5.6 E. T. Tsui and R. Y. Ibaraki, "Third-Order Loop Filter Design for Acceleration-Rate," *IEEE Trans. Aerosp. & Electron. Syst.* ***AES-13***, 200–204, Mar. 1977.

5.7 S. C. Gupta, "Transient Analysis of a Phase-Locked Loop Optimized for a Frequency Ramp Input," *IEEE Trans. Space Electron. & Telem.* ***SET-10***, 79–83, June 1964.

5.8 A. J. Viterbi, *Acquisition and Tracking Behavior of Phase-Locked Loops*, External Publ. 673, Jet Propulsion Laboratory, Pasadena, CA, July 1959.

5.9 A. J. Viterbi, *Principles of Coherent Communication*, McGraw-Hill, New York, 1966, Chap. 3.

5.10 A. Blanchard, *Phase-Locked Loops*, Wiley, New York, 1976, Sec. 10.2.1.

5.11 F. M. Gardner and J. F. Heck, "Angle Modulation Limits of a Noise-Free Phase Lock Loop," *IEEE Trans. Commun.* ***COM-26***, 1129–1136, Aug. 1978.

5.12 F. M. Gardner and J. F. Heck, "Phaselock Loop Cycle Slipping Caused by Excessive Angle Modulation," *IEEE Trans. Commun.* ***COM-26***, 1307–1309, Aug. 1978.

CHAPTER 6

EFFECTS OF ADDITIVE NOISE

Its ability to cope with high levels of noise is a major strength of a phaselock loop. This chapter scrutinizes the effects of additive stationary Gaussian noise. Although white noise is the most important example of additive noise and accordingly given the most attention, techniques for analyzing the effects of colored noise are also included. Additive noise causes tracking errors; small amounts of noise cause small errors and large amounts of noise cause large errors. Performance in small noise is amenable to linear analysis via transfer functions, but large noise drives a PLL into nonlinear operation for which transfer functions do not apply; more onerous nonlinear methods are required instead. Both noise regimes are reported in the following pages.

6.1 LINEAR OPERATION

There are two parts to the linear analysis: First, a noise model of the phase detector is developed, and then that model is placed into the PLL feedback loop. The analysis leads to the invaluable concepts of noise bandwidth and of signal-to-noise ratio in the loop.

6.1.1 Noise Model of a Phase Detector

Consider the phase detector to be a perfect multiplier to which is applied two inputs, denoted $v_i(t)$ and $v_o(t)$, each with dimensions of volts. Its output is the

Phaselock Techniques, Third Edition, by Floyd M. Gardner

product $K_m v_i v_o$, where K_m is a constant with dimensions of (volts)$^{-1}$. [**Comment**: Phase detectors are frequently modeled as multipliers, partly for analytical convenience and partly because many practical phase detectors are good approximations to multipliers. Moreover, multiplier-type phase detectors are the best choice in the presence of large additive noise. See Chapter 10.]

Inputs to PD One input to the multiplier $v_i(t)$ consists of the sum of a sinusoidal signal plus additive noise $n(t)$ that is real (as opposed to complex), stationary, Gaussian, bandpass, and zero mean:

$$v_i(t) = V_s \sin(\omega_i t + \theta_i) + n(t) \tag{6.1}$$

The other input to the multiplier comes from the VCO and has the form

$$v_o(t) = V_o \cos(\omega_i t + \theta_o) \tag{6.2}$$

[**Comments**: (1) Note that v_i and v_o are really 90° out of phase with one another; the input signal has been written as a sine and the VCO voltage has been written as a cosine. The two phases θ_i and θ_o are based on these quadrature references. It is typical of multiplier-type phase detectors that the VCO locks in quadrature to the incoming signal, so the notation is arranged in anticipation of that fact. (2) Observe that equation (6.2) describing the VCO output has the same frequency ω_i as the input signal in (6.1). Equality of frequencies implies that the loop is phaselocked, a necessary condition for applying linear analysis.]

For purposes of this chapter the input phase θ_i is assumed to be time invariant. Treatment of θ_o is less straightforward. Temporarily, assume θ_o to be time invariant, but that condition clearly does not occur in reality. Noise accompanying the signal causes the VCO phase to fluctuate; determining the statistics of those fluctuations is the objective of the linear analysis. To proceed, assume a fictitious open-loop condition such that noise does not reach the VCO. In essence, attention is restricted to the phase detector alone for the first part of the analysis. Later, the loop is closed and time-dependent θ_o is admitted.

Phase-Detector Products Bandpass input noise $n(t)$ can be decomposed into two quadrature, independent components [6.1, Sec. 8-5] in the form

$$n(t) = n_c(t) \cos \omega_i t - n_s(t) \sin \omega_i t \tag{6.3}$$

whereupon the output of the multiplier is found to be

$$\begin{aligned} v_d(t) &= K_m v_i(t) v_o(t) \\ &= \tfrac{1}{2} K_m V_s V_o \sin(\theta_i - \theta_o) + \tfrac{1}{2} K_m n_c V_o \cos\theta_o + \tfrac{1}{2} K_m n_s V_o \sin\theta_o \\ &\quad + \tfrac{1}{2} K_m V_s V_o \sin(2\omega_i t + \theta_i + \theta_o) + \tfrac{1}{2} K_m n_c V_o \cos(2\omega_i t + \theta_o) \\ &\quad - \tfrac{1}{2} K_m n_s V_o \sin(2\omega_i t + \theta_o) \end{aligned} \tag{6.4}$$

The multiplier product consists of three low-frequency terms and three terms at $2\omega_i$, twice the input frequency. Our interest is in the difference-frequency terms, so the double-frequency *ripple* terms are discarded for this analysis. In practice, filtering or other expedients must be applied to suppress the double-frequency ripple. Otherwise ignored in this chapter, ripple is a serious disturbance in many applications and substantial effort is often needed for its suppression. Chapter 10 contains further information on phase-detector ripple.

Now define phase-detector gain,

$$K_d = \frac{K_m V_s V_o}{2} \tag{6.5}$$

so the multiplier output, after the ripple has been discarded, becomes

$$v_d = K_d \sin(\theta_i - \theta_o) + \frac{n_c K_d}{V_s} \cos\theta_o + \frac{n_s K_d}{V_s} \sin\theta_o \tag{6.6}$$

Equivalent Noise Next define $n'(t)$ as

$$n'(t) = \frac{n_c}{V_s} \cos\theta_o + \frac{n_s}{V_s} \sin\theta_o \tag{6.7}$$

which is a dimensionless quantity, as opposed to $n(t)$, which has dimensions of volts. The output of the phase detector is thereby simplified to

$$v_d = K_d[\sin(\theta_i - \theta_o) + n'(t)] \tag{6.8}$$

An exact nonlinear equivalent circuit of the phase detector is shown in Fig. 6.1. No linearizing approximation has yet been applied. The phase-detector output consists of the linear superposition of a signal term $K_d \sin(\theta_i - \theta_o)$ and a noise term $K_d n'(t)$.

Note from (6.5) that K_d is proportional to the input-signal level. Therefore, if the input-signal amplitude varies, K_d and all loop parameters that depend on loop gain also vary. (Phase-detector gain K_d is also proportional to the VCO amplitude

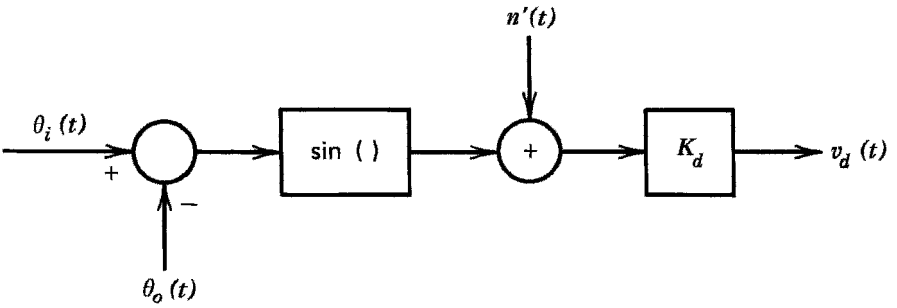


Figure 6.1 Nonlinear noise equivalent circuit of a phase detector.

V_o in the ideal-multiplier model. Dependence on VCO amplitude is less critical than dependence on signal amplitude, partly because VCO amplitude tends to be much more constant than that of the signal and also because many practical phase detectors effectively clip the VCO input, thereby removing variability of VCO amplitude.)

Properties of n′(t) Let's develop some of the statistical properties of $n'(t)$. From the bandpass zero-mean definition of $n(t)$, it can be concluded that n' also has zero mean. (Explicit dependence on time is dropped for notational convenience.) If θ_o is assumed to be time invariant, although arbitrary, the variance of n' is

$$\sigma_{n'}^2 = \frac{1}{V_s^2}\{E[n_c^2]\cos^2\theta_o + E[n_s^2]\sin^2\theta_o + 2E[n_c n_s]\sin\theta_o\cos\theta_o\} \tag{6.9}$$

where $E[\cdot]$ indicates statistical expectation. As is well known [6.1, Sec. 8-5], bandpass Gaussian noise has the properties that $E[n_c^2] = E[n_s^2] = E[n^2] = \sigma_n^2$ and $E[n_c n_s] = 0$. Moreover, $\cos^2\theta_o + \sin^2\theta_o = 1$, so

$$\sigma_{n'}^2 = \frac{\sigma_n^2}{V_s^2} \tag{6.10}$$

The intensity of the equivalent noise is rotationally invariant; it does not depend on the value of θ_o.

Now revise the model for θ_o; consider it to be a random variable, uniformly distributed over $[0, 2\pi)$ and independent of the noise. The spectrum of n' is obtained by finding its autocorrelation function and taking the Fourier transform. From (6.7), the autocorrelation of n' is

$$\begin{aligned} E[n'(t_1)n'(t_2)] = \frac{1}{V_s^2}\{&E[n_c(t_1)n_c(t_2)]E[\cos^2\theta_o] + E[n_s(t_1)n_s(t_2)]E[\sin^2\theta_o] \\ &+ (E[n_c(t_1)n_s(t_2)] + E[n_s(t_1)n_c(t_2)])E[\sin\theta_o\cos\theta_o]\} \end{aligned} \tag{6.11}$$

Expectations of the trigonometric factors are $E[\cos^2\theta_o] = 0.5 = E[\sin^2\theta_o]$ and $E[\sin\theta_o\cos\theta_o] = 0$. Since the noises are stationary, the noise autocorrelations depend only on the time difference $\tau = t_1 - t_2$. Denoting autocorrelation by $R(\tau)$, the autocorrelation of n' is found to be

$$\begin{aligned} R_{n'}(\tau) &= \frac{1}{2V_s^2}[R_{nc}(\tau) + R_{ns}(\tau)] \\ &= \frac{1}{V_s^2}R_{nc}(\tau) \end{aligned} \tag{6.12}$$

since $R_{nc} = R_{ns}$ [6.1, p. 162].

Noise Spectrum The *two-sided* spectrum of n' is

$$S_{n'}(f) = \frac{S_{nc}(f)}{V_s^2} = \frac{S_{ns}(f)}{V_s^2} \tag{6.13}$$

where $S_x(f)$ is the Fourier transform of $R_x(\tau)$. To obtain the two-sided baseband spectra S_{nc} or S_{ns} from the passband spectrum $S_n(f)$, slide the negative-frequency portion of $S_n(f)$ to the right by an amount $f_i = \omega_i/2\pi$, slide the positive-frequency portion to the left by the same amount $f_i = \omega_i/2\pi$ and add the translated portions. In equation form,

$$S_{nc}(f) = S_{ns}(f) = u(f + f_i)S_n(f + f_i) + u(f_i - f)S_n(f - f_i) \tag{6.14}$$

The frequency-domain unit step $u(f)$ selects the positive-frequency portion of the spectrum $S_n(f)$; it is defined as $u(f) = 0$ if $f < 0$ and $u(f) = 1$ if $f \geq 0$. The negative-frequency portion of the spectrum is selected by $u(-f)$.

One- and Two-Sided Spectra The formal definition of spectral density as the Fourier transform of the autocorrelation function $R(\tau)$ produces a two-sided spectrum $S(f)$ that is defined for both positive and negative frequencies. Theoreticians prefer a two-sided spectrum because of its orderly mathematical properties. But over many years, electronic engineering practice in general and the PLL literature in particular have used one-sided spectra instead. Provided that the time-domain signal or noise underlying the spectrum is real, the one-sided spectrum is related to the two-sided spectrum by

$$W(f) = 2S(f), \qquad f \geq 0 \tag{6.15}$$

Negative frequencies are deemed nonexistent in one-sided spectra.

A one-sided definition of spectrum can be employed if the underlying signal or noise is real, because then the two-sided spectrum is real and even-symmetric. A one-sided spectrum cannot be used for complex signals and it is awkward in the presence of spectral folding or aliasing, even with real signals. Nonetheless, since most signals in the past have been real, much of the existing PLL literature has traditionally used one-sided spectra; this book follows that tradition.

[**Notation**: Most textbooks on noise and stochastic processes, such as [6.1], use $S(f)$ as the symbol for a two-sided spectrum. No similar consensus exists for notation for one-sided spectra. This book uses $W(f)$.]

[**Dimensions**: If the underlying signal or noise is measured in volts, the dimensions of $R(\tau)$ are $(\text{volts})^2$ and the dimensions of $S(f)$ or $W(f)$ are V^2/Hz. If the underlying signal or noise, such as $n'(t)$, is dimensionless, $R_{n'}(\tau)$ is also dimensionless and the dimensions of $S_{n'}(f)$ or $W_{n'}(f)$ are Hz^{-1}.]

The one-sided baseband spectra $W_{nc}(f) = W_{ns}(f)$ are obtained from the one-sided passband spectrum $W_n(f)$ by the relation

$$W_{nc}(f) = W_{ns}(f) = [W_n(f_i + f) + W_n(f_i - f)], \qquad f \geq 0 \tag{6.16}$$

For the special case of white noise, in which $W_n(f) = N_0$ V^2/Hz, the spectrum of n' is

$$W_{n'}(f) = \frac{2N_0}{V_s^2} \tag{6.17}$$

Equivalent Phase Jitter Noise output (6.8) of the phase detector is $K_d n'(t)$. Such an output could be caused by the additive noise, as described, or it could be caused by an input phase disturbance $\theta_{ni}(t)$ such that $\sin\theta_{ni}(t) = n'(t)$. If θ_{ni} is small enough, the sinusoidal nonlinearity can be neglected and the variance of the fictitious input phase disturbance is $E[\theta_{ni}^2] = \sigma_{\theta ni}^2 = \sigma_{n'}^2 = \sigma_n^2/V_s^2$. The input signal-to-noise ratio is $\text{SNR}_i = V_s^2/2\sigma_n^2$, so the input phase variance is approximated by $\sigma_{\theta ni}^2 = 1/(2\text{SNR}_i)$ rad^2. This is the phase jitter to be expected from measurement of the phase difference between a clean signal and one corrupted by noise under conditions of a large signal-to-noise ratio. This relation is used later in establishing a definition for the signal-to-noise ratio of the phaselock loop. In light of the duality between n' and θ_{ni}, it is sometimes useful to consider n' to be an angle disturbance with units of radians, a dimensionless quantity. Then the spectral density $W_{n'}$ can be considered to have units of rad^2/Hz.

Linearization Except for the definition of θ_{ni} as a phase modulation equivalent to the additive input noise (a side issue in the overall development), no linearizing approximation has been made in the results obtained so far. The assumption of time-invariant θ_o has implied an open loop; if the loop were closed, the noise would angle-modulate the VCO, causing θ_o to fluctuate in random fashion. The phase detector is nonlinear, so the fed-back fluctuations intermodulate with the incoming signal plus noise. Any simple analysis is blocked by this nonlinearity; to apply transfer-function analysis requires simplifying approximations.

The most common simplification assumes that noise is small enough that phase error $(\theta_i - \theta_o)$ remains small and that the PD can be regarded as linear. Under these conditions, the intermodulation may be neglected and a linearized phaselock loop with simple, additive noise $n'(t)$, may be considered, as shown in Fig. 6.2. Transfer-function analysis is applicable to this linearized loop.

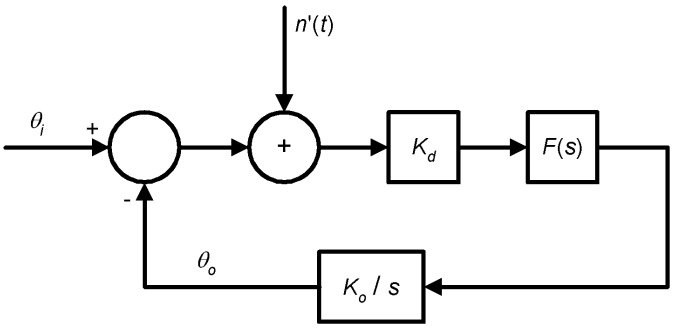


Figure 6.2 Block diagram of a linearized PLL.

6.1.2 Noise Transfer Function

Figure 6.2 shows that $n'(t)$ in a linearized loop is directly additive to the input-signal phase θ_i. The system closed-loop transfer function $H(s)$ relating θ_o to θ_i was derived in Chapter 2. Because n' is additive to θ_i, the same transfer function relates θ_o to $n'(t)$ by the principle of superposition. Spectral density $W_{\theta no}$ of the fluctuations of VCO phase is related to the spectrum of n' by

$$\begin{aligned} W_{\theta no}(f) &= W_{n'}(f)|H(f)|^2 \\ &= \frac{1}{V_s^2}[W_n(f_i - f) + W_n(f_i + f)]|H(f)|^2, \qquad f \geq 0 \end{aligned} \tag{6.18}$$

[**Notation**: The symbol $H(f)$ is a shorthand version of $H(s)|_{s=j2\pi f}$. Purists object, rightly, that this expedient constitutes abuse of notation and should be condemned. Nonetheless, the abbreviation is well accepted by engineers, it is convenient, and it is employed throughout this book.]

The variance of the VCO phase is the integral of (6.18):

$$\sigma_{\theta no}^2 = \int_0^\infty W_{n'}(f)|H(f)|^2\, df \qquad \text{rad}^2 \tag{6.19}$$

6.1.3 Noise Bandwidth

In general, the integral (6.19) is cumbersome to evaluate. However, it simplifies radically in the important special case of white input noise. If $W_n(f) = N_0$ V^2/Hz for all frequencies of interest, then (6.19) reduces to

$$\sigma_{\theta no}^2 = \frac{2N_0}{V_s^2}\int_0^\infty |H(f)|^2\, df \qquad \text{rad}^2 \tag{6.20}$$

The integral in (6.20) defines the *noise bandwidth* B_L of the loop as

$$B_L = \int_0^\infty |H(f)|^2\, df \qquad \text{Hz} \tag{6.21}$$

A similar integral is applicable to time-discrete (digital) PLLs:

$$2B_L t_s = \frac{1}{2\pi j}\int_{|z|=1} H(z)H(1/z)\frac{dz}{z} = \frac{1}{2\pi}\int_{-\pi}^{\pi} H(e^{j\omega t_s})H(e^{-j\omega t_s})\, d\omega t_s \tag{6.22}$$

Noise bandwidth is a meaningful concept only if the loop is stable.

The integrals of (6.21) and (6.22) have been evaluated explicitly for several PLLs of importance, and the resulting expressions for noise bandwidth are shown in Table 6.1 (refer to Chapters 2 and 4 for definitions of notation). Notice that B_L has dimensions of hertz, despite the fact that the other parameters with dimensions

TABLE 6.1 Noise Bandwidths of Common PLLs

PLL Description	Equations[a]	Noise Bandwidth, B_L (Hz)
Type 1, order 1	(2.32)	$K/4$
Type 1, order 2 (Lag filter)	(2.33)	$K/4$
Type 2, order 2	(2.16)	$\frac{\omega_n}{2}\left(\zeta + \frac{1}{4\zeta}\right)$
Type 2, order 2	(2.19)	$\frac{K}{4}\left(1 + \frac{1}{K\tau_2}\right) = \frac{K}{4}\left(1 + \frac{1}{4\zeta^2}\right)$
Type 2, order 2	(2.14), (2.17)	$\frac{K}{4}\left(1 + \frac{K_2}{KK_1}\right)$
Type 2, order 3	(2.39)	$\frac{K}{4}\frac{1 + 1/K\tau_2}{1 - 1/b}$
Type 3, order 3	(2.45)	$\frac{K}{4}\frac{1 + \frac{K_2}{KK_1} + \frac{K_1K_3}{KK_2^2}}{1 - \frac{K_1K_3}{KK_2^2}}$
Type 3, order 3	(3B.2)	$\frac{K}{4}\frac{1 + \frac{1}{K}\left(\frac{1}{\tau_A} + \frac{1}{\tau_B} - \frac{1}{\tau_A + \tau_B}\right)}{1 - \frac{1}{K(\tau_A + \tau_B)}}$
DPLL, type 2, $D = 1$	(4.17)	$\frac{\kappa}{4t_s}\frac{1 + \frac{\kappa_2}{\kappa} - \frac{\kappa_2}{2}(3 - \kappa_2)}{1 - \kappa_2 - \frac{\kappa}{4}(2 - \kappa_2 + \kappa_2^2)}$

[a]Equation numbers for transfer functions of each PLL.

of frequency, such as K and ω_n, are given in radians per second. The table exposes several notable features:

- Loop gain K (analog PLLs) or κ (digital PLLs) plays a central role in establishing noise bandwidth.
- Adding a simple lag filter to a first-order PLL does not affect the noise bandwidth.
- Adding a third pole to a second-order type 2 PLL has little effect for practical values of the third-pole parameter b (e.g., $b \geq 9$).
- The noise bandwidth formula for a type 2 digital PLL approaches $\kappa(1 + \kappa_2/\kappa)/4t_s$ if $\kappa_2 < \kappa \ll 1$. This form agrees with the time-continuous approximation of Section 4.7. Figure 6.3 shows an example of the ratio of the noise bandwidth given by the formula in the table to the approximation as a function of the loop gain κ. Actual noise bandwidth always exceeds the approximation and exceeds it greatly for large κ, but the approximation is quite reasonable for κ as large as 0.2.

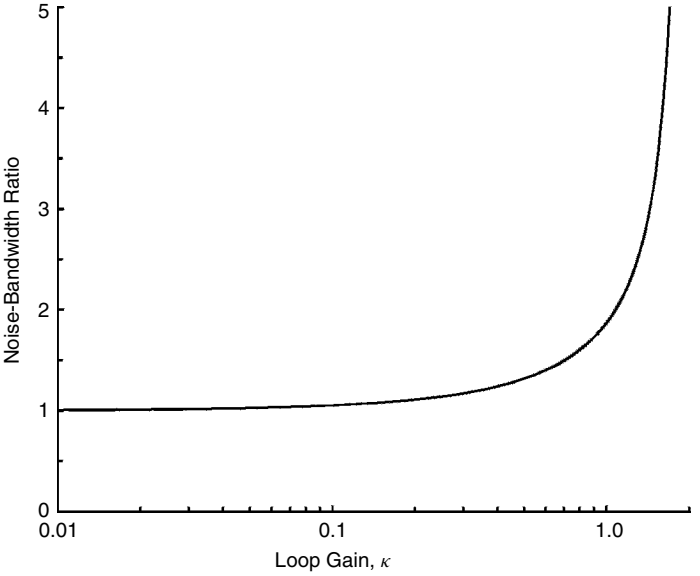


Figure 6.3 Comparison of the noise bandwidth of the type 2 digital PLL of Table 6.1 to the time-continuous approximation of Section 4.7 ($D = 1$, $\kappa_2 = \kappa/2$, equivalent to $\zeta = 0.707$).

6.1.4 Signal-to-Noise Ratio in a PLL

Signal-to-noise ratio (SNR) is a useful engineering concept and it is often helpful to define one for the phaselock loop. Definition of the input signal-to-noise ratio SNR_i is straightforward; it is merely the ratio of input signal power to input noise power as delivered to the phase detector. By contrast, there is no "signal" internal to the PLL; for example, normal tracking is about a null output of the phase detector. Also, loop "noise" is a function of the location in the loop at which the measurements are performed—there is no unique definition. As a result, the loop signal-to-noise ratio SNR_L must be defined arbitrarily and is a fictitious quantity without firm physical meaning.

In this book SNR_L is defined for white-noise input by analogy between phase jitters. If the input noise applied to the loop is white so that (6.20) is applicable, the phase variance of the VCO is given by the simple formula

$$\sigma_{\theta no}^2 = \frac{2N_0 B_L}{V_s^2} = \frac{W_0 B_L}{P_s} \qquad \text{rad}^2 \tag{6.23}$$

where P_s is the input signal power in watts and W_0 is the input noise power spectral density in W/Hz.

The input phase jitter for large SNR_i was determined above to be

$$\sigma_{\theta ni}^2 = \frac{1}{2\text{SNR}_i} \qquad \text{rad}^2 \tag{6.24}$$

By analogy, define SNR_L as

$$\sigma_{\theta no}^2 = \frac{1}{2\mathrm{SNR}_L} \qquad \mathrm{rad}^2 \tag{6.25}$$

Using (6.23) as the definition of VCO phase jitter caused by input additive white noise produces

$$\mathrm{SNR}_L = \frac{P_s}{2B_L W_0} = \frac{V_s^2/2}{2B_L N_0} \tag{6.26}$$

Equation (6.26) is taken as the arbitrary definition of the loop signal-to-noise ratio for all values of SNR_L, large or small. However, (6.23) and (6.24), which were used in generating (6.26), are valid only for large SNR_L. Nonlinear operation (at small SNR_L) is considered later in the chapter. [**Note**: Another common definition of loop signal-to-noise ratio is P_s/W_0B_L; it and (6.26) are equally valid and equally arbitrary. Care must be taken in reading phaselock literature to ascertain which definition is used by the author.]

Despite its arbitrary definition, SNR_L can be endowed with a useful conceptual meaning. Consider that the PLL acts as a bandpass filter on the received signal. The filter is centered at the frequency of the signal and has a noise bandwidth of B_L on each side of center for a total equivalent input bandwidth of $2B_L$. Thus, for white noise of spectral density W_0, the total noise power that enters the loop is $2B_LW_0$ watts. The ratio of signal power to this value of noise power is definition (6.26) of SNR_L.

Like all bandwidths in this book, B_L is a *one-sided* bandwidth. One might reasonably and properly call $2B_L$ the *double sideband* noise bandwidth of the loop, but $2B_L$ is *one-sided*. The distressing, though common practice of describing $2B_L$ as the *two-sided* noise bandwidth is wrong, especially in conjunction with one-sided noise spectra.

6.1.5 Optimality

Equation (6.23) for the phase variance in a PLL caused by additive white noise can be shown to be equal to the Cramér–Rao lower bound (CRB). That is, no unbiased phase estimator with the same signal, noise, and bandwidth can deliver a smaller variance. For further information on the CRB, see [6.22]. The variance of a linear-operating PLL meets the CRB but the PLL performance deteriorates when it is driven into nonlinear operation.

6.2 NONLINEAR OPERATION

A linear approximation led to a handful of simple equations, (6.19) to (6.26), that are sufficient for analysis of PLLs at reasonably large SNR_L. The approximation is satisfactory for the vast majority of present-day applications of PLLs. However, there remain several applications (such as narrowband phaselocked receivers for

deep-space missions) in which performance at low SNR_L is critical. The linear approximation is inadequate at low SNR_L; nonlinear methods are needed.

Analysis of a nonlinear PLL at low SNR_L is far from simple. The problem has been attacked by some of the brightest minds in our profession (see the references) with imposing displays of mathematical virtuosity. Unfortunately, only the simplest of PLL circuits and signal formats have been amenable to the nonlinear analysis; in many practical situations, engineers still must rely upon time-consuming simulations or unproven empirical extrapolations of known results. This section provides a brief summary of the nonlinear problem; the literature is far more extensive.

6.2.1 Observed Behavior

When operation of a PLL is monitored in the laboratory, the phase jitter of the VCO is observed to be more than is predicted by (6.23) and (6.25) as SNR_L is reduced below about 4 dB (see curve *a* in Fig. 6.4). The discrepancy should

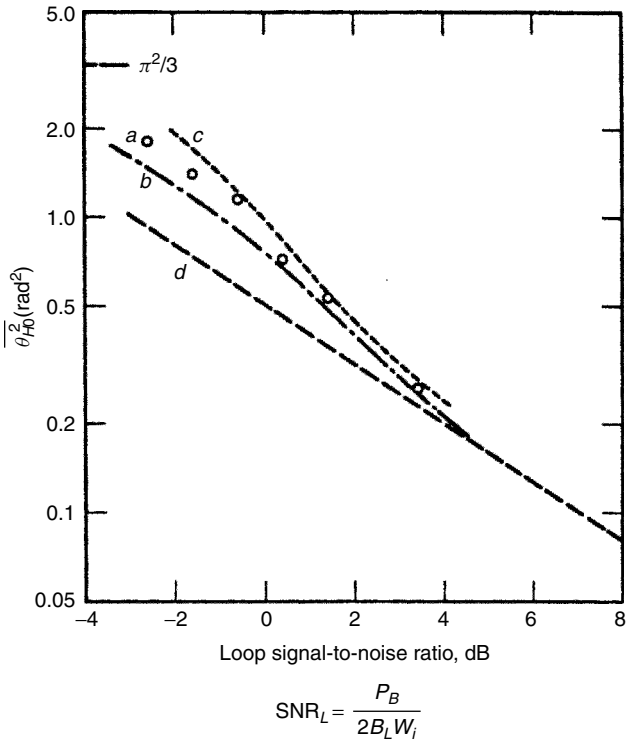


Figure 6.4 Phase-error variance. (*a*) Experimental data [6.4] for a second-order type 2 PLL with $\zeta = 0.707$. (*b*) Exact nonlinear analysis [6.2] for a first-order PLL. (*c*) Approximate nonlinear analysis [6.3, 6.8] for a type 2 PLL, $\zeta = 0.707$. (*d*) Linear approximation, eq. (6.25).

cause no surprise, since the linear analysis was based on an assumption of small phase error in the loop, but the actual error at low SNR_L is not small. The linear analysis fails when its underlying assumption is violated.

Another phenomenon appears at low SNR_L; the oscillator phase occasionally slips one or more cycles with respect to the signal. A large noise event, in effect, knocks the loop temporarily out of lock and tracking returns to equilibrium n cycles away from its original condition ($n = \pm 1, \pm 2$, etc.). Frequency of slipping is a very steep function of SNR_L, as shown in Fig. 6.5. Cycle slips are particularly destructive to operations in which every cycle matters, such as Doppler velocity

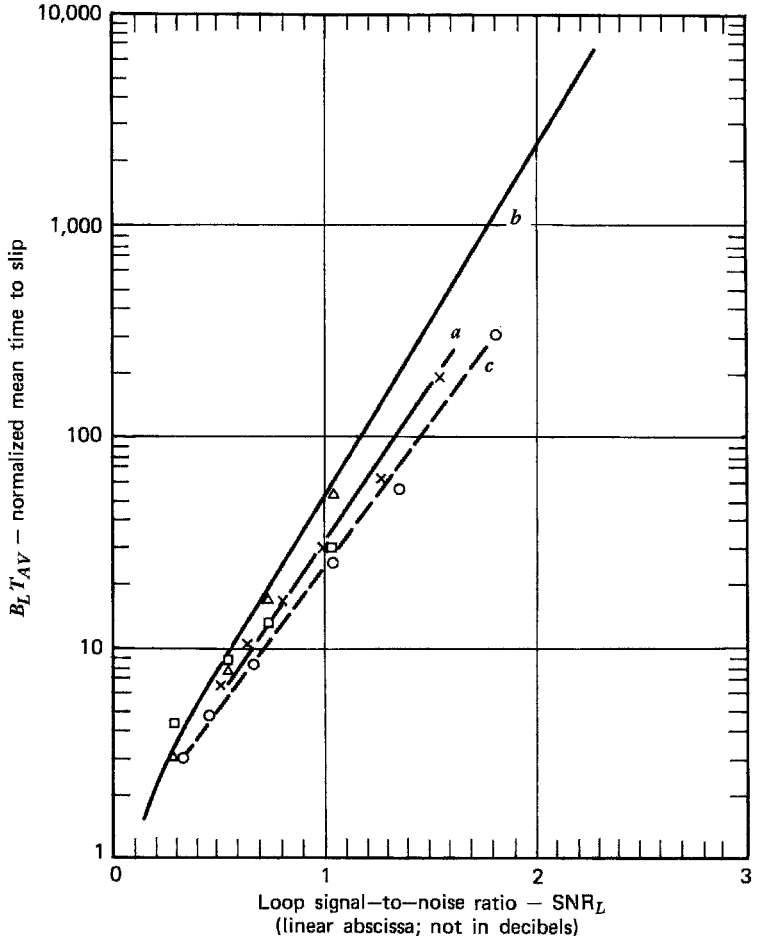


Figure 6.5 Mean time to first slip. (*a*) Experimental data [6.4] for a second-order type 2 PLL with $\zeta = 0.707$, x-marker. (*b*) Exact result for a first-order PLL, eq. (6.30). (*c*) Simulation results [6.10] for a second-order type 2 PLL with $\zeta = 0.707$, circle-marker. (*d*) Simulation results [6.9] for a second-order type 2 PLL with $\zeta = 1.4$, triangle-marker. (*e*) Simulation results [6.9] for a second-order type 2 PLL with $\zeta = 0.35$, square-marker.

measurements or recovery of digital clock timing. Slips are also important to an understanding of phaselock FM demodulators (Chapter 16).

A third phenomenon emerges if SNR_L is reduced sufficiently; the loop drops out of lock and stays out. Control of the VCO is lost; its frequency wanders off from the signal frequency. Although both phenomena have often been lumped together under the same name, drop lock is qualitatively distinct from repeated cycle slips. The drop-lock SNR_L typically is in the vicinity of 0 dB, although extreme care with loop components may extend the dropout point 1 to 2 dB lower. An observer gets the impression that the loop at low SNR_L is staggering (because of repeated cycle slips), and eventually everything seems to collapse as lock is lost completely. Reacquisition of lock is nearly impossible after dropout unless SNR_L is raised substantially (to about 3 to 6 dB).

Experience of the drop-lock phenomenon led to the concept of a noise threshold of the PLL; that is, the loop falls out of lock if SNR_L is below the "threshold" level. Later, it gradually became apparent that well-built loops could hold lock below the analytical threshold, whereupon it was realized that the predicted threshold was a feature of the approximations in the analysis and not of a physical PLL.

Presently accepted nonlinear analyses do not reveal a noise threshold (see the next section). Current opinion holds that drop lock arises from a complicated nonlinear interaction between the noise-caused phase jitter and small biases, drifts, and DC offsets arising from imperfections of the loop components, especially the phase detector (see Chapter 10). The imperfections are circuit specific and are usually unpredictable, even after the circuit is built. In general, drop lock would be very difficult to analyze and any analysis would be difficult to apply.

Analytical difficulties aside, this viewpoint sees drop lock as a technological problem and not inherent to the PLL as such. If the viewpoint is correct, drop lock can, in principle, be pushed to lower signal-to-noise ratios by improvements in the loop components. Nonetheless, measurement of the drop-lock SNR_L provides a valuable indication of the quality of implementation of a PLL, even though the information can be evaluated only in comparison to other PLL implementations; there is no theoretical bound for comparison. By contrast, analyses described in the sequel predict cycle-slipping characteristics very well, at least for the simpler PLLs to which the analyses apply. Moreover, the cycle-slip predictions are for ideal loops, so no relief can be obtained from improvements in loop components.

6.2.2 Nonlinear Analysis of Phase Error

In a linear system, a Gaussian input gives rise to a Gaussian output. Therefore, earlier assumptions of linear operation of the loop and of Gaussian input noise imply that the VCO phase jitter would be Gaussian. A Gaussian process is defined completely by its autocorrelation function or, equivalently, its spectral density as derived in (6.18). Variance is readily found from either one. Response of a nonlinear system to a Gaussian stimulus is generally non-Gaussian and the second-order statistics do not define the process completely. Nonlinear analysis

of a PLL has been concerned with deriving the non-Gaussian probability density function (pdf) of the phase error, computing the phase variance from the pdf, and investigating the statistics of cycle slipping.

The analytical simplicity of transfer functions is lost in a nonlinear system. Analysis of a nonlinear system is much more difficult and demands a higher level of mathematical sophistication than linear analysis. The treatment here presents a summary of the results of the various nonlinear analyses, a treatment that is more than sufficient for most engineering purposes. (In fact, linear analysis will suffice for the great bulk of engineering design problems.) References are provided for those who are interested in the detailed mathematics.

6.2.3 Probability Density and Variance

Viterbi's groundbreaking exact analysis [6.2, 6.21] of the first-order loop has provided much insight and many useful tools for understanding nonlinear operation. First, one must recognize that cycle slips cause the phase error $\theta_e = (\theta_i - \theta_o)$ to be a growing quantity and ultimately unbounded; that is, phase error is nonstationary in the presence of cycle slips, so the well-honed tools of stationary analysis are not applicable directly. To avoid this problem, Viterbi defined a new phase variable,

$$\phi = (\theta_i - \theta_o) \quad \text{mod-}2\pi \qquad \text{rad} \tag{6.27}$$

so that although $(\theta_i - \theta_o)$ can take on any value from $-\infty$ to $+\infty$, the value of ϕ is bounded since the mod-2π notation in (6.27) means $\theta_i - \theta_o = \phi + 2n\pi$, where n is chosen to cause ϕ to lie in the interval $[-\pi, \pi)$.

This definition of ϕ implies that all cycles of a sine wave look alike and cannot readily be distinguished from one another. Cycle slips are neglected by this definition and they must be treated separately. Most laboratory instruments operate modulo 2π and therefore yield ϕ rather than $(\theta_i - \theta_o)$; the concept agrees well with normal practice, despite any initial impressions of peculiarity.

It turns out that ϕ is stationary (in the steady state, after any transients have died out), which allows application of stationary statistics. Denote the probability density function of ϕ as $p(\phi)$; this is found as the steady-state solution of a nonlinear, stochastic partial differential equation known as the *Fokker–Planck equation*. Bypassing details [6.2], the pdf turns out to be the Tikhonov density,

$$p(\phi) = \frac{\exp(\rho \cos \phi)}{2\pi I_0(\rho)}, \qquad |\phi| \le \pi \tag{6.28}$$

where $\rho = 2\text{SNR}_L$ and $I_0(\rho)$ is the modified Bessel function of the first kind and zero order. [**Comment**: Equation (6.28) is valid only if $E[\phi]$, the static phase error, is zero. See Chapter 5 for an explanation of static phase error. See [6.3, Chap. 9] if static phase error (loop stress) is not zero.] The density (6.28) approaches Gaussian for large SNR_L, thereby agreeing with the linear analysis. At very small SNR_L, $p(\phi)$ approaches a uniform density over $(-\pi, \pi]$, which is characteristic of the phase of random noise.

The variance of the phase error can be found by numerical evaluation of

$$\sigma_\phi^2 = \int_{-\pi}^{\pi} \phi^2 p(\phi)\, d\phi \qquad \text{rad}^2 \tag{6.29}$$

The result, phase variance reduced modulo-2π, is plotted in curve b of Fig. 6.4. The exact variance agrees with the linear analysis for large SNR_L and approaches $\pi^2/3$ rad^2 for very small SNR_L. [The variance of a random variable uniformly distributed over $(-\pi, \pi)$ is $\pi^2/3$.]

6.2.4 Cycle Slips

Knowledge of variance is useful but is insufficient by itself because cycle slipping is ignored in its computation. Statistics of cycle slips are an important attribute of PLL operation at low SNR_L, even more important than the phase variance. By means of manipulations on the Fokker–Planck equations, Viterbi [6.2] derived an expression for the average time T_{AV} between cycle slips. From an initial condition of zero phase error, T_{AV} is the average time required for the loop phase error to reach $\pm 2\pi$ for the first time. If slips occur primarily as single, isolated events, the frequency of cycle slips is $1/T_{\text{AV}}$. If slips occur in clusters, as may happen in a type 2 or higher loop, T_{AV} and the slip rate are not related simply.

For a first-order loop with zero static phase error,

$$T_{\text{AV}} = \frac{\pi^2 \rho I_0^2(\rho)}{2B_L} \qquad \text{sec} \tag{6.30}$$

which is approximated for large ρ by

$$T_{\text{AV}} \approx \frac{\pi}{4B_L} \exp(2\rho) \qquad \text{sec} \tag{6.31}$$

A plot of T_{AV} from (6.30) is shown as curve b in Fig. 6.5; the straightness shows that (6.31) is acceptable for all practical SNR_L. In addition, time between slips is exponentially distributed; the probability that the loop has not slipped T seconds after starting from zero error is

$$P(T) = \exp(-T/T_{\text{AV}}) \tag{6.32}$$

This distribution is well confirmed by computer simulations and laboratory measurements on both first- and second-order loops.

Viterbi's results apply exactly in a first-order loop with sinusoidal phase detector, zero static phase error, and additive white Gaussian noise. The first-order pdf (6.28) and variance also apply without modification if a simple lag filter of the form $F(s) = (s\tau + 1)^{-1}$ is inserted into the loop [6.4, 6.5], provided that static phase error is zero.

6.2.5 Experimental and Simulation Results

Fokker–Planck equations can be written for a type 2 PLL, but exact closed-form solutions have been unattainable. The second-order type 2 PLL is technologically the most important configuration, so there is a strong motivation to determine its statistics for low SNR_L. Experimental measurements of $p(\phi)$, σ_ϕ^2, and slip statistics were reported by Charles and Lindsey in [6.4]. Their measured variance is shown as experimental points (curve *a*) of Fig. 6.4 and cycle-slip results are shown in curve *a* of Fig. 6.5. Three significant conclusions can be drawn:

1. Phase variance found by exact nonlinear analysis of the first-order loop is in close agreement with measured variance of the type 2 loop of the same noise bandwidth for SNR_L in excess of 0 dB, that is, for any useful value of SNR_L.
2. Approximate linear analysis yields good accuracy of variance if SNR_L exceeds 5 to 6 dB.
3. Average time to first slip is shorter in a type 2 PLL than in a first-order PLL, especially for smaller values of damping factor.

Meyr and Ascheid [6.6, Chap. 6], [6.7] performed an intensive simulation study of cycle slips in second-order type 2 PLLs. They found that slips are likely to occur in extended bursts if the damping factor is less than $\zeta \approx 0.9$ and are likely to be isolated if damping exceeds 0.9. Their findings corroborate those of previous authors [6.3, 6.8–6.13], but Ascheid and Meyr provide a physical explanation for the observed behavior, an explanation that was previously lacking. In brief: Occurrence of a slip causes a disturbance of the stored voltage in the integrator of the loop filter. That stored voltage controls the average frequency of the VCO. An error in stored frequency worsens the slip behavior of the PLL, so slips tend to repeat until feedback corrects the integrator-stored voltage. Smaller damping implies larger gain in the integral path of the loop filter and consequently, greater susceptibility of the integrator to noise disturbances. The lesson for designers faced with the prospect of cycle slips is to employ somewhat larger damping (e.g., $\zeta = 1$ or greater) rather than the widely touted $\zeta = 0.707$ that had so much attention in the earlier literature.

6.2.6 Approximate Analyses

Because of the practical importance of a type 2 PLL, numerous approximate analyses have been devised [6.3–6.5, 6.8, 6.14–6.16]. These analyses generally involve clever assumptions and heroic mathematics. Among the several methods, [6.3] and [6.8] start with the Fokker–Planck equation and so yield approximations to the pdf and slip statistics as well as phase variance. They provide substantial detail in the form of charts and formulas. Their prediction of variance is shown as curve *c* of Fig. 6.4. The approximation is clearly very close to the measured results, albeit slightly pessimistic. Similar agreement is found in comparing the measured pdf against the predicted value.

The analyses of [6.3] and [6.8] predict that the phase variance will have a weak inverse dependence on damping factor ζ; that is, jitter is slightly worse for small damping given the same loop bandwidth B_L and same SNR_L. The prediction is borne out by simulation results [6.17]. However, if SNR_L exceeds unity—as it must in a useful loop—the variance spread between light damping ($\zeta = 0.35$) and a first-order loop ($\zeta = \infty$) is small and may be neglected for most purposes.

Several investigators have studied cycle slips in a second-order type 2 PLL by means of computer simulation and by measurements on physical loops in the laboratory. A summary of their published results is given in Fig. 6.5. It is evident that slipping worsens if damping is small. The first-order loop has infinite damping, so its time to first slip T_{AV} is greater than that of any type 2 loop of the same noise bandwidth. (The experimental curves suffer from statistical fluctuations due to a finite number of samples in the measurements and from ad hoc redefinitions of the meaning of "slip" [6.9]. Some caution must be exercised in applying the data.)

Predictions of T_{AV} for a second-order type 2 PLL are given by formulas developed (at great effort) by Lindsey [6.3], Lindsey and Simon [6.8], and Tausworthe [6.10, 6.11], predictions in fairly close agreement with the experiments. Since the formulas are cumbersome and since their derivation necessarily involved approximations, the practicing engineer will usually find the curves of Fig. 6.5 to be a more convenient guide to slip behavior.

Inspection of the data points of Fig. 6.5 shows a reasonably good fit to a straight line when they are plotted on a logarithmic ordinate. This means that T_{AV} is approximately exponentially dependent on SNR_L. Curves b and c of the figure represent boundary limits that encompass all the configurations that have been explored. Curve b is the exact result for a first-order loop and is described by (6.30) and (6.31). Curve c is from simulation of a type 2 PLL with damping 0.707; its level is thought to be somewhat pessimistic. The points of c are well fitted by an empirical relation (valid only for zero static-phase error)

$$B_L T_{\text{AV}} = \exp(\pi \cdot \text{SNR}_L) \tag{6.33}$$

It appears to be reasonable to use (6.31) and (6.33) as upper and lower bounds on T_{AV}. Some numbers are of interest. Let $B_L = 20$ Hz at $\text{SNR}_L = 1$ (0 dB), from which (6.33) predicts $T_{\text{AV}} = 1.16$ sec: very poor performance indeed. Use of (6.31) predicts $T_{\text{AV}} = 2.1$ sec, which is also very poor. Now consider $\text{SNR}_L = 10$ for which the lower bound on T_{AV} is predicted to be 2.2×10^{12} sec or about 70,000 years (assuming, without experimental verification, that the exponential relation can be extrapolated to large SNR). A notable later nonlinear analysis by Meyr and Ascheid [6.6, Pt. 4] is suggested for advanced students.

6.2.7 Miscellaneous Features

Effect of Loop Stress The preceding results have been presented for a loop that is unstressed by any other phase error, such as may be caused by steady

phase error or angle modulation. Chapter 5 discusses the origins of such phase errors and how they may be reduced. The presence of a static phase error (i.e., $E[\phi] \neq 0$) causes the phase variance to increase, and the presence of noise causes any static phase error to increase from its no-noise level. An appealing physical insight is provided by the approximate analysis of Blanchard [6.18]. Other analyses are given in [6.3] and [6.8]. As might well be imagined, presence of a phase error increases the propensity to cycle slipping. The effect of a static phase error on slips is expounded in [6.3], [6.8–6.13], and [6.23].

Effect of PD s-Curves All results given above are applicable only for phase detector s-curves that are continuous sinusoids with period 2π. Phase detectors with other s-curve shapes are often implemented; some examples appear in Fig. 5.13 and others in Chapter 10. Chie [6.19] has analyzed the influence of the s-curve shape on slip statistics. He found that the integrated area under the s-curve from the tracking point (i.e., the s-curve zero crossing if loop stress is absent) to the cycle-slip boundary is a key property; T_{AV} is an increasing function of that area. His paper provides convenient summary equations for several common s-curves. Also included is the effect of loop stress. Chie cautions that nonsinusoidal s-curves deteriorate toward sinusoidal when input SNR at the phase detector becomes small. Deterioration of s-curve shape is pursued further in Chapter 10.

Narrowband Noise All of the foregoing nonlinear analysis, simulation, and measurement applies only to white noise. In practice, "white" means that the bandwidth of the input noise is large compared to the noise bandwidth $2B_L$ of the PLL. The Fokker–Planck approach is not applicable to analysis with narrowband noise. In consequence, it has not been generally feasible to predict cycle slipping for narrowband noise inputs. Hess [6.20] has devised an approximate analysis of cycle slipping in a first-order loop exposed to bandlimited noise. His formulas are confirmed by measurements of cycle slipping in laboratory PLLs.

Noise pdf All of the foregoing has assumed that Gaussian noise was applied to the PLL; different noise statistics require modified analysis. A limiter is often used in front of a phase detector, making the noise statistics decidedly non-Gaussian. Discussion of the effect of a limiter in the nonlinear region of loop operation may be found in [6.8] and [6.20]. The effect of a limiter in the linear regime is examined in Chapter 10.

Higher-Order PLLs Finally, the known information on nonlinear operation is confined to first- and second-order loops; there is almost nothing published on higher-order loops. Since third-order and type 3 loops have practical importance, this lack of data is a barrier to fully understood design. The only present expedient is to assume that a third-order loop behaves much the same as a second-order loop of the same noise bandwidth.

REFERENCES

6.1 W. B. Davenport and W. L. Root, *Random Signals and Noise*, McGraw-Hill, New York, 1958.

6.2 A. J. Viterbi, *Principles of Coherent Communications*, McGraw-Hill, New York, 1966, Part 1.

6.3 W. C. Lindsey, *Synchronization Systems in Communication and Control*, Prentice Hall, Englewood Cliffs, NJ, 1972.

6.4 F. J. Charles and W. C. Lindsey, "Some Analytical and Experimental Phaselocked Loop Results for Low Signal-to-Noise Ratios," *Proc. IEEE* ***54***, 1152–1166, Sept. 1966.

6.5 A. Blanchard, *Phase Locked Loops*, Wiley, New York, 1976, Chap. 12.

6.6 H. Meyr and G. Ascheid, *Synchronization in Digital Communications*, Wiley, New York, 1990.

6.7 G. Ascheid and H. Meyr, "Cycle-Slips in Phase-Locked Loops: A Tutorial Survey," *IEEE Trans. Commun.* ***COM-30***, 2228–2241, Oct. 1982.

6.8 W. C. Lindsey and M. K. Simon, *Telecommunication Systems Engineering*, Prentice Hall, Englewood Cliffs, NJ, 1973, Chap. 2.

6.9 R. W. Sanneman and J. R. Rowbotham, "Unlock Characteristics of the Optimum Type II Phase-Locked Loop," *IEEE Trans. Aerosp. Navig. Electron.* ***ANE-11***, 15–24, Mar. 1964.

6.10 R. C. Tausworthe, "Cycle Slipping in Phase-Locked Loops," *IEEE Trans. Commun.* ***COM-15***, 417–421, June 1967.

6.11 R. C. Tausworthe, "Simplified Formula for Mean Cycle-Slip Time of Phase-Locked Loops with Steady-State Phase Error," *IEEE Trans. Commun.* ***COM-20***, 331–337, June 1972.

6.12 E. A. Bozzoni, G. Marchetti, U. Mengali, and F. Russo, "An Extension of Viterbi's Analysis of Cycle Slipping in a First-Order Phase-Locked Loop," *IEEE Trans. Aerosp. Electron. Syst.* ***AES-6***, 484–490, July 1970.

6.13 J. K. Holmes, "First Slip Times Versus Static Phase Error Offset for the First- and Passive Second-Order Phase-Locked Loop," *IEEE Trans. Commun.* ***COM-19***, 234, Apr. 1971.

6.14 H. L. Van Trees, "Functional Techniques for the Analysis of the Nonlinear Behavior of Phase-Locked Loops," *Proc. IEEE* ***52***, 894–911, Aug. 1964.

6.15 J. K. Holmes, "On a Solution to the Second-Order Phase-Locked Loop," *IEEE Trans. Commun.* ***COM-18***, 119–126, Apr. 1970.

6.16 H. Meyr, "Nonlinear Analysis of Correlative Tracking Systems Using Renewal Process Theory," *IEEE Trans. Commun.* ***COM-23***, 192–203, Feb. 1975.

6.17 J. R. Rowbotham and R. W. Sanneman, "Random Characteristics of the Type II Phase-Locked Loop," *IEEE Trans. Aerosp. Electron. Syst.* ***AES-3***, 604–612, July 1967.

6.18 A. Blanchard, "Phase-Locked Loop Behavior near Threshold," *IEEE Trans. Aerosp. Electron. Syst.* ***AES-12***, 628–638, Sept. 1976; corrections: ***AES-12***, 823, Nov. 1976.

6.19 C. M. Chie, "New Results on Mean Time-to-First-Slip for a First-Order Loop," *IEEE Trans. Commun.* ***COM-33***, 897–903, Sept. 1985.

6.20 D. T. Hess, "Cycle-Slipping in a First-Order Phase-Locked Loop," *IEEE Trans. Commun.* ***COM-16***, 255–260, Apr. 1968.

6.21 A. J. Viterbi, "Phase-Locked Loop Dynamics in the Presence of Noise by Fokker–Planck Techniques," *Proc. IEEE* ***51***, 1737–1753, Dec. 1963.

6.22 U. Mengali and A. N. D'Andrea, *Synchronization Techniques for Digital Receivers*, Plenum Press, New York, 1997, Secs. 2.4 and 5.3.6.

6.23 W. C. Lindsey and M. K. Simon, "The Effect of Loop Stress on the Performance of Phase-Coherent Communication Systems," *IEEE Trans. Commun.* ***COM-18***, 569–588, Oct. 1970.

CHAPTER 7

EFFECTS OF PHASE NOISE

Chapter 6 treated the effects of additive stationary Gaussian noise, concentrating mainly on noise with a constant (white) spectrum. Additive noise is a dominant concern in high-sensitivity phaselock receivers such as those employed in deep-space communication links. The behavior of a PLL in additive noise has a firm theoretical basis, even if the mathematical particulars become overwhelming for high noise levels in other than the simplest PLL configurations.

Phase noise has emerged as the paramount concern in phaselock frequency synthesizers and in local oscillators of transmitters and receivers. This chapter treats some of the effects of phase noise; further coverage is provided in Chapters 9 and 15. Unlike the additive noise of Chapter 6, phase noise is multiplicative, nonstationary, and its spectrum is not white. The probability distribution of phase noise seems to be an open question; numerous publications assume a Gaussian distribution, but hard evidence is lacking for some components of phase noise. Most strikingly, mathematical analysis of phase noise does not rest on a firm theoretical foundation; one goal of this chapter is to identify anomalous areas and to offer a guide around them. Phase noise is still an open subject for theoretical research. Earlier background articles and extensive references on the subject may be found in [7.1].

In Chapter 6 it was shown that the effects of additive noise in a PLL are reduced by using a small noise bandwidth. Analysis in this chapter demonstrates that ill effects of phase noise in a PLL are reduced by using a large bandwidth. A designer trades off between the two kinds of noise to find a compromise bandwidth that minimizes the total phase jitter.

Phaselock Techniques, Third Edition, by Floyd M. Gardner

7.1 PROPERTIES OF PHASE NOISE

This section introduces elementary properties of phase noise.

7.1.1 Oscillator Model

The dominant source of phase noise in a well-designed system should be from the system oscillators. That is not to say that other hardware elements make no significant contributions to phase noise but that oscillators have characteristics that make them special. Consider a generic oscillator whose output voltage $v_o(t)$ has a sinusoidal waveshape and a nominal oscillation frequency f_o hertz:

$$v_o(t) = [A + a(t)]\cos[2\pi f_o t + \phi(t)] \tag{7.1}$$

where A is the mean amplitude of the oscillator output, $a(t)$ is the zero-mean amplitude noise, and $\phi(t)$ contains all phase and frequency departures from the nominal oscillation frequency f_o and phase $2\pi f_o t$. Phase disturbance $\phi(t)$ (in radians) includes random zero-mean phase noise, initial phase, and integrated effects of frequency offset and drift.

7.1.2 Neglect of Amplitude Noise

Standard analyses of phase noise neglect amplitude noise $a(t)$. Oscillators contain an amplitude-control mechanism that largely suppresses amplitude fluctuations. They also contain an inherent mechanism that accumulates phase fluctuations [7.2, 7.3]. As a result, the effects of phase noise far overshadow the effects of amplitude noise in most situations. Amplitude noise is neglected in this book.

7.1.3 Variance

In the absence of amplitude noise, the variance of oscillator output $v_o(t)$ is simply $A^2/2$, irrespective of any conceivable fluctuations or other disturbances in $\phi(t)$. That is, the variance is bounded, stationary, and well behaved. Docility of its nature is pointed out now to contrast it to the unruly properties of $\phi(t)$ to be revealed subsequently.

7.1.4 Nonstationarity

Consider the autocorrelation function of $v_o(t)$. Or rather, to avoid irrelevant mathematical clutter, work with the complex envelope $z_o(t)$ instead of $v_o(t)$:

$$z_o(t) = A\exp[j(2\pi f_o t + \phi(t))] \tag{7.2}$$

whose autocorrelation function is

$$E[z_o(t_1)z_o^*(t_2)] = A^2\exp[j2\pi f_o(t_1 - t_2)]E\{\exp[j(\phi(t_1) - \phi(t_2))]\} \tag{7.3}$$

This expectation is wide-sense stationary only if the first increment $\phi(t_1) - \phi(t_2)$ of the phase process $\phi(t)$ is stationary; otherwise, it is nonstationary. [A process is wide-sense stationary if its autocorrelation function depends only on the time difference $(t_1 - t_2)$ and on no other function of time.] For example, consider an oscillator frequency that drifts at a rate Λ rad/sec^2 so that $\phi(t) = \Lambda t^2/2$. Regard Λ as a random variable of unknown and irrelevant probability distribution. Also assume that phase disturbances are otherwise absent. Then the autocorrelation function of (7.3) becomes

$$E[z_o(t_1)z_o^*(t_2)] = A^2 \exp[j2\pi f_o(t_1 - t_2)]E\{\exp[j\Lambda(t_1 - t_2)(t_1 + t_2)]\} \quad (7.4)$$

which depends on $t_1 + t_2$ as well as $t_1 - t_2$. Therefore, the autocorrelation function of an oscillator with frequency drift is nonstationary.

Why is stationarity important enough to warrant attention? The spectrum of a signal is of high practical importance as a vital tool for understanding the properties of the signal. Since spectral density is defined formally as the Fourier transform of the autocorrelation function, and since a one-dimensional Fourier transform is defined only for a stationary autocorrelation function, the standard definition of spectral density is meaningful only if the autocorrelation function is stationary.

Hold on, you will say, all oscillators drift. Yet oscillator spectra are displayed in laboratories every day on spectrum analyzers. How can that be if the spectra do not exist? As a practical matter, a spectrum analyzer displays an approximation of a spectrum that is reasonable for engineering purposes if the frequency drift $\Lambda T_m/2\pi$ accumulated during a measurement interval T_m is much smaller than the resolution bandwidth of the analyzer. In a swept-frequency spectrum analyzer, T_m is the time interval during which the signal is within the resolution bandwidth.

Frequency drift is not ordinarily regarded as a constituent of phase noise, nor is it usually a critical problem in PLLs. (See Section 5.1 for PLL response to a frequency ramp.) Moreover, other mathematical tools (outside the scope of this book) are available for analysis of signals with changing frequency. Frequency drift has been introduced for three purposes: (1) to serve as an easily understood example of a nonstationary process, (2) to show that nonstationarity undermines the standard formulation of spectrum (a grave matter inasmuch as spectral analysis is such a valuable tool), and (3) to demonstrate that engineering approximations can circumvent the breakdown of the standard formulation, even in the absence of rigorous theoretical support.

Most components of phase noise are nonstationary, so the exercise with frequency drift gives warning that theoretical contradictions are to be anticipated in spectral representations of phase noise. Fortunately, engineering approximations that are consistent with widely observed actual behavior of PLLs allow sensible design procedures to be successful in most instances despite the shaky theoretical basis.

7.2 SPECTRA OF PHASE NOISE

A number of different spectral density functions are commonly used to characterize phase noise [7.4, 7.5]:

- $W_{vo}(f)$: theoretical passband spectrum of the oscillator signal $v_o(t)$
- $\mathcal{L}(\Delta f)$: normalized version of $W_{vo}(f)$
- $W_{\mathrm{RF}}(f)$, $P_{\mathrm{RF}}(f)$: approximate spectra of the oscillator signal $v_o(t)$ as observed on an RF spectrum analyzer
- $W_{\phi}(f)$: baseband spectrum of the phase noise $\phi(t)$
- $W_{\omega}(f)$: spectrum of the frequency noise $\omega(t) = d\phi(t)/dt$

[**Comments**: (1) These are all *one-sided* spectra, as defined in Section 6.1.1. (2) Questions of existence of these spectra, as raised by nonstationarity of phase noise, are deferred until later.]

7.2.1 Theoretical Spectrum $W_{vo}(f)$

This bandpass spectrum is the Fourier transform of the autocorrelation function of the random process $v_o(t)$ of (7.1). The definition requires that the autocorrelation function be stationary. From (7.3), the autocorrelation of $v_o(t)$ is stationary if the first increment of $\phi(t)$ is stationary. Unfortunately, one component always present in $\phi(t)$ has a nonstationary first increment, so the standard definition of spectrum is not applicable. For now, just ignore the true nature of the autocorrelation and pretend that the standard spectrum does exist.

Figure 7.1 provides a qualitative view of the character of the theoretical spectrum of $v_o(t)$. In the absence of phase noise, the spectrum is a single line—a delta function—at $f = f_o$. The presence of phase noise spreads the spectrum: Small amounts of noise cause small spreading; larger amounts of noise cause greater spreading. Irrespective of the amount or other character of phase noise, the variance of $v_o(t)$—the integral of $W_{vo}(f)$ over all frequencies $f = 0$ to ∞—is equal to $A^2/2$ volts^2, a defined and bounded number. Moreover, except for the unattainable condition of $\phi(t) \equiv 0$, the theoretical spectrum $W_{vo}(f)$ is everywhere finite. The units of $W_{vo}(f)$ are V^2/Hz.

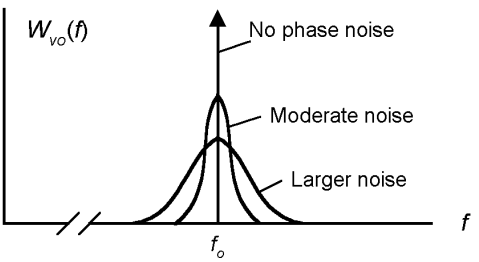


Figure 7.1 Theoretical spectrum $W_{vo}(f)$ of oscillator output $v_o(t)$.

7.2.2 Normalized Spectrum $\mathcal{L}(\Delta f)$

Another one-sided spectral description [7.5] is $\mathcal{L}(\Delta f)$, a normalized version of the theoretical spectrum $W_{vo}(f)$. It is defined as

$$\mathcal{L}(\Delta f) = \frac{W_{vo}(f_o + \Delta f)}{A^2/2} \tag{7.5}$$

In words: $\mathcal{L}(\Delta f)$ is the noise power, relative to the total power in the signal, in a bandwidth of 1 Hz in a single sideband at a frequency offset of Δf from the carrier frequency f_o. Numerically, the value of $\mathcal{L}(\Delta f)$ is commonly expressed in decibel format as $10 \log[\mathcal{L}(\Delta f)]$ dBc/Hz. [**Notation**: dBc means "dB relative to carrier," where the term *carrier* actually means total power in the signal; "per Hz" refers to a bandwidth of 1 Hz.] As defined here, $W_{vo}(f)$ and $\mathcal{L}(\Delta f)$ are entirely legitimate spectral representations of the narrowband RF random process $v_o(t)$, stationarity questions aside. Unfortunately, the idea of $\mathcal{L}(\Delta f)$ and its notation are extensively misused in ways described later.

7.2.3 RF Spectra $W_{RF}(f)$ and $P_{RF}(f)$

A theoretical spectrum is an ensemble property of a random process and can never be observed. Only sample functions of the random process are available. A spectrum analyzer is a laboratory instrument that measures a signal and displays an approximation to its theoretical spectrum. Figure 7.2 is a simplified block diagram of one kind of spectrum analyzer. The signal frequency f_o is mixed against the frequency f_{LO} of a swept local oscillator. The difference frequency $f_o - f_{\text{LO}}$ is applied to a bandpass filter that has center frequency f_{IF} and resolution bandwidth RBW. The output of the bandpass filter is applied to a square-law detector whose output is applied to a lowpass smoothing filter with video bandwidth VBW. Smoothing-filter output either goes directly to a display that shows power or else, through a logarithmic converter, to a display that shows power on a dB scale. Examples measured on an actual oscillator are shown in Figs. 7.3 and 7.4.

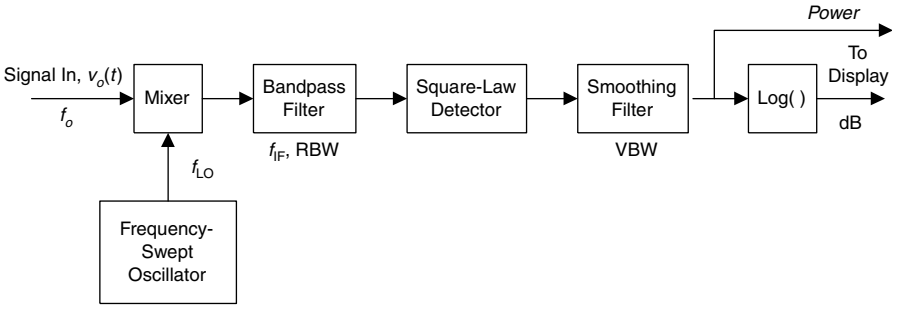


Figure 7.2 Simplified block diagram of a spectrum analyzer.

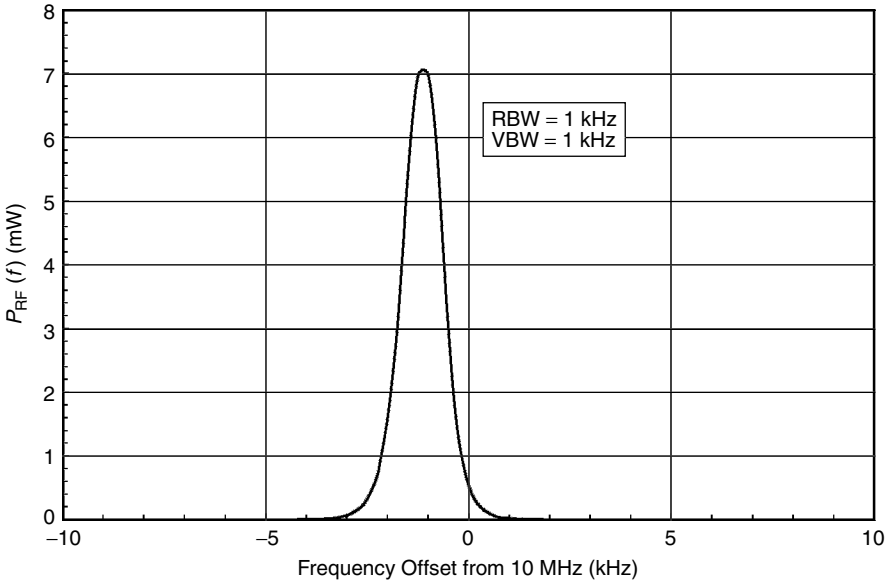


Figure 7.3 Measured RF power spectral density $P_{RF}(f)$ of a 10-MHz crystal oscillator (ordinate shows power in mW in a bandwidth of RBW hertz).

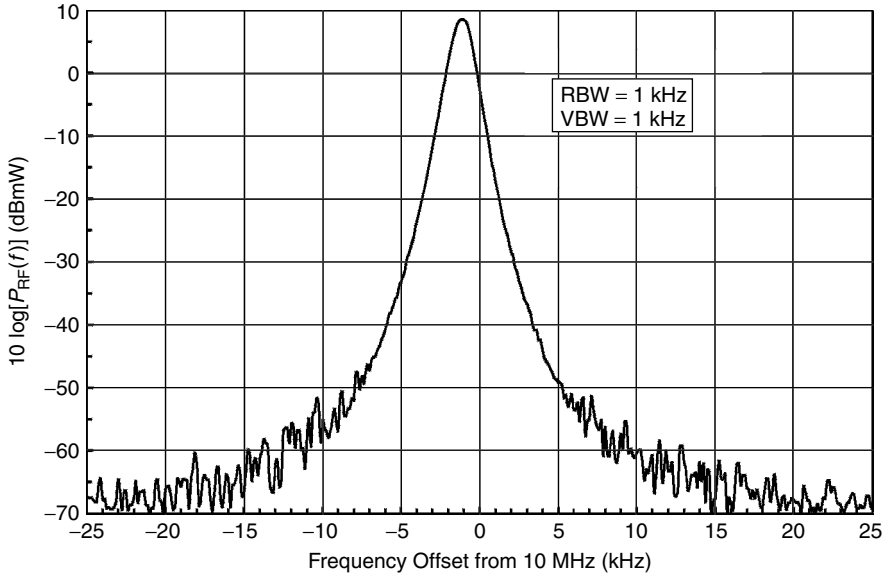


Figure 7.4 Measured RF power spectral density on dB scale (same oscillator as in Fig. 7.3; the ordinate shows power in a bandwidth RBW in dB relative to 1 mW).

Raw measurements do not display spectral density $W_{\mathrm{RF}}(f)$, at least not without appreciable processing of the measured data. The vertical axis of a spectral display represents signal power lying within the resolution bandwidth RBW (actual power in watts, rather than variance in volts^2, because the analyzer input connector has an accurate resistance termination, typically 50 ohms). Spectral density is the power in a 1-Hz bandwidth. Since RBW is usually substantially larger than 1 Hz in most RF spectrum analyzers, the display is not the true spectral density. For that reason, the vertical axes in Figs. 7.3 and 7.4 have been labeled "$P_{\mathrm{RF}}(f)$," not "$W_{\mathrm{RF}}(f)$." To convert from P to W in Fig. 7.3, divide the P scale by RBW; to convert on the logarithmic ordinate in Fig. 7.4, subtract 10 log(RBW) dB. (Some analyzers might have built-in capability for adjusting the axis scales.) Of course, adjusting the vertical scale does not give the 1-Hz resolution implied in $W_{\mathrm{RF}}(f)$; the resolution is still RBW. Finer resolution is possible only by reducing RBW and reducing the scan speed correspondingly.

Suppose that the signal source and the spectrum analyzer itself are both free of phase noise so that the spectrum of the signal applied to the analyzer is a single line. In that case, the display traces out the frequency response of the bandpass filter in the analyzer; the display is not a single line. For realistic signals, the analyzer displays a frequency-domain convolution between the signal spectrum and the frequency response of the bandpass filter. Unless RBW is very narrow compared to the signal spectrum, the analyzer will smear the displayed spectrum. Desmearing (deconvolution) is a nontrivial operation, one not likely to be built into the analyzer.

Attempts are often made to estimate the normalized $\mathcal{L}(\Delta f)$ from $P_{\mathrm{RF}}(f)$. That is a plausible action if the correct normalizing power is employed in the estimation procedure. In terms of the display of the spectrum analyzer, the proper normalizing power is the integral of the power spectrum as shown in Fig. 7.3—an integral that is rarely calculated. More commonly, the peak of a spectrum like that in Fig. 7.4 is taken as the normalizing power. The peak can only be an approximation to the correct power and the approximation can approach good accuracy only if substantially all of the signal power falls within the resolution bandwidth. In that case, the apparent sidebands close in are determined more by the shape of the analyzer's passband than by the signal being analyzed. Cautious interpretation of an RF spectrum is always advisable.

An RF spectrum is a good tool to use in searching for spurious outputs of a signal source or for exposing potential noise-sideband interference to adjacent-channel signals. The latter condition is illustrated in Fig. 7.5. Keep in mind that an RF spectrum analyzer does not distinguish between phase noise and amplitude noise but shows the total power entering the filter passband from all sources.

7.2.4 Phase-Noise Spectrum $W_\phi(f)$

The quantities $W_{vo}(f)$, $\mathcal{L}(\Delta f)$, $W_{\mathrm{RF}}(f)$, and $P_{\mathrm{RF}}(f)$ are all spectra of the physical RF signal $v_o(t)$. Their peaks are at the carrier frequency f_o, and they have sidebands to either side of the peaks. Since resolution of the display is the same

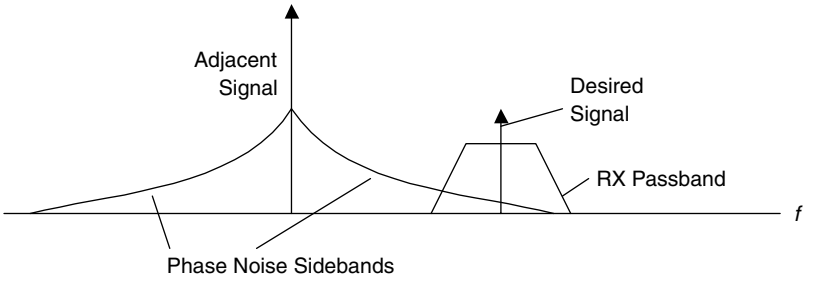


Figure 7.5 Interference from the sidebands of an adjacent-channel signal.

for all frequency offsets, details close in to f_o are sacrificed if more distant sidebands are to be included. Moreover, any practical RF analyzer has dynamic-range requirements; it must accommodate the total power of the signal without overload and yet be able to display weak sidebands, a difficult prescription to fill. Most significantly, there is no easy analytical relationship between recognizable noise sources within a PLL and the ultimate RF spectrum that is produced.

These shortcomings of RF spectra have led to the widespread use of $W_\phi(f)$—the lowpass, one-sided spectrum of the phase-noise modulation $\phi(t)$—as a better tool for dealing with PLLs. A conceptual block diagram for the measurement of $W_\phi(f)$ is shown in Fig. 7.6. The measurement instrument consists of a phase demodulator which reproduces a magnitude-scaled version of $\phi(t)$, a low-frequency spectrum analyzer to produce $W_\phi(f)$, and a logarithmic converter for display purposes. Display is typically 10 log $W_\phi(f)$ vs. frequency on a logarithmic scale. Units of $W_\phi(f)$ are rad^2/Hz and the dB format should be interpreted as dB re 1 rad^2/Hz.

In practice, although the actual measurement is of $W_\phi(f)$, the ordinate regrettably is almost always displayed as 10 log $\mathcal{L}(\Delta f)$, which is supposed to be the relative noise in a 1-Hz bandwidth in a single sideband at frequency offset Δf. If the phase-noise amplitude is small enough, it can be shown that $\mathcal{L}(\Delta f) \approx W_\phi(f)/2$, so the supposed $\mathcal{L}(\Delta f)$ display is really $W_\phi(f)/2$. But the phase noise amplitude is *never* small enough at frequencies close in to the carrier frequency, so $W_\phi(f)$ is *never* a good representation of the close-in $\mathcal{L}(\Delta f)$, the close-in RF sidebands. It is the label $\mathcal{L}(\Delta f)$ applied to the baseband spectrum of phase-noise modulation $\phi(t)$ that is at fault, not the measurements themselves. Just add 3 dB to any lowpass spectrum that purports to be 10 log$[\mathcal{L}(\Delta f)]$ to obtain the proper 10 log$[W_\phi(f)]$.

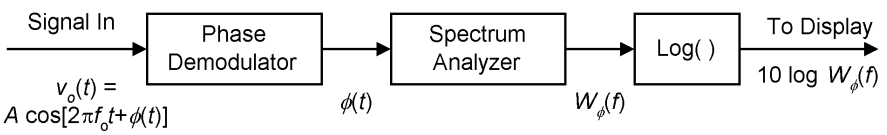


Figure 7.6 Block diagram of a generic phase-noise analyzer.

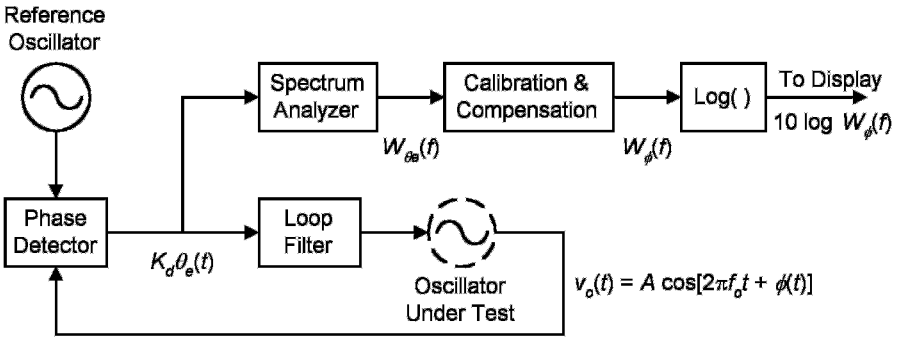


Figure 7.7 Block diagram of a phase-noise analyzer based on a PLL.

The heart of a phase-noise analyzer is the phase demodulator. One prominent realization is shown, greatly simplified, in Fig. 7.7. A PLL is employed as the phase demodulator; see Chapter 16 for further explanation of PLLs as modulators and demodulators. The oscillator to be tested is connected as the VCO in a PLL and its phase is compared against that of a suitable reference source. Phase-error fluctuations $\theta_e(t)$, scaled by the phase-detector gain K_d, are applied to a spectrum analyzer that produces a spectrum $W_{\theta e}(f)$ which is an approximation to the desired spectrum $W_\phi(f)$.

Raw output of the spectrum analyzer usually is not a good-enough approximation to $W_\phi(f)$; various elaborate calibrations and compensations have to be applied, such as:

- The scale factor K_d has to be calibrated.
- The closed-loop error response $E(f)$ of the PLL (see Chapter 2) has to be determined and the measured spectrum compensated accordingly. Error response $E(f)$ serves as a highpass filter between oscillator noise $\phi(t)$ and PLL phase error $\theta_e(t)$.
- Native resolution of the spectrum analyzer is not usually 1 Hz; the measured spectrum has to be scaled to represent a 1-Hz resolution.

Other tedious operations are also necessary. Fortunately, since phase-noise modulation is a low-frequency (as compared to a carrier frequency) matter, the spectrum analysis and calibration/compensation operations can be carried out digitally. The spectrum analyzer typically would be a filter bank implemented by a fast Fourier transform algorithm instead of the single swept-frequency analysis filter typical of many RF spectrum analyzers. Digital operations of immense complexity and flexibility are employed to impart great power to instruments of this kind.

Additional sources of noise besides the phase noise in the oscillator under test exist in the instrument of Fig. 7.7. One obvious source is the phase noise of the reference oscillator itself. A common expedient, if the reference oscillator cannot

be made far quieter than the oscillator being tested, is to employ a reference oscillator that is nearly identical to the tested oscillator and to ascribe half of the measured noise spectrum to each oscillator.

All the various noise sources within the measuring instrument combine to establish a noise floor, below which measurements on the tested oscillator are impractical. Phase-noise spectra typically are large at low frequencies and fall off toward higher frequencies. In consequence, the phase-noise spectrum of the tested oscillator usually is well above the noise floor at low frequencies but may fall below the floor at higher frequencies.

7.2.5 Frequency-Noise Spectrum $W_\omega(f)$

Instantaneous radian frequency $\omega(t)$ is the time derivative of phase $\phi(t)$. If phase $\phi(t)$ has a Fourier transform $\Phi(f)$, the Fourier transform of its derivative is $\Omega(f) = j2\pi f\Phi(f)$. The Fourier transform of an infinite-energy random process does not exist, but the derivative transformation nonetheless can be applied to the spectral densities to give the relation

$$W_\omega(f) = 4\pi^2 f^2 W_\phi(f) \qquad (\text{rad/sec})^2/\text{Hz} \tag{7.6}$$

Equation (7.6) is the basis of another arrangement for measuring the phase-noise spectrum of $\phi(t)$. The signal $v_o(t)$ is applied to a frequency discriminator whose output is a scaled version of the frequency modulation $\omega(t) = d\phi(t)/dt$. That recovered frequency modulation is delivered to a spectrum analyzer whose output is $W_\omega(f)$. The frequency spectrum is then weighted by $1/4\pi^2 f^2$ to give the desired $W_\phi(f)$.

Calibration and compensation are also needed for a phase-noise analyzer based on a frequency discriminator, much like that described above for a PLL-based analyzer. The spectrum of frequency noise will be weighted more strongly toward higher frequencies, so the lower-frequency portions of $W_\omega(f)$ might fall below the instrument's noise floor, but the higher-frequency portions are emphasized. Thus, the two kinds of phase-noise analyzers are complementary; some instruments might contain both kinds to span a larger range of Fourier frequencies.

7.2.6 Example Phase-Noise Spectrum

Figure 7.8 shows the phase-noise spectrum of the same oscillator whose RF spectrum was shown in Figs. 7.3 and 7.4. The vertical axis is labeled "10 $\log[\mathcal{L}(f)]$ (dBc/Hz)," the format reported by the analyzer. If you want to show 10 $\log[W_\phi(f)]$ instead, just add 3 dB and change the units to "re 1 rad^2/Hz."

Several features of the chart are worthy of note:

- A number of large spikes are prominent. These are located at harmonics of 60 Hz, an indication of interference from the power supply. Greater care in hardware implementation and possibly the test setup should reduce the spikes substantially.

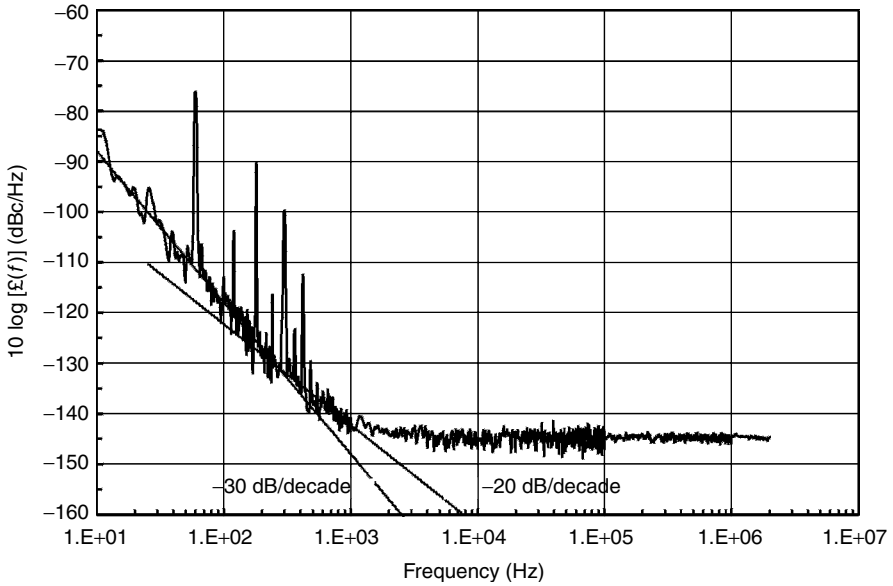


Figure 7.8 Measured phase-noise spectral density $W_\phi(f)/2$ on log-log scales (the same oscillator as in Figs. 7.3 and 7.4).

- Considerable hash is visible on the plot. Hash comes about because the noise waveform in the test is a sample function from an infinite ensemble, and as such has random discrepancies from the ensemble spectrum. Longer observation times along with appropriate smoothing of the spectrum analyzer output would reduce the hash.
- Character of the hash changes abruptly at several frequency-decade boundaries. These changes suggest that internal parameters of the analyzer are altered automatically for different frequency ranges.
- Notice the two straight lines that indicate the log-log slopes of portions of the phase-noise graph. These slope values appear in the spectra of most oscillators; they arise from different spectral components of the phase noise, as explained later in this chapter and in Chapter 9.

7.3 PROPERTIES OF PHASE-NOISE SPECTRA

The phase-noise spectrum consists of a continuous part due to the random phase noise plus discrete spectral lines that arise from periodic interference from such sources as AC-power residues in the supply voltages, incompletely suppressed ripple from the phase detector in a PLL, or other ingress from the environment. Continuous parts and discrete parts are considered separately below.

7.3.1 Typical Continuous Spectra

Numerous measurements have shown consistently that the continuous phase-noise spectra of oscillators tend to be well approximated by [7.6]

$$W_\phi(f) \approx \frac{h_4}{f^4} + \frac{h_3}{f^3} + \frac{h_2}{f^2} + \frac{h_1}{f} + h_0 \qquad \text{rad}^2/\text{Hz} \tag{7.7}$$

where the h_ν are coefficients that are particular to each individual device. Dimensions of h_ν are $\text{rad}^2 \cdot \text{Hz}^{\nu-1}$. On log-log scales, (7.7) will plot approximately as connected straight-line segments, as sketched in Fig. 7.9. Each segment is labeled with its h_ν/f^ν term plus the log-log slope in dB/decade. The resemblance to the example of Fig. 7.8 is evident.

The h_4/f^4 term appears mainly in the spectra of precision frequency standards (e.g., cesium clocks) at frequencies well below 1 Hz and normally is not an issue for oscillators in PLLs. It is not considered further in this book; accordingly, its line segment is shown dashed. The other terms are all significant. Each term arises from a different source of phase noise, as explained further in Chapter 9. The phase-noise terms h_3/f^3 and h_2/f^2 arise out of flicker $(1/f)$ and white noises within the oscillator that cause flicker and white fluctuations of frequency, as discussed in Chapter 9. Those frequency fluctuations are integrated to phase

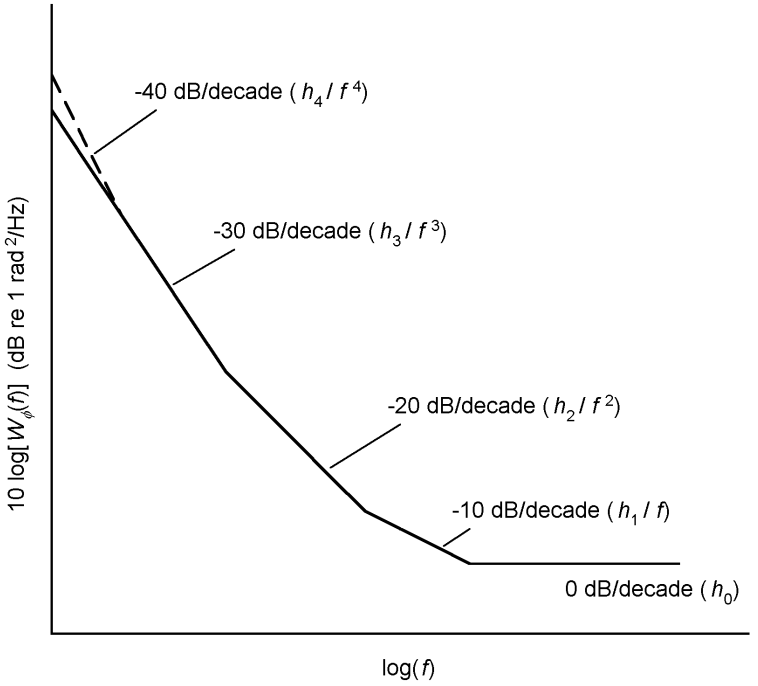


Figure 7.9 Typical spectral components of oscillator phase noise.

in the oscillator to produce the $1/f^3$ and $1/f^2$ spectral components. Ordinary nonintegrating circuit elements do not exhibit $1/f^3$ or $1/f^2$ noise spectra, but these terms are prevalent in oscillators.

7.3.2 Meaning of $W_\phi(f)$

Observe that each term in (7.7) is infinite at $f = 0$ if $\nu > 0$. These singularities cause the integral of $W_\phi(f)$ from $f = 0$ to any nonzero upper frequency to fail to converge, implying that the variance of $\phi(t)$ is infinite. You might speculate that the spectrum has to level out at some extremely low frequency, but measurements [7.7] have not revealed such leveling nor does any theoretical basis exist for spectral leveling. Infinite variance of phase noise appears to be a fact of nature.

An infinite variance is intuitively disturbing at first but appears more reasonable after the accumulative nature of phase noise is recognized. There is no restoring force or fading memory for the phase deviations of an oscillator; any incremental deviation persists forever. Instantaneous phase noise $\phi(t)$ is the accumulation of all of the phase deviation that has occurred since the oscillator first started operation. But infinite deviation accumulates only after an infinite time and so is never observed. More subtly, spectral density of a random process is defined formally as the Fourier transform of the autocorrelation function of the process, provided that the autocorrelation function is stationary. But the autocorrelation function of an h_2/f^2 process is nonstationary and the autocorrelation of an h_3/f^3 process is worse than nonstationary—it does not even exist. Strictly speaking, then, the standard definition of spectral density is inapplicable to phase noise.

So, exactly what is meant by the *spectrum* of phase noise? First, let it be said that a spectrum may not be a sound representation for phase noise and that perhaps some other representation ought to be used instead. (Examples: Wornell [7.8, 7.9] argues in favor of a wavelet representation; Flandrin [7.10] explores Wigner–Ville spectra as well as wavelets.) The future may bring an improved representation but "spectrum" is what is used today.

Despite the shaky theoretical foundation, laboratory phase-noise analyzers deliver data purported to describe a spectrum and PLLs are designed and evaluated successfully on the basis of those data. Bandpass analysis filters in phase-noise analyzers have transmission zeros at $f = 0$ and $f = \infty$, zeros that help suppress the effects of singularities associated with the extreme frequencies. Also, any physical measurements necessarily extend over only a finite time interval, whereas the infinite variance implied in the spectral model requires an infinite time interval to accumulate. A physical time-limited, amplitude-bounded waveform, therefore a finite-energy waveform such as is delivered by the phase demodulator in a phase-noise analyzer, has a well-defined Fourier transform that also has finite energy and is free of other peculiarities. An analyzer presumably constructs something related to the squared magnitude of that well-behaved Fourier transform and places a label "spectrum" on that something. With the

help of transmission zeros in the filters and the necessarily finite time interval for measurement, a phase-noise analyzer never confronts the infinities in the spectral model of (7.7). In light of that success in spectrum analyzers, a practitioner's answer to the question above is: The spectrum of phase noise consists of the data delivered by a phase-noise spectrum analyzer.

7.3.3 Interpretation of Spectral Displays

Figure 7.8 shows a mixture of a continuous spectrum and individual lines of a discrete spectrum. These have to be interpreted differently: first the continuous spectrum. Denote by $W_\phi(f)$ the underlying continuous phase-noise spectrum of the random process $\phi(t)$ and assume that this spectrum has a meaningful existence, even in the absence of a rigorous definition. Assume a phase-noise analyzer as in Fig. 7.6. Let the analysis filter for the analysis frequency f_m have a bandpass shape that is centered on f_m, and denote the frequency response of the filter as $Y(f; f_m)$. Assume that the filter's frequency response is concentrated in the vicinity of f_m and has adequate selectivity to suppress the zero-frequency singularities from its output. Each analysis frequency f_m has its own analysis filter, either because of frequency sweeping or because of a filter bank in the spectrum analyzer.

The spectral density of the filter output is $W_\phi(f)|Y(f; f_m)|^2$. The operations of squaring and smoothing are represented by

$$P_c(f_m) = \int_0^\infty W_\phi(f)|Y(f; f_m)|^2\, df \qquad \text{rad}^2 \tag{7.8}$$

where $P_c(f_m)$ is the intensity of the f_m filter output and the subscript c indicates a continuous spectrum. If $W_\phi(f)$ is nearly flat within the effective bandwidth of the analysis filter, (7.8) can be approximated as

$$P_c(f_m) \approx W_\phi(f_m) \int_0^\infty |Y(f; f_m)|^2\, df \tag{7.9}$$

The integral in (7.9) is simply $|Y(f_m; f_m)|^2 B_N(f_m)$, where $|Y(f_m; f_m)|$ is the magnitude of the analysis filter's frequency response at frequency f_m and $B_N(f_m)$ is the noise bandwidth of that filter. The analyzer's estimate of the spectral density at frequency f_m therefore is

$$W_\phi(f_m) \approx \frac{P_c(f_m)}{|Y(f_m; f_m)|^2 B_N(f_m)} \tag{7.10}$$

where the numerator of (7.10) is a measured number and the denominator is a hardware parameter known to the designer of the analyzer. The division operation is part of the calibration and compensation needed and can be trivial in a digital system.

[**Comments:** (1) The operations shown above, especially the integration, are simplified and idealized. Nonetheless, they indicate the approximate nature of the operations that have to be conducted in the analyzer. (2) Actual calibration may involve a characteristic bandwidth of the filter other than the noise bandwidth without changing the principles of the calibration. (3) A more sophisticated analyzer may be able to take the nonflat shape of $W_\phi(f)$ into account and thereby obtain a more accurate estimate of $W_\phi(f_m)$. Calibration then is somewhat more complicated, but the underlying principle is the same.]

Now consider the treatment of discrete spectral components. Let $\phi_d(t) = \beta\cos(2\pi f_m t)$ where the subscript d indicates discrete and β is the peak phase deviation in radians. The one-sided power spectral density of this component is $(\beta^2/2)\delta(f - f_m)$: that is, a discrete line of infinite height, zero width, and area $\beta^2/2$ located at $f = f_m$. The combination of passing $\phi_d(t)$ through the filter, squaring the filter output, and smoothing the squared output is approximated by

$$P_d(f_m) = \int_0^\infty \frac{\beta^2}{2}\delta(f - f_m)|Y(f, f_m)|^2\, df = \frac{\beta^2}{2}|Y(f_m, f_m)|^2 \qquad (7.11)$$

The power $\beta^2/2$ in the discrete component is estimated by dividing the measurement $P_d(f_m)$ by the squared magnitude of the filter response at the frequency f_m. Observe that this measurement is independent of the bandwidth of the analysis filter, unlike the measurement of a continuous-spectrum phase noise. It would be seriously wrong to apply the calibration of (7.10) to a discrete spectral component.

If an analyzer can reliably distinguish a discrete-spectrum component automatically from a continuous-spectrum component and if the analyzer applies the correct calibration, there is no problem for the user. But if the analyzer does not distinguish between discrete and continuous spectral components (the instruction manual presumably tells if it has the capability) and applies the same calibration to both classes, the results for one of the classes is seriously wrong. A user can often distinguish between discrete- and continuous-spectrum components from inspection of the display. Then the user can rescale the spectral component that has been wrongly calibrated if the analyzer bandwidth and the calibration principle are known.

In summary: Be aware that one or the other class of spectral components could be displayed at an incorrect scale. You will not be able to correct the display unless you know the calibration principles of the analyzer and the bandwidth of the analysis filter.

7.3.4 Relationship Between $W_\phi(f)$ and $\mathcal{L}(\Delta f)$

Because the spectrum $W_\phi(f)$ of the baseband phase noise $\phi(t)$ and the normalized spectrum $\mathcal{L}(\Delta f)$ of the passband signal $v_o(t)$ are just different aspects of the same signal, it follows that there must be a relationship between the two spectra. An engineer would like, for example, to take a measured $W_\phi(f)$ and calculate

the corresponding $\mathcal{L}(\Delta f)$ therefrom. No broadly applicable calculation methods are known. This section briefly surveys some piecemeal approaches.

Sinusoidal Modulation As is well known, if a carrier is phase-modulated by a single sinusoid, the spectrum of the modulated signal contains a residual carrier component plus an infinity of sidebands, spaced from one another by the modulating frequency. Amplitudes of the residual carrier and of the ith sideband are proportional to the Bessel function $J_i(\Delta\theta)$, where $\Delta\theta$ is the peak phase deviation of the modulation. Inherent nonlinearity of phase modulation generates the plethora of sidebands. If $\Delta\theta$ is small enough, only the first-order sidebands, proportional to $J_1(\Delta\theta) \approx \Delta\theta/2$, have significant amplitude, while the other sidebands can be neglected.

If the baseband modulating signal contains more than one sine wave, the modulated spectrum contains an infinity of sidebands corresponding to each of the baseband sine waves plus multiple infinities of intermodulation products between the baseband components, all with amplitudes based on Bessel functions. The intermodulation arises from the inherent nonlinearity of phase modulation. If the total phase deviation is small enough, only the first-order sidebands of each baseband signal will be significant.

Discrete Approximation of $W_\phi(f)$ One approach to analysis of a continuous spectrum such as $W_\phi(f)$ is to divide it up into a large number of equal contiguous frequency increments and to substitute a discrete spectral line of the same variance for each such increment. This is more or less equivalent to performing a Fourier series analysis of a finite segment of the time-domain process such as $\phi(t)$, a procedure described in texts on stochastic processes, for example [7.11].

The idea of spectral discretization applied to phase noise is to employ the known Bessel function relationship for sinusoidal phase modulation. In particular, if each discrete spectral line has a small-enough phase deviation, its contribution to the modulated spectrum is supposed to be significant only at the corresponding pair of first-order sidebands. When $W_\phi(f)$ is discretized in this manner, the result is $\mathcal{L}(\Delta f) \approx W_\phi(f)/2$ for offset frequencies sufficiently removed from the carrier. That is, the far-out sidebands of the RF signal can be approximated from knowledge of the spectrum $W_\phi(f)$ of the baseband phase noise $\phi(t)$.

But $W_\phi(f)$ always includes components of the form h_ν/f^ν as in (7.7). These components grow without bound as f approaches zero. The small-deviation approximation is *never* valid for small-enough frequency offset. The simple relationship between $W_\phi(f)$ and $\mathcal{L}(\Delta f)$ *always* breaks down for sideband frequencies close enough to the carrier. The relationship has to break down since $\mathcal{L}(\Delta f)$ is always finite for all f, whereas $W_\phi(f)$ is unbounded for small-enough f.

Special Cases The h_2/f^2 term in W_ϕ arises from white noise in the oscillator circuit (see Chapter 9) and its relation to $\mathcal{L}(\Delta f)$ has been well studied [7.3, 7.12–7.14]. It is a special case in which a simple formula for $\mathcal{L}(\Delta f)$ is known

and can be related to $W_\phi(f)$. Following the development in [7.14], if $W_\phi(f) = h_2/f^2$, the expression for $\mathcal{L}(\Delta f)$ is the *Lorentzian* shape

$$\mathcal{L}(\Delta f) = \frac{h_2/2}{(\pi h_2/2)^2 + \Delta f^2} \tag{7.12}$$

a shape long known in optical spectroscopy. Equation (7.12) has the following properties:

- $\mathcal{L}(0) = 2/h_2\pi^2$, a finite value in contrast to the infinite value of $W_\phi(0)$.
- $\int_{-\infty}^{\infty} \mathcal{L}(\Delta f)\, d\Delta f = 1$.
- The half-power bandwidth (full width, half maximum: FWHM) is πh_2 hertz.
- If $\Delta f^2 \gg (\pi h_2/2)^2$, then $\mathcal{L}(\Delta f) \approx h_2/2\Delta f^2 = W_\phi(f)/2$.

For other spectral shapes: An expression for $\mathcal{L}(\Delta f)$ corresponding to the h_3/f^3 term of $W_\phi(f)$ has been derived in [7.2]. Propagation of h_2/f^2 phase noise through a PLL is analyzed in [7.15] to determine $\mathcal{L}(\Delta f)$ at the PLL output. Be aware that although the h_ν/f^ν terms are additive in $W_\phi(f)$, they combine nonlinearly in $\mathcal{L}(\Delta f)$ in a fashion that has so far withstood a more general analysis.

7.4 PROPAGATION OF PHASE NOISE

This section tells how phase noise propagates through various devices that are commonly found in electronic circuits. Propagation is summarized first for certain auxiliary devices and then for PLLs. Except as noted explicitly for the PLL, only propagation of phase noise is considered here, not noise generated internally to the devices.

7.4.1 Phase-Noise Propagation in Auxiliary Devices

The auxiliary devices of interest are frequency multipliers, frequency dividers, mixers, and hard limiters. These are all nonlinear devices. Their approximate effects on phase noise are summarized in Fig. 7.10. A times-N frequency multiplier magnifies input phase noise by a factor of N, that is, $20\log(N)$ in decibels. Similarly, a divide-by-M frequency divider reduces phase noise by a factor of M, that is, $-20\log(M)$ in decibels. Frequency multipliers and dividers preserve the time jitter of their inputs.

Phase noise is carried on each of the two inputs to a mixer. Output phase noise is the sum or difference of the input phase noises, depending on which mixer product is selected for the output. If the input noises are uncorrelated, the output noise spectrum is the sum of the two input noise spectra, translated to the output carrier frequency. If the input noises are correlated, a condition that can arise in some systems, a difference–frequency output product might cancel some of the

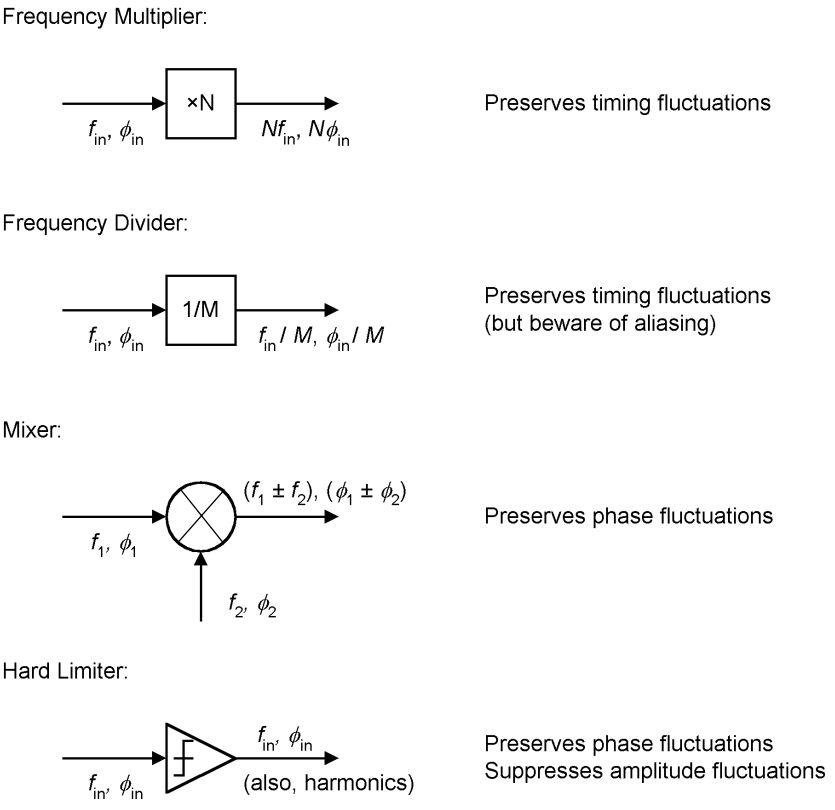


Figure 7.10 Phase-noise propagation through elements often included in PLLs.

phase noise. A limiter preserves the phase noise of its input and suppresses the amplitude noise.

These rules are all first order. Implicit in them is a narrowband assumption that harmonics of the device outputs can be neglected and that there is no spectral folding of significance. Frequency dividers might very well cause substantial spectral folding (examined further in Chapter 15), and limiters will have appreciable output harmonics unless those are suppressed by filters. In the absence of spectral folding and if unwanted harmonics can be neglected, the simple rules all imply that these devices do not alter the spectral shape of the input phase noise; they only scale the overall magnitude.

Although a limiter does not change the input phase noise, it converts an additive disturbance—such as additive noise or a narrowband interferer—into output phase noise. If the desired signal input to the limiter has amplitude A and the interferer has amplitude $B \ll A$, the limiter output will have unwanted phase modulation of approximate amplitude B/A radians. If the interferer frequency is offset by Δf from that of the desired signal, the spectrum of the limiter output will contain a pair of interference sidebands at $\pm\Delta f$ from the desired signal, with

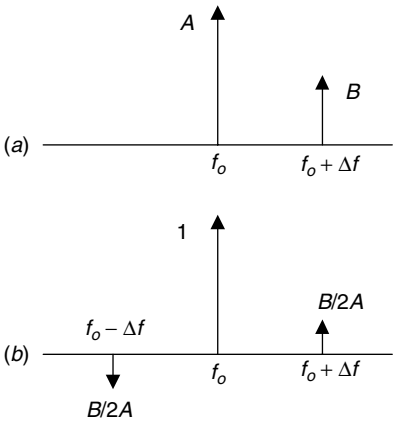


Figure 7.11 Interference-to-phase conversion in a limiter.

amplitudes $B/2A$ relative to the desired signal output. Figure 7.11 illustrates the effect for an isolated interferer, and Appendix 7A contains an analysis. A limiter is of interest by itself (see Chapter 10) and also because one is implicitly included within a digital frequency divider, a device widely used in PLL-based frequency synthesizers (see Chapter 15).

7.4.2 Phase-Noise Propagation in PLLs

Let the input signal to the PLL have phase-noise spectral density $W_{\phi i}(f)$ rad^2/Hz. The *tracked* phase-noise spectrum appearing on the VCO output in response to the input phase noise is

$$W_{\theta o \phi i}(f) = W_{\phi_i}(f)|H(f)|^2 \tag{7.13}$$

where $H(f)$ is the system frequency response of the closed-loop PLL (see Section 2.1.2). Tracked phase noise is simply the input phase noise as transmitted through the lowpass filter with response $H(f)$. The *untracked* phase-noise spectrum—the phase-error spectrum—caused by input phase noise is

$$W_{\theta e \phi i}(f) = W_{\phi i}(f)|E(f)|^2 \tag{7.14}$$

where $E(f)$ is the error response of the PLL. [**Notation**: $W_{\theta e \phi i}(f)$ means the phase-noise spectral density in the phase error θ_e due to input phase noise ϕ_i.]

Another important source of PLL phase noise originates within the VCO as illustrated in Fig. 7.12. The dashed box in the figure encloses the physical VCO, which consists of a fictitious noise-free VCO delivering output phase θ_v, plus an internal phase noise source ϕ_o of spectral density $W_{\phi o}(f)$ rad^2/Hz. From circuit analysis, the transfer function from ϕ_o to θ_o is found to be the error

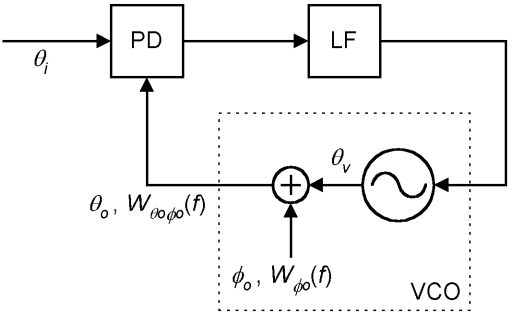


Figure 7.12 Model of oscillator phase noise in a PLL.

response $E(f)$, so the phase-noise spectral density at the output of the physical phase-locked VCO caused by the internal phase noise of the VCO is

$$W_{\theta o \phi o}(f) = W_{\phi o}(f)|E(f)|^2 \tag{7.15}$$

This is the spectrum of untracked phase jitter from ϕ_o. Observe that the phase error θ_e due to ϕ_o is $-\theta_o$, so (7.15) also specifies the spectral density $W_{\theta e \phi o}$.

Equations (7.14) and (7.15) for untracked jitter have the same format, so they can be combined into a single expression for untracked jitter:

$$W_{u\phi}(f) = W_{\phi}(f)|E(f)|^2 \tag{7.16}$$

where $W_{\phi}(f) = W_{\phi i}(f) + W_{\phi o}(f)$ and the subscript u indicates *untracked* phase noise.

7.5 INTEGRATED PHASE NOISE IN PLLs

Knowledge of the shape of phase-noise spectra, as described in Section 7.4, is useful for understanding the nature of the phase noise, for identifying its sources, and for design guidance. Another key datum is the *integrated phase noise*: the phase-noise spectral density integrated over all frequencies. Several features of integrated phase noise are covered in this section.

7.5.1 Basic Formulas

The integral of *tracked* phase noise does not converge if the spectrum of the phase-noise source is like that of (7.7); the integrated tracked phase noise is always infinite. In practical terms, a locked PLL tracks slow enough accumulation of phase deviation, no matter how large it grows. Infinite phase deviation accumulates only after infinite time. Because of the infinity, integrated tracked

phase noise is not a useful concept. Integrated *untracked* phase noise θ_u is a much more useful concept:

$$\sigma_{\theta u}^2 = \int_0^\infty W_{u\phi}(f)\,df = \int_0^\infty W_\phi(f)|E(f)|^2\,df \qquad \text{rad}^2 \tag{7.17}$$

Equation (7.17) describes the mean-squared tracking error caused by phase noise on the input signal and on the VCO with a combined spectrum $W_\phi(f)$.

7.5.2 Excessive Phase Noise

Untracked phase noise causes a stress upon PLL tracking. Excessive untracked phase noise will cause cycle slips or even complete loss of lock. Not much information exists about permissible bounds on untracked phase noise. A very rough idea of the boundaries of excessive stress may be inferred from the better known results for additive white noise as laid out in Chapter 6. Equation (6.25), derived on the basis of linear operations, related phase variance to the signal-to-noise ratio SNR_L by $\sigma_{\theta no}^2 = 1/(2\text{SNR}_L)$ rad^2. Then Section 6.2 stated that the phase variance is noticeably larger than predicted by this formula if $\text{SNR}_L < 2.5$ (i.e., 4 dB) and a PLL typically loses lock for $\text{SNR}_L \approx 1$ (0 dB). Formula (6.25) gives phase variances of 0.2 rad^2 (26° rms) for an SNR_L value of 4 dB and 0.5 rad^2 (40° rms) at an SNR_L value of 0 dB. Similar values from (7.17) for untracked phase jitter due to phase noise should be taken as warnings of highly unacceptable operation.

7.5.3 Effect on Coherent Demodulation

Phase noise is a serious problem in numerous communications systems. The problem worsens as ever-higher carrier frequencies are employed in an effort to find an unoccupied radio spectrum. Stress on the PLL is not the only ill effect of untracked phase noise or even the most important effect. Amounts of untracked phase noise that are tolerable from the standpoint of loop stress might induce unacceptably high rates of decision errors in a receiver with a coherent demodulator. Effects of phase noise are especially damaging in larger closely packed signal constellations. Phase noise needs an allowance in a system's performance budget and has to be evaluated to assure that it is within the budget.

7.5.4 Bandwidth Trade-off

Observe that $E(f)$ has a highpass frequency response. An increase in PLL bandwidth K shifts the highpass corner to a higher frequency and reduces the integrated untracked phase jitter. This is opposite to the effect of bandwidth on phase jitter caused by additive noise as given by (6.19), wherein additive noise

is suppressed by the lowpass filter action of $H(f)$. The total phase jitter caused by additive noise and phase noise together takes the form

$$\sigma_{\theta o}^2 = \int_0^\infty W_{n'}(f)|H(f)|^2\,df + \int_0^\infty W_\phi(f)|E(f)|^2\,df \qquad (7.18)$$

There is a choice of loop parameters that minimizes the total phase jitter of (7.18). For a specialized example, consider a narrowband second-order type 2 PLL subjected to white additive noise and predominantly h_3/f^3 phase noise. Under these restricted (but realistic in some applications) conditions, analysis [7.16, 7.17] shows that the contribution of phase noise to the integrated phase jitter is minimized for any noise bandwidth B_L if ζ is set at 1.14. Since the contribution due to the additive white noise depends solely on B_L [see (6.23)] and not on ζ, selection of $\zeta = 1.14$ is the optimum damping for the stated conditions. To find the optimum B_L, substitute the following into (7.18): The phase-jitter contribution due to additive white noise in terms of SNR and B_L as defined in (6.23); the square-root approximation to (7B.4) in Appendix 7B, defining the contribution of h_3/f^3 phase noise to integrated phase jitter; the expression from Table 6.1 relating B_L to K and ζ; and $\zeta = 1.14$. Differentiate on B_L, set the derivative to zero, and solve to obtain the optimum noise bandwidth as $B_L \approx (15h_3P_s/W_0)^{1/3}$ Hz.

The phase-noise spectrum is rarely as simple in most other applications. Analytical optimization is usually infeasible, so numerical integration and search for the minimum are normally required. A spreadsheet is a convenient tool for assembling the requisite data and finding the optimum parameters by trial and error.

7.5.5 Integration

Integrated untracked phase noise can be determined formally by inserting the terms of (7.7) and an expression for $|E(f)|^2$ into (7.17) and evaluating the integral. Appendix 7B provides an example for a second-order type 2 PLL for each individual term of (7.7). The results, although not necessarily directly applicable to real-world situations in which the phase-noise spectrum is more complicated, provide useful insight into the tracking capabilities of PLLs and the dependence of integrated untracked phase noise on PLL parameters. Approximate integration of phase noise as specified in actual hardware can be carried out with the aid of spreadsheets, using equations described in Appendixes 7C and 7D.

Adjacent-channel interference caused by combined phase- and amplitude-noise sidebands is best evaluated from the RF spectrum of the interferer $W_I(f)$ and the RF-referred frequency response $X(f)$ of the victim receiver. The interfering power in the passband of the receiver is

$$P_I = \int_0^\infty W_I(f)|X(f)|^2\,df \qquad \text{watts} \qquad (7.19)$$

where the subscript I refers to the interferer. If the carrier frequencies of the interferer and victim are sufficiently far apart, amplitude noise might be comparable to phase noise in the interfering sidebands so that evaluation of phase noise alone would be overly optimistic.

In addition, phase noise originating in the local oscillators of a receiver will spread the spectrum of an adjacent-channel signal, thereby causing interference to a desired signal, even if the transmitted adjacent-channel signal has a spectrum entirely confined outside the receiver's passband. If the normalized passband spectrum of receiver phase noise is $\mathcal{L}_R(\Delta f)$, the resulting interfering power in the receiver's passband is

$$P_I = \int_0^\infty [W_I(f) \otimes \mathcal{L}_R(f - f_I)]|X(f)|^2 \, df \qquad \text{watts} \qquad (7.20)$$

where $\otimes$ denotes convolution, f_I is the carrier frequency of the interferer, and the subscript R refers to the receiver.

7.5.6 A Paradox

Provided that a high-frequency cutoff is applied for the h_1 and h_0 spectral terms (see Appendix 7B), the integral of (7.17) converges for all terms of (7.7) for any type 2 or higher PLL. However, the integral diverges on the h_3 and h_4 terms for any type 1 PLL; the formal procedure predicts that type 1 PLLs will have infinite untracked phase jitter and so will lose lock. Since a perfect integrator cannot be built in analog circuits, all analog PLLs are type 1. Experience of many years and innumerable successful PLLs has confirmed that analog PLLs phaselock very well and do not lose lock except under seriously adverse conditions, despite the theory. What is one to think when theory and observed behavior diverge so drastically?

Caveats regarding nonstationarity of phase noise and the meaning of phase-noise spectra were brought out in previous sections. [All components of phase noise in (7.7), except that for h_0, are nonstationary. In addition, the first increments of h_4/f^4 and h_3/f^3 are nonstationary.] Therefore, one possibility is that the theory developed by ignoring these caveats may be wrong and the predicted inability to lock of a type 1 PLL may be solely a consequence of a deficient theory.

Another possibility is that the theory is correct but its interpretation is wrong. Since phase noise is accumulative, the theory may merely be saying that the mean-squared integrated phase error in a type 1 PLL is not stationary but grows over time. The growth rate in practical PLLs may be so slow that the time to unlock is too long to be observed. This explanation was suggested by Gray and Tausworthe [7.16] and expanded on by Egan [7.18].

Neither of the explanations advanced above give much comfort to a practicing engineer who has to design a working PLL. Several expedients have been employed to circumvent the paradox:

- If the PLL has an imperfect integrator in its loop filter, just pretend that it is perfect. The integrals then converge and all is well. This expedient has

served rather well since so many practical PLLs contain an approximate integrator in their loop filters.

- Ignore the h_3 and h_4 constituents of phase noise; the integrals converge for the remaining constituents (provided that a high-frequency cutoff is applied for the h_1 and h_0 terms). It is reasonable to ignore the h_4 term for most PLLs, but it is hazardous to ignore the h_3 term, particularly in PLLs with small bandwidths.
- Apply a low-frequency cutoff to the problem; that is, set the lower limit of integration at some low frequency greater than zero. This is a common, though hazardous expedient, especially tempting when specifications or experimental data have a lower-frequency limit. It is particularly hazardous for PLLs with small bandwidths. This issue is addressed further as part of Appendix 7C. Recognize that there is no theoretical or experimental evidence for existence of a low-frequency easing of the $1/f^3$ shape in the spectra of most oscillators [7.7].

7.5.7 Integration of Spectral Lines

The integration formulas above for phase-noise modulation were developed on the basis of a continuous phase-noise spectral density $W_\phi(f)$. Section 7.3.3 warned that a phase-noise analyzer may not be able to distinguish a discrete spectral line from the continuous spectrum and so may report a wrong value for the power in the discrete line because of improper calibration. Remarkably enough, if the analyzer fails to distinguish the discrete line from the continuous spectrum, integration of the reported spectral density, including the mischaracterized calibration, might yield the correct value for integrated phase noise. See Appendix 7D for details.

Power in discrete lines in an RF spectrum generally are reported correctly, provided that resolution bandwidth is broad enough to contain substantially all of the power in a line. (To check whether the bandwidth is large enough, double RBW and see if the power ascribed to the line increases significantly.) It is the continuous RF spectrum that usually has to be scaled by the user, as explained in Section 7.2.3. To determine integrated sideband power (which includes both phase noise and amplitude noise), first separate out all of the discrete lines and just add their powers. (Add mW, not dBmW.) Then integrate the remaining continuous spectrum, taking proper calibration into account. For either kind of spectrum, RF or demodulated phase noise, the integral has to include a weighting function, such as $|E(f)|^2$ as in (7.17) or $|X(f)|^2$ as in (7.19) or (7.20), to confine the integration to the frequencies of interest.

7.5.8 Phase-Noise Specifications

Constraints on phase noise often have to be incorporated into a formal specification. One common approach is to specify $10\log[W_\phi(f)]$ in dBc/Hz for a single value of frequency offset f. (Actually, the specification is usually

stated in terms of $10\log[\mathcal{L}(\Delta f)]$ but that ordinarily should be interpreted as $10\log[W_\phi(f)] - 3$ dB.) This scheme of specification is risky. Equipment could meet this specification but still not provide satisfactory performance. The scheme places no restrictions on the shape of the phase-noise spectrum and makes no allowance whatever for discrete spectral components. It is an underspecification and generally should be avoided, despite its widespread use.

Another approach is to specify a phase-noise mask on $W_\phi(f)$, a format appropriate for catalog data on oscillators and synthesizers. That is much safer than the single-point specification, but it too has a couple of deficiencies for specifications of systems: (1) a mask is often an overspecification in that it may place unnecessarily stringent constraints on a supplier, thereby increasing costs, and (2) it does not readily accommodate discrete spectral components.

A specification of integrated phase noise avoids these perils of under- and overspecification. All pertinent phase noise is taken into account, not just the spectral density at only one offset frequency. Details of the spectral shape are immaterial and so are omitted from the specification. Discrete spectral lines are included automatically. The specification consists of a limit on integrated phase noise variance plus the characteristics of a weighting filter in which the phase noise is measured. In the case of adjacent-channel interference, the frequency spacing between the desired signal and the interferer is also specified. Phase-noise contributions from different blocks of a system usually combine as the sums of the individual variances; the integrated phase noise of each block can be specified individually when that condition holds. For example, specification for a communications system might call out the allowed mean-squared phase noise plus the type, noise bandwidth, and damping factor of a PLL as well as an upper frequency limit for the integration. Characteristics of the PLL along with the upper frequency limit establish the properties of the integration-weighting filter. Specifications of this kind have been employed over many years for phaselock receivers in space communications systems.

7.6 TIMING JITTER

Rather than phase fluctuations, one often needs to characterize timing fluctuations, commonly called *timing jitter*. Phase and timing fluctuations are closely related, as explained in Appendix 7E, but description of timing exhibits subtleties that have not appeared thus far in description of phase fluctuations. These subtleties, plus the many sources of timing jitter and a diversity of applications, have led to a certain amount of confusion in the literature on the subject. The account in Appendix 7E is restricted to timing jitter arising in oscillators and processed in PLLs. The definitions and results in that appendix closely follow Lee [7.19], with notation altered to conform to that established elsewhere in this book and with simplifications and abridgements. See references in Lee's paper for earlier publications on the subject.

Do not conclude that oscillators are the main sources of timing jitter. Other sources rarely include the h_2/f^2, h_3/f^3, and h_4/f^4 spectral terms arising from

phase accumulation in an oscillator, but there are numerous sources that introduce far greater jitter than any decent oscillator. Some are:

- Jitter in digital telecommunications land lines, caused by multiplexers and demultiplexers that stuff pulses or adjust pointers [7.20–7.25]
- Jitter caused by additive noise (Chapter 6)
- Self-noise, a form of intersymbol interference [7.26, 7.27]
- Buildup of jitter in a chain of data repeaters [7.28–7.33]
- Additive interference (co-channel or adjacent channel); crosstalk
- Internal pickup from other nearby circuits in the same system, especially switching of digital circuits

APPENDIX 7A: ANALYSIS OF INTERFERENCE IN A HARD LIMITER

Input signal plus interference:

$$\begin{aligned} x(t) &= A\cos 2\pi f_o t + B\cos 2\pi (f_o + \Delta f)t \\ &= A\cos 2\pi f_o t + B[\cos 2\pi f_o t \cos 2\pi \Delta f t - \sin 2\pi f_o t \sin 2\pi \Delta f t] \\ &= (A + B\cos 2\pi \Delta f t)\cos 2\pi f_o t - B\sin 2\pi \Delta f t \sin 2\pi f_o t \end{aligned}$$

which will be recognized as combining amplitude and phase modulations. A hard limiter wipes out amplitude modulation and leaves only phase modulation:

$$\begin{aligned} \phi(t) &= \tan^{-1}\frac{B\sin 2\pi \Delta f t}{A + B\cos 2\pi \Delta f t} \\ &\approx \frac{B}{A}\sin 2\pi \Delta f t \qquad \text{if} B \ll A \end{aligned}$$

Denote limiter output by $y(t)$:

$$\begin{aligned} y(t) &= \cos[2\pi f_o t + \phi(t)] \approx \cos\left(2\pi f_o t + \frac{B}{A}\sin 2\pi \Delta f t\right) \\ &= \cos 2\pi f_o t \cos\left(\frac{B}{A}\sin 2\pi \Delta f t\right) - \sin 2\pi f_o t \sin\left(\frac{B}{A}\sin 2\pi \Delta f t\right) \\ &\approx \cos 2\pi f_o t - \sin 2\pi f_o t\left[2\sum_{n=1}^{\infty} J_{2n-1}\left(\frac{B}{A}\right)\sin 2\pi(2n-1)\Delta f t\right] \\ &\approx \cos 2\pi f_o t - \frac{B}{A}\sin 2\pi f_o t \sin 2\pi \Delta f t \\ &= \cos 2\pi f_o t + \frac{B}{2A}\cos 2\pi (f_o + \Delta f)t - \frac{B}{2A}\cos 2\pi (f_o - \Delta f)t \end{aligned}$$

APPENDIX 7B: INTEGRALS OF UNTRACKED PHASE NOISE

A continuous spectrum of phase noise is often well approximated by a sum of terms of the form h_ν/f^ν, as in (7.7), where $\nu = 0$, 1, 2, 3, or 4. The contribution from the νth term to integrated untracked phase noise is given by

$$\sigma_\nu^2 = h_\nu \int_0^\infty \frac{1}{f^\nu} |E(f)|^2 \, df \qquad \text{rad}^2 \tag{7B.1}$$

This appendix lists those integrals for a second-order type 2 PLL for $\nu = 0$ to 4. All integrals converge for a type 2 PLL, provided that a high-frequency cutoff is applied for $\nu = 1$ and 0. [**Comment**: The term for $\nu = 4$ is not often an issue for PLLs; it is listed here for information.]

7B.1 Integration Procedures

An error transfer function with parameters K and ζ was used, as in (2.21). Manipulation of (2.21) yielded

$$|E(f)|^2 = \frac{(4\pi f\zeta)^4}{(8\pi f\zeta^2)^2(K^2 + 4\pi^2 f^2) - 2(4\pi K f\zeta)^2 + K^4} \tag{7B.2}$$

This expression was multiplied by h_ν/f^ν and integrated for each ν by a computer-algebra program. The expressions for $\nu = 1$ and 0 require a high-frequency cutoff to force convergence at the upper limit of integration. Two different cutoffs were applied for these terms: an abrupt cutoff at $f = B$ Hz or a single-pole rolloff with magnitude-squared frequency response $1/(1 + f^2/B^2)$. Results of both are listed.

7B.2 Results of Integrations

Details of the integrations were hidden within the program, so only the results are listed below.

- h_4/f^4 term:

$$\sigma_4^2 = h_4 \frac{16\pi^4\zeta^2}{K^3} \tag{7B.3}$$

- h_3/f^3 term:

$$\sigma_3^2 = \begin{cases} \dfrac{h_3}{K^2} \dfrac{2\pi^2\zeta[\pi - 2\sin^{-1}(2\zeta^2 - 1)]}{\sqrt{1-\zeta^2}}, & \zeta < 1 \\ \dfrac{h_3}{K^2} 8\pi^2, & \zeta = 1 \\ \dfrac{h_3}{K^2} \dfrac{2\pi^2\zeta \ln[(2\zeta\sqrt{\zeta^2-1} + 2\zeta^2 - 1)^2]}{\sqrt{\zeta^2-1}}, & \zeta > 1 \end{cases} \tag{7B.4}$$

Two simple approximations have been found for $\sigma_3{}^2$; they can be used in numerical calculations in place of the complicated (7B.4).

- Square-root approximation:

$$\sigma_3^2 \approx \frac{8\pi^2 h_3}{K^2}\sqrt{\zeta} \qquad \text{rad}^2$$

- Quadratic approximation:

$$\sigma_3^2 \approx \frac{4\pi h_3}{K^2}(1 + 2\pi\zeta - \zeta^2) \qquad \text{rad}^2$$

Both approximations are exact at $\zeta = 1$; the quadratic approximation is also exact at $\zeta \approx 0.75$ and 2. The error in the square-root approximation does not exceed $\pm 7.5\%$ (0.3 dB) for any $\zeta \geq 0.7$, whereas the quadratic approximation is within $\pm 1\%$ for $0.6 < \zeta < 2.25$ and within $\pm 10\%$ for $0.32 < \zeta < 3.2$. Accuracy of the quadratic approximation deteriorates sharply for $\zeta > 3.2$, whereas accuracy of the square-root approximation gradually improves for larger ζ.

- h_2/f^2 term:

$$\sigma_2^2 = \frac{h_2\pi^2}{K} \qquad \text{independent of } \zeta \tag{7B.5}$$

- h_1/f term ($\zeta = 1$, abrupt cutoff):

$$\begin{aligned}\sigma_1^2 &= \frac{h_1}{2}\frac{(16\pi^2B^2 + K^2)\ln\left(\dfrac{16\pi^2B^2 + K^2}{K^2}\right) - 16\pi^2B^2}{16\pi^2B^2 + K^2} \\ &= \frac{h_1}{2}\left\{2\ln\left(\frac{4\pi B}{K}\right) + \ln\left[1 + \left(\frac{K}{4\pi B}\right)^2\right] - \frac{1}{1 + (K/4\pi B)^2}\right\} \\ &\approx h_1\left[\ln\left(\frac{4\pi B}{K}\right) - 1/2\right], \qquad K \ll 4\pi B \end{aligned} \tag{7B.6}$$

- h_1/f term ($\zeta = 1$, one-pole rolloff):

$$\begin{aligned}\sigma_1^2 &= h_1\frac{8\pi^2B^2\left[32\pi^2B^2\ln\left(\dfrac{4\pi B}{K}\right) - 16\pi^2B^2 + K^2\right]}{(16\pi^2B^2 - K^2)^2} \\ &= h_1\left[\frac{1}{1 - (K/4\pi B)^2}\right]\left[\frac{1}{1 - (K/4\pi B)^2}\ln\left(\frac{4\pi B}{K}\right) - 1/2\right] \\ &\approx h_1\left[\ln\left(\frac{4\pi B}{K}\right) - 1/2\right], \qquad K \ll 4\pi B \end{aligned} \tag{7B.7}$$

- h_0 term ($\zeta = 1$, abrupt cutoff):

$$\sigma_0^2 = h_0 \frac{4\pi B\left[2 + 3\left(\frac{K}{4\pi B}\right)^2\right] - 3K\left[1 + \left(\frac{K}{4\pi B}\right)^2\right]\tan^{-1}\left(\frac{4\pi B}{K}\right)}{8\pi\left[1 + \left(\frac{K}{4\pi B}\right)^2\right]}$$

$$\approx h_0 \frac{8\pi B - 3K\pi/2}{8\pi} = h_0\left(B - \frac{3K}{16}\right), \qquad K \ll 4\pi B$$

$$\approx h_0 B, \qquad K \ll 16B/3 \tag{7B.8}$$

- h_0 term (one-pole rolloff):

$$\sigma_0^2 = h_0 \frac{\pi^2 B^2(8\pi B\zeta^2 + K)}{8\pi B\zeta^2(2\pi B + K) + K^2} \tag{7B.9}$$

If $\zeta = 1$:

$$\sigma_0^2 = h_0 \pi B \frac{1 + \frac{K}{8\pi B}}{2 + \frac{K}{\pi B} + \frac{1}{8}\left(\frac{K}{\pi B}\right)^2} \tag{7B.10}$$

$$\approx \frac{h_0 \pi B}{2}, \qquad K \ll \pi B$$

7B.3 Discussion

Results in (7B.6) to (7B.8) are shown only for $\zeta = 1$ because the results for arbitrary ζ are much too elaborate and tangled for display or for ready understanding. Since a damping of $\zeta \approx 1$ is often employed, the expressions for $\zeta = 1$ should be good approximations for many PLLs. The inequalities specified for justification of the approximate results in (7B.6) to (7B.10) will be applicable in most practical situations. In regard to the approximate results in (7B.8) and (7B.10), observe that B is the noise bandwidth (B_N) of an abrupt cutoff lowpass filter and $\pi B/2$ is the noise bandwidth of a one-pole lowpass filter. Moreover, the abrupt cutoff and the one-pole rolloff are extreme instances of lowpass filters; almost any other practical lowpass filter will have properties in between the extremes. Those observations suggest that the integrated untracked phase noise variance due to white phase noise might be well approximated by $h_0 B_N$ without concern for the other characteristics of the lowpass filter.

APPENDIX 7C: NUMERICAL INTEGRATION OF PLL PHASE NOISE

This appendix shows how to calculate the integrated untracked phase noise of a PLL from numerical data obtained from measurements or specifications of

baseband phase-noise spectra. The model is of a PLL with specified transfer function that is beset by phase noise either on the incoming signal or in its own VCO; the calculations are the same for either source of phase noise.

7C.1 Definition and Application of Integrated Phase Noise

Integrated untracked phase noise was defined in (7.17) as

$$\sigma_{\theta u}^2 = \int_0^\infty W_\phi(f)|E(f)|^2\,df \qquad \text{rad}^2 \tag{7C.1}$$

where $W_\phi(f)$ is the one-sided spectral density of the phase-noise source and $E(f)$ is the error response of the PLL. Equation (7C.1) is the integral that is to be approximated numerically. The method described here assumes that error response is described by an algebraic formula and that the phase-noise spectral density is provided as tabular data.

A physical system is likely to have multiple sources of phase noise. The contribution from each source can be calculated individually, to the extent that the system performs only linear operations on the phase modulation (such as linear filtering or such as multiplication or division by a scalar) and to the extent that all sources are uncorrelated. Total integrated phase variance is readily calculated as the sum of variances of the individual contributions. Accordingly, the treatment here deals with a single source. Furthermore, the phase variance caused by additive noise, evaluated separately by formulas (6.19), (6.21) to (6.23), or (6.26), can also be included in the sum. These simple calculations for additive noise are not described further at this point. The minimum-attainable total phase variance (under the constraint of specified phase-noise sources, input signal-to-noise density ratio, and PLL transfer function) can be found from a search over the parameters of the PLL. All calculations are well suited for spreadsheets.

7C.2 Data Formats

Baseband spectral data can represent either the correct phase noise spectrum $W_\phi(f)$ or the mislabeled $\mathcal{L}(\Delta f) = W_\phi(f)/2$. In actuality, phase-noise data almost always are delivered in dB format as $10\log[W_\phi(f)]$ or $10\log[\mathcal{L}(\Delta f)] = 10\log[W_\phi(f)] - 3$ dB. The processing method makes provision for both. Input data are provided as a finite number of entries at discrete frequencies designated f_i and spectral-density data designated $\mathrm{D}(f_i) = \mathrm{D}_i$, where i is a symbolic index on frequency. Datum D_i is either $10\log[W_\phi(f_i)]$ or $10\log[\mathcal{L}(f_i)]$ as dictated by the nomenclature of the origin of the data.

The lowest frequency in the data set is designated f_a and the highest frequency in the set is designated f_b. A data set from a manufacturer's specification might contain only a handful of points, whereas a data set from a phase-noise analyzer might contain thousands of points. There is no implication that frequency points are equally spaced; to the contrary, spacing almost always is nonuniform. Error response of the PLL is most conveniently provided as $\mathrm{E}d_i = 10\log|E(f_i)|^2$ dB

(d is short for dB), a quantity to be evaluated for each frequency f_i involved in the calculations.

7C.3 Data Adjustments

Calculations often require adjustments of spectral data at each data point or of the number of points in the data set.

Spectrum Adjustments Denote an adjusted spectral datum as $\mathrm{W}d_i = \mathrm{D}_i + \mathrm{A}$, where A in dB is an adjustment to be applied to the spectral data; it is the same at all data points. There are several possible constituents of A:

- $\mathrm{A} = 0$ if no adjustment is needed.
- Add 3 dB to A if D_i as furnished is designated $\mathcal{L}(f)$ instead of $W(f)$.
- Add (subtract) 20 log(N) dB to A if the phase-noise source comes from the Nth harmonic (subharmonic) of an oscillator whose phase noise is specified at its fundamental frequency.
- Add 3 dB to A if the system has two uncorrelated phase-noise sources with the same spectra. This situation often arises in communications systems that employ similar local oscillators in both the transmitter and receiver.

Data Set Adjustments Several other adjustments are laid out in the next few paragraphs. The integral (7C.1) has limits of zero and infinity, but the data set does not extend to either limit. A feasible lower limit f_3 and upper limit f_h have to be established for numerical integration. The number of points in the data set may need enhancement or pruning.

Choose an Upper Limit In digital-data communications systems, f_h is commonly chosen as half of the symbol rate. If $f_h < f_b$, simply truncate the data set appropriately. If $f_h > f_b$, the data set has to be extended. A pessimistic extension simply adds one additional data point at frequency f_h with spectral datum $\mathrm{D}(f_h) = \mathrm{D}(f_b)$, that is, a flat extension. Alternatively, an extension that is usually more optimistic evaluates the dB-spectral slope at f_b and extends that to f_h. Extensions of upper-frequency limits are likely to be accompanied by considerable uncertainty regarding the true spectral data in the extension. It is prudent to try to establish best and worst cases to estimate the gravity of the uncertainty.

Treatment of the Lower Limit The lowest frequency f_a in the data set always exceeds zero, but there is no justification for a lower limit of integration other than zero. Moreover, all oscillators exhibit a phase-noise spectrum of h_3/f^3 that predominates over all other spectral components at low-enough frequencies. A reasonable treatment of low frequencies is to set the lower limit of numerical integration at a frequency f_3, the corner frequency below which h_3/f^3 noise predominates. Then the contribution of h_3/f^3 is evaluated analytically according

to the methods of Appendix 7B and simply added to the phase variance resulting from the numerical integration for frequencies starting at f_3.

If the corner frequency f_3 is in the data set, the frequencies below f_3 are truncated from the numerical integration. If no h_3/f^3 slope is evident in the low-frequency region of the data set, the most conservative action is to assign $f_3 = f_a$. If $\mathrm{W}d(f_3)$ is the adjusted spectral datum for the f_3 point, the value of the h_3 coefficient for the analytical integration is calculated as

$$h_3 = f_3^3 \cdot 10^{\mathrm{W}d(f_3)/10} \tag{7C.2}$$

The lower limit of analytical integration is $f = 0$. For simplicity, the analytical integration might as well be carried to an upper limit of infinity since the integral converges at $f = \infty$ and since the other spectral components predominate at frequencies above f_3. For these reasons, extension of analytical integration above f_3 will not have much effect on the total of analytical and numerical calculations.

[**Comment**: A spectrum might contain more than one region of $1/f^3$ behavior. Only the $1/f^3$ region from $f = 0$ to f_3 is to be integrated analytically. Any separate $1/f^3$ region beyond f_3 is to be integrated numerically along with the other spectral components.]

Interpolated Points Spacing of data frequencies is sometimes much too sparse, particularly in data sets extracted from brief specifications. It may be necessary to interpolate additional points into the data set to obtain acceptable results from the numerical integration. Density for interpolation is a matter of judgment for which no firm rules can be laid down. Keep in mind the following considerations: (1) The numerical integration is most accurate in spectral regions that display as straight lines on a log-log graph. Regions of appreciable curvature or sharp changes of slope should be filled in with extra data points. (2) Spectral regions close to the highpass corner frequency of $E(f)$ make an inordinate contribution to the integrated phase noise. (The spectrum of highpass-filtered phase noise commonly has a peak near that corner.) You would do well to provide ample density of data points in the vicinity of the corner frequency.

7C.4 Data Filtering

The effect of filtering is readily calculated by adding $\mathrm{E}d_i$ to the adjusted dB-spectral data at each frequency f_i. Symbolically:

$$\mathrm{U}_i = \mathrm{W}d_i + \mathrm{E}d_i = \mathrm{D}_i + \mathrm{A} + \mathrm{E}d_i \tag{7C.3}$$

7C.5 Numerical Integration

If data points are provided in sufficient density, the graphs of the data on log-log scales will appear as nearly straight lines from one point to the next, as illustrated in Fig. 7C.1. That is, a plot of $\mathrm{U}(f_i) = \mathrm{U}_i$ will closely approximate

$$\mathrm{U}(f) \approx 10[r_i \log(f) + q_i] \qquad \mathrm{dB} \tag{7C.4}$$

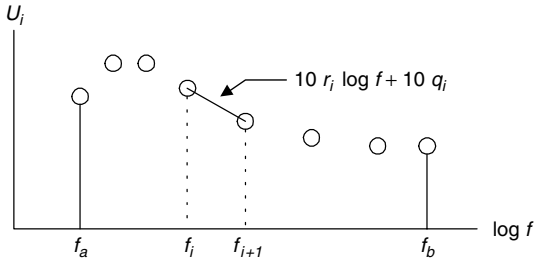


Figure 7C.1 Nomenclature of adjusted data.

where r_i is the log-log slope and q_i is the log-log intercept of the approximation at $f = f_i$.

Integration has to be on linear scales, not logarithmic. To that end, define

$$\mathrm{V}(f) = 10^{\mathrm{U}(f)/10} \approx f^{r_i} \cdot 10^{q_i} \tag{7C.5}$$

Thus, $\mathrm{V}(f)$—the quantity to be integrated—follows a power law in f. The integral in the ith interval is

$$\begin{aligned} \mathrm{I}_i &= \int_{f_i}^{f_{i+1}} \mathrm{V}(f)\, df \\ &\approx \frac{10^{q_i}}{1 + r_i}(f_{i+1}^{r_i+1} - f_i^{r_i+1}), \quad r_i \neq -1 \\ &\approx 10^{q_i} \ln(10) \log \frac{f_{i+1}}{f_i}, \qquad r_i = -1 \end{aligned} \tag{7C.6}$$

Slope for the ith interval is approximated as the forward-divided difference

$$r_i \approx \frac{1}{10} \frac{\mathrm{U}(f_{i+1}) - \mathrm{U}(f_i)}{\log(f_{i+1}/f_i)} \tag{7C.7}$$

and the associated intercept is

$$10^{q_i} \approx \frac{10^{\mathrm{U}(f_i)/10}}{f_i^{r_i}} \tag{7C.8}$$

The total integral I_N from f_3 to f_h is the sum of I_i from $f_i = f_3$ to f_{h-1}. To that must be added the analytically evaluated I_3 for the h_3/f^3 noise plus contributions from additive noise and from integrals of other sources of phase noise.

APPENDIX 7D: INTEGRATION OF DISCRETE LINES IN THE PHASE-NOISE SPECTRUM

Represent one-sided single-frequency phase modulation as the cisoid

$$\phi_d(t) = \frac{\beta}{\sqrt{2}} e^{j2\pi f_d t} \qquad \text{rad} \tag{7D.1}$$

The Fourier transform of this complex waveform is a delta function at $f = f_d$ with area $\beta/\sqrt{2}$. Intensity of the modulation is $|\phi_d(t)|^2 = \beta^2/2$ rad^2, the same intensity as that of a real cosine wave (which has a two-sided Fourier transform). Regard (7D.1) as the output of a phase demodulator in a phase-noise analyzer. That output is applied to a one-sided complex analysis filter whose measurement frequency f_m is swept over $-\infty$ to ∞. Denote the frequency response of the filter as $Y(f - f_m)$. Filter response has its single peak at $f = f_m$ and the filter skirts have properties desirable in a spectrum analyzer. Sweep is assumed to be slow enough that it can be deemed quasistationary; that is, measured filter output is the steady-state value, uncontaminated by transients. Filter response can be expressed in polar format as

$$Y(f - f_m) = |Y(f - f_m)|e^{j\psi(f-f_m)} \tag{7D.2}$$

where ψ is the phase shift of the filter. The bandwidth and shape of the analysis filter is assumed to be the same for all f_m; only its location in frequency changes as f_m is swept. In particular, its peak response $|Y(0)|$ and its noise bandwidth B_N are assumed to be independent of f_m.

If $\phi_d(t)$ is applied as input to the filter, the output is

$$x(t) = \frac{\beta}{\sqrt{2}}|Y(f_d - f_m)|e^{j[2\pi f_d t + \psi(f_d - f_m)]} \tag{7D.3}$$

The final output of the analyzer, following subsequent square-law detection and smoothing, is

$$P_d(f_m) = |x(t)|^2 = \frac{\beta^2}{2}|Y(f_d - f_m)|^2 \tag{7D.4}$$

Integrate $P_d(f_m)$ over all f_m to obtain a measure of the integrated phase noise due to $\phi_d(t)$:

$$\begin{aligned}\int_{-\infty}^{\infty} P_d(f_d - f_m)\,df_m &= \frac{\beta^2}{2}\int_{-\infty}^{\infty}|Y(f_d - f_m)|^2\,df_m \\ &= \frac{\beta^2}{2}|Y(0)|^2 B_N = \sigma_d^2|Y(0)|^2 B_N\end{aligned} \tag{7D.5}$$

where $\sigma_d^2 = \beta^2/2$ rad^2 is the contribution of the line spectrum to the integrated phase noise and $|Y(0)|^2 B_N$ is the calibration factor of the analyzer.

The indication in a phase-noise analyzer resulting from a continuous spectrum $W_\phi(f_m)$ was found in (7.9) to be (with slight adjustment of notation to fit the model of this appendix)

$$P_c(f_m) \approx W_\phi(f_m)|Y(0)|^2 B_N \tag{7D.6}$$

and its integral over all frequencies is

$$\int_{-\infty}^{\infty} P_c(f_m)\,df_m = |Y(0)|^2 B_N \int_0^{\infty} W_\phi(f_m)\,df_m = \sigma_c^2|Y(0)|^2 B_N \tag{7D.7}$$

where σ_c^2 rad^2 is the contribution of the continuous phase-noise spectrum to the integrated phase noise.

Observe that (7D.5) and (7D.7) have the same format—the same calibration factor $|Y(0)|^2 B_N$. Therefore, although the indications for continuous and discrete components in the spectrum display of the analyzer have different calibration factors, integrals of those indications will have the same calibration factors. A user need not be concerned about calibration factor when integrating measured phase noise; the analyzer has taken care of that already if, as is likely, the analyzer does not distinguish between discrete and continuous spectral components.

[**Comments**: (1) A phase-noise analyzer reports its results at a set of discrete frequencies, not a continuum as derived above. The data set has to be dense enough in the vicinity of f_d to provide good definition of $|Y(f_d - f_m)|$. (2) Integration of untracked phase noise in a PLL has to be weighted by $|E(f)|^2$, a factor omitted from the equations above.]

APPENDIX 7E: TIMING JITTER

7E.1 Jitter Definitions

Consider an oscillator with nominal period $T_o = 1/f_o$ whose nth cycle should end at time $t = nT_o$ but the cycle end is actually displaced by jitter to time $t = t_n$. (Think of a cycle end as, for example, a positive-going zero crossing of the oscillator voltage.)

Absolute Jitter Lee [7.19] defines *absolute jitter* by the sequence

$$\{t_n - nT_o\} \tag{7E.1}$$

Absolute phase jitter is the sequence $\{\phi_n = 2\pi f_o(t_n - nT_o)\}$, a sampled version of the phase fluctuation $\phi(t)$ considered heretofore. The variance of the absolute timing jitter is

$$\sigma_A^2 \approx \frac{1}{(2\pi f_o)^2} \int_0^\infty W_\phi(f)\, df \qquad \text{sec}^2 \tag{7E.2}$$

[**Comments**: (1) Lee shows $f_o/2$ as the upper limit of integration since ϕ_n, strictly speaking, is a sequence of discrete values sampled at a rate f_o. The infinite upper limit in (7E.2) is a convenient approximation. (2) Integration of the h_1/f and h_0 spectral components of $W_\phi(f)$ yield an infinite variance if the upper limit of integration is infinite. A finite upper limit is required for these two components; $f_o/2$ is plausible in the absence of any other candidate. Besides, measurements of phase-noise spectrum rarely extend to frequencies higher than a small fraction of $f_o/2$, so the actual spectrum tends to be unknown at higher frequencies. (3) Contributions to the integral of the h_1/f, h_2/f^2, h_3/f^3, and h_4/f^4 spectral components are infinite; the variance of absolute timing jitter of a free-running oscillator is infinite.]

Period Jitter Lee defines *period jitter* by the sequence

$$\{J_n = t_{n+1} - t_n - T_o\} \tag{7E.3}$$

which is the first increment of the absolute jitter. Other names appearing in the literature are *cycle jitter, cycle-to-cycle jitter*, and *edge-to-edge jitter*. No name has yet been standardized at this writing (October 2003). Period jitter is of great concern in high-speed digital circuits, such as computers, where timing margins are critical. Jitter over multiple periods is defined by the sequence

$$\{J_n(kT_o) = t_{n+k} - t_n - kT_o\} \tag{7E.4}$$

Variance of the period jitter is

$$\sigma_J^2(kT_o) \approx \frac{1}{(\pi f_o)^2} \int_0^\infty \sin^2(\pi f k T_o) W_\phi(f)\, df \qquad \text{sec}^2 \tag{7E.5}$$

where the difference $t_{n+k} - t_n$ has inserted the $\sin^2$ factor into the integrand, thereby providing convergence at $f = 0$ for the h_2/f^2 spectral component of $W_\phi(f)$. The zero-frequency singularities of the h_3/f^3 and h_4/f^4 components still make an infinite contribution to the variance of the period jitter of a free-running oscillator; the $\sin^2$ factor is not sufficient to cancel their singularities. The h_1/f and h_0 components still require a finite upper limit of integration for a finite contribution.

Equation (7E.5) was evaluated for the spectral components whose integrals converge, with the following results:

- For white phase noise; upper integration limit $= f_o/2$:

$$\sigma_{J0}^2 = \frac{h_0 T_o}{4\pi^2} \qquad \text{sec}^2 \tag{7E.6}$$

- For $1/f^2$ phase noise; upper integration limit $= \infty$:

$$\sigma_{J2}^2 = \frac{h_2}{2f_o^2} |kT_o| \qquad \text{sec}^2 \tag{7E.7}$$

This last result is often invoked but it applies only to the $1/f^2$ phase-noise component. No simple closed-form result could be found for the period jitter caused by the $1/f$ phase-noise component.

Another Paradox Manufacturers' brochures on crystal oscillators often specify jitter as some number of picoseconds rms. Presumably, this specification is a measurement of period jitter over one period (i.e., $k = 1$). Typically, the measurement is performed with a sampling digital oscilloscope that extracts the statistics of the

first positive zero crossings following the positive zero crossings that trigger the horizontal sweeps. Various measurement setups and results are presented in [7.34].

Measured jitter invariably is very small for a properly functioning high-quality oscillator, but the integral (7E.5) yields infinity for the omnipresent h_3/f^3 spectral component. Why is theory so far from practice? One explanation is the same as that for the purported failure of a type 1 PLL to track the h_3/f^3 spectral component of phase noise as advanced in Section 7.5.6. In short: the h_3/f^3 component is slowly varying in the time domain, so that it contributes very little to the jitter measured over any small time interval. Conversely, the jitter measured should increase if measurements are conducted over long intervals, but the simple theory (based on fictitious stationarity) does not cope with nonstationary behavior.

7E.2 Jitter in PLLs

Integrated timing jitter in a PLL is found in the same manner as in (7.17) for integrated phase noise. That is, insert $|E(f)|^2$ as a factor into the integrand of (7E.2) for absolute jitter:

$$\sigma_A^2 \approx \frac{1}{(2\pi f_o)^2} \int_0^\infty |E(f)^2| W_\phi(f)\, df \qquad \text{sec}^2 \tag{7E.8}$$

or into (7E.5) for period jitter of a locked PLL:

$$\sigma_J^2(kT_o) \approx \frac{1}{(\pi f_o)^2} \int_0^\infty |E(f)|^2 \sin^2(\pi f k T_o) W_\phi(f)\, df \qquad \text{sec}^2 \tag{7E.9}$$

Observe that (7E.8) for absolute timing jitter is the same as expression (7.17) for untracked phase jitter, except that it is divided by $(2\pi f_o)^2$ to convert to time variance from phase variance. Appendix 7B has evaluations of (7.17) for the h_v/f^v spectral components of phase noise in a second-order type 2 PLL; divide these results by $(2\pi f_o)^2$ to find absolute timing jitter and substitute $B = f_o/2$ in the expressions for h_0 and h_1/f noise components. Absolute jitter in seconds squared for a first-order PLL with loop gain K rad/sec is as follows:

- h_0 term, upper limit $= f_o/2$:

$$\begin{aligned} \sigma_{A0}^2 &= \frac{h_0 T_o}{8\pi^2}\left(1 - \frac{K}{\pi f_o}\tan^{-1}\frac{\pi f_o}{K}\right) \approx \frac{h_0 T_o}{8\pi^2}\left(1 - \frac{K}{2f_o}\right) \\ &\approx \frac{h_0 T_o}{8\pi^2} \end{aligned} \tag{7E.10}$$

- h_1/f term, upper limit $= f_o/2$:

$$\sigma_{A1}^2 = \frac{h_1}{8\pi^2 f_o^2} \ln\left[1 + \left(\frac{\pi f_o}{K}\right)^2\right] \tag{7E.11}$$

- h_2/f^2 term, upper limit $= \infty$:

$$\sigma_{A2}^2 = \frac{h_2}{4 f_o^2 K} \tag{7E.12}$$

- h_3/f^3 term:

$$\sigma_{A3}^2 = \infty \tag{7E.13}$$

Result (7E.13) comes about because $|E(f)|^2$ of a first-order PLL does not adequately cancel the zero-frequency singularity of h_3/f^3. This finding is exactly equivalent to the previous finding of infinite integrated untracked phase jitter in a first-order PLL exposed to h_3/f^3 phase noise and is subject to the same skepticism as that expressed in Section 7.5.6.

Substitution of expressions for $|E(f)|^2$ for first-order or for second-order type 2 PLLs leads to integral forms that could not be evaluated readily. Lee [7.19] reports that the period jitter in a first-order PLL subjected to h_2/f^2 phase noise is

$$\begin{aligned} \sigma_{J2}^2 &= 2\sigma_{A2}^2(1 - e^{-K|kT_o|}) \\ &\approx 2\sigma_{A2}^2 K|kT_o|, \qquad |kT_o| \ll 1/K \\ &\approx 2\sigma_{A2}^2, \qquad\quad |kT_o| \gg 1/K \end{aligned} \tag{7E.14}$$

He found similar, although somewhat more complicated behavior for period jitter in a second-order type 2 PLL subjected to h_2/f^2 phase noise.

Notice that the $\sin^2$ factor in (7E.9) inserts two zeros at $f = 0$, and the $|E(f)|^2$ factor inserts at least another two, even for a type 1 PLL. That minimum of four zeros in the weighting functions is sufficient to cancel all of the singularities in h_3/f^3 or h_4/f^4 phase noise, so that the integral in (7E.9) should converge at $f = 0$ for all PLLs and all spectral components of phase noise. So here is a corollary of the paradoxes identified earlier; existing theory predicts infinite absolute timing jitter for a type 1 PLL, but finite period jitter.

REFERENCES

7.1 V. F. Kroupa, ed., *Frequency Stability: Fundamentals and Measurement*, Reprint Volume, IEEE Press, New York, 1983.

7.2 A. Demir, "Phase Noise and Timing Jitter in Oscillators with Colored-Noise Sources," *IEEE Trans. Circuits Syst. I* ***49***, 1782–1791, Dec. 2002.

7.3 D. Ham and A. Hajimiri, "Virtual Damping and Einstein Relation in Oscillators," *IEEE J. Solid-State Circuits* ***38***, 407–418, Mar. 2003.

7.4 L. S. Cutler and C. L. Searle, "Some Aspects of the Theory and Measurement of Frequency Fluctuations in Frequency Standards," *Proc. IEEE* ***54***, 136–154, Feb. 1966. Reprinted in [7.1].

7.5 D. W. Allan, J. H. Shoaf, and D. Halford, "Statistics of Time and Frequency Data Analysis," in B. E. Blair, ed., *Time and Frequency: Theory and Fundamentals*, Natl.

Bur. Std. Monogr. 140, U.S. Department of Commerce, Washington, DC, 1974, Chap 8.

7.6 J. A. Barnes et al., "Characterization of Frequency Stability," *IEEE Trans. Instrum. Meas.* ***IM-20***, 105–120, May 1971. Reprinted in [7.1].

7.7 W. R. Attkinson, L. Fey, and J. Newman, "Spectrum Analysis of Extremely Low Frequency Variations of Quartz Oscillators," *Proc. IEEE* ***51***, 379, Feb. 1963. Reprinted in [7.1].

7.8 G. W. Wornell, "Wavelet-Based Representations for the $1/f$ Family of Fractal Processes," *Proc. IEEE* ***81***, 1428–1450, Oct. 1993.

7.9 G. W. Wornell, *Signal Processing with Fractals: A Wavelet-Based Approach*, Prentice Hall, Upper Saddle River, NJ, 1996.

7.10 P. Flandrin, "On the Spectrum of Fractional Brownian Motion," *IEEE Trans. Inf. Theory* ***IT-35***, 197–199, Jan. 1989.

7.11 W. B. Davenport and W. L. Root, *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill, New York, 1958, Secs. 6-4 and 8-5.

7.12 J. Salz, "Coherent Lightwave Communication," *AT&T Tech. J.* ***64***, 2153–2209, Dec. 1985.

7.13 J. R. Barry and E. A. Lee, "Performance of Coherent Optical Receivers," *Proc. IEEE* ***78***, 1369–1394, Aug. 1990.

7.14 A. Demir, A. Mehrotra, and J. Roychowdhury, "Phase Noise in Oscillators: A Unifying Theory and Numerical Methods for Characterization," *IEEE Trans. Circuits Syst. I* ***47***, 655–674, May 2000.

7.15 A. Mehrotra, "Noise Analysis of Phase-Locked Loops," *IEEE Trans. Circuits Syst. I* ***49***, 1309–1316, Sept. 2002.

7.16 R. M. Gray and R. C. Tausworthe, "Frequency-Counted Measurements and Phase Locking to Noisy Oscillators," *IEEE Trans. Commun.* ***COM-19***, 21–30, Feb. 1971.

7.17 F. M. Gardner, *Phaselock Techniques*, 2nd ed., Wiley, New York, 1979, p. 104.

7.18 W. F. Egan, *Phase-Lock Basics*, Wiley, New York, 1998, Sec. 11.5.

7.19 D. C. Lee, "Analysis of Jitter in Phase-Locked Loops," *IEEE Trans. Circuits Syst. II* ***49***, 704–711, Nov. 2002.

7.20 D. L. Duttweiler, "Waiting Time Jitter," *Bell Syst. Tech. J.* ***51***, 165–208, Jan. 1972.

7.21 D. Choi, "Waiting Time Jitter Reduction," *IEEE Trans. Commun.* ***37***, 1231–1236, Nov. 1989.

7.22 H. Sari and G. Karam, "Cancellation of Pointer Adjustment Jitter in SDH Networks," *IEEE Trans. Commun.* ***42***, 3200–3207, Dec. 1994.

7.23 K. Murakami, "Jitter in Synchronous Residual Time Stamp," *IEEE Trans. Commun.* ***44***, 742–748, June 1996.

7.24 K. Murakami, "Waveform Analysis of Jitter in SRTs Using Continued Fractions," *IEEE Trans. Commun.* ***46***, 819–825, June 1998.

7.25 S. Bregni, *Synchronization of Digital Telecommunications Networks*, Wiley, Chichester, West Sussex, England, 2002, Chap. 3.

7.26 L. E. Franks and J. P. Bubrouski, "Statistical Properties of Timing Jitter in a PAM Timing Recovery Scheme," *IEEE Trans. Commun.* ***22***, 913–920, July 1974.

7.27 F. M. Gardner, "Self-Noise in Synchronizers," *IEEE Trans. Commun.* ***28***, 1159–1163, Aug. 1980.

7.28 C. J. Byrne, B. J. Karafin, and D. B. Robinson, "Systematic Jitter in a Chain of Digital Regenerators," *Bell Syst. Tech. J.* ***42***, 2679–2714, Nov. 1963.

7.29 U. Mengali and G. Pirani, "Jitter Accumulation in PAM Systems," *IEEE Trans. Commun.* ***28***, 1172–1183, Aug. 1980.

7.30 Y. Takasaki, *Digital Transmission Design and Jitter Analysis*, Artech House, Norwood, MA, 1991.

7.31 P. R. Trischitta and E. L. Varma, *Jitter in Digital Transmission Systems*, Artech House, Norwood, MA, 1989.

7.32 H. Meyr, L. Popken, and H. R. Mueller, "Synchronization Failures in a Chain of PLL Synchronizers," *IEEE Trans. Commun.* ***34***, 436–445, May 1986.

7.33 M. Moeneclaey, S. Starzak, and H. Meyr, "Cycle Slips in Synchronizers Subject to Smooth Narrow-Band Loop Noise," *IEEE Trans. Commun.* ***COM-36***, 867–874, July 1988.

7.34 J. A. McNeil, "Jitter in Ring Oscillators," *IEEE J. Solid-State Circuits* ***32***, 870–879, June 1997.

CHAPTER 8

ACQUISITION OF PHASELOCK

In all the preceding chapters, it was assumed that the loop was already in lock. But a loop starts out in an unlocked condition and must be brought into lock, either by its own natural actions or with the help of auxiliary circuits. The process of bringing a loop into lock, called *acquisition*, is the subject of this chapter.

8.1 CHARACTERIZATION

If the loop acquires lock by itself, the process is called *self-acquisition* and if it is assisted by auxiliary circuits, the process is called *aided acquisition*. Self-acquisition can be a slow and unreliable process. Although a PLL is an excellent tracking device, it tends to be rather clumsy in acquisition. Therefore, acquisition-aid circuits are commonly used and it is not unusual to find them constituting half of the total circuitry in representative PLLs.

A type n PLL contains n integrators. Each integrator is either perfect, as in the VCO or a digital integrator, or imperfect, as in an analog integrator. With each integrator there is associated a state variable of the loop: phase, frequency, frequency rate, and so on. To bring the loop into lock, it is necessary to set each of the state variables (i.e., each of the integrators) into close agreement with the corresponding conditions of the input signal. Therefore, a designer must plan for phase acquisition, frequency acquisition, and so on, up to n forms of acquisition for a type n loop. Frequency acquisition has received the most attention, but the other state variables are also important, sometimes critically so. Acquisition

Phaselock Techniques, Third Edition, by Floyd M. Gardner

is inherently a nonlinear phenomenon; nonlinear analysis is needed generally, without easy help from linear approximations.

8.2 PHASE ACQUISITION

Phase ordinarily is self-acquired. Study of phase acquisition leads to better understanding of the overall acquisition problem and provides guidance if aided phase acquisition is needed.

8.2.1 First-Order Loop

It is instructive to begin with analysis of a first-order loop. To show performance, the nonlinear differential equation of the loop is derived and its meaning examined. Let ω_i be the input frequency (assumed constant) to the PLL and let ω_o be the free-running frequency of the VCO so that the instantaneous frequency of the VCO is $\omega_o + K_o v_d$. Voltage $v_d = K_d \sin\theta_e$ is the error voltage out of the phase detector, applied directly to the VCO without intervening filtering. A phase detector with a sinusoidal s-curve is assumed; somewhat different results obtain with other shapes of s-curve.

The input phase is $\omega_i t$ and the oscillator phase is

$$\begin{aligned}\theta_o(t) &= \omega_o t + \int_0^t K_o v_d(\tau)\, d\tau + \theta_o(0) \\ &= \omega_o t + \int_0^t K_o K_d \sin\theta_e(\tau)\, d\tau + \theta_o(0)\end{aligned} \tag{8.1}$$

where the loop gain in rad/sec is $K_o K_d = K$ and the phase error θ_e is

$$\theta_e = \theta_i - \theta_o = (\omega_i - \omega_o)t - \int_0^t K \sin\theta_e(\tau)\, d\tau - \theta_o(0) \tag{8.2}$$

Let $\Delta\omega = \omega_i - \omega_o$ and differentiate (8.2) to obtain

$$\frac{d\theta_e(t)}{dt} = \Delta\omega - K \sin\theta_e(t) \tag{8.3}$$

This is the nonlinear differential equation of the first-order phaselock loop. By definition, $d\theta_e/dt$ will be zero if the loop is in phaselock equilibrium. However, is the converse true? That is, is the loop necessarily phaselocked correctly if $d\theta_e/dt = 0$? That question is pursued in the next few paragraphs.

Before proceeding, note that the hold-in limit (see Section 5.2.2) is obtained directly from (8.3); if $d\theta_e/dt = 0$; then $\sin\theta_e = \Delta\omega/K$. Because $\sin\theta_e$ cannot exceed unity, the loop can lock only if $|\Delta\omega| < K$.

To continue on the question of correct phaselocking, it is useful to divide (8.3) by K and then plot the normalized (8.3) as in Fig. 8.1. (This analysis follows a similar one by Viterbi [8.1, 8.2]. Figure 8.1 is a degenerate phase-plane portrait.)

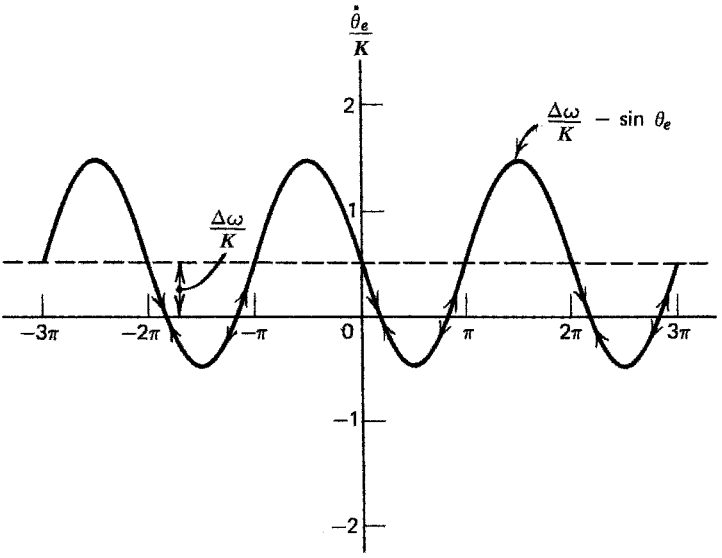


Figure 8.1 Phase-plane plot of a first-order PLL ($\Delta\omega/K = 0.5$).

From the figure it may be seen that if $|\Delta\omega| < K$, there are two points (nulls) in each interval of 2π for which $d\theta_e/dt$ goes to zero. The frequency difference between input and VCO is zero at a null.

Adjacent nulls are of opposite slope. To analyze the behavior of the loop, consider the operating point as slightly displaced from one of the nulls. For a null of negative slope the sign of $d\theta_e/dt$ drives θ_e toward the null. (As an example, if phase displacement is slightly to the left of a negative-slope null, the sign of $d\theta_e/dt$ is positive and θ_e must necessarily increase, that is, move toward the null.) Conversely, a displacement from one of the positive-slope nulls will drive the state of the loop away from the null. Thus, negative-slope nulls are stable and positive-slope nulls are unstable. Arrows in Fig. 8.1 show the direction of phase change.

Prior to lock $d\theta_e/dt$ is nonzero, which means that θ_e must change (increase or decrease) monotonically. For this reason, θ_e must eventually take on the value of one of the stable nulls (provided, of course, that $|\Delta\omega| < K$). When θ_e reaches a stable null, the loop is locked and θ_e remains fixed at the static error. Because every cycle has a stable null, θ_e cannot change by more than one cycle before locking. Thus, there is no cycle skipping in the lock-up process. The time required to approach a null depends on the initial values of phase and frequency, but as a rough rule of thumb, it will be on the order of $3/K$ sec.

The exact settling time can be found [8.3] by integration of the differential equation (8.3). (Exact closed-form integration is possible for a first-order loop but not for second or higher orders.) Some example phase transients are shown in Fig. 8.2 for $\Delta\omega = 0$ and several values of $\theta_e(0)$. If θ_e is small, the loop operation is almost linear and the phase-error waveforms are nearly exponentials with time

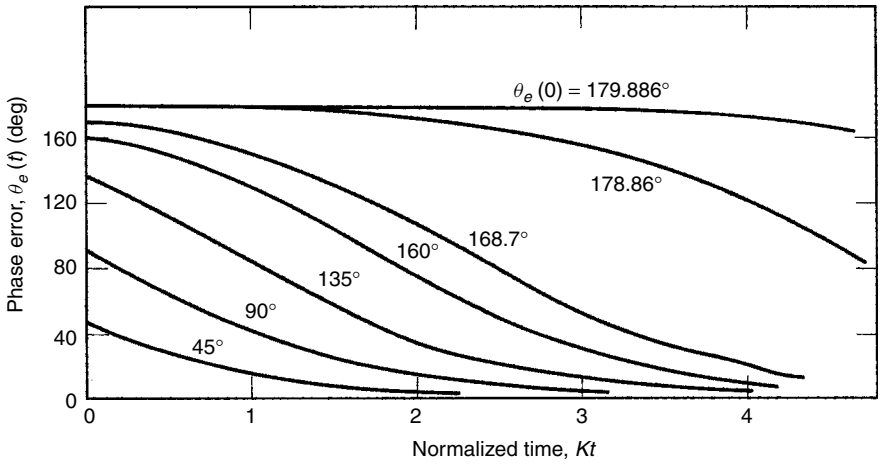


Figure 8.2 Transient phase errors in a first-order PLL, illustrating hang-up.

constant $1/K$. If θ_e is large, the waveforms diverge substantially from a simple exponential, and settling times increase from those attained by an exponential of the same initial phase error.

8.2.2 Hang-up

If the initial phase error is very close to an unstable null, the phase can dwell near the null for an extended time, as illustrated by the two upper curves of Fig. 8.2. This dwell phenomenon, dubbed the *hang-up effect* [8.4], can be extremely troublesome in applications where rapid acquisition is needed with high reliability. Hang-up is illustrated in Fig. 8.2 for a noise-free first-order loop with sinusoidal phase detector and zero frequency error. Despite intuitive notions to the contrary, changing any or all of these conditions does not eliminate hang-up. Specifically, hang-up is aggravated by noise or other disturbances; second- or higher-order loops are equally subject to hang-up; using an extended phase-detector characteristic (e.g., sawtooth) can alleviate hang-up but not necessarily eliminate it; and offsetting the frequency merely shifts the location of the unstable null, as shown in Fig. 8.1. One hang-up-free phase detector is known: the phase-frequency detector (PFD) of Chapter 10. The full causes of hang-up, its statistics, and some anti-hang-up proposals are presented in [8.4], [8.5], [8.6, Chap. 4], and [8.7].

8.2.3 Lock-in

If signal frequency is close enough to VCO frequency, a PLL locks up with just a phase transient; there is no cycle slipping prior to lock. The frequency range over which the loop acquires phase without slips is called the *lock-in range* of the PLL. In a first-order loop, the lock-in range is equal to the hold-in range; the loop self-acquires any signal that it can hold. The same is not true

of type 2 or higher loops; the lock-in range is invariably less than the hold-in range. Moreover, there is a frequency interval, smaller than the hold-in interval and larger than the lock-in interval, over which the loop will acquire lock after slipping cycles for awhile. This intermediate interval, called the *pull-in range*, is discussed in Section 8.3.

Lock-in, self-acquisition of phase by a PLL, is the subject of this section. The proportional-plus-integral loop filter for the familiar second-order type 2 PLL has one pole (at $s = 0$) and one zero (at $s = -1/\tau_2$). The filter's amplitude response is asymptotically flat at high frequencies, as sketched in Fig. 8.3. Denote the high-frequency asymptotic response of the filter by $F(\infty)$. At high frequencies the loop is indistinguishable from a first-order loop with gain $K = K_d K_o |F(\infty)|$. As a fair approximation, the type 2 loop has the same lock-in range as a first-order loop with the same gain K.

The lock-in limit of a first-order loop is equal to K. The argument here is that the same limit,

$$|\Delta\omega_L| = K \tag{8.4}$$

is a useful, though crude engineering approximation for the lock-in range for a PLL of higher order and type. The lock-in limit (8.4) is obtained under the assumption of a sinusoidal phase-detector characteristic. An extended PD characteristic (as in Fig. 5.13) would extend the lock limit.

The argument leading to the approximate lock-in range is a simplification of the real behavior of a PLL. In a higher-order or higher-type PLL, it is not possible to determine whether the loop will or will not slip cycles, before locking, on the basis of initial frequency error alone; all initial state variables must be examined. In a second-order type 2 loop, the variables are frequency and phase; they are studied with the aid of a phase-plane portrait.

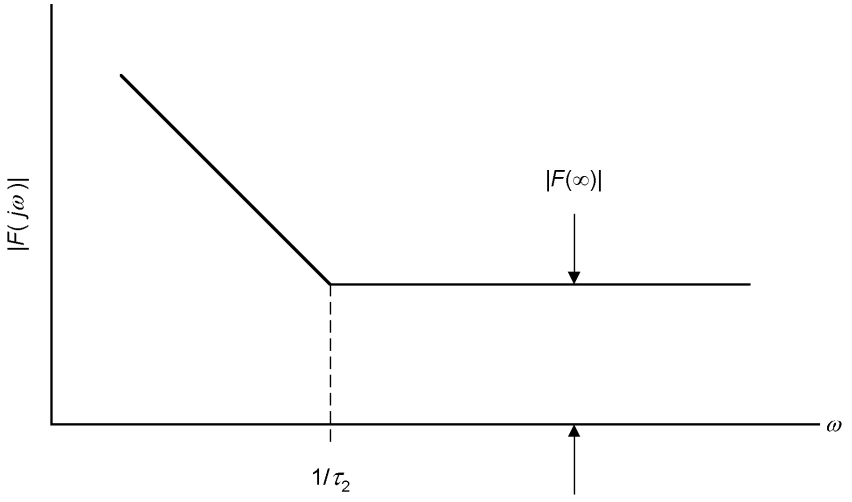


Figure 8.3 Amplitude response of a proportional-plus-integral loop filter.

When inspecting the phase plane (e.g., Fig. 5.14), it is immediately apparent that the entire concept of lock-in is oversimplified. A second-order loop locks without slips if the initial state falls between the separatrices. Since a separatrix is a sinuous boundary, there is no natural way to define exactly any unique lock-in frequency. One might arbitrarily define the average ordinate of the positive separatrix as the lock-in frequency; or, the definition might be the separatrix ordinate at $\theta_e = 0$ or $\theta_e = -180^\circ$. Examination of Fig. 5.14, or the more numerous set of portraits in [8.1], suggests that (8.4) is a conservative estimate of lock-in range. Despite its vague reality, lock-in range is a useful concept for engineering calculations and in analyses presented in later paragraphs.

8.2.4 Aided Phase Acquisition

Unless hang-up is a problem, phase usually is self-acquired if the phase detector has any of the usual characteristics (e.g., Fig. 5.13). However, there are some signal types for which the phase-detector characteristic has only a small active region; over most of the phase-error interval, the s-curve is zero. An example is shown in Fig. 8.4. A pseudorandom noise (PRN) signal is one kind that yields such a PD characteristic [8.9]; a gated pulse train is another. The phase detector for the latter might be a radar range gate.

A loop of this sort can acquire phase only if the initial phase error falls into the active region of the PD. If the initial error lies in the dead portion of the PD characteristic, no error information of any kind is available to the loop, so acquisition can occur only by accident of phase drift. Likelihood of acquisition would be very poor if the PRN code were long or if the pulse duty cycle were short. To acquire the signal, the equipment performs a phase search over all phases. When the active region of the PD is encountered, the loop is supposed to lock and the search is supposed to be discontinued. Application of a phase search constitutes aided acquisition of phase.

A continuous phase sweep is the same as a frequency offset in the VCO and is usually an easy way to implement a phase search. If the phase rate (frequency offset) is too large, the search will sweep right through the active region without stopping and go on into the next dead region. There is a rate limit that must not be exceeded if acquisition is to be successful. Acquisition with a second-order type 2 loop is analyzed by means of a phase-plane portrait. Gilchriest [8.10]

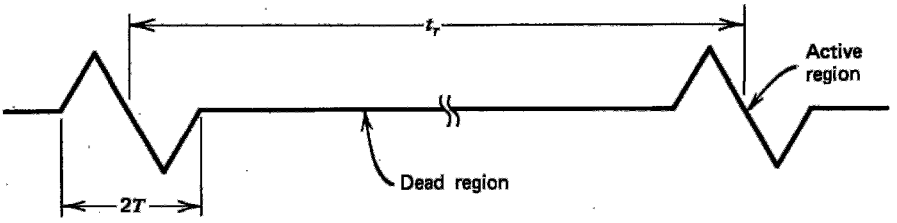


Figure 8.4 Phase-detector s-curve for a short pulse or a PRN signal.

has investigated the PRN signal and Gardner (unpublished) has examined the gated pulse train. For a PD characteristic of the kind shown in Fig. 8.4 and for damping factors of 0.75 or greater, they have found that the maximum phase rate is given by

$$\Delta f \approx B_L \delta \qquad \text{Hz} \tag{8.5}$$

where B_L is the noise bandwidth defined in Chapter 6 and δ is the duty ratio for a pulse system or the chip/code-period ratio of a PRN signal. As might be expected, changing the shape of the PD characteristic has substantial influence on the allowable phase sweep rate. Moreover, smaller damping reduces the allowable rate.

8.3 FREQUENCY ACQUISITION

Acquisition of frequency ordinarily is more difficult, is slower, and requires more design attention than does phase acquisition. In consequence, the literature has concentrated largely on frequency acquisition, to the point that "acquisition" is almost synonymous with "frequency acquisition." Furthermore, the study of frequency acquisition has been devoted mainly to the second-order type 2 loop, partly because of its technological importance, but also because of the greater difficulties of analyzing higher-type loops. Discussion in this section concentrates mostly on second-order type 2 loops.

Self-acquisition of frequency is known as *frequency pull-in*, or simply, *pull-in*. Pull-in tends to be slow and often unreliable, so a number of aided frequency-acquisition techniques have been devised, including frequency sweeping, frequency discriminators, and bandwidth-widening methods.

8.3.1 Frequency Pull-in

Pull-in, particularly in a loop with very narrow bandwidth, is fascinating to watch. When the signal is first applied, the loop is not locked and only a beat note at frequency $\Delta\omega = \omega_i - \omega_o$ appears at the output of the PD, where ω_i is the frequency of the input signal and ω_o is the frequency of the VCO. The frequency of the beat note decreases slowly—the VCO frequency slowly approaches that of the signal—until the lock limit is reached, whereupon the loop snaps into lock without further cycle slipping.

Description of Pull-in Pull-in behavior may be understood by recognizing that the beat note is reduced in amplitude by the loop filter but is not suppressed completely. An attenuated beat note with peak amplitude $K_d|F(j\Delta\omega)|$ is applied to the VCO control terminal, causing the VCO to be frequency modulated at the beat frequency. (Throughout this analysis it is assumed that the PD is a multiplier with a sinusoidal s-curve and the loop filter has constant response at high frequencies, as in Fig. 8.3.) Therefore, the PD output is the low-frequency multiplier product of a sine wave and a frequency-modulated wave. Since the

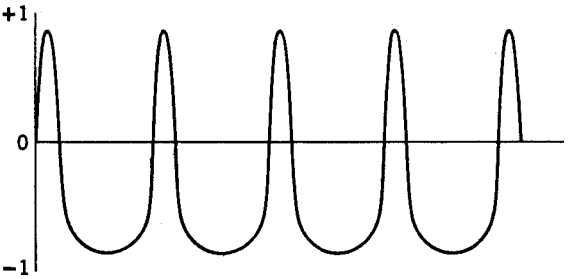


Figure 8.5 Typical beat-note waveshape, first-order PLL, $\Delta\omega/K = 1.10$.

modulating frequency is equal to the beat frequency, the beat-note waveform could hardly be sinusoidal.

Richman [8.3] has derived the waveform of the beat note for a first-order loop by integrating the differential equation (8.3) of the loop. The explicit equation describing the waveform is cumbersome and does not provide much insight into the problem. However, a plot of the waveform is very revealing, as in Fig. 8.5; the nonsinusoidal character of the beat note is evident. Moreover, and vitally important, the positive and negative excursions are obviously unequal in area; therefore, the phase-detector output must contain a DC component even before lock is obtained. It is the presence of this component that allows pull-in to occur.

Once the existence of a DC component is recognized, an alternative explanation of its presence aids understanding; that is, the beat-note frequency modulates the VCO. This modulation generates FM sidebands in the VCO output at frequencies $\omega_k = \omega_o + k\,\Delta\omega$, where k takes on all integer values. Modulated VCO output is multiplied in the phase detector by the sinusoidal input with frequency ω_i.

The difference signal out of the phase detector consists of individual signals at all the frequencies $\omega_i - \omega_k = \omega_i - \omega_o - k\,\Delta\omega = (1 - k)\,\Delta\omega$. The individual signal corresponding to $k = 1$ has a frequency of zero; that is, $k = 1$ corresponds to a DC component. Relevant spectra are shown in Fig. 8.6. Give this DC component the name *pull-in voltage* and denote it by the symbol v_p. The effect is not of much value in a first-order loop; if the initial difference frequency exceeds the lock-in frequency the magnitude of the DC component is insufficient to pull into lock. However, the average difference frequency is reduced; the first-order loop pulls toward lock even if it cannot reach lock.

A type 2 loop includes an integrator in its loop filter. This integrator builds up an increasing output in response to a DC input. As the integrator output builds up, the VCO frequency is adjusted toward lock. If the initial difference frequency is not too great, the loop will eventually lock up.

Analysis of Pull-in Approximate formulas for pull-in time and pull-in limits may be obtained by following a method originated by Richman [8.3]. Represent the loop as in Fig. 8.7. There is a high-frequency path from PD to VCO with a flat gain of $|F(\infty)| = \tau_2/\tau_1$ and a low-frequency path that contains an integrator. Regard the integrator as perfect. The output of the phase detector consists of an

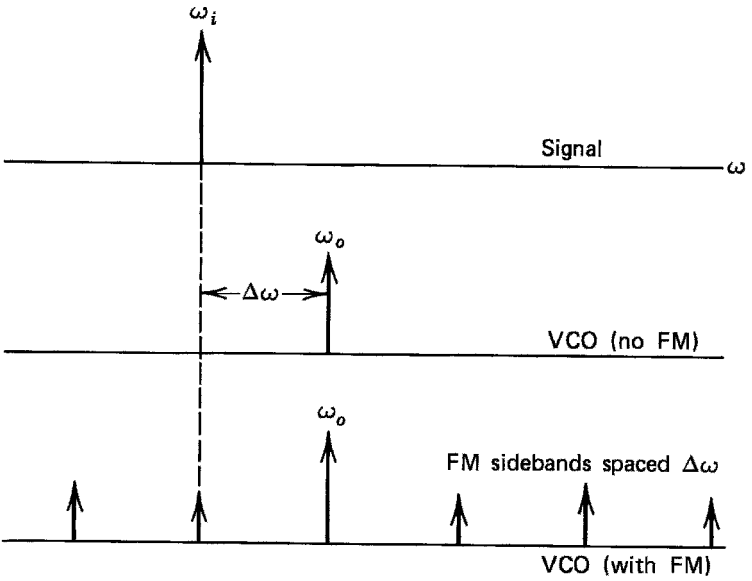


Figure 8.6 Signal and VCO spectra, illustrating pull-in.

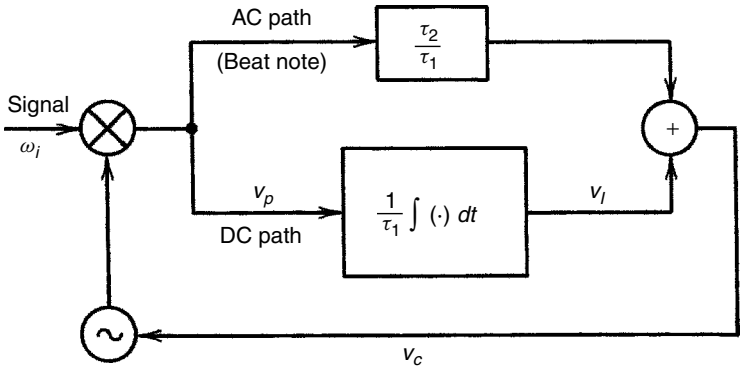


Figure 8.7 Model of a second-order type 2 PLL for pull-in analysis.

AC beat note and the DC pull-in voltage v_p. For analysis purposes, pretend that the AC portion passes only through the high-frequency path and is suppressed completely in the integrator path. (There is little pretense involved for high-enough beat frequencies.) Similarly, assume that the DC pull-in voltage is passed mainly by the integrator and only a negligible portion goes through the high-frequency path. This is an accurate approximation for time intervals appreciably larger than the time constant τ_2.

Input frequency is ω_i, initial frequency of the VCO is ω_o, and initial frequency difference is $\Delta\omega = \omega_i - \omega_o$. If the loop is to pull in slowly rather than lock in quickly, the relation $|\Delta\omega| > K$ must apply. The average frequency (average over

a beat cycle) of the VCO during pull-in is $\Omega_o(t) = \omega_o + K_o v_I(t)$, where v_I is the output of the integrator. Any change in v_I or Ω_o is negligible over the time of a single beat cycle. The average frequency error over a short time is $\Omega = \omega_i - \Omega_o$.

Pull-in voltage varies as a function of Ω. Richman has integrated the differential equation of a first-order loop and found its pull-in voltage to be

$$v_p = K_d \left[\frac{\Omega}{K} - \sqrt{\left(\frac{\Omega}{K}\right)^2 - 1} \right] \tag{8.6}$$

for $|\Omega| > K$. Use the same formula for pull-in voltage of a type 2 loop, a reasonable expedient under the assumptions that have been imposed. Combining the various equations around the low-frequency loop gives

$$\Omega(t) = \Delta\omega - \frac{K_o}{\tau_1} \int_0^t v_p(\tau)\, d\tau \tag{8.7}$$

which is differentiated to give the equation

$$\frac{d\Omega}{dt} = -\frac{K_o v_p(t)}{\tau_1} \tag{8.8}$$

Substitute (8.6) for v_p and solve for dt to obtain

$$dt = -\frac{\tau_2\, d\Omega}{K\left[(\Omega/K) - \sqrt{(\Omega/K)^2 - 1}\right]} \tag{8.9}$$

Pull-in Time Pull-in time T_p is defined as the time required for the average frequency error to change from the initial condition $\Omega = \Delta\omega$ to the lock limit $\Omega = K$. Find T_p by integrating (8.9) between the limits of $\Delta\omega$ and K. Assuming that $|\Delta\omega| \gg K$, the pull-in time is

$$T_p \approx \frac{(\Delta\omega)^2 \tau_2}{K^2} = \frac{4\zeta^2(\Delta\omega)^2}{K^3} = \frac{(\Delta\omega)^2}{2\zeta\omega_n^3} = \frac{(\Delta\omega)^2 \zeta^2 (1 + 1/4\zeta^2)^3}{16 B_L^3} \tag{8.10}$$

Because of the approximations, this formula should not be applied if $|\Delta\omega|$ is either very large (near $\Delta\omega_p$, the pull-in limit, to be defined shortly) or very small (near K). It is best applied in the midrange and should be considered as the time required to pull in from the initial offset to a beat frequency equal to K (at which time the loop quickly locks in). A narrowband loop can take a very long time to pull in. For example, if $\Delta\omega/2\pi = 1$ kHz and $B_L = 10$ Hz, pull-in time would be 1 hour and 10 minutes, which is intolerably long for almost any application.

Pull-in Limits If the loop filter contains a perfect integrator, pull-in will be accomplished no matter how large the initial frequency error. (This statement neglects clipping limits; the loop clearly cannot pull in a signal that requires excessive control voltage to the VCO. Also, it is assumed that there are no

unwanted DC offsets within the loop that would counteract the pull-in voltage and cause the VCO frequency to be pushed out instead.) In an analog loop filter, the integrator is imperfect and the DC gain is some finite number $F(0)$. If v_p is small enough—if the initial frequency error is large enough—the loop cannot pull in. The largest frequency for which the loop can still pull into lock is called the *pull-in limit* and is represented by $\Delta\omega_p$.

To derive the pull-in limit, replace the perfect integrator in Fig. 8.7 by an imperfect integrator with DC gain $F(0) - F(\infty)$. [The DC gain of the entire loop filter is $F(0)$, and the DC gain of the high-frequency path is $F(\infty)$. Therefore, the DC gain of the low-frequency path must be $F(0) - F(\infty)$.] Assume that $|\Delta\omega|$ is so large that the loop cannot pull in. The phase detector still generates a pull-in voltage v_p which is amplified by the factor $F(0) - F(\infty)$ and is applied to the VCO, where it causes a steady-state frequency change of $K_o[F(0) - F(\infty)]v_p$. The steady-state frequency error is

$$\Omega = \Delta\omega - K_o[F(0) - F(\infty)]v_p \tag{8.11}$$

Upon substituting (8.6) for v_p and remembering from Chapter 2 that $K_{\mathrm{DC}} = K_o K_d F(0)$ and $K = K_o K_d F(\infty)$, the average frequency error in the unlocked steady state is

$$\Omega = \Delta\omega - (K_{\mathrm{DC}} - K)\left[\frac{\Omega}{K} - \sqrt{\left(\frac{\Omega}{K}\right)^2 - 1}\right] \tag{8.12}$$

Equation (8.12) can be solved for the steady-state frequency error. A real solution is found if $|\Delta\omega| \geq K(2K_{\mathrm{DC}}/K - 1)^{1/2}$. A smaller value of $|\Delta\omega|$ leads to complex roots of (8.12), which means that no real final frequency error satisfies the equation; the loop pulls in for smaller values of $|\Delta\omega|$.

Because of the many approximations that have been made, the boundary is accurate only if $K_{\mathrm{DC}} \gg K$. Therefore, an approximate formula for the pull-in limit is

$$\Delta\omega_p \approx \sqrt{2K_{\mathrm{DC}}K} \tag{8.13}$$

In principle, the pull-in range can be made as large as may be needed simply by using a large DC gain K_{DC}. Moreover, the large pull-in can be achieved with as narrow a noise bandwidth as necessary; the parameters K and K_{DC} are independent.

Formula (8.10) for pull-in time is valid only if the initial frequency error is substantially larger than the loop gain K and substantially smaller than the pull-in limit. Richman [8.3] developed improved formulas that describe the pull-in time for all conditions, including initial frequency error near either of the bounds. The results are much more cumbersome than (8.10).

Pull-in for Other Conditions Many investigators have investigated pull-in. Viterbi [8.1, 8.2] examined the problem through limit cycles in the phase plane

and arrived at essentially the same results given here. The foregoing results apply only to loops with sinusoidal phase detectors. Mengali [8.11] summarizes work by other authors on extended PD characteristics and arrives at general formulas for pull-in time and range that take the PD characteristic into account. As might be expected, an extended PD characteristic provides an extended pull-in range and faster pull-in time. Meer [8.12] investigated extended PD characteristics and higher-order loops. He derived the pull-in voltages associated with triangular and sawtooth PDs and observed that these are larger than for sinusoidal PDs.

Pull-in for Type 3 PLLs In a type 3 loop, there are two integrators in the low-frequency path; the double-integrated pull-in voltage has parabolic rather than linear growth. As a result, pull-in is faster in a type 3 loop than it is for a type 2 loop. Assume that both integrators are ideal and that $|\Delta\omega| \gg K$. Following Meer's analysis and specializing to a loop filter with its two zeros coincident at $s = -1/\tau_2$, pull-in time for a type 3 loop was found to be

$$T_p \approx \frac{|\Delta\omega|\tau_2\sqrt{\pi}}{K} \tag{8.14}$$

Pull-in time for the type 3 PLL varies as the first power of initial frequency error rather than $|\Delta\omega|^2$ as in the type 2 PLL, as shown in (8.10).

Unfortunately, frequency pull-in to zero beat does not assure rapid phaselocking in a type 3 PLL. Equation (8.14) indicates the time needed to accumulate the correct tracking charge on the frequency integrator in the loop filter, but the charge on the frequency-rate integrator will be wrong at that time. It is entirely likely that the charge stored on the first integrator will force the second integrator to continue to charge rather than stop at the proper frequency. If that should happen, the VCO frequency overshoots the correct equilibrium, pull-in voltage reverses polarity, and the pull-in action heads for equilibrium from the opposite direction.

In other words, the approach to lock can be oscillatory and (8.14) only tells the time to the first passage through zero frequency error, not to phaselocking. Lock is not possible until the charge on the frequency-rate integrator settles to the correct value needed for equilibrium tracking. On the other hand, in the vicinity of zero frequency error, the high-frequency path through the loop filter has a strong locking action. If that locking force can overcome the frequency-slewing force from the first integrator, the loop will lock at first passage and will not oscillate about frequency equilibrium. Tausworthe and Crow [8.13, 8.14] found that lock occurs on first passage if the closed-loop poles are overdamped and oscillatory acquisition occurs if the poles are underdamped. Additional information on pull-in of type 3 PLLs is contained in [8.15].

Practical Limitations of Pull-in The analyses and references presented above deal only with quasi-type 2 or higher-type PLLs having loop filters with equal numbers of poles and zeros. The analyses fail badly if these conditions are violated. Section 14.4 describes some unfortunate consequences of additional poles

within the loop. From the many papers on the subject, a casual reader might get the impression that pull-in is the dominant applied method of frequency acquisition. Actually, one could argue that pull-in is more interesting than it is practical. Besides its slowness, pull-in can be defeated by unwanted but unavoidable DC offsets arising in the phase detector (Chapter 10) or active loop filter (Chapter 11), or it can be converted to push-out or false locking by excess poles or delays within the loop (Chapter 14). There is little information available on pull-in behavior in the presence of significant noise.

In the author's experience, pull-in is practical only in a comparatively benign environment: where noise is small, bandwidth is large enough and initial frequency offset is small enough to permit rapid action, and the loop circuits are simple so that extra poles are avoided. In more challenging applications, pull-in is almost always found to be unsatisfactory or unusable, and some form of aided acquisition is needed. Forms of aided frequency acquisition are discussed on the following pages.

8.3.2 Frequency Sweeping

Faster, more-reliable frequency acquisition can be attained by sweeping the frequency of a VCO in a search for the signal frequency. If the search is applied correctly, the loop will lock up as the VCO frequency sweeps into coincidence with the signal. Lock-up inhibits further change of VCO frequency, so the sweep process is self-terminating. Sweep acquisition is a blind search that is just about the only practicable method when the signal is deeply immersed in noise.

Sweep-Rate Limitations: Noise-free From the earlier discussion on hold-in in the presence of a frequency ramp, it should be evident that the sweep rate must not be excessive. Section 5.2.2 showed that the loop with sinusoidal PD cannot hold lock if the sweep rate Λ exceeds ω_n^2 rad/sec^2. If a loop cannot hold lock on a signal, it certainly will be unable to acquire lock. Therefore, an absolute maximum limit on the allowable sweep rate is ω_n^2 (for a PD with sinusoidal s-curve).

Viterbi [8.1, 8.2] has investigated frequency-acquisition problems by means of phase-plane trajectories. He discovered that acquisition is not certain even if $\Lambda < \omega_n^2$ and the loop is noise-free. If Λ becomes somewhat larger than $\omega_n^2/2$, there is a possibility that the VCO can sweep right through the input frequency without locking. The chance of locking or nonlocking depends on the random initial conditions of frequency and phase. Viterbi's phase-plane graphs were used to estimate the probability of locking, and results are plotted against sweep rate in Fig. 8.8. These results apply directly only to the special case of a second-order type 2 PLL with a sinusoidal PD and $\zeta = 0.707$. However, qualitatively similar behavior should be expected for other damping factors and PD s-curves.

Further qualitative information on sweep acquisition behavior is available from the simulation study [8.16] by Frazier and Page. [**Comment**: There is a numerical error by a factor of 1.4 the runs throughout their paper, making quantitative interpretation difficult.] Their paper indicates that for fixed natural frequency and

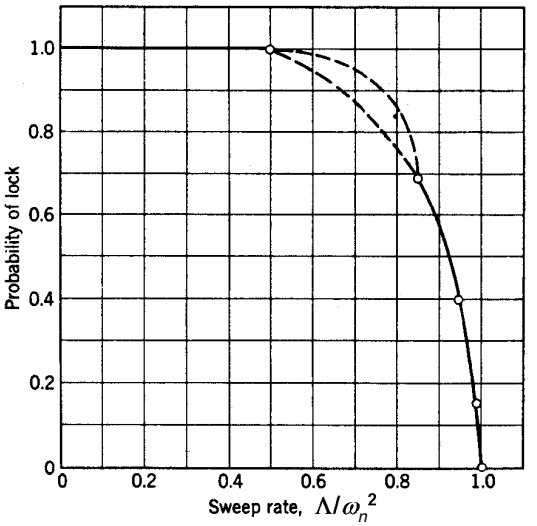


Figure 8.8 Probability of sweep acquisition in a noise-free second-order type 2 PLL with $\zeta = 0.707$.

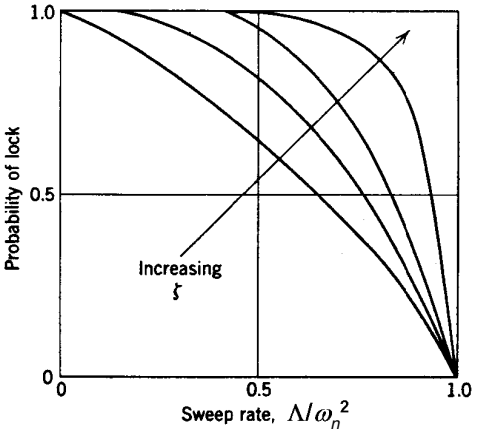


Figure 8.9 Probability of sweep acquisition, showing the effect of damping.

sweep rate, the probability of lock improves as damping increases. See Fig. 8.9, which seems to imply that the loop should be heavily damped, at least until it is locked. Such a conclusion is premature; noise bandwidth of the PLL varies with damping even though natural frequency is fixed. On the basis of fixed-noise bandwidth, the largest value of ω_n (and therefore the largest maximum sweep rate) occurs for $\zeta = 0.5$. Yet the probability of acquiring lock at sweep rates less than ω_n^2 improves as damping increases. There is some value of ζ that provides best acquisition performance; the exact value is not known, but it probably lies between 0.7 and 1.0.

Sweep-Rate Limitations in Noise So far the loop has been assumed to be noise-free. In actuality, noise is always present and must be taken into account. Simple intuition leads one to expect that noise will make it more difficult to acquire a signal; it would be useful if this difficulty could be expressed by a number. Frazier and Page's experiments provide empirical data that suggest that sweep rate should be reduced by a factor of $[1 - (\mathrm{SNR}_L)^{-1/2}]$ if an acceptably high probability of acquisition is to be maintained in the presence of noise. This expression predicts that acquisition becomes impossible at 0-dB signal-to-noise ratio in the loop. Experience suggests this conclusion to be optimistic.

Combining disparate fragments of information and the author's experience, a better preliminary design value for sweep rate might be

$$\Lambda = \frac{1}{2}\omega_n^2 \left(1 - \frac{2}{\sqrt{\mathrm{SNR}_L}}\right) \qquad \mathrm{rad/sec}^2 \tag{8.15}$$

in combination with $\zeta = 0.7$ to 1.0. This choice implies that sweep acquisition is impossible below 6 dB SNR_L, which is a somewhat conservative statement but not drastically wrong. Experimental adjustment from these values can provide refinement, if needed.

Because of the nonlinearity, sweep acquisition has defied satisfactory analysis in the presence of significant noise. Meyr and Ascheid [8.6, Chap. 5] provide a more nuanced approach than the rule of thumb of (8.15), based on the probability of cycle slips after the loop has locked. Blanchard [8.17] reports on an extensive series of laboratory measurements that relate sweep speed, signal-to-noise ratio, and probabilities of correct acquisition and false alarm.

The results given here apply to a loop with sinusoidal phase detector. A different PD characteristic can be expected to produce different sweep capabilities. The matter does not appear to have been investigated, probably due to the fact that blind sweep acquisition has been applied mainly to systems for which the SNR is low at the PD input. Chapter 10 shows that a sinusoidal s-curve is the only shape possible if the input SNR is very small.

Sweep Implementation Sweep can be applied to a type 2 PLL in a very simple and elegant manner. Some workers have built separate sawtooth generators that add a sweep voltage directly into the VCO, but that approach is unnecessarily complicated and arises from inadequate understanding of the state variables of the loop. A far better approach is to insert a constant slew current I_s into the integrator of the loop filter. Integrated output is a ramp that is applied to the VCO, causing the frequency to sweep. The slope of the ramp is determined by the time constant of the integrator and the magnitude of the current. Circuit details are shown in Fig. 8.10.

The slew current is inserted at the junction of R_2 and C, not directly into the summing junction of the operational amplifier. If the current were applied directly to the op amp, there would be an output step component (in addition to the desired ramp) of $I_s R_2$ whenever the slew current was turned on or off. The step could cause the loop to jump out of lock, depending on the circuit parameters. When

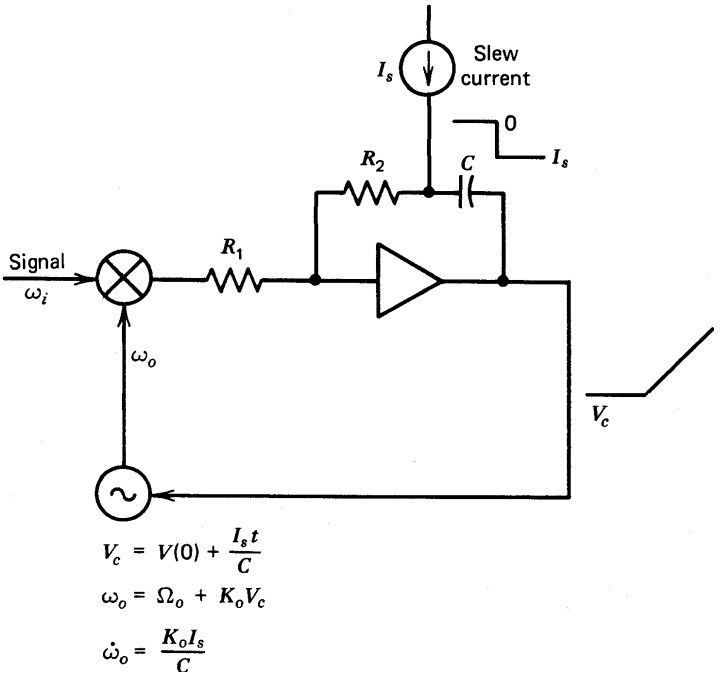


Figure 8.10 Frequency-sweep circuit for a type 2 PLL.

the loop locks, the integrator has exactly the right charge needed to hold the VCO at the signal frequency. The loop overcomes the injected slew current by means of a DC output from the PD, which, in turn, is produced by a dynamic-lag phase error (Section 5.1.1).

After lock has been achieved, the phase error constitutes a loop stress that impairs the tracking capability in the presence of noise or other disturbances. It is advisable to shut off the slew current once lock has been verified. (Lock detectors are described later in this chapter.) Slew shutoff is particularly necessary if the signal is subject to fast fading; the sweep circuit could carry off the VCO frequency in the event of a fade and the entire sweep range might have to be searched before the signal could be reacquired. However, the decision to shut off the slew does not have to be particularly fast. The loop does hold lock with the slew applied, so a sufficient time can be taken for lock verification to assure a reliable decision.

The simplicity described so far and the freedom to perform a leisurely lock verification is offered only with a closed-loop sweep. One could also perform an open-loop sweep [8.17] but it then becomes necessary to detect frequency agreement very rapidly and then quickly shut off the sweep and close the loop. In principle, the sweep rate is no longer restricted by the ramp tracking limits of the loop, but the need for reliable measurement of frequency coincidence in the presence of noise still places limits on rates.

Sweep can also be applied to type 3 PLLs. Since the type 3 PLL is supposed to be better able to track a frequency ramp, one might expect that a faster sweep should be possible. Unfortunately, the extra complexity of the type 3 loop has so far prevented discovery of a practical method of achieving the supposed improvement. On the contrary, there is fear among designers that acquisition with a closed-loop type 3 PLL might be unstable; various expedients to avoid instability are often employed. One solution is to employ open-loop search, as mentioned earlier. This search requires fast recognition of zero beat and immediate closing of the loop; these are tricky operations although they have been accomplished successfully.

Another solution is to search with a closed type 2 loop and then insert the additional loop integrator after lock has been achieved. The search rate cannot be any greater than allowed for the type 2 loop. Tausworthe and Crow [8.13, 8.14] show that the third-order poles should be overdamped to assure retaining lock through the loop-switching operation. No public recognition has been given to the fact that a type 3 loop must acquire three variables: phase, frequency, and frequency rate. It may be necessary to engage in a two-dimensional search for both frequency and frequency rate. (Phase presumably is self-acquired.) This subject needs more investigation.

8.3.3 Discriminator-Aided Frequency Acquisition

If the input signal-to-noise ratio is large enough, a frequency discriminator can be used in a conventional automatic frequency control loop to bring the VCO frequency close to that of the signal. Phaselocking occurs when the frequency error is brought within the lock limit.

Discriminators with Linear s-Curves Linear analysis can be applied to a discriminator whose s-curve is approximately a linear function of the frequency error. A typical block diagram of a combined phase- and frequency-lock loop and its linearized loop equations are shown in Fig. 8.11. The phase loop has little effect when out of lock; the VCO is controlled almost exclusively by the frequency loop. After locking, the phase loop dominates because it has much larger DC gain (infinite, in fact, because of the phase-integrating property of the VCO) and the discriminator can then be disconnected if desired.

If a type 2 transfer function is an appropriate choice for the PLL, a type 1 transfer function is appropriate for the frequency loop; the loop filter for the frequency loop would be a simple integrator, without any lead zero. The two loops could share the same operational integrator, as shown in Fig. 8.12. After phaselock has occurred, the closed-loop system transfer function of the PLL in Fig. 8.12 is

$$H(s) = \frac{sK_o\left(\dfrac{K_d\tau_2}{\tau_1} + \dfrac{K_f}{\tau_f}\right) + \dfrac{K_oK_d}{\tau_1}}{s^2 + sK_o\left(\dfrac{K_d\tau_2}{\tau_1} + \dfrac{K_f}{\tau_f}\right) + \dfrac{K_oK_d}{\tau_1}} \tag{8.16}$$

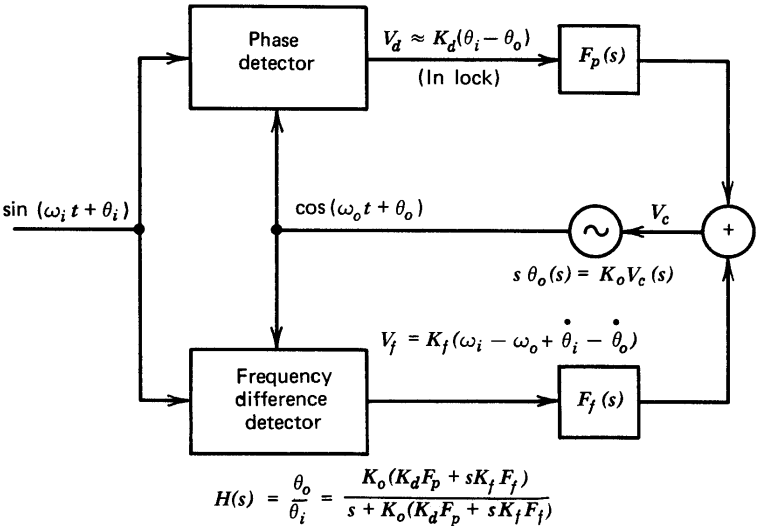


Figure 8.11 Discriminator-aided frequency acquisition.

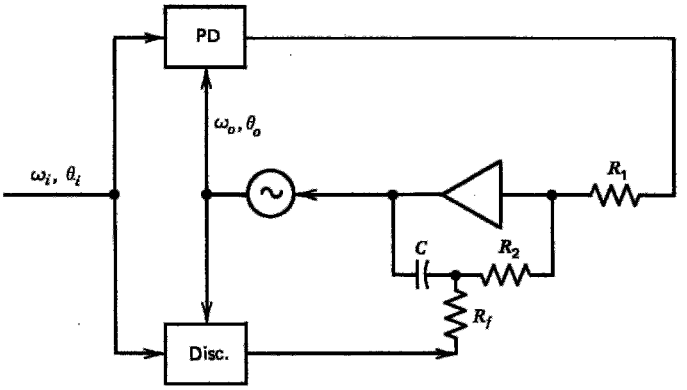


Figure 8.12 Discriminator-aided frequency acquisition in a type 2 PLL.

an expression with the same form as (2.15) that defines the transfer function of a second-order type 2 PLL. In (8.16), the coefficient K_f is the gain of a linear discriminator in volts per rad/sec, $\tau_f = R_f C$, $\tau_1 = R_1 C$, and $\tau_2 = R_2 C$. Comparing (8.16) to (2.15) yields

$$\begin{aligned} \omega_n^2 &= \frac{K_o K_d}{\tau_1} \\ \zeta &= \frac{K_o}{2\omega_n}\left(\frac{K_d \tau_2}{\tau_1} + \frac{K_f}{\tau_f}\right) \\ K &= K_o\left(\frac{K_d \tau_2}{\tau_1} + \frac{K_f}{\tau_f}\right) \end{aligned} \tag{8.17}$$

It is evident that the presence of the discriminator branch has no effect on ω_n but increases the values of ζ and K. In principle, the loop-filter zero necessary for stability of a type 2 PLL could be omitted entirely and all damping could be supplied by the discriminator branch. One could leave the discriminator connected permanently and merely weight the relative contributions of phase and frequency detectors so as to obtain the desired damping.

Permanent connection of a discriminator may not provide good performance. A sweep operation can proceed satisfactorily with a poor input signal-to-noise ratio because the PLL is a coherent device and can recover a signal buried in noise. By contrast, a discriminator is an incoherent device and cannot distinguish between signal and noise. Its average output tends to be the average frequency of signal plus noise—approximately the frequency of the centroid of the spectrum applied to the discriminator. If noise dominates, the discriminator output is determined almost entirely by the noise properties and the signal is suppressed. A discriminator can be used only under conditions where it provides useful information on signal frequency, which ordinarily means that the input signal must exceed the noise. As a rule of thumb, one should be cautious if the input SNR is less than 6 to 10 dB and should be very concerned—to the point of abandoning the discriminator—if input SNR is less than 0 dB.

For a type 2 loop, the preceding analysis has shown that pull-in time is proportional to the square of the initial frequency difference and that sweep-search time is linearly proportional to the search range. If a linear discriminator is used, it can be shown that the frequency-acquisition time is proportional to the logarithm of the initial frequency error. Where applicable, discriminator aiding is a fast method of frequency acquisition.

Nonlinear Discriminators The foregoing analysis of discriminator aiding applies to linear discriminators. Several techniques employing nonlinear discriminators are worthy of notice. The average output of a nonlinear discriminator is a substantially nonlinear function of the frequency error. An important example is the phase/frequency detector (PFD) of Chapter 10, a very popular device. When the PLL is out of lock, the PFD delivers DC output, averaged over many cycles, of approximately $K_d/2$. (This DC level occurs if the frequency error is small compared to the signal frequency. It grows toward K_d as the frequency error increases.) Chapter 10 has a detailed account of the PFD.

In effect, the DC out of the PFD applies a constant slew to the integrator in the loop filter, so the VCO is swept in a search for the correct frequency. Analysis shows that a DC output of $K_d/2$ from the PFD generates a sweep rate of $\pi\omega_n^2$ rad/sec^2, which is 2π times as fast as can be tolerated by a PLL with sinusoidal PD. This increased rate comes about for two reasons: (1) the s-curve for the PFD is linear over 4π rather than sinusoidal over 2π, and (2) the PFD smoothly converts itself from a nonlinear frequency detector to a linear phase detector as frequency error is reduced to within its lock-in range.

One other feature contributes to the enhanced speed of frequency acquisition with a PFD (or any other device that indicates the correct sign of frequency error). An ordinary sweep is blind; there is an equal chance of starting in the wrong direction as in the right direction. By contrast, a PFD indicates the correct direction, causing the sweep to follow the shortest path.

For even faster acquisition, consider a system in which the VCO can be switched to discrete frequencies at nominally uniform spacing. A binary search among the discrete frequencies (halving the range of frequency uncertainty at each step), along with a direction-indicating frequency detector, will achieve frequency acquisition in a time proportional to the logarithm of the number of switched frequencies, in contrast to a continuous sweep, whose time is linearly proportional to the frequency range. An example of binary search was disclosed in [8.18].

Open-Loop Frequency Acquisition The ability to switch the VCO to known frequencies can be used to achieve even faster acquisition if the desired frequency is known (e.g., as in a synthesizer) and if the switched frequencies are themselves well calibrated. Just switch the VCO to the desired frequency, without any search and without any discriminator aiding. The desired frequency is not known ordinarily (except in synthesizers) and accurate, stable calibration of switched frequencies is seriously difficult with analog VCOs. Search methods are useful since they do not require knowledge of either the desired frequency or the switched frequencies. Digital number-controlled oscillators afford excellent knowledge of switched frequencies, but they cannot yet substitute for analog VCOs in many applications.

One way to implement switchable frequency in a VCO is to apply a frequency-control voltage from a digital-to-analog converter (DAC). This switchable voltage is in addition to the control voltage from the loop filter. But DACs can be noisy and might induce excessive phase noise in the VCO. In an alternative scheme, proposed in [8.19], an accurately controlled amount of charge is metered rapidly into the integrator of the loop filter. Since integrator output controls the VCO frequency, that is a way of changing VCO frequency quickly while avoiding the noise contamination from a DAC. Excellent calibration of the VCO is needed to make this scheme work well.

Another approach, sometimes used in systems with substantial digital capabilities, is to compute the spectrum of the region of frequency uncertainty with the aid of a fast Fourier transform (FFT) and to select the frequency of the peak of the computed power spectrum as the best estimate of the signal frequency being sought. The VCO of the PLL is then set to that frequency. This scheme, where it is workable, has the advantage that it can operate well down into the noise level in the bandwidth of uncertainty, an operational capability that otherwise is possible only with blind sweep. The frequency of the VCO has to be settable with sufficient accuracy to take full advantage of this method.

These schemes of switching to known frequencies are open-loop methods that do not employ discriminators or feedback. Open-loop methods generically are

faster than closed-loop methods but generically must be well calibrated since they do not have the benefit of feedback to correct for inaccuracies.

8.3.4 Implementation of Frequency Discriminators

Conventional circuits for frequency discriminators, as found in radio-engineering textbooks, can be used, but better alternatives exist. Instead of a measurement of absolute frequency, the acquisition discriminator should provide an indication of the frequency difference between the incoming signal and the VCO; a frequency-difference discriminator is needed. Richman [8.3] describes a frequency-difference discriminator which he calls the *quadricorrelator*. A block diagram and pertinent equations are shown in Fig. 8.13. The input bandpass signal is translated to two quadrature baseband components by the pair of multipliers (mixers, phase detectors) driven by the oscillator. Baseband lowpass filters establish the frequency-difference range over which the circuit will operate. (Richman also included highpass sections in the baseband filters to disconnect the quadricorrelator automatically for very small frequency differences at which the PLL takes control.)

One of the filtered baseband channels is differentiated and then multiplied by the other channel. The product contains a DC component proportional to the frequency difference between signal and oscillator, including the proper sign. It provides an excellent frequency-difference indication. (There is also a sinusoidal ripple component of equal amplitude at double the difference frequency. This can be a serious nuisance if the quadricorrelator were to be used as an FM demodulator, but the difference frequency goes to zero when the phase loop locks, so the ripple vanishes when the quadricorrelator is used as an acquisition aid.)

Additional information on quadricorrelators and quadricorrelator-like structures may be found in [8.20–8.24]. For other discriminator techniques: Natali [8.24] has proposed an AFC feedback loop using the FFT algorithm; Alberty and Hespelt

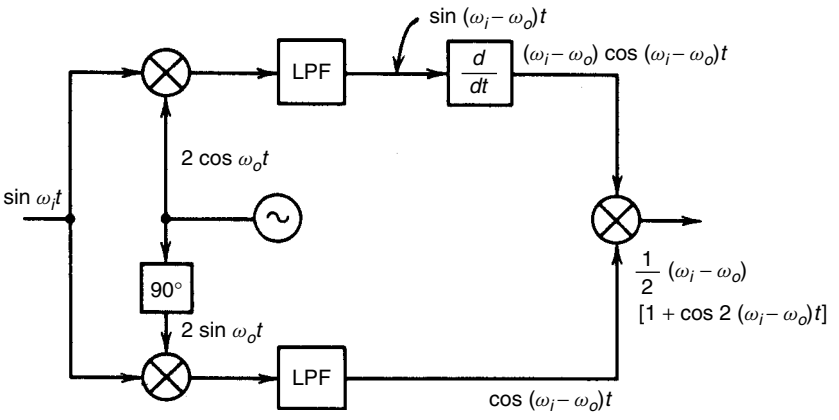


Figure 8.13 Quadricorrelator.

[8.25] disclose a frequency discriminator for data-modulated signals that avoids self-noise, a serious problem in ordinary quadricorrelators.

Messerschmitt [8.26] originated the concept of a *rotational* frequency detector. Picture a vector rotating at an angular rate that is the frequency difference between an input signal and the feedback from a VCO. Define the vector angle at phaselock as zero degrees and label the quadrants counterclockwise from zero around the circle as I, II, III, and IV. Passage of the vector from quadrant II to III indicates that the signal frequency is greater than the VCO frequency, whereas passage from III to II indicates that the signal frequency is the lesser. The number of such passages per second is an indication of the magnitude of frequency difference. A linear frequency discriminator can be implemented by circuits that (1) detect the occurrence and direction of II–III passages and (2) apply a metered charge of correct polarity to the integrator in the PLL loop filter on each such passage.

Outputs from a pair of phase detectors driven in quadrature from the VCO provide enough information to identify the quadrant of phase error at any instant. One PD delivers output dependent on the sine of the phase error, and the other's output is dependent on the cosine. It is sufficient to examine the signs of the two PD outputs. Both PDs have positive output if phase error is in the first quadrant, a sine PD has positive output but a cosine PD has negative output for phase in the second quadrant, and so on. Detection of a passage from II to III, or the reverse, requires memory of the previous quadrant measurement and comparison of that memory to the current measurement. A passage is detected if output from the cosine PD is negative on adjacent measurements *and* if the sine PD output changes sign on adjacent measurements. The direction of passage is indicated by the sign of the sine PD at each passage detected. Examples of rotational detectors may be found in [8.27–8.30].

8.4 DIVERSE MATTERS

There are several topics, such as lock indicators, variable-bandwidth methods, and loop memories that are more or less associated with the subject of acquisition but do not fit into a neat heading of their own. They are grouped together in this section.

8.4.1 Lock Indicators

An often-employed method of lock indication is the *quadrature phase detector*, also known as the *auxiliary phase detector* or the *coherent amplitude detector*, as shown in Figure 8.14. The quadrature phase detector has the received signal applied as one input and a 90° phase-shifted version of the VCO as the other. The main phase detector has an output voltage proportional to $\sin\theta_e$ and quadrature output proportional to $\cos\theta_e$. In the locked condition θ_e is small, so $\cos\theta_e \approx 1$. When the loop is unlocked, the outputs from both phase detectors are beat notes at the difference frequency, so the DC output is almost zero. Thus, the filtered output of the quadrature detector provides a useful indication of lock. The magnitude of

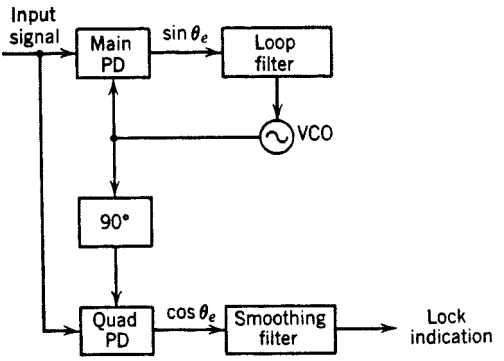


Figure 8.14 Lock detection with a quadrature phase detector.

the output voltage relative to that obtained from a noise-free stable input provides a measure of the quality of lock. When used in this manner, the smoothed voltage is sometimes known as the *correlation* output. It is also possible to use the same voltage as a source of coherent AGC control voltage. The AGC topic is well covered by Meyr and Ascheid [8.6, Sec. 7.2].

An output-smoothing filter is a vital part of a practical lock indicator. Without smoothing, the indication will flicker on and off because of noise, giving false indications of lock or loss of lock. If there is excessive smoothing, the lock or unlock indication is delayed unduly from the time of its actual occurrence. A compromise amount of smoothing is required. Tausworthe [8.31] has performed a detailed analysis of the problem and has produced design curves.

An entirely different principle of lock detection (explained in Section 10.3.8) is commonly employed with phase/frequency detectors.

8.4.2 Wide-Bandwidth Methods

Speed of acquisition—by pull-in, sweep, or discriminator aiding—is improved by widening the loop bandwidth. A loop can be built to have a large bandwidth for rapid acquisition and a much narrower bandwidth for good tracking in the presence of noise. It should be apparent that increase in bandwidth can be successful only if the loop signal-to-noise ratio is sufficiently large and the loop is stable at the wide bandwidth. Acquisition is not likely if the bandwidth change brings the loop close to noise threshold or instability.

Bandwidth is changed by altering the loop gain. That can be accomplished by switching in different resistor values in a circuit, such as that of Fig. 2.2, by changing the pump current in a charge-pump PLL (Chapter 12), or by controlling the amplitude of the input signal applied to a multiplier-class phase detector (Chapter 10). Do not switch capacitors or otherwise disturb the integrator in the loop filter; any disturbance to the integrator ruins the frequency memory and is likely to cause loss of lock.

The signal to command switching of bandwidth can be the lock indication voltage from the quadrature phase detector. When the loop is out of lock, the absence

of indication voltage sets the switches to their wideband position. When the loop locks, the indication voltage appears and sets the switches into their narrowband position. Switched adaptive bandwidth has been explored over many years, but it has rarely lived up to the benefits that it seems to offer. In consequence, few practical PLLs incorporate bandwidth switching as an acquisition aid.

If coherent AGC is employed, the same effect of adaptive bandwidth can be obtained without switches. In the unlocked condition there is no coherent AGC voltage and the signal level at the phase detector is large. When the loop locks, AGC voltage appears and reduces the applied signal voltage. Since the gain of a multiplier-type phase detector—and therefore, loop gain—is proportional to signal level, the loop bandwidth and damping both decrease automatically when the loop locks; no switches are needed.

8.4.3 Memory

In the absence of disturbance, the VCO of a type 2 PLL tends to remain close to its locked frequency in the event of signal dropout because of the charge stored on the integrator in the loop filter. When the signal returns, reacquisition by lock-in or pull-in should be very rapid. The loop has a *frequency memory* in the integrator. When signal drops out, the loop opens and the integrator drifts off slowly at a rate determined by its open-loop time constant and any DC offsets that may be present. Furthermore, zero-mean noise delivered to the integrator input is converted to a random walk at the output, so even a perfect integrator is volatile, even in the absence of DC offsets. Memory persistence is improved if the integrator input can be disconnected from disturbances upon detection of loss of lock.

A first-order loop has a volatile phase memory. Upon signal dropout, the VCO phase immediately begins to drift off from its locked condition at a rate equal to the frequency difference between the signal and the free-running VCO. In other words, the VCO reverts instantly to its free-running frequency when the signal disappears. Because it has frequency memory, a type 2 PLL retains its phase information much better than does a first-order PLL. A type 3 PLL has frequency-rate memory in addition to frequency and phase memories. The third memory can be helpful if input frequency is changing during a signal fade.

REFERENCES

8.1 A. J. Viterbi, *Acquisition and Tracking Behavior of Phase-Locked Loops*, External Publ. 673, Jet Propulsion Laboratory, Pasadena, CA, July 1959.

8.2 A. J. Viterbi, *Principles of Coherent Communication*, McGraw-Hill, New York, 1966, Chap. 3.

8.3 D. Richman, "Color Carrier Reference Phase Synchronization Accuracy in NTSC Color Television," *Proc. IRE* ***42***, 106–133, Jan. 1954.

8.4 F. M. Gardner, "Hangup in Phase-Lock Loops," *IEEE Trans. Commun.* ***COM-25***, 1210–1214, Oct. 1977. Reprinted in [8.8].

8.5 H. Meyr and L. Popken, "Phase Acquisition Statistics for Phase-Locked Loops," *IEEE Trans. Commun.* ***COM-28***, 1365–1372, Aug. 1980.

8.6 H. Meyr and G. Ascheid, *Synchronization in Digital Communications*, Wiley, New York, 1990.

8.7 F. M. Gardner, "Equivocation as a Cause of PLL Hangup," *IEEE Trans. Commun.* ***COM-30***, 2242–2243, Oct. 1982. Reprinted in [8.8].

8.8 W. C. Lindsey and C. M. Chie, eds., *Phase-Locked Loops*, Reprint Volume, IEEE Press, New York, 1986.

8.9 J. J. Spilker, "Delay-Lock Tracking of Binary Signals," *IEEE Trans. Space Electron. Telem.* ***SET-9***, 1–8, Mar. 1963.

8.10 C. E. Gilchriest, *Pseudonoise System Lock-in*, Research Summary 36–9, Vol. I, pp. 51–54, Jet Propulsion Laboratory, Pasadena, CA, July 1, 1961.

8.11 U. Mengali, "Acquisition Behavior of Generalized Tracking Systems in the Absence of Noise," *IEEE Trans. Commun.* ***COM-21***, 820–826, July 1973.

8.12 S. A. Meer, "Analysis of Phase-Locked Loop Acquisition: A Quasi-Stationary Approach," *IEEE 1966 Conv. Rec.*, Vol. 14, Pt. 7, pp. 85–106, 1966.

8.13 R. C. Tausworthe and R. B. Crow, "Improvements in Deep-Space Tracking by the Use of Third-Order Loops," *IEEE Int. Conf. Commun.*, 1972, pp. 577–583.

8.14 R. C. Tausworthe, "Improvements in Deep-Space Tracking by the Use of Third-Order Loops," *JPL Q. Tech. Rev.* ***1***, 96–106, July 1971.

8.15 F. Russo and L. Verranzzani, "Pull-in Behavior of Third-Order Generalized Phase-Locked Loops," *IEEE Trans. Aerosp. Electron. Syst.*, ***AES-12***, 213–218, Mar. 1976.

8.16 J. P. Frazier and J. Page, "Phase-Lock Loop Frequency Acquisition Study," *IRE Trans. Space Electron. Telem.* ***8***, 210–227, Sept. 1962.

8.17 A. Blanchard, *Phase-Locked Loops*, Wiley, New York, 1976, Chap. 11.

8.18 C.-C. Chung and C.-Y. Lee, "An All-Digital Phase-Locked Loop for High-Speed Clock Generation," *IEEE J. Solid-State Circuits* ***38***, 347–351, Feb. 2003.

8.19 J. Hakkinen and J. Kostamovaara, "Speeding Up an Integer-*N* PLL by Controlling the Loop Filter Charge," *IEEE Trans. Circuits Syst. II* ***50***, 343–354, July 2003.

8.20 F. M. Gardner, "Properties of Frequency Difference Detectors," *IEEE Trans. Commun.* ***COM-33***, 131–138, Feb. 1985. Reprinted in [8.21]. Also, extended version in [8.8].

8.21 B. Razavi, ed., *Monolithic Phase-Locked Loops and Clock Recovery Circuits*, Reprint Volume, IEEE Press, New York, 1996.

8.22 J. A. Bellisio, "A New Phase-Locked Timing Recovery Method for Digital Regenerators," *IEEE Intl. Commun. Conf. Rec.*, Vol. 1, pp. 10–17 to 10–20, June 1976. Reprinted in [8.21].

8.23 R. R. Cordell, J. B. Forney, C. N. Dunn, and W. G. Garrett, "A 50 MHz Phase- and Frequency-Locked Loop," *IEEE J. Solid-State Circuits* ***SC-14***, 1003–1009, Dec. 1979. Reprinted in [8.21].

8.24 F. D. Natali, "AFC Tracking Algorithms," *IEEE Trans. Commun.* ***COM-32***, 935–947, Aug. 1984. Reprinted in [8.8].

8.25 T. Alberty and V. Hespelt, "A New Jitter Free Frequency Error Detector," *IEEE Trans. Commun.* ***COM-37***, 159–163, Feb. 1989.

8.26 D. G. Messerschmitt, "Frequency Detectors for PLL Acquisition in Timing and Carrier Recovery," *IEEE Trans. Commun.* ***COM-27***, 1288–1295, Sept. 1979. Reprinted in [8.21].

8.27 F. M. Gardner, "A Cycle-Slip Detector for Phase-Locked Demodulators," *IEEE Trans. Instrum. Meas.* ***IM-26***, 251–254, Sept. 1977.

8.28 J. A. Afonso, A. J. Quiterio, and D. S. Arantes, "A Phase-Locked Loop with Digital Frequency Comparator for Timing Signal Recovery," *IEEE Natl. Telecommun. Conf. Rec.*, Vol. 1, pp. 14.4.1–14.4.5, 1979. Reprinted in [8.21].

8.29 A. Pottbäcker, U. Langmann, and H.-U. Schreiber, "A Si Bipolar Phase and Frequency Detector IC for Clock Extraction up to 8 Gb/s," *IEEE J. Solid-State Circuits* ***SC-27***, 1747–1751, Dec. 1992. Reprinted in [8.21].

8.30 L. M. DeVito, "A Versatile Clock Recovery Architecture and Monolithic Implementation," pp. 405–420 in [8.21].

8.31 R. C. Tausworthe, *Design of Lock Detectors*, JPL Space Programs Summary 37–43, Vol. III, pp. 71–75, Jet Propulsion Laboratory, Pasadena, CA, Jan. 31, 1967.

CHAPTER 9

OSCILLATORS

An oscillator with controllable frequency is an essential element of a phaselock loop. This chapter provides an overview of various classes of oscillators, with emphasis on phase-noise issues.

9.1 DESIRED PROPERTIES

Many requirements are placed on oscillators, requirements usually in conflict with one another; compromise is needed almost always. Important requirements include:

1. Low phase noise
2. Frequency accuracy
3. Wide tuning range
4. Tuning linearity
5. Wideband (i.e., fast) modulation capability
6. Low power consumption
7. Small size
8. Integration on a chip

Phase-noise performance must inevitably be sacrificed to achieve any of the other features.

Phaselock Techniques, Third Edition, by Floyd M. Gardner

9.2 CLASSES OF OSCILLATORS

Two classes of analog oscillator are important in PLLs: those that employ a frequency-selective resonator and those that operate on a relaxation principle. Resonator types include quartz crystals, surface-acoustic wave (SAW) devices, microwave or optical cavities, dielectric cylinders (DROs: dielectric resonator oscillators), transmission lines, inductance/capacitance (LC) tanks, ceramic filters, electromechanical filters, and YIG (yttrium–iron garnet) spheres. A ring oscillator is the most prominent relaxation oscillator at the time this is being written; it has largely replaced the earlier astable multivibrator.

Great efforts are being expended currently (2004) to incorporate all PLL components onto a single integrated-circuit (IC) chip, along with all other analog and digital circuits in a system. Ring oscillators are popular in part because they are readily integrated on-chip using the same IC processes employed for digital logic devices. Another circuit popular for integration is a push-pull LC oscillator. These have better phase noise than ring oscillators and inherently narrower tuning range. Numerous papers on integrated ring and LC oscillators are in Razavi's collections [9.1, 9.2], along with papers on the challenging design of on-chip inductors.

9.3 PHASE NOISE IN OSCILLATORS: SIMPLIFIED APPROACH

Oscillators were constructed for many years before the importance of phase noise was recognized. Numerous articles and books describe oscillator circuits without mention of phase noise. As the ill effects of phase noise gradually became evident, substantial intellectual effort was devoted to its formulation. Kroupa's [9.3] reprint volume contains samples of these early phase-noise efforts and many references to others, along with additional papers well worth the attention of present-day designers of PLLs.

Phase noise can be formulated in the time domain (via the Allan variance [9.4]) or the frequency domain. A frequency-domain formulation is preferred for PLL design and applications. In this book, phase noise is described in the frequency domain by its one-sided baseband spectral density $W_\phi(f)$ in rad^2/sec (see Chapter 7 for definitions).

9.3.1 Leeson's Model

In 1966, Leeson [9.5] published a famous paper that has been a landmark ever since. He proposed a simple model of an oscillator consisting of an amplifier and a resonator connected in a positive-feedback loop together with two sources of noise. His analytical results are qualitatively close to the shapes of phase-noise spectra measured on almost all physical oscillators. Figure 9.1 shows a slightly specialized version of the model. The amplifier was assumed to be linear. One source of noise was assumed to be additive and white; in Fig. 9.1 it is shown as an equivalent noise at the input of the amplifier. The other source of noise was assumed to have a flicker ($1/f$) spectral shape and to cause phase modulation of the signal.

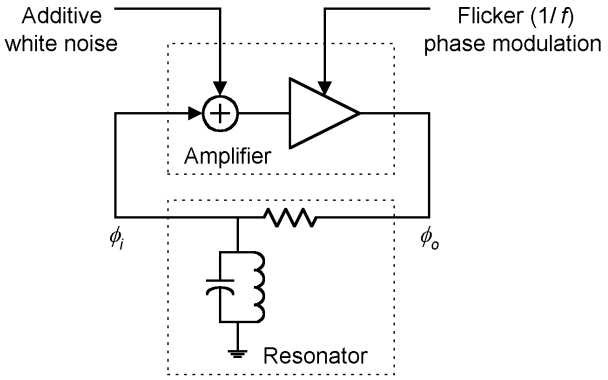


Figure 9.1 Oscillator block diagram illustrating Leeson's model.

Reasoning Leeson invoked the Barkhausen criterion [9.12, Chap. 1; 9.13, Sec. 6.1] of oscillation to explain observed phase-noise spectra. According to this criterion, phase shift around a stable oscillating feedback circuit must be an integer multiple of 2π. Additive noise can be resolved into amplitude and phase components. The phase component appears to alter the phase around the loop from its stable condition, so the oscillator shifts its frequency to try bring around-the-loop phase back to its correct integer multiple of 2π. Phase correction is afforded by the phase vs. frequency characteristic of the resonator; the frequency shift is that which provides the phase alteration required. A similar argument applies to the flicker-induced phase-noise modulation. Thus, the *frequency fluctuations* of the oscillator follow the *phase* component of noise fluctuations in the amplifier. The oscillator frequency-noise spectrum $W_{\omega o}(f)$ will have the same shape as the spectrum of the noise source, so the oscillator phase-noise spectrum has the form $W_{\phi o}(f) = W_{\omega o}(f)/4\pi^2 f^2$ (Section 7.2.5).

The argument of the preceding paragraph applies only for modulation frequencies f well within the half-bandwidth $f_o/2Q$ of the resonator, where f_o is the frequency of oscillation. Noise frequencies well outside that bandwidth are substantially attenuated by the resonator and so do not propagate around the feedback loop. Thus, the phase-noise spectrum for those frequencies is the same as the spectra of the noise sources: a combination of white noise and perhaps $1/f$ noise. As a further refinement, the phase ϕ_o at the output of the amplifier will be different from the phase ϕ_i at the output of the resonator; the resonator acts as a filter on the amplifier's phase noise.

Result Taking all of these conditions into account leads to a pair of equations that approximate the phase-noise spectra in the oscillator configuration of Fig. 9.1:

$$\begin{aligned} W_{\phi o}(f) &= \frac{W_0}{P_s}\left[1+\left(\frac{f_o}{2Qf}\right)^2\right]\left(1+\frac{f_3}{f}\right) \\ W_{\phi i}(f) &= \frac{W_0}{P_s}\left(\frac{f_o}{2Qf}\right)^2\left(1+\frac{f_3}{f}\right) \end{aligned} \tag{9.1}$$

where W_0 is the noise spectral density of the white noise, P_s is the oscillator power, Q is the loaded quality factor of the resonator, and f_3 is a corner frequency related to the flicker noise. Sauvage [9.6] performed a transfer-function analysis that came to the same format as (9.1), without invoking the Barkhausen criterion.

Spectral Shapes Several features of the spectrum can be deduced from (9.1):

- $W_{\phi i}(f)$ can only have $1/f^3$ and $1/f^2$ frequency regions because the idealized resonator shown in Fig. 9.1 cannot support a $1/f$ or white-noise spectrum at its output. The corner frequency between the $1/f^3$ and $1/f^2$ regions is f_3.
- $W_{\phi o}(f)$ always has a $1/f^3$ region.
- $W_{\phi o}(f)$ has a $1/f^2$ region extending from f_3 to $f_o/2Q$, but only if $f_3 < f_o/2Q$.
- $W_{\phi o}(f)$ has a $1/f$ region extending from $f_o/2Q$ to f_3, but only if $f_3 > f_o/2Q$.
- $W_{\phi o}(f)$ has a white-noise region starting at the larger of $f = f_o/2Q$ or $f = f_3$.

9.3.2 Guides for Oscillator Design

Leeson's model and (9.1) provide several valuable design guides to low-phase-noise oscillators:

- White-noise spectral density W_0 should be small.
- Oscillator power P_s should be large.
- Resonator Q should be large.
- Flicker corner f_3 should be small.
- If phase noise at large f is important, oscillator output should be taken from the resonator output rather than the amplifier output, to take advantage of the extra filtering in the resonator. Be aware, though, that post-oscillator circuits (e.g., buffer amplifiers, frequency multipliers, frequency dividers, or phase-noise analyzers) often contribute $1/f$ or white noise sufficient to mask oscillator behavior.

The configuration of Fig. 9.1 was chosen for discussion because phase-noise output of its resonator rolls off asymptotically at -20 dB/decade rather than flattening as at the amplifier output. If resonator attenuation flattens out at higher frequencies, $W_{\phi i}(f)$ also flattens to the detriment of phase-noise performance [9.6].

A quartz crystal is an example of a resonator with flattened asymptotic response [9.7]. Equivalent circuits of quartz crystals consist of a series *RLC* combination shunted by a parallel capacitance C_p. Many crystal-oscillator circuits run at or near series resonance, but the parallel capacitance provides a bypass path that places a floor on the attenuation of noise frequencies well removed from the frequency of oscillation. As a result, phase-noise spectra of crystal oscillators typically have a white-noise floor starting at relatively small values of offset frequency f. This feature suggests merit in a half bridge, with the crystal in one

arm and a capacitance equal to C_p in the other, to cancel the bypass path of the crystal's shunt capacitance and thereby improve the phase-noise spectrum at larger f. Radio engineers knew this technique as *neutralization*, a widespread practice at one time.

Oscillator $1/f^3$ phase noise is tracked out in a wideband PLL, so it may not be of much concern. But $1/f^3$ phase noise can be dominant if the PLL must have narrow bandwidth; it then behooves one to minimize the flicker modulation (i.e., to minimize the f_3 corner frequency). A short note by Halford, Wainwright, and Barnes [9.9] reports that local RF feedback in amplifiers and frequency multipliers (e.g., by means of unbypassed resistors in emitter or source circuits of amplifiers) can produce drastic improvements of flicker phase modulation. I do not know whether this technique has been applied to oscillators. Figure 9.1 shows flicker noise arising in the amplifier. However, Walls and Wainwright [9.8] discovered that flicker noise also arises in quartz crystals. High-Q crystals tend to have lower flicker noise than that of crystals with low Q.

9.3.3 Example Phase-Noise Spectra

Figure 9.2 displays phase-noise spectra that were measured or specified for several oscillators of different configurations. Observe that the ordinate is not $W_\phi(f)$; it is not fair to directly compare the phase noise of, say, a ring oscillator running at 1 MHz against that of a DRO at 20 GHz. Instead, the spectrum is normalized [9.10] to the oscillation frequency by dividing $W_\phi(f)$ by f_o^2 and the quantity plotted is $10 \log[W_\phi(f)/f_o^2]$.

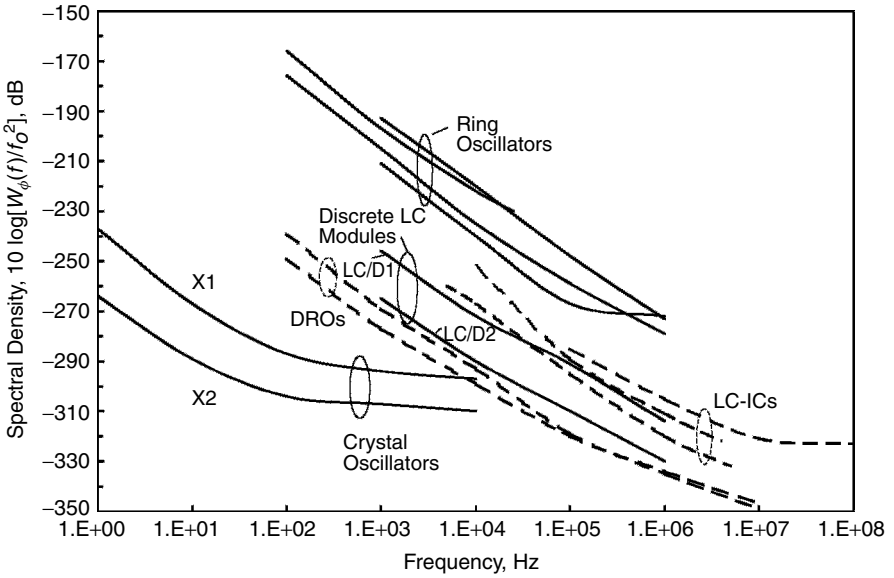


Figure 9.2 Phase-noise spectra of representative oscillators, normalized to a common oscillation frequency.

Why should this normalized measure be used? Signal from an oscillator running in any particular frequency band can be multiplied, divided, or synthesized to deliver output in any other desired frequency band. If the multipliers, dividers, or synthesizers are assumed to be free of phase noise (a fantasy, but a starting point for further evaluation), performance of different oscillators can be compared at a desired frequency and not at disparate individual frequencies. Normalization on oscillation frequency reduces phase-noise spectra from all oscillators to a common basis.

Figure 9.2 shows that oscillators of widely different configurations have phase-noise spectra of similar shapes, and these shapes closely resemble the predictions from Leeson's model. Ring oscillators exhibit the worst phase noise, LC oscillators occupy a broad middle range, dielectric-resonator oscillators (DROs) are somewhat better yet, and quartz-crystal oscillators are far better than all others, particularly at low offset frequencies. These results would be expected on the basis of the relative Q values of the various resonators. These plots illustrate the basis of phaselock frequency synthesizers, as expounded further in Chapter 15. A crystal reference has the best phase noise at close-in offset frequencies, but higher-frequency oscillators usually are better at farther-out offset frequencies. The bandwidth of a synthesizer PLL is chosen such that crystal performance is achieved close-in, while the performance of a controlled oscillator dominates farther out.

Two modular LC oscillators (labeled LC/D1 and LC/D2), manufactured with discrete components, are included in Figure 9.2. These two are packaged in identical cases, operate in the same frequency region, and presumably have inductors of comparable size (implying comparable Q factors). Yet their phase-noise spectral densities differ by nearly 20 dB. The oscillator with the lesser noise density (LC/D2) has a very much smaller tuning range, illustrating the statement at the beginning of the chapter that large tuning range and low phase noise are in conflict with one another.

The modular LC oscillators in the figure exhibit better phase noise than the IC LC oscillators. That advantage can be attributed to Q in the discrete inductors higher than in typical IC inductors and to the higher power levels in the modular oscillators. Even better performance can be expected from LC (or cavity, or transmission-line) oscillators with larger, higher-Q resonators and with higher power levels. Of the two plots for crystal oscillators, that labeled X1 reports the catalog specification on a high-quality commercial unit, thought to be of conventional design, whereas plot X2 shows the noise performance predicted for an unusual proposed design [9.11]. Its principles might well be valuable in applications demanding exceptionally low phase noise.

9.3.4 Shortcomings of Leeson's Model

Leeson's model is based on oscillators with resonators. Figure 9.2 shows ring oscillators that have the spectral shapes predicted by Leeson, but a ring oscillator does not have a resonator; the model cannot be applied for analysis. Leeson's

model is based on a linear amplifier and is supposed to predict the phase-noise spectrum caused by additive white noise. However, none of the oscillators shown in Fig. 9.2 is linear; the simple model does not apply. More to the point, operations of most oscillators are highly nonlinear; the linear model rarely applies.

Leeson's model acknowledges the existence of flicker noise and rightly observes that low-frequency flicker can affect the spectrum of a high-frequency oscillator through angle-modulation effects. But the physical operation and magnitude of the modulation are left open. The model is incomplete; it cannot be used to predict the spectral components caused by flicker noise. More exactly, the flicker corner frequency f_3 is not predictable from the model. In short, Leeson's model provides a base for qualitative thinking and rules of thumb for improved phase noise, but it is not sufficient for critical designs. Something more is needed.

9.4 CLASSIFICATIONS OF OSCILLATORS

As a prelude to a deeper explanation of phase noise, this section describes relevant properties of oscillators. Most oscillators of interest in PLLs are positive-feedback networks, containing at a minimum a frequency-determining circuit (such as a resonator) and an oscillation-sustaining amplifier. Stable oscillations satisfy the Barkhausen criterion—in equilibrium, the magnitude of effective gain around the feedback loop is 1 and the phase around the loop is 0 modulo-2π.

To assure fast, reliable start of oscillations, it is usual to provide more than the minimum required gain: typically by a factor of 3 to 5. Therefore, the oscillator has to include a mechanism that reduces the effective loop gain to unity once oscillations have reached steady-state equilibrium. The amplifier then runs in a linear or nonlinear region, depending on how the amplitude control is arranged. Linear operation is achieved through an element that detects the amplitude of oscillation and adjusts the gain around the loop to hold the amplitude at a desired level within the linear range of the amplifier. The gain adjustment can be in a separate element [9.13, p. 215], or more commonly, the gain of the sustaining amplifier itself is controlled by, for example, adjusting a bias current [9.14–9.17]. [**Comment**: Only [9.16] and possibly [9.14] employ linear sustaining amplifiers. The other papers illustrate amplitude control of nonlinear oscillators.] Harmonics are ideally absent (or very small) in a linear oscillator because of the linearity of the amplifier and the selectivity of the resonator. Class A biasing and drive of a single-transistor amplifier is the most common method for establishing linear operation. Class A oscillators are not often encountered in PLLs.

Level-control loops contain filtering (typically, an integrator) between the sensor and the controlled element. The filter is cascaded with the dynamic envelope response of the oscillator. Envelope response will be slow if the resonator bandwidth is narrow. Care must be taken to assure stable operation [9.18] of the control loop.

Two classes of nonlinear oscillators can be distinguished: those that rely on limiters to set the amplitude of oscillation and those that adjust the conduction

angle of the amplifier to control the amplitude. Additional classes might also exist. It is not always easy to determine the class of a particular nonlinear oscillator; the distinctions are rarely examined in the literature. Examples of limiter-controlled oscillators have not been common in the past; back-to-back diodes in parallel are advocated in [9.19] and a voltage comparator (a hard limiter) in [9.20]. A limiter always conducts current (i.e., always delivers power to the resonator) throughout the oscillation cycle, never shutting off. If the limiter characteristic is symmetric, only odd harmonics will be present in its output. The push-pull LC IC oscillators [9.2] that have become popular recently can be construed as being in the limiter class, as can ring oscillators.

Members of another class of nonlinear oscillators are often called *harmonic oscillators* or—confusingly—*self-limiting oscillators*. In these, the amplifier works in class C operation; current flows in the amplifier for substantially less than half of an oscillation cycle. Initially, before the amplitude builds up, current flows continually. Bias will change during the buildup transient to reduce current and the effective gain. At equilibrium, the oscillation has built up to a large amplitude but so has the bias. In consequence, the amplifier is shut off during most of the cycle and current flows only in short pulses. These pulses drive the resonator only briefly on each cycle; otherwise, the amplifier is disconnected from the resonator during most of the cycle. Trains of short pulses have high harmonic content; nearly sinusoidal voltages appear only because of the narrowband filtering of the resonator. Most of the best known oscillators operate nonlinearly in the class C regime.

An unresolved question has been debated in the electronics community for many years: Which class of oscillator has the lesser phase noise: class A or class C? Earlier literature [9.12, Chap. 7; 9.13, Sec. 6.9] favors class A operation, especially bridge oscillators, with claims of greatly improved frequency stability. Proponents of class C oscillators cite the short duty cycle of amplifier current such that the amplifier delivers noise to the circuit over only a small portion of each cycle. The resonator free-wheels during the cutoff intervals, without any noise disturbance from the amplifier.

The foregoing descriptions of oscillators have been in terms of feedback loops. For many oscillators, the description can be recast as a negative resistance (or conductance) generated by an active circuit and connected to a resonator. In equilibrium, the negative resistance exactly cancels the positive resistance of the losses in the resonator, resulting in stable oscillations. Push-pull LC oscillators are conveniently analyzed from the negative-resistance viewpoint. Some oscillators, such as those using Gunn diodes or IMPATT diodes as the active element, entail no feedback; the active element in fact does generate a negative resistance. These oscillators can only be analyzed as negative-resistance circuits.

All of the foregoing oscillators have included resonators, but other circuits do not; a ring oscillator is an important practical example of the latter. A resonator establishes a narrow bandwidth in the feedback loop and has a phase that changes steeply with frequency. Narrow bandwidth and steepness of phase shift goes with a high Q value in the resonator. Since oscillation takes place at a frequency where

the phase around the loop is an integer multiple of 360°, a steep phase slope in the resonator confines the oscillation frequency to a small range, irrespective of likely phase fluctuations in the amplifier.

A ring oscillator might be regarded variously as a *phase-shift oscillator* or as a *relaxation oscillator*. In a phase-shift oscillator, electrical networks (usually resistance–capacitance circuits) establish a phase shift (typically 180°), and a phase inversion in active elements establishes another 180°, for a total of 360°, as needed for oscillation. The phase slope of an RC phase-shift network is very shallow compared to that of most resonators (its Q value is very low compared to that of any reasonable resonator), so phase fluctuations in the active circuits have a greater influence on oscillation frequency.

A relaxation oscillator has one or more time bases that are established by ramps generated by charging and discharging of capacitors, plus switches that are actuated by the crossing of thresholds by the ramp voltages on the capacitors. Time intervals between threshold crossings determine the period of oscillation.

The attention devoted to classifying oscillators is not for its own sake but because the various linear and nonlinear types react to noise in different ways. The classifications are important in the analysis of phase noise; a procedure valid for one class may not be applicable for another. For example, Leeson's model is supposed to apply only to linear oscillators with resonators. Class distinctions appear again in the next section.

9.5 PHASE NOISE IN OSCILLATORS: ADVANCED ANALYSIS

The number of publications on oscillator phase noise exploded in the 1990s. This heightened interest came about for several reasons:

1. Oscillators were being integrated on-chip, along with all other circuits of a system. A circuit designer has to understand oscillators to be able to design them; that task can no longer be passed off to specialized manufacturers of oscillators, as previously.
2. Phase noise was becoming better recognized as a critical source of system degradation and thus had to be better controlled.
3. IC oscillators, so far, have worse phase noise than that of discrete-circuit oscillators, thereby needing more attention than in past applications.
4. Phase noise is more troublesome at the higher frequencies to which communications links are trending.
5. Understanding of device and circuit models has improved dramatically over the years, making intelligent analysis more feasible.
6. Circuit-simulation programs put extremely powerful tools into the hands of the design engineers, offering glimpses into circuit operation that were difficult or impossible to gain from measurements on operating circuits.
7. Mathematical sophistication has advanced in the engineering community; there are those among us who better understand the mathematics of nonlinear circuits, and they have been writing papers on oscillator phase noise.

Notable publications, listed chronologically, include [9.21–9.32]. Practical application of these papers is highly computer intensive; they do not offer simple formulas. These publications are in contentious disagreement with one another but each makes a contribution worthy of consideration. One hopes that the future will bring a refinement and harmonious synthesis of their disparate theses.

Two questions should be answered by an analysis of oscillators:

1. How does the oscillator transform additive white noise in the vicinity of the oscillator frequency f_o (or, as it turns out, near its harmonics, too) into phase noise whose baseband spectral shape $W_\phi(f)$ is proportional to $1/f^2$? Equivalently, how does the oscillator transform additive sinusoidal interference at frequency $f_o + \Delta f$ into a pair of equal-amplitude, opposite-phase sidebands at $f_o \pm \Delta f$ (neglecting weaker sidebands at $f_o \pm n\Delta f$, $|n| > 1$), each of whose power is proportional to $1/\Delta f^2$?
2. How does the oscillator transform additive low-frequency flicker noise (with spectral density proportional to $1/f$) into phase noise whose spectrum $W_\phi(f)$ is proportional to $1/f^3$? Equivalently, how does the oscillator transform additive sinusoidal interference at a low frequency f_a (typically, audio range) into a pair of equal-amplitude, opposite-phase sidebands at $f_o \pm f_a$, each of whose power is proportional to $1/f_a^3$?

9.5.1 Impulse Sensitivity Function

Hajimiri and Lee have devised a tool for phase-noise analysis of oscillators to which they gave the name *impulse sensitivity function* (ISF). Their most detailed explanation (as of early 2004) is given in [9.21], a shorter introduction is given in [9.27], and ISF information together with summaries of their numerous other publications are collected in their book [9.26]. An ISF quantifies the phase disturbances caused by a noise impulse originating in a specific location in the oscillator at a particular instant in the oscillation cycle. The ISF gives answers to the questions above regarding the transformation of additive interference into phase noise. Each category of additive interference listed in the questions is handled in essentially the same fashion by the ISF procedure; no special treatment is needed for any of them. The ISF method is applicable to all categories of oscillators: linear or nonlinear, resonator-based or not.

An *effective ISF* is defined with regard to each noise source within the oscillator; it is a function of the oscillation waveforms and is periodic. The ISF is most conveniently determined through injecting impulses into a simulation of the oscillator circuit. The periodic effective ISF can be expanded in a Fourier series, and a separate contribution of phase noise is associated with each term of the series.

White additive noise is transformed into $1/f^2$ phase noise. The $1/f^2$ spectral shape is attributed to the accumulation of phase disturbances that is inherent to all

oscillators, without invoking resonator selectivity. The $1/f^3$ phase noise caused by low-frequency $1/f$ flicker interference is determined solely by the intensity of the noise source and the coefficient of the DC term of the Fourier series expansion of the ISF. Low $1/f^3$ phase noise can be attained if that coefficient can be made small. A small coefficient can be achieved if the oscillator has symmetric rise and fall waveshapes, even if the low-frequency flicker noise is large, as in MOS transistors.

Suppression of flicker noise by means of differential oscillator circuits was proposed long ago in [9.34] but was disputed later in [9.35]. Hajimiri and Lee agree with the counterclaim: Differential circuits alone do not suffice to suppress flicker upconversion; symmetrical waveform transitions are required on each device of a differential circuit. An oscillator with a linear amplifier but a separate limiter was proposed in [9.19], the limiter to be implemented by a pair of Schottky diodes connected back to back in parallel. It is unclear whether this arrangement would provide the desired waveform symmetry; further investigation (via ISF analysis) might be fruitful.

Accuracy of prediction of phase noise depends on accurate calculation of the ISF plus accurate knowledge of noise sources within the oscillator circuit. Hajimiri and Lee in [9.21] recount several experiments to test the accuracy of prediction. They show prediction within a fraction of a dB of measurement in the $1/f^2$ region of the phase-noise spectrum, based entirely on theoretically based characterizations of the ISF and the thermal- and shot-noise sources within the oscillators. To predict $1/f^3$ phase noise they had to measure the flicker noise on a sample of the active devices; presumably, it was not feasible to determine flicker noise with sufficient accuracy from theoretical characterization.

Ou et al. [9.33] conducted simulations and measurements on several oscillator circuits. They performed an ISF analysis to predict phase noise, and evaluated phase noise independently with two different commercial simulation programs. Agreement was fair between the programs and the ISF method—discrepancies of fractions of a dB to several dB—but the measured phase noise was consistently off by 3 to 4 dB (in either direction, depending on the particular oscillator) from all predictions. Reasons for the discrepancies were not proffered. Even this much discrepancy is encouraging, as it is closer to reality than earlier analyses.

9.5.2 Nonlinear Analyses for Phase Noise

Several papers approach the phase-noise problem through rigorous nonlinear analyses. Huang [9.29] restricts his analysis to a class C Colpitts oscillator. He derives an amplitude of oscillation based on limitations imposed by the bias current and then attacks question 1 above. In a lengthy derivation, he shows mathematically how a sinusoidal additive disturbance at $f_o + \Delta f$ causes equal-amplitude, opposite-phase sidebands in the RF spectrum at $f_o \pm \Delta f$ and how the power in each sideband is proportional to $1/f^2$. Measured results agree well with predictions of phase-noise spectrum in the $1/f^2$ region. The published analysis does not appear to deal with upconversion of flicker noise or with oscillators that differ substantially from the class C model.

Samori et al. [9.22] analyze a differential LC oscillator using bipolar transistors. They observe that the instantaneous transconductance $dI_{\text{out}}(t)/dV_{\text{in}}(t)$ of the differential stage is an even function of the instantaneous voltage $V_{\text{in}}(t)$ across the LC resonator. The voltage is a periodic near-sinusoid at oscillation frequency f_o, so the instantaneous transconductance can be expanded in a Fourier series with nonzero terms existing for frequencies $2nf_o$, $n = -\infty$ to ∞, including $n = 0$.

Introduce an additive sinusoidal interferer differentially into the amplifier pair at frequency $f_o - \Delta f$, with small amplitude compared to that of the oscillation. That interferer intermodulates with the instantaneous transconductance term at frequency $2f_o$ to generate a third-order product at $f_o + \Delta f$. The intermodulation contains both AM and PM components. Limiting is invoked to suppress the AM, while the $1/\Delta f^2$ dependence is attributed to selectivity of the tuned circuit (not the accumulation of phase fluctuation inherent in oscillators). They also show that noise at $\pm\Delta f$ away from odd harmonics of f_o is folded into noise sidebands at $f_o \pm \Delta f$; this folded noise also consists of AM and PM components.

Considering noise generated in the current source in the tail, the analysis of [9.22] finds noise originating around even harmonics $2nf_o \pm \Delta f$ folding into AM and PM sidebands at $f_o \pm \Delta f$. Contrary to experience, the analysis predicts low-frequency noise at $f = \Delta f$ (in practice, dominated by flicker noise) contributing AM sidebands to the oscillation but not PM. No noise measurements are presented for comparison to the predictions.

Leeson's model predicts that phase noise is inversely proportional to signal power (or the square of signal voltage). Reference [9.37] reports that this ideal dependence on oscillation amplitude is valid only up to a point; in actuality, phase noise increases for large-enough amplitudes. The reference attributes the increase to current-related changes in delay through bipolar transistors (e.g., modulation of base spreading resistance), but the phenomenon can also be attributed to AM–PM conversion in nonlinear reactive elements (e.g., voltage-sensitive capacitances of transistors or varactor diodes). The reference also claims that the upconversion to phase noise of low-frequency additive noise from the tail current source in a push-pull LC oscillator arises from AM–PM conversion and not from the nonlinear effects treated in [9.22].

Tail current noise in push-pull oscillators has received attention as a source of phase noise in several other papers [9.24, 9.38–9.40], which suggest ways to reduce the problem. Levantino et al. [9.40] go so far as to recommend elimination of the tail-current transistor. A paper by Ham and Hajimiri [9.41] deals in considerable detail with trade-offs in the design of differential LC oscillators via ISF concepts. Additional papers on differential LC oscillators are provided in [9.2].

Demir, Mehrotra, and Roychowdhury [9.28] present a nonlinear analysis applicable to any oscillator that can be described by a system of nonlinear differential equations. The published analysis is formal, rigorous, and abstract, employing Floquet theory and stochastic differential equations in the proof. The outcome of an analysis is the RF spectrum $W_v(f)$ or its normalization $\mathcal{L}(f)$, a challenging problem in itself, as discussed in Section 7.3.4. If the spectrum of the

additive noise is white, Demir et al. show that the RF spectrum of the oscillation is Lorentzian (see Section 7.3.4). They briefly discuss two numerical methods for evaluating the nonlinear differential equations and state that these methods are several orders of magnitude faster to compute than Monte Carlo methods. Several examples of noise prediction are shown but no quantitative comparison between prediction and measurement is offered in the article.

A later paper by Vanassche, Gielen, and Sansen [9.86] follows up on [9.28], providing details to assist a practicing engineer. A key feature in [9.86] is the analytical separation of low-frequency envelope processes (most significantly, phase noise) from a high-frequency carrier. Separation permits enormous speed-up of computer simulations. Coram [9.30] examines some technical aspects of the limit cycle that underlies the work of Hajimiri and Lee as well as that of Demir, Mehrotra, and Roychowdhury. He concludes that the latter's approach is rigorously correct, whereas an approximation employed by Hajimiri and Lee, although often acceptable, could be troublesome under some circumstances.

Nonlinear oscillators can misbehave in unexpected ways. A phenomenon known as *squegging* has long been known [9.13, Sec. 6.8]; the time constant of a self-bias network compounds with the time constant of response of the resonator to cause large-amplitude, low-frequency oscillations on top of the high-frequency oscillation desired. The account in [9.13] tells how to guard against squegging. A paper by Maggio, DeFeo, and Kennedy [9.36] reveals that a nonlinear oscillator can become chaotic if designed improperly. The chances of chaos are greatest if the Q value of the resonator is low and if the starting gain of the amplifier is much larger than necessary to assure oscillation.

The foregoing references have dealt with oscillators incorporating tuned resonators; the worst of them have $Q \approx 5$ or 6, whereas the best crystal oscillators have Q values approaching 10^6. By contrast, ring oscillators, which have no tuned circuits, have large bandwidths and thus worse to much worse internally generated phase noise than that of most resonator oscillators. Principles of ring oscillators may be found in [9.1], [9.25], [9.42–9.48], and [9.85].

9.6 OTHER DISTURBANCES

The papers on phase noise cited in Section 9.5 all deal with additive noise sources inside an oscillator with white or flicker spectrum. Although not nearly as well documented, other disturbances also afflict oscillators. External noise sources in close physical proximity to an oscillator can easily cause far more phase noise than internal sources. The worst external noise comes from switching of digital circuits on the same printed-circuit board or—worse yet—the same IC chip as the oscillator. Ring oscillators typically are the most vulnerable to external noise because of their lack of frequency selectivity and their typical high sensitivity to small disturbances. Herzel and Razavi [9.49] provide information on externally caused phase noise. In their analysis of ring oscillators, Hajimiri, Limotyrakis, and Lee [9.25] show that single-ended stages have appreciably less phase noise

from internal sources but recommend differential stages if external noise sources have to be rejected. Heydari [9.87] has analyzed the problem of power, ground, and substrate noise on an IC chip and provided a mathematical model.

External noise conducted on power-supply lines can be combated by separate power lines or by isolation regulators for the vulnerable components. External noise is also transferred through inductive or capacitive coupling; these are layout and isolation issues. The worst external noise coupling is generated in ground leads and substrates that are shared by analog and digital circuits [9.50]. Differential analog circuits by themselves are not sufficient; the circuits need high common-mode rejection to disturbances and have to be designed such that the external disturbances are entirely common mode while avoiding any conversion to differential mode. It also helps to reduce the external disturbances, such as by using differential current-mode circuits in the digital logic.

The physical environment—temperature, pressure, vibration, gravity, supply voltage—is also important, although seldom treated in the literature. Environmentally caused effects on the frequency of high-quality crystal oscillators have been reported in [9.51].

Another disturbance is that of unexplained frequency jumps. These typically are sudden, small jumps in frequency occurring for no obvious reason. Significant frequency jumps occur so rarely that they do not show up in the usual measurements of phase noise and so have escaped wide notice. A PLL responds to a frequency step with a transient in phase error as described in Chapter 5. If the jump is large enough compared to the loop bandwidth, the PLL loses lock and skips cycles until it can reacquire. Such behavior can be extremely disruptive to a system.

Specialists [9.52] have been aware of the phenomenon for a long time, but the problem has not made its way into the broader literature. It does not appear that satisfactory explanations have yet been found. Various explanations have been advanced, as listed below. None are proven; more than one could be correct; something thus far unimagined could be the right explanation. At this time, be aware that jumps really do exist, even if suitable countermeasures are not obvious.

- Frequency jumps might be an artifact of additive Gaussian noise [9.52], an unlikely explanation.
- Frequency jumps might be an extreme and rare artifact of flicker noise [9.52]. Since flicker noise is not understood nearly as well as thermal or shot noise, this explanation is not quite so readily dismissed. If true, devices with high flicker noise (e.g., MOS transistors) should be avoided.
- Frequency jumps might be a part of the normal aging processes of oscillator components. This explanation could be especially applicable to quartz crystals, whose mechanical vibration continually sheds adsorbed gas molecules or tiny fragments of metal electrodes and quartz. A change of mass loading affects frequency.
- Frequency jumps might arise from popcorn noise in the active elements. Popcorn noise is often characterized by a jump in one direction, followed

somewhat later by a jump back to the original condition. Forward-biased PN junctions conduct current in hot spots. Popcorn noise might arise from a sudden jump in the location of the hot spot. That change of location makes small changes in the device properties, which are then reflected into the frequency of the oscillator. If true, bipolar transistors with their forward-biased base–emitter junctions should be avoided in oscillators.

- Changes in temperature might build up stresses in the resonator (or elsewhere in the oscillator circuit) that are relieved suddenly, causing a small jump in frequency. If true, careful thermal-mechanical design is required.
- Alpha particles or similar radiation might strike vulnerable portions of the oscillator, thus causing frequency jumps.

9.7 TYPES OF OSCILLATOR TUNING

Several popular implementations of oscillators have been mentioned in preceding sections of this chapter. This section remarks on selected additional types, divided into two classes: (1) continuous-tuning oscillators and (2) discrete-tuning oscillators.

9.7.1 Continuous-Tuning Oscillators

A continuous-tuning oscillator can be tuned to any frequency within its tuning range. All traditional analog oscillators fall into this class. Issues of continuous tuning are treated further in Section 9.8. Preceding sections have mentioned ring oscillators, differential LC oscillators, Colpitts LC oscillators, and crystal oscillators. Many other oscillator circuits have been devised over the years; for examples, see Edson [9.12]. This section comments on assorted continuous-tuning oscillators of interest to PLL engineers. [**Comment**: Edson's book, although long out of print and dealing solely with vacuum tubes, is one of the most thorough books ever published on oscillators. It is well worthwhile to read if you can get your hands on it; transistors can be substituted for the vacuum tubes.]

Crystal Oscillators The Pierce crystal oscillator [9.12, Sec. 9.7; 9.53, 9.54] has a reputation for simplicity and frequency stability. The references give equivalent circuits and detailed analyses of operation. Oscillation occurs at a frequency where the resonator impedance is inductive, supplied either by running slightly above the crystal's series-resonance frequency or introducing additional inductance in series with the crystal.

LC Oscillators Engineers should be aware of Clapp oscillators [9.55] in addition to the differential and Colpitts LC oscillators mentioned above. The Clapp circuit has been regarded either as a modified Colpitts circuit or as an LC version of a crystal oscillator. It is supposed to have better stability (implying lower

phase noise) than a Colpitts oscillator. Several incompatible explanations for the improved performance have been offered in [9.12, Sec. 8.9; 9.56], and the latest in [9.23]. For a simplistic explanation, intuition suggests that the resonator in a Clapp circuit is better isolated from the noisy and unstable active element than in other LC oscillators.

Quadrature Oscillators Many modern receivers and transmitters make use of quadrature mixers, also known as IQ mixers, for image rejection mixers, for single-sideband generation or recovery, or for downconversion to I and Q baseband signals. The local oscillator for such mixers has to deliver two outputs at the same frequency, closely 90° apart in phase, and of closely the same amplitudes. One way of producing such signals is by means of *quadrature oscillators*.

A promising technique is to put two identical differential LC oscillators on the same IC chip, a placement conducive to close matching between them. The two oscillators are tuned with the same control voltage and would run at nearly the same nominal frequency if they were isolated from one another. Instead, they are cross-coupled in such a manner as to injection lock to one another 90° out of phase. References (listed chronologically) include [9.57–9.62]. A warning is issued in [9.62]: The two oscillators can start up in two different modes in which either oscillator at random might have the leading phase. Typical systems require that the leading–lagging phase relation be rigidly defined, not arbitrary. A method for suppressing the unwanted mode is described in [9.62].

A ring oscillator can be used to generate quadrature outputs; use any multiple of four stages and tap off at the quadrature phases or use differential circuits in a two-stage ring. A ring oscillator is too noisy for many applications. To combat the noise, Kinget et al. [9.63] use a two-stage ring oscillator as a quadrature generator and injection lock it to a low-noise reference. Injection locking is similar to phaselocking with a first-order wideband PLL [9.64–9.67], so the noise of the ring oscillator is tracked out within the effective bandwidth of the injection.

9.7.2 Discrete-Tuning Oscillators

A digital or hybrid PLL is able to deliver only a discrete set of frequencies over its tuning range. Two varieties of discrete-frequency oscillators can be distinguished: those that deliver a time-continuous (i.e., "analog") output signal and those that deliver true digital output (i.e., a sample sequence of numbers). Because the frequency of an input signal to the PLL almost always takes on a continuum of values, the output frequency of a discrete-tuning oscillator almost never can be exactly the same as that of the input. A well-behaved PLL with quantized output frequency switches back and forth between the two adjacent quantized frequencies closest to that of the input. Loop feedback forces the average output frequency to be equal to that of the input. The unavoidable equivocation introduces phase jitter as explained in Chapter 13.

Discrete-Tuned Analog Oscillators Digital-to-analog converters (DACs) are often used to convert a control word from a digitally implemented loop

filter into an analog voltage for tuning an analog VCO. Digital loop filters are attractive in PLLs that require extremely narrow bandwidths, PLLs whose loop filters need large time constants. Such hybrid-digital PLLs are widespread in the telecommunications network, where loop bandwidths of a small fraction of 1 Hz are employed.

Various manufacturers produce numerous DACs, so DAC-controlled tuning is an obvious and successful way to implement discrete-frequency tuning of a VCO. Be aware, though, that DACs can be noisy and that the noise of a DAC is rarely specified by the manufacturer. Also, the combination of required tuning range and required fine spacing of the discrete frequencies may require finer resolution (more bits) than that available in a practical DAC.

A recent development [9.68] has eliminated DACs and, instead, employs switched tuning capacitors within the oscillator itself. The tuning capacitors and associated switches are MOS devices. This arrangement is in line with larger trends: of incorporating all elements onto a single IC chip, of including large numbers of components onto a chip, of the great complexity made feasible by integration, and of finding methods to deal with fine geometry and low voltages. The reference not only describes the circuit innovations but also puts forward a method of analysis and discusses architectural issues that do not arise in conventional VCOs.

Discrete-Frequency Digital Oscillators Number-controlled oscillators (NCOs) were introduced in Section 4.2 and are illustrated in Fig. 9.3. The core of an NCO consists of a register that accumulates a frequency-control word $u_c[n]$ delivered by the loop filter of the PLL. The difference equation for the core is

$$\varepsilon_o[n] = \{u_c[n-1] + \varepsilon_o[n-1]\} \quad \text{mod-1} \qquad \text{cycles} \tag{9.2}$$

where n is the sample index and mod-1 indicates that the accumulator content $\varepsilon_o \in [0, 1)$ is regarded as a phase in fractions of one cycle. The NCO is

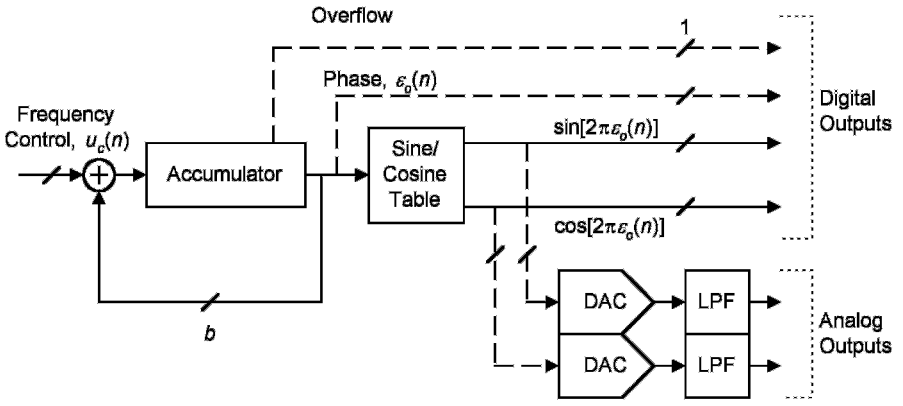


Figure 9.3 Number-controlled oscillator (NCO), illustrating various output options.

clocked at a frequency f_{ck}. Output frequencies f_o in the range $\pm f_{ck}/2$ can be generated without aliasing ambiguities. External considerations might impose a narrower range. Frequency spacing is $f_{ck}/2^b$, where b is the number of bits in the phase register.

Various outputs are possible.

1. An overflow output, implemented, for example, by the MSB in the register, offers a coarse output with a phase irregularity of f_o/f_{ck} cycle. That is usually unsatisfactory.
2. The register content $\varepsilon_o[n]$ is available as a phase with a resolution as fine as $1/2^b$ cycles, actual resolution depending on the number of bits preserved in the output. Since b typically might be anywhere from 24 to 64 bits, the full resolution is not often needed outside the accumulator.
3. Sine and cosine digital outputs can be taken from a look-up table or even calculated. These functions find use in I/Q (complex-signal) frequency conversions within transmitters or receivers, including complex-signal phase detectors.
4. Digital sines and cosines can be converted to an analog staircase in DACs whose outputs are filtered to suppress unwanted artifacts and deliver relatively clean analog sinusoids. The analog configuration is known as a *direct digital synthesizer* (DDS) and has had wide exposure in the literature [9.69, 9.70].

Number-controlled oscillators are versatile, can deliver a wide range of frequencies, can provide extremely small, uniform frequency increments, and have been studied extensively. They grow unwieldy as the number of bits increases and may pose delay and synchronization issues.

A *recursive digital sinusoidal* oscillator (RDSO) [9.71] is an alternative worth considering if a wide tuning range is not required and if a degree of nonuniformity of frequency increments is acceptable. An RDSO is a second-order digital feedback network arranged to provide two outputs that are samples of sinusoids. Turner [9.71] tells how to force the two outputs to be equal in amplitude and in phase quadrature. A level-control mechanism is needed to maintain oscillations in the presence of the inevitable round-off errors. An RDSO might be particularly attractive when the external system requires quadrature sinusoids from the oscillator. The RDSO literature is scant compared to the NCO/DDS literature. Turner [9.71] provides a valuable mathematical foundation but touches only lightly on digital engineering matters. Consequently, more analysis is required when designing an RDSO, and greater engineering risk is involved.

9.8 TUNING OF ANALOG VCOs

A designer of phaselock loops has to deal with the tuning characteristics of a VCO. Much has been written on various oscillators but not much at all on tuning characteristics. This section is a synopsis of my experiences with tuning.

9.8.1 Tuning Curve

Figure 9.4 shows a typical (although artificial) plot of VCO frequency vs. tuning voltage v_c. A decided curvature is evident. Slope at the low-frequency end of the curve is much steeper than at the high-frequency end, a characteristic often encountered. Suppliers of VCOs often specify the curvature of the tuning curve as the maximum departure from the best straight line as a percentage of the full-scale tuning range. The "best" straight line is defined as the line for which the extremes of departure are all equal, as illustrated by the dashed line in the figure. By that definition, the frequency curve in Fig. 9.4 is linear within $\pm 7.3\%$—somewhat better than the $\pm 10\%$ claimed in many data sheets.

This definition of curvature is useless to a PLL engineer despite its pervasive citation in catalogs. Much more important is the varying slope of the frequency curve; this slope is the gain K_v of the VCO, shown as MHz/V in the figure. Because the slope varies, the VCO gain and therefore the loop gain also vary, depending on the frequency to which the VCO is tuned. Gain varies by 5.7 to 1 in Fig. 9.4: a sizable variation but not unusual. An engineer must take the varying gain into account in designing the PLL.

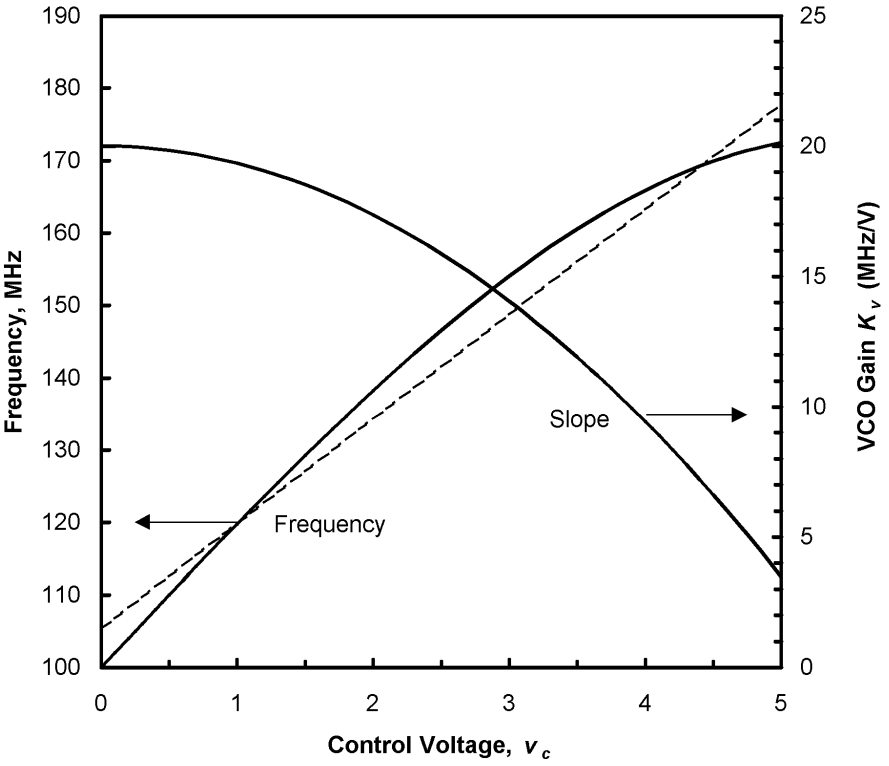


Figure 9.4 Tuning curve, showing nonlinearity, best straight-line approximation (dashed curve), and slope variation.

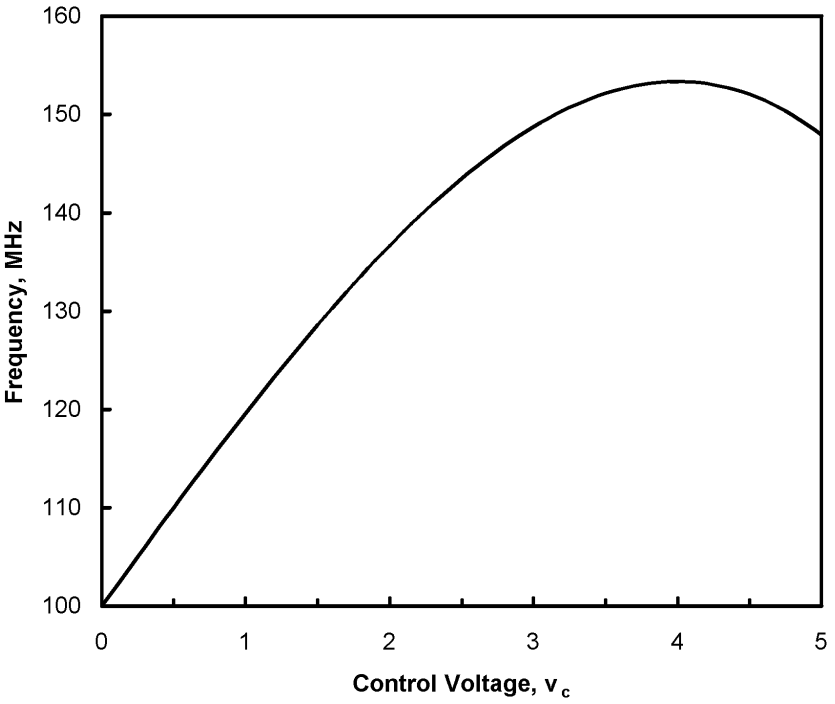


Figure 9.5 Nonmonotonic tuning curve: beware!

Figure 9.5 shows another trap for the unwary. The tuning curve has an extremum inside the control range and the slope reverses sign beyond the extremum. If the control voltage should ever reach the reverse-slope region, the PLL could run to the far limit of control voltage and latch there. Avoid tuning curves with reverse-slope regions.

9.8.2 Tuning Methods

Electrical tuning in resonator oscillators is accomplished most commonly by means of varactors: voltage-sensitive capacitors. Varactor tuning is pursued further after brief mention of other schemes.

Miscellaneous Tuning Methods Reactance modulators—active circuits that present an artificial controllable reactive impedance to the resonator—were employed before the advent of varactors and still appear in some designs [9.72, 9.73]. It is also possible to change oscillation frequency by inserting a voltage-controlled phase shift into the feedback loop of an oscillator circuit, thereby forcing the oscillator to change frequency to maintain exactly 360° phase shift around the loop. The frequency of some oscillators, especially at microwave frequencies, is controlled by varying the bias conditions on the transistor amplifier. The mechanisms involved in such an expedient are not clear

but probably involve, among others, a change of transistor capacitances resulting from a change of bias.

Magnetic tuning also has seen service. Saturable inductors have been used at low frequencies. At microwave frequencies, YIG resonators are tuned by altering the magnetic field in which they are immersed. Magnetic fields are adjusted by control of the current applied to an electromagnet. Magnetic tuning typically offers a large tuning range (an octave or more) but current can only be changed slowly in the high-inductance coils of electromagnets. Often, a magnetically tuned oscillator will have two tuning coils: one large, slow coil for covering the full frequency range, and one small coil, of limited frequency range, suitable for the fast tuning needed for wide loop bandwidths. Iron cores employed in electromagnets are vulnerable to pickup of stray magnetic fields and may also be sources of magnetic fluctuations, both of which cause phase noise.

A relaxation oscillator contains capacitors (one or more) that are charged and discharged by controllable currents. When the voltage across a capacitor crosses a threshold value, a switch is activated to alter the charging conditions in such fashion that oscillations are sustained. Such relaxation oscillators are really CCOs (current-controlled oscillators) but they are usually preceded by a voltage-to-current converter, so that the overall package is a VCO (voltage-controlled oscillator). Ring oscillators might be current-controlled, as above, or might operate by controlling the resistance or even the capacitance in an RC circuit. Elementary principles of ring oscillators are explained in [9.1]. Some low-frequency VCOs [9.74–9.76] employ integrators and multipliers in an analog active-circuit version of the RDSOs described in Section 9.7.2.

Varactors Tuning of LC-VCOs and crystal VCXOs has long been implemented with voltage-controlled capacitors. Reverse-biased pn-junction diodes were employed for many years but are now being displaced by MOS devices. There are several advantages to MOS devices: (1) pn diodes can be driven into forward conduction, thereby ruining capacitance properties, whereas MOS devices remain capacitive for all voltage conditions short of insulation breakdown; (2) capacitance of MOS devices can be adjusted with smaller ranges of control voltage; (3) MOS devices are easy to include on an IC chip.

Diode varactors typically are procured as discrete devices. Their properties are extensively available from manufacturers' data sheets. Since MOS varactors are ordinarily included on-chip, the chip engineer is responsible for their design. Articles on the configurations and properties of MOS varactors include [9.77–9.79], among others.

Varactor Connections Single varactors are sometimes employed when one side of the varactor can be operated at RF ground, an option not always available. As a better arrangement, a matched pair of varactors are used together in a series-opposing connection. For diode varactors, the series connection has an advantage that one diode is always reverse biased, even if the other has been driven into forward conduction; forward current tends to be blocked by the reverse-biased diode.

[**Comment**: The RF current through the capacitor of a parallel-resonant circuit is Q times the terminal RF current of the LC resonator. Voltage across the capacitor of a series-resonant circuit is Q times the terminal RF voltage of the LC resonator. Be alert to high voltages and currents in high-Q resonators. Also, points within the resonator circuit might be at a high impedance level and therefore susceptible to ill effects from unintended loading by external circuits. Consider these factors whenever designing with varactors.]

In a differential oscillator, the use of two opposing varactors places their common point near a virtual RF ground, thereby easing the introduction of control voltage. In most other oscillators, all three terminals of the opposing varactors may have significant RF voltages on them. The RF cannot be allowed to escape into the control-drive circuits, the bias and control circuits should not load the resonator significantly, correct bias has to be delivered to all varactor terminals, and the combination of bias circuits, control circuits, and varactors should not place an undue restriction on the bandwidth of the VCO control port.

Isolation between the varactors and the control circuit is often provided by a resistor. If resonator Q is large, the resistor value has to be very large to avoid unwanted loading. A large resistance produces more thermal noise than a small resistance. If diode varactors are employed, forward current at RF peaks and leakage current cause a DC voltage drop across the resistor, so that the actual control voltage on the varactors is not the same as that delivered by the control circuit. A resistor in conjunction with the capacitance it has to drive (varactor capacitance, plus that of any bypass capacitors that might be present if a single varactor is employed) constitutes a lowpass filter that has to be taken into account in the frequency response and stability of the PLL.

A more elegant connection, where feasible, is an inductor instead of a resistor to isolate the varactors from the control circuits. The inductor can be self-resonant at the oscillator frequency so as to present a high impedance to the RF voltages but a low impedance to the low-frequency control voltage. An ideal inductor is lossless and thus noiseless; a practical inductor generates noise only through its unavoidable losses. The lowpass bandwidth of the inductor in combination with the capacitance of two varactors (assuming that a pair of opposing varactors are employed) is much larger than that of the same varactors with a large resistor instead of the inductor. A lowpass filter consisting solely of a series inductor and shunt capacitor will exhibit large peaking in its frequency response around the series resonance of these two elements. A small resistance is needed in series with the inductor to damp that resonance.

All of the foregoing tacitly assumes that a voltage-variable tuning capacitor will be located at only one place in an oscillator circuit. Winch [9.83] explores combinations of three capacitors whereby tuning range can be increased and substantial improvements gained in tuning linearity.

Varactor Nonlinearity Conventionally, oscillation frequency has been calculated as $f_o = 1/2\pi\sqrt{LC(v_c)}$, where $C(v_c)$ is the resonator capacitance (including the varactor plus all other pertinent capacitors) obtained with a bias voltage v_c on

the varactor. Implied in this calculation is the assumption that the biased varactor responds in a linear manner to the RF voltage, so that the static capacitance determines the tuning. In reality, a varactor is highly nonlinear, especially those that have large relative changes in capacitance in response to small changes of control voltage. Also, the RF voltage typically is large—RF swings might well exceed the bias voltage. The conventional calculation of oscillation frequency is not correct. Improved analyses [9.79, 9.80] take into account harmonics generated by the nonlinear capacitors and arrive at an effective tuning capacitance. The nonlinearities are found to reduce the slope of the tuning curve from that predicted by the conventional calculation.

Also of importance, any additive noise (including low-frequency flicker noise) applied to a varactor is indistinguishable from control voltage and so causes fluctuations of oscillator frequency that show up in phase noise. Furthermore, any amplitude fluctuations of the RF oscillation (amplitude multiplicative noise, not additive noise) are converted to frequency fluctuations because the capacitance alteration caused by the varactor nonlinearity depends on the RF amplitude. These sources of phase noise furnish examples of why variable-frequency oscillators are noisier than fixed-frequency oscillators. See [9.84] for additional information.

9.8.3 Speed of Tuning

One might surmise that the selectivity of a resonator would have a major influence on the speed at which an oscillator could change frequency. I found that surmise to be incorrect (unpublished experiment, 1959). A fast square wave was applied as control voltage to a voltage-controlled crystal oscillator, and its output was monitored on an FM receiver. Oscillator frequency was found to have rise and fall times under 2 μs, whereas the bandwidth of the crystal would lead to expectations of $\sim$2 ms. The main conclusion from the experiment is that the oscillator frequency changed just as quickly as the bias voltage on the varactors could be changed. Speed of change of bias voltage depends solely on the filtering interposed between the source of control voltage and the varactors.

That conclusion has to be modified, depending on the detailed circuitry of the oscillator. Shibutani et al. [9.81] report on the transient of oscillation frequency in response to a step change of tuning capacitance in a Colpitts oscillator. Analysis therein, motivated by [9.82], predicts that instead of the instantaneous change of frequency envisaged by the experiment with the crystal oscillator, the frequency transient in a Colpitts oscillator is prolonged and highly underdamped. This behavior is related to the fact that the resonator in a Colpitts oscillator is a third-order network. A shift in frequency causes a shift in bias; frequency and bias settle together according to the time constants of the circuit. The analysis concludes that no such behavior would arise in a circuit with a second-order resonator circuit. The article develops an equivalent baseband filter circuit to be included in the transfer-function analysis for a PLL containing a VCO exhibiting this behavior.

REFERENCES

9.1 B. Razavi, ed., *Monolithic Phase-Locked Loops and Clock Recovery Circuits*, Reprint Volume, IEEE Press, New York, 1996.

9.2 B. Razavi, ed., *Phase-Locking in High-Performance Systems*, Reprint Volume, IEEE Press, New York, 2003.

9.3 V. F. Kroupa, ed., *Frequency Stability: Fundamentals and Measurement*, Reprint Volume, IEEE Press, New York, 1983.

9.4 D. W. Allan, "Time and Frequency (Time Domain) Characterization, Estimation, and Prediction of Precision Clocks and Oscillators," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* ***UFFC-34***, Nov. 1987.

9.5 D. B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum," *Proc. IEEE* ***54***, 329–330, Feb. 1966. Reprinted in [9.1] and [9.3].

9.6 G. Sauvage, "Phase Noise in Oscillators: A Mathematical Analysis of Leeson's Model," *IEEE Trans. Instrum. Meas.* ***IM-26***, 408–410, Dec. 1977.

9.7 R. Brendel, M. Olivier, and G. Marianneau, "Analysis of the Internal Noise of Quartz Crystal Oscillators," *IEEE Trans. Instrum. Meas.* ***IM-24***, 160–170, June 1975.

9.8 F. L. Walls and A. E. Wainwright, "Measurement of the Short-Term Stability of Quartz Crystal Resonators and the Implications for Crystal Oscillator Design and Applications," *IEEE Trans. Instrum. Meas.* ***IM-24***, 15–20, Mar. 1975. Reprinted in [9.3].

9.9 D. Halford, A. E. Wainwright, and J. A. Barnes, "Flicker Noise of Phase in RF Amplifiers and Frequency Multipliers: Characterization, Cause, and Cure," *Proc. 22nd Annu. Symp. Freq. Control*, 1968, pp. 340–341. Reprinted in [9.3].

9.10 V. F. Kroupa, "Noise Properties of PLL Systems," *IEEE Trans. Commun.* ***COM-30***, 2244–2252, Oct. 1982. Reprinted in [9.3].

9.11 F. L. Walls and S. R. Stein, "A Frequency Lock System for Improved Quartz Crystal Oscillator Performance," *IEEE Trans. Instrum. Meas.* ***IM-27***, 249–252, Sept. 1978.

9.12 W. A. Edson, *Vacuum-Tube Oscillators*, Wiley, New York, 1953.

9.13 K. K. Clarke and D. T. Hess, *Communication Circuits: Analysis and Design*, Addison-Wesley, Reading MA, 1971, Chap. 6.

9.14 D. Aebischer, H. Oguey, and V. R. von Kaenel, "A 2.1-MHz Crystal Oscillator Time Base with a Current Consumption Under 500 nA," *IEEE J. Solid-State Circuits* ***32***, 999–1005, July 1997.

9.15 M. A. Margarit, J. L. Tham, R. G. Meyer, and M. J. Deen, "A Low-Noise, Low-Power VCO with Automatic Amplitude Control for Wireless Applications," *IEEE J. Solid-State Circuits* ***34***, 761–771, June 1999.

9.16 R. A. Bianchi, J. M. Karam, and B. Courtois, "Analog ALC Crystal Oscillators for High-Temperature Applications," *IEEE J. Solid-State Circuits* ***35***, 2–13, Jan. 2000.

9.17 A. Zanchi, C. Samori, A. L. Lacaita, and S. Levantino, "Impact of AAC Design on Phase Noise Performance of VCOs," *IEEE Trans. Circuits Syst. II* ***48***, 537–547, June 2001.

9.18 D. Li and Y. P. Tsividis, "A Loss-Control Feedback Loop for VCO Indirect Tuning of RF Integrated Filters," *IEEE Trans. Circuits Syst. II* ***47***, Mar. 2000.

9.19 P. Grivet and A. Blaquiere, "Non-Linear Effects of Noise in Electronic Clocks," *Proc. IEEE* ***51***, 1606–1614, Nov. 1963.

9.20 S. Pavan and Y. P. Tsividis, "An Analytical Solution for a Class of Oscillators and Its Application to Filter Tuning," *IEEE Trans. Circuits Syst. I* ***45***, 547–556, May 1998.

9.21 A. Hajimiri and T. H. Lee, "A General Theory of Phase Noise in Electrical Oscillators," *IEEE J. Solid-State Circuits* ***33***, 179–194, Feb. 1998. Reprinted in [9.2]. Corrections: 928, June 1998.

9.22 C. Samori, A. L. Lacaita, F. Villa, and F. Zappa, "Spectrum Folding and Phase Noise in LC Tuned Oscillators," *IEEE Trans. Circuits Syst. II* ***45***, 781–790, July 1998.

9.23 A. L. Lacaita and C. Samori, "Phase Noise Performance of Crystal-like LC Tanks," *IEEE Trans. Circuits Syst. II* ***45***, 898–900, July 1998.

9.24 A. Hajimiri and T. H. Lee, "Design Issues in CMOS Differential LC Oscillators," *IEEE J. Solid-State Circuits* ***34***, 717–724, May 1999.

9.25 A. Hajimiri, S. Limotyrakis, and T. H. Lee, "Jitter and Phase Noise in Ring Oscillators," *IEEE J. Solid-State Circuits* ***34***, 790–804, June 1999. Reprinted in [9.2].

9.26 A. Hajimiri and T. H. Lee, *The Design of Low Noise Oscillators*, Kluwer Academic, Norwell, MA, 1999.

9.27 T. H. Lee and A. Hajimiri, "Oscillator Phase Noise: A Tutorial," *IEEE J. Solid-State Circuits* ***35***, 326–336, Mar. 2000.

9.28 A. Demir, A. Mehrotra, and J. Roychowdhury, "Phase Noise in Oscillators: A Unifying Theory and Numerical Methods for Characterization," *IEEE Trans. Circuits Syst. I* ***47***, 655–674, May 2000.

9.29 Q. Huang, "Phase Noise to Carrier Ratio in LC Oscillators," *IEEE Trans. Circuits Syst. I* ***47***, 965–980, July 2000.

9.30 G. J. Coram, "A Simple 2-D Oscillator to Determine the Correct Decomposition of Perturbations into Amplitude and Phase Noise," *IEEE Trans. Circuits Syst. I* ***48***, 896–898, July 2001.

9.31 A. Demir, "Phase Noise and Timing Jitter in Oscillators with Colored-Noise Sources," *IEEE Trans. Circuits Syst. I* ***49***, 1782–1791, Dec. 2002.

9.32 D. Ham and A. Hajimiri, "Virtual Damping and Einstein Relation in Oscillators," *IEEE J. Solid-State Circuits* ***38***, 407–418, Mar. 2003.

9.33 Y. Ou, N. Barten, R. Fetche, N. Seshan, T. Fiez, U.-K. Moon, and K. Mayaram, "Phase Noise Simulation and Estimation Methods: A Comparative Study," *IEEE Trans. Circuits Syst. II* ***49***, 635–638, Sept. 2002.

9.34 H. B. Chen, A. van der Ziel, and K. Amberiadis, "Oscillator with Odd-Symmetrical Characteristics Eliminates Low-Frequency Noise Sidebands," *IEEE Trans. Circuits Syst.* ***CAS-31***, 807–809, Sept. 1984.

9.35 C. P. Hearn, Comments on [9.34], *IEEE Trans. Circuits Syst.* ***CAS-34***, 324–331, Mar. 1987.

9.36 G. M. Maggio, O. DeFeo, and M. P. Kennedy, "Nonlinear Analysis of the Colpitts Oscillator and Applications to Design," *IEEE Trans. Circuits Syst. I* ***46***, 1118–1130, Sept. 1999.

9.37 C. Samori, A. L. Lacaita, A. Zanchi, S. Levantino, and G. Cali, "Phase Noise Degradation at High Oscillation Amplitudes in LC-Tuned VCO's," *IEEE J. Solid-State Circuits* ***35***, 96–99, Jan. 2000.

9.38 B. De Muer, M. Borremans, M. Steyaert, and G. Li Puma, "A 2-GHz Low-Phase-Noise Integrated *LC*-VCO Set with Flicker-Noise Upconversion Minimization," *IEEE J. Solid-State Circuits* ***35***, 1034–1038, July 2000.

9.39 E. Hegazi, H. Sjöland, and A. A. Abidi, "A Filtering Technique to Lower LC Oscillator Phase Noise," *IEEE J. Solid-State Circuits* ***36***, 1921–1930, Dec. 2001.

9.40 S. Levantino, C. Samori, A. Bonfanti, S. L. J. Gierkink, A. L. Lacaita, and V. Boccuzzi, "Frequency Dependence on Bias Current in 5-GHz VCOs: Impact on Tuning Range and Flicker Noise Upconversion," *IEEE J. Solid-State Circuits* ***37***, 1003–1011, Aug. 2002.

9.41 D. Ham and A. Hajimiri, "Concepts and Methods in Optimization of Integrated LC VCO's," *IEEE J. Solid-State Circuits* ***36***, 896–909, June 2001.

9.42 B. Razavi, "A Study of Phase Noise in CMOS Oscillators," *IEEE J. Solid-State Circuits* ***31***, 331–343, Mar. 1996. Reprinted in [9.2].

9.43 J. A. McNeil, "Jitter in Ring Oscillators," *IEEE J. Solid-State Circuits* ***32***, 870–879, June 1997. Reprinted in [9.2].

9.44 S. L. J. Gierkink, E. A. M. Klumperink, A. P. van der Wel, G. Hoogzaad, A. J. M. van Tuijl, and B. Nauta, "Intrinsic $1/f$ Device Noise Reduction and Its Effect on Phase Noise in CMOS Ring Oscillators," *IEEE J. Solid-State Circuits* ***34***, 1022–1025, July 1999.

9.45 L. Sun and T. A. Kwasniewski, "A 1.25-GHz 0.35-μm Monolithic CMOS PLL Based on a Multiphase Ring Oscillator," *IEEE J. Solid-State Circuits* ***36***, 910–916, June 2001.

9.46 O. T.-C. Chen and R. R.-B. Sheen, "A Power-Efficient Wide-Range Phase-Locked Loop," *IEEE J. Solid-State Circuits* ***37***, 51–62, Jan. 2002.

9.47 L. Dai and R. Harjani, "Design of Low-Phase-Noise CMOS Ring Oscillators," *IEEE Trans. Circuits Syst. II* ***49***, 328–338, May 2002.

9.48 S. Docking and M. Sachdev, "A Method to Derive an Equation for the Oscillation Frequency of a Ring Oscillator," *IEEE Trans. Circuits Syst. II* ***50***, 259–263, Feb. 2003.

9.49 F. Herzel and B. Razavi, "A Study of Oscillator Jitter Due to Supply and Substrate Noise," *IEEE Trans. Circuits Syst. II* ***46***, 56–62, Jan. 1999. Reprinted in [9.2].

9.50 P. Larsson, "Measurements and Analysis of PLL Jitter Caused by Digital Switching Noise," *IEEE J. Solid-State Circuits* ***36***, 1113–1119, July 2001. Reprinted in [9.2].

9.51 H. Hellwig, "Environmental Sensitivities of Precision Frequency Sources," *IEEE Trans. Instrum. Meas.* ***IM-39***, 301–306, Apr. 1990.

9.52 J. A. Barnes, *Models for the Interpretation of Frequency Stability Measurements*, NBS Tech. Note 683, National Bureau of Standards, U.S. Department of Commerce, Washington, DC, 1976, Sec. 5.

9.53 E. P. Felch and J. O. Israel, "A Simple Circuit for Frequency Standards Employing Overtone Crystals," *Proc. IRE* ***43***, 596–603, May 1955.

9.54 W. L. Smith, "Miniature Transistorized Crystal-Controlled Oscillators," *IRE Trans. Instrum.* ***I-9***, 141–148, Sept. 1960.

9.55 J. K. Clapp, "An Inductance–Capacitance Oscillator of Unusual Frequency Stability," *Proc. IRE* ***36***, 356–358, Mar. 1948.

9.56 J. K. Clapp, "Frequency Stable *LC* Oscillators," *Proc. IRE* ***42***, 1295–1300, Aug. 1954.

9.57 M. Tiebout, "Low-Power Low-Phase-Noise Differentially Tuned Quadrature VCO Design in Standard CMOS," *IEEE J. Solid-State Circuits* ***36***, 1018–1024, July 2001.

9.58 P. Vancorenland and M. S. J. Steyaert, "A 1.57 GHz Fully Integrated Very Low-Phase-Noise Quadrature VCO," *IEEE J. Solid-State Circuits* ***37***, 653–656, May 2002.

9.59 J. van den Tang, P. van de Ven, D. Kasperovitz, and A. van Roermund, "Analysis and Design of an Optimally Coupled 5-GHz Quadrature LC Oscillator," *IEEE J. Solid-State Circuits* ***37***, 657–661, May 2002.

9.60 P. Andreani, A. Bonfanti, L. Romano, and C. Samori, "Analysis and Design of a 1.8-GHz CMOS *LC* Quadrature VCO," *IEEE J. Solid-State Circuits* ***37***, 1737–1747, Dec. 2002.

9.61 S. L. J. Gierkink, S. Levantino, R. C. Frye, C. Samori, and V. Boccuzzi, "A Low-Phase-Noise 5-GHz CMOS Quadrature VCO Using Superharmonic Coupling," *IEEE J. Solid-State Circuits* ***38***, 1148–1154, July 2003.

9.62 S. Li, I. Kipness, and M. Ismael, "A 10-GHz CMOS Quadrature *LC*-VCO for Multirate Optical Applications," *IEEE J. Solid-State Circuits* ***38***, 1626–1634, Oct. 2003.

9.63 P. Kinget, R. Melville, D. Long, and V. Gopinathan, "An Injection-Locking Scheme for Precision Quadrature Generation," *IEEE J. Solid-State Circuits* ***37***, 845–851, July 2002.

9.64 R. Adler, "A Study of Locking Phenomena in Oscillators," *Proc. IRE* ***34***, 351–357, June 1946.

9.65 K. Kurokawa, "Noise in Synchronized Oscillators," *IEEE Trans. Microwave Theory. Tech.* ***MTT-16***, 234–240, Apr. 1968. Reprinted in [9.3].

9.66 R. Adler, "A Study of Locking Phenomena in Oscillators," *Proc. IEEE* ***61***, 1380–1385, Oct. 1973.

9.67 K. Kurokawa, "Injection Locking of Microwave Solid-State Oscillators," *Proc. IEEE* ***61***, 1386–1410, Oct. 1973.

9.68 R. B. Staszewski, D. Leipold, K. Muhammad, and P. T. Balsara, "Digitally Controlled Oscillator (DCO)-Based Architecture for RF Frequency Synthesis in a Deep-Submicrometer CMOS Process," *IEEE Trans. Circuits Syst. II* ***50***, 815–828, Nov. 2003.

9.69 J. Tierny, C. M. Rader, and B. Gold, "A Digital Frequency Synthesizer," *IEEE Trans. Audio Electroacoust.* ***AU-19***, 48–57, Mar. 1971. Reprinted in [9.70].

9.70 V. F. Kroupa, ed., *Direct Digital Frequency Synthesizers*, Reprint Volume, IEEE Press, New York, 1999.

9.71 C. S. Turner, "Recursive Discrete-Time Sinusoidal Oscillators," *IEEE Signal Process. Mag.*, 103–111, May 2003.

9.72 J. F. Parker, K. W. Current, and S. H. Lewis, "A CMOS Continuous-Time NTSC-to-Color-Difference Decoder," *IEEE J. Solid-State Circuits* ***30***, 1524–1532, Dec. 1995.

9.73 W.-Z. Chen and J.-T. Wu, "A 2-V, 1.8 GHz BJT Phase-Locked Loop," *IEEE J. Solid-State Circuits* ***34***, 784–789, June 1999.

9.74 S. K. Saha, "Linear VCO with Sine Wave Output," *IEEE Trans. Instrum. Meas.* ***IM-35***, 152–155, June 1986.

9.75 S. K. Saha and L. C. Jain, "Linear Voltage Controlled Oscillator," *IEEE Trans. Instrum. Meas.* ***IM-37***, 148–150, Mar. 1988.

9.76 V. P. Singh and S. K. Saha, "Voltage Controlled Oscillator with Sine-Wave Output," *IEEE Trans. Instrum. Meas.* ***IM-37***, 151–153, Mar. 1988.

9.77 A.-S. Porret, T. Melly, C. C. Enz, and E. A. Vittoz, "Design of High-Q Varactors for Low-Power Wireless Applications Using a Standard CMOS Process," *IEEE J. Solid-State Circuits* ***35***, 337–345, Mar. 2000. Reprinted in [9.2].

9.78 P. Andreani and S. Mattisson, "On the Use of MOS Varactors in RF VCO's," *IEEE J. Solid-State Circuits* ***35***, 905–910, June 2000. Reprinted in [9.2].

9.79 R. L. Bunch and S. Raman, "Large-Signal Analysis of MOS Varactors in CMOS-$G_m LC$ VCOs," *IEEE J. Solid-State Circuits* ***38***, 1325–1332, Aug. 2003.

9.80 E. Hegazi and A. A. Abidi, "Varactor Characteristics, Oscillator Tuning Curves, and AM–FM Conversion," *IEEE J. Solid-State Circuits* ***38***, 1033–1039, June 2003.

9.81 A. Shibutani, T. Saba, S. Moro, and S. Mori, "Transient Response of Colpitts-VCO and Its Effect on Performance of PLL System," *IEEE Trans. Circuits Syst. I* ***45***, 717–725, July 1998.

9.82 G. Sarafian and B. Z. Kaplan, "A New Approach to the Modeling of the Dynamics of RF VCO's and Some of Its Practical Implications," *IEEE Trans. Circuits Syst. I* ***40***, 895–901, Dec. 1993.

9.83 R. G. Winch, "Wide-Band Varactor-Tuned Oscillators," *IEEE J. Solid-State Circuits* ***17***, 1214–1219, Dec. 1982.

9.84 S. Levantino, C. Samori, A. Zanchi, and A. L. Lacaita, "AM-to-PM Conversion in Varactor-Tuned Oscillators," *IEEE Trans. Circuits Syst.* ***CS-49***, 509–512, July 2002.

9.85 S. Docking and M. Sachdev, "An Analytical Equation for the Oscillation Frequency of High-Frequency Ring Oscillators," *IEEE J. Solid-State Circuits* ***39***, 533–537, Mar. 2004.

9.86 P. Vanassche, G. Gielen, and W. Sansen, "Efficient Analysis of Slow-Varying Oscillator Dynamics," *IEEE Trans. Circuits Syst. I* ***51***, 1457–1467, Aug. 2004.

9.87 P. Heydari, "Analysis of the PLL Jitter Due to Power/Ground and Substrate Noise," *IEEE Trans. Circuits Syst. I* ***51***, 2404–2416, Dec. 2004.

CHAPTER 10

PHASE DETECTORS

Two broad classes of phase detectors can be distinguished: *multiplier* (or *combinatorial*) devices and *sequential* devices. Multipliers generate their useful DC error output as the average product of the input-signal waveform times the local-oscillator waveform. Multipliers are zero memory devices. A properly designed multiplier is capable of operation on an input signal deeply buried in noise.

A sequential phase detector generates a useful error-output voltage that depends solely on the time interval between a transition of the signal waveform and a transition of the VCO waveform. Other details of the waveform do not contribute to the output. Sequential phase detectors contain the memory of past transitions. They can generate PD characteristics that are difficult or impossible to obtain with multiplier circuits. Because a sequential circuit operates on transitions, it can be intolerant of missing or extra transitions; in consequence, its noise-handling capability is inferior to that of a multiplier.

Sequential PDs are usually built from digital logic circuits (flip-flops, gates) and operate with binary rectangular input waveforms. Accordingly, they are often called "digital" phase detectors and the PLLs that contain them are often called "digital" phaselock loops. This terminology is incorrect; the outputs of most sequential PD are analog quantities and their PLLs are analog circuits. See Chapter 13 for examples of digital PDs and PLLs.

10.1 MULTIPLIER PHASE DETECTORS

If both inputs to an ideal multiplier are sinusoidal, the useful DC output is proportional to the product of the amplitudes of the two inputs and to the cosine

Phaselock Techniques, Third Edition, by Floyd M. Gardner

of the phase difference between them. (The phase error is zero when the phase difference is 90°.) Equations of an ideal multiplier were described in Chapter 6. In addition to the useful output, there is also an unwanted sinusoidal ripple at double the input frequency with amplitude equal to the maximum available DC output level. Ripple must be suppressed to prevent unwanted sidebands from appearing on the VCO. Appendix 10A examines ripple in greater depth. Multiplication can be implemented physically by means of a four-quadrant analog multiplier such as the Gilbert cell [10.1]. Such devices are available as monolithic integrated circuits. Good performance can be obtained in today's (2004) technology at frequencies into the hundreds of megahertz. Needs arise for which true multipliers are the best solution; for an example, see Section 10.5.

10.1.1 Switching Phase Detectors: Principles

A true multiplier provides a useful analytical model for a phase detector, but it is not ordinarily found in actual equipment. Instead, *switching phase detectors* are far more prevalent. Suppose that the sinusoidal VCO drive to a multiplier phase detector is replaced by a square wave of the form

$$v_o(t) = \text{sgn}[\cos(\omega_i t + \theta_o)] \tag{10.1}$$

where the signum function is defined as $\text{sgn}(x) = 1$ if $x > 0$ and $\text{sgn}(x) = -1$ if $x < 0$. (Observe that the frequency of the VCO is shown as ω_i, the same as the frequency of the input signal. Unless stated otherwise, all explanations in this chapter deal with a locked loop.) The square wave is periodic and can be expanded in a Fourier series as

$$v_o(t) = \frac{4}{\pi}\left[\cos(\omega_i t + \theta_o) - \frac{1}{3}\cos 3(\omega_i t + \theta_o) + \frac{1}{5}\cos 5(\omega_i t + \theta_o) + \cdots\right] \tag{10.2}$$

The output of the multiplier is the sum of each individual term of the Fourier series multiplied by the input signal.

Very often, the input signal and noise are bandlimited to a narrow spectrum around the carrier frequency; no harmonics are present at the input. In this case, it is easy to show that the only multiplier product containing a low-frequency (near-DC) component is the one associated with the fundamental frequency of the square wave. All other products only contribute high-frequency ripple. [**Caution**: This property is applicable only if the input signal is free of harmonics.]

Let the input signal be $v_i(t) = V_s \sin(\omega_i t + \theta_i)$. The average value (the DC component) of the product $v_i v_o$ is

$$v_d(t) = \frac{2}{\pi} V_s \sin(\theta_i - \theta_o) \tag{10.3}$$

Using the notation of Chapters 2 and 6, the phase detector gain is $K_d = 2V_s/\pi$ V/rad. In other words, the useful output is identical to that which would have

been obtained if the VCO drive were a sinusoid with amplitude $4/\pi$. The circuit produces exactly the same DC signal and exactly the same low-frequency noise as the equivalent phase detector with sinusoidal drive from the VCO.

But multiplication by a unit-amplitude square wave is exactly equivalent to periodic switching of the polarity of the input; the multiplier can be replaced, without penalty (except for ripple waveform), by a polarity switch. Since a switch is often much simpler and less expensive to build than a linear multiplier, the most common multiplier type of phase detector is really a switching device. Output amplitude and PD gain K_d for a true switching PD is proportional to the input signal amplitude V_s but independent of the amplitude of switching voltage.

The foregoing describes a full-wave switching PD; it generates output on both halves of the switching cycle. Half-wave circuits pass the input signal $v_i(t)$ on, say, the positive halves of the switching cycle and block the input signal on the negative halves. Waveforms of half-wave and full-wave PDs are shown in Fig. 10.1. Average output of a half-wave PD is exactly half the output obtained from a full-wave PD, so gain of a half-wave PD is V_s/π V/rad. Examination of the half-wave waveforms of Fig. 10.1 reveals that the fundamental ripple

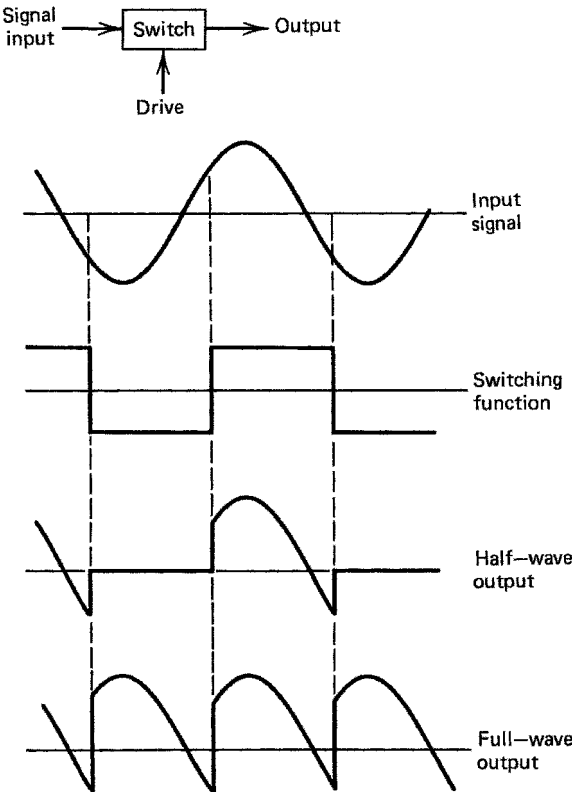


Figure 10.1 Waveforms in switching phase detectors.

frequency is at the signal frequency; ripple is more difficult to suppress than in a full-wave PD, where the fundamental ripple frequency is twice the signal frequency, as shown at the bottom of Fig. 10.1. Therefore, one technique to reduce PD ripple is to employ full-wave circuits rather than half-wave circuits.

10.1.2 Switching Phase Detectors: Examples

Many different device types have been employed as switches in switching phase detectors, including transistors of all kinds, diodes, vacuum tubes, electromagnetic relays, and optoelectronic devices.

Modulators or Mixers Many circuits characterized as modulators or mixers also serve well as phase detectors. (A phase detector can be regarded as a mixer that translates the signal to zero frequency.) These devices can be distinguished as *active* (containing amplifying elements requiring power from a DC supply) or passive (no amplification; no connection to a DC supply).

Active Modulators One popular kind of switching PD is based on balanced modulators (equivalently, mixers). Both single-balanced and double-balanced configurations are used as exemplified in Fig. 10.2. The figure shows bipolar junction transistors in the circuit, but MOS transistors are often used instead, with equivalent operation. Figure 10.3 shows waveforms for the single-balanced circuit of Fig. 10.2*a*. In addition to the ripple typical of a full-wave switching phase detector, there is also a square wave of peak amplitude I_E at the signal frequency, which is much worse than the normal ripple. The square-wave portion of the ripple must be filtered before it enters the output amplifier.

A double-balanced circuit is equivalent to two single-balanced circuits connected together. The polarity of connections is such that the collector-current gaps evident in Fig. 10.3 are filled in and the ripple becomes identical to that of a full-wave switching PD, as in Fig. 10.1. Phase-detector gain of the double-balanced circuit is

$$V_d = \frac{4V_s}{\pi R_E} \sin(\theta_i - \theta_o) \frac{R_B R_C}{R_A + R_C} \tag{10.4}$$

whereas that of the single-balanced circuit is exactly half as great. [Equation (10.4) was obtained under the assumptions that the input signal plus noise does not overload the input transistors; that current gain of all transistors is very large, that internal emitter resistances are included in R_E, and that symmetrically located circuit components are perfectly matched.]

Prior to the advent of well-balanced integrated circuits, use of an active phase detector was precluded by problems of DC offsets. The circuits of Fig. 10.2 would never have been successful using discrete components. Excellent matching of like components on a single IC chip achieves balances that are unimaginable with separate active components. Even so, great care must be taken to achieve close balance between external components and to have a low-impedance, well-balanced input drive if DC offset is to be held to small values. Balanced-modulator PDs have differential, balanced outputs with a common-mode DC

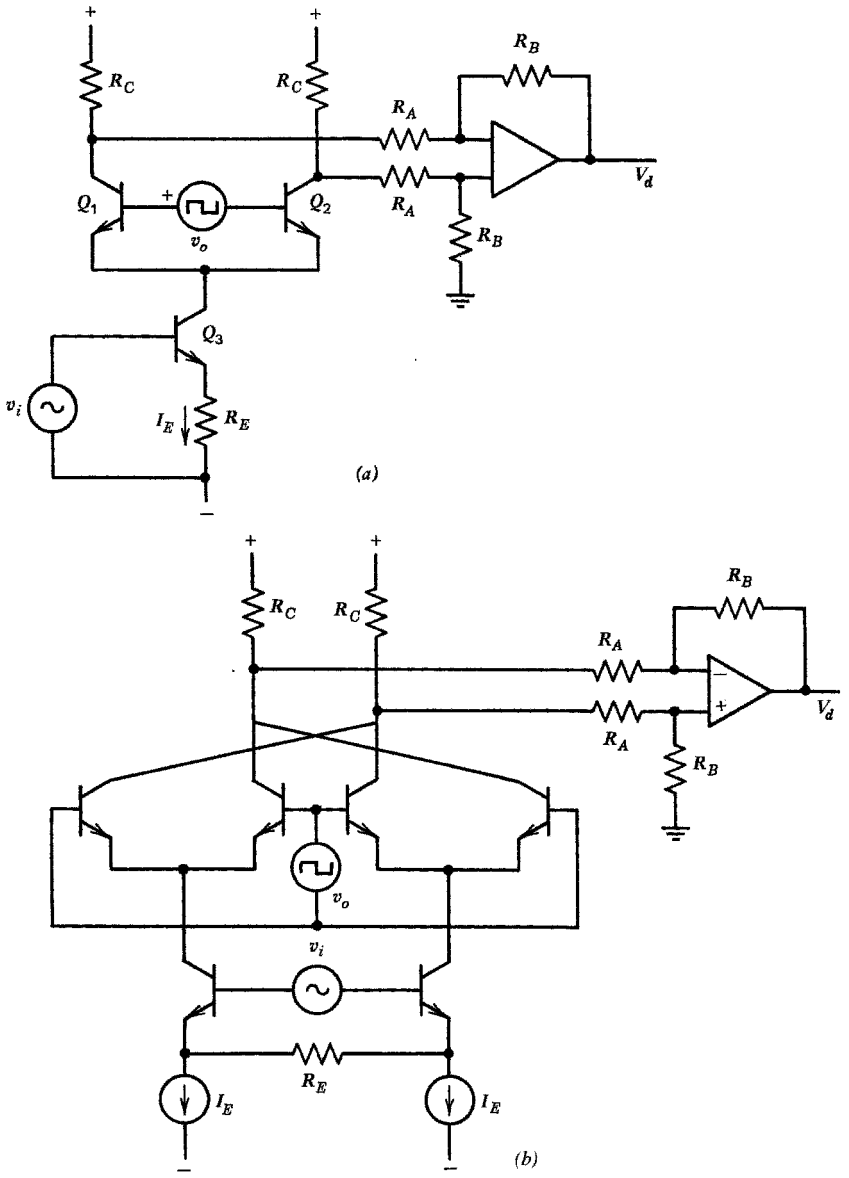


Figure 10.2 Active balanced-modulator phase detectors: (*a*) single-balanced; (*b*) double-balanced.

offset. Most loop-filter circuits have required a single-ended input with zero offset. Figure 10.2 shows differential-to-single-ended conversion by a separate operational amplifier, a technique often applied.

Knowledgeable readers will object that close matching is required among the like-designated resistor pairs in Fig. 10.2 and that the DC amplifier is required to

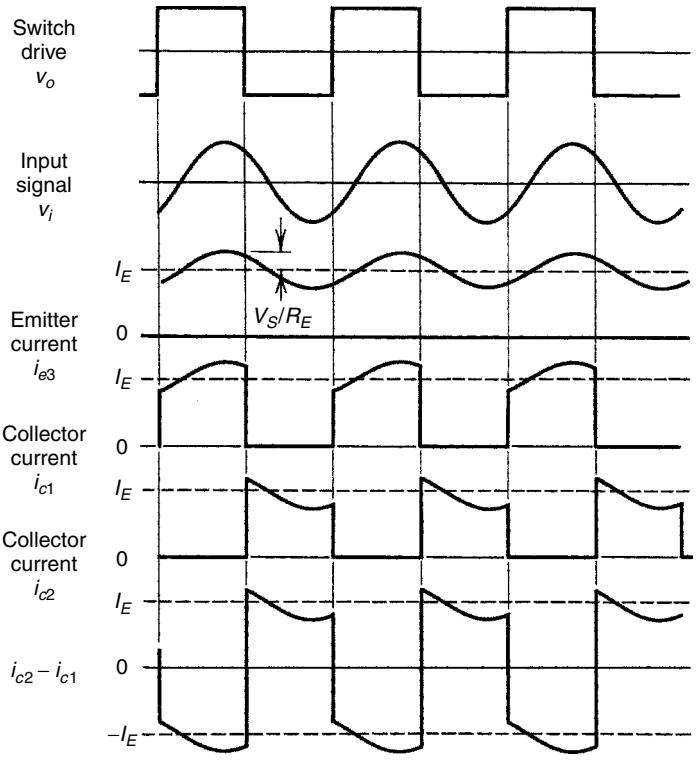


Figure 10.3 Waveforms in an active single-balanced-modulator phase detector.

have low offset voltages and currents. Furthermore, the balanced-to-unbalanced scheme seems overly complicated compared to a current-mirror circuit. These objections are valid. Unfortunately, though, an experiment (unpublished) discovered that current mirrors employing bipolar PNP transistors afford very poor balance, even if the transistors themselves are closely matched. The problem is that the collector voltages at the current-mirror transistors differ greatly on opposite sides of the differential connections, causing the mirror's current gain to depart far from the ideal value of unity, thereby destroying the fine balance that is sought. Therefore, balanced-modulator PDs require op-amps for differential-to single-ended conversions, not PNP current mirrors. I have not investigated the performance of current mirrors using MOS transistors.

Diode Mixers Another popular circuit is the diode ring of Fig. 10.4. These units are sold in large quantities at low cost under the name *double-balanced mixers*. They have wide bandwidths, are available over an extremely large frequency range (they operate well at frequencies far above the capabilities of transistor PDs), impose little burden on the PLL designer, and provide good performance. Accurate analysis is elaborate. If the diodes are assumed to be ideal and if the signal voltage is much smaller than the switching voltage, operation is closely

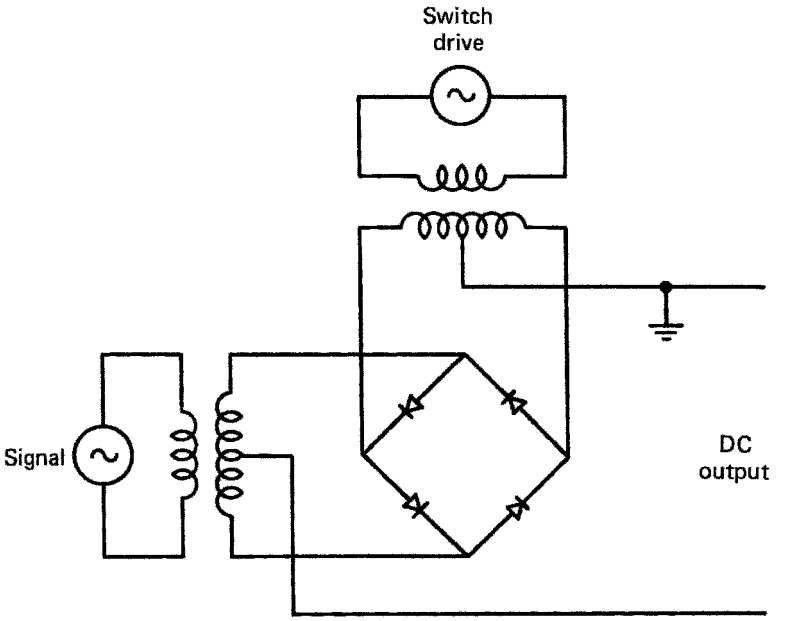


Figure 10.4 Diode-ring phase detector.

the same as that of any full-wave switching PD [10.2, Chap.2]. These conditions are often violated, so the existing analyses are approximations.

Provided that the signal voltage is substantially smaller than the switching voltage, the PD s-curve takes the form $V_d = V_m \sin(\theta_i - \theta_o)$, where V_m is proportional to the signal amplitude. Usual values of V_m range up to 0.3 to 0.4 V. A DC offset on the order of 1 mV is typical. If signal and switching amplitudes are nearly-equal, the PD characteristic becomes triangular instead of sinusoidal (see Section 10.1.4).

A standard diode ring is usually specified for 5 mW of sinusoidal drive from a 50-ohm source. Since the diodes are a nonlinear load, and since the time-averaged load is not necessarily matched, the specification is the available power, not the actual power delivered. If the signal is immersed in noise, the total signal plus noise must be well below the switching drive level if clipping is to be avoided. "High-level" circuits, in which two or more diodes are connected in series in each arm of the ring, can accept larger switching drive and therefore larger input signals. The maximum possible output voltage V_m is proportional to the number of series diodes.

A diode ring is not really characterized very well for phase-detector service. It is fortunate that the circuit is tolerant of a wide range of operating conditions.

Sample-and-Hold PDs Sample-and-hold phase detectors are sometimes encountered [10.4]. A sampler is merely a switch that is driven by a short pulse. Signal value at the instant of the pulse is stored on a capacitor until the next

sample is taken. If the signal is sinusoidal, the PD characteristic is also sinusoidal, with maximum DC output equal to the peak signal amplitude. Sample-and-hold PDs are used to lock to harmonics of the sampling rate, to suppress ripple, or in applications where the signal appears in short bursts. Harmonic operation is discussed in Sections 14.2 and 17.3.2.

If noise is absent and if the input signal is not modulated, the sampling always occurs at the same point of the input waveform from one cycle to the next. The DC value (near zero for equilibrium tracking) does not change. Except for possible sharp spikes at sampling times, due to switch imperfections, the voltage on the storage capacitor remains constant. Ripple is suppressed completely, a valuable property. Analysis of a sampled loop is not accomplished quite accurately by the Laplace transform methods of Chapter 2; it is better to use z-transforms instead. Response and stability of a sampled loop [10.5, 10.6] differ from the time-continuous behavior presented Chapters 2 and 3.

10.1.3 Hybrid-Transformer PD

A PD circuit that was once very common—it was considered *the* phase-detector circuit—is shown in Fig. 10.5. A hybrid transformer forms the vector sum and vector difference of the two input signals; these are converted to DC signals by the diode rectifiers. The useful output is the difference between the two rectified voltages. Analysis shows that the output is proportional to the sine of the phase error and is a function of the two input amplitudes [10.3]. If $V_o \gg V_s$, then V_d is proportional to V_s and is nearly independent of V_o. This insensitivity to the larger input voltage is found in many different PD circuits, including the diode ring presented above. Ripple is reduced from that encountered in the previous PDs because of the nonlinear filtering action in the RC loads of the peak detectors.

The popularity of this circuit has declined with the advent of good ICs and packaged rings. Since output is a small difference between two large DC voltages, balance is a critical adjustment if DC offset is to be avoided. It is much easier to buy a well-balanced integrated or modular circuit than to build one from discrete components. However, the basic circuit should not be dismissed entirely. It has the potential of operating over a frequency range from audio to light. The transformer could be replaced by a coaxial hybrid junction, or a waveguide

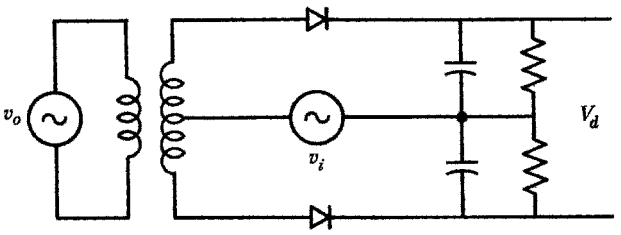


Figure 10.5 Hybrid-based phase detector.

magic-T, or even an optical device. Detectors need not be diode rectifiers; they could also be bolometers, thermocouples, or photodiodes. There is still a niche for the circuit at frequencies above the capability of diode rings or any transistors.

10.1.4 Nonsinusoidal *s*-Curves

Several quite different PD circuits have been examined above, and in each case a sinusoidal *s*-curve (DC error voltage vs. phase error) was found. One might think that a sinusoid is a common property of the various circuits. Actually, the shape of the *s*-curve depends on the applied waveforms, not necessarily on the circuit. For example, the *s*-curve becomes triangular if rectangular waveforms are applied at both inputs of any true multiplier or switching PD. This result comes out of exactly the same circuit that produces a sinusoidal *s*-curve when presented with a sinusoidal input.

If waveforms are rectangular, digital logic gates can be used in place of analog circuits. The digital-circuit equivalent of a switching phase detector is an exclusive-OR gate. Average DC output is a triangular function of the phase error and the ripple waveform is rectangular with a duty cycle that depends on phase error. At zero phase error, the ripple is a square wave at twice the signal frequency, with 50% duty cycle. Note that the output is an analog quantity despite the fact that a digital circuit and digital input waveforms are used.

The *s*-curve of a sampled PD is exactly the waveform of the sampled signal. Almost any desired characteristic can be obtained by appropriate shaping of the waveform to be sampled. For example, a rectangular PD characteristic occurs if the waveform is rectangular. A rectangular PD characteristic has infinite slope at zero phase error, which implies infinite loop gain. Nonlinear analysis as a bang-bang sampled loop is thereby required. A nonlinear PLL can be very useful despite the analytical difficulties. A sawtooth *s*-curve results from sampling a sawtooth waveform, but a sawtooth is more readily obtained from sequential PDs, as is explained in Section 10.2.

Some phase-detector circuits produce nonsinusoidal *s*-curves even when fed with sinusoidal inputs. Example circuits that try to extend the linear range of the *s*-curve are provided in [10.7–10.10]. These circuits are rarely used. One reason is that a sawtooth characteristic (which they try to approximate) is readily obtained with simple sequential circuits, as described in Sections 10.2 and 10.3. Another reason, explored in Section 10.4.3, is that noise degrades any extended characteristic. If the signal is immersed in noise, the *s*-curve of the PD approaches a sinusoid irrespective of its shape for signal alone.

A sinusoidal *s*-curve has the same magnitude of slope at its unstable null at 180° as it does at the stable null at 0° (see Fig. 8.1). The same is true for a triangular or rectangular *s*-curve or any PD with even symmetry about its peak output. Feedback polarity with this type of PD ordinarily is immaterial; the loop selects automatically whichever of the two nulls provides negative feedback. An extended PD characteristic (such as a sawtooth) has unequal magnitudes of slope at the two nulls. To assure stable tracking about the desired null, the polarity

around the entire loop must be correct. Reverse polarity of feedback forces the loop to try to track about the wrong null, usually with unacceptable behavior such as instability. Make sure that the feedback has the correct polarity.

10.2 SEQUENTIAL PHASE DETECTORS

Sequential PDs operate on the transitions of the signal and local oscillator waveforms; any other characteristics of the waveforms are ignored. For reliable operation of the circuits, the waveforms are usually clipped to a rectangular shape. Average output is proportional to the time interval between a transition of the signal and a transition of the VCO waveform. The circuit must incorporate memory to be able to measure the time difference.

Flip-Flop PD The simplest sequential PD is an ordinary RS flip-flop [10.11]. Transitions (say, negative-going) on one input set the flip-flop to a true state and transitions on the other input reset it to the false state. Typical waveforms are shown in Fig. 10.6, and the s-curve—a sawtooth—is shown in Fig. 10.7. This kind of PD has been used in laboratory phase meters and has also had service in the telecommunications network [10.11].

Denote the phase difference between input signal and VCO output by θ_d. Useful output is the DC average V_d on one output terminal of the flip-flop. For

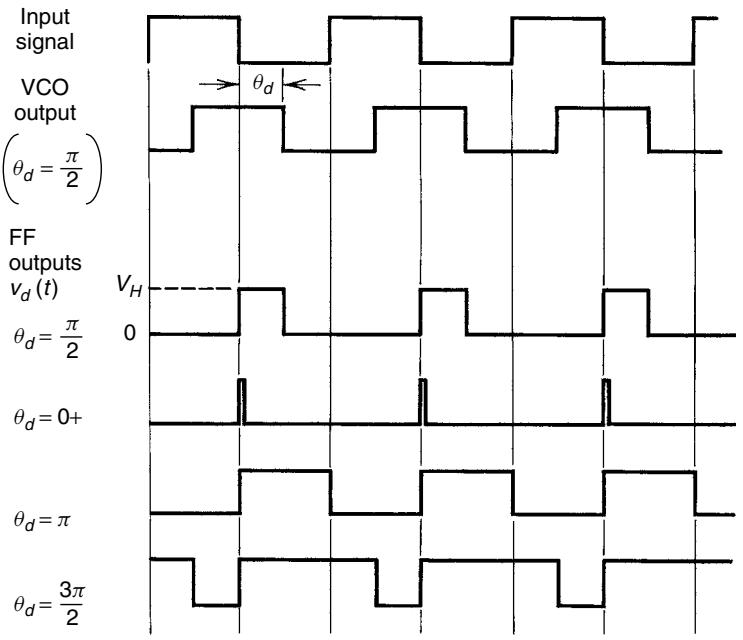


Figure 10.6 Waveforms in an RS flip-flop phase detector.

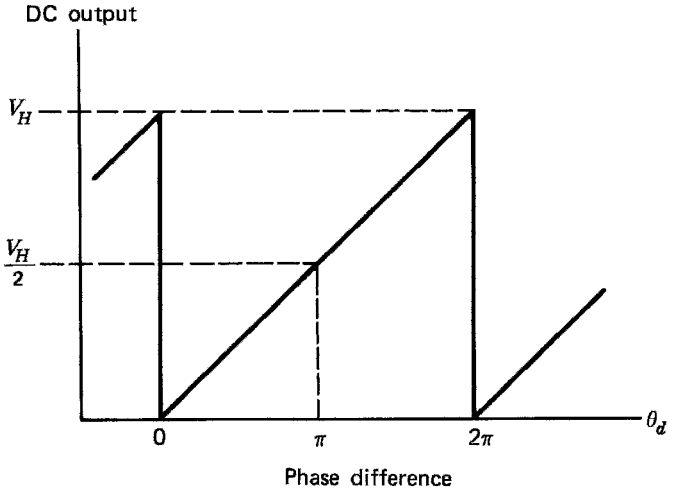


Figure 10.7 s-Curve of a flip-flop phase detector.

$0 < \theta_d < 2\pi$, that output is

$$V_d = \frac{V_H \theta_d}{2\pi} \qquad \text{V/rad} \tag{10.5}$$

where V_H is as defined in Fig. 10.6. The linear range is centered at $\theta_d = 180^\circ$ rather than at 90° as in multiplier PDs. Phase-detector gain is $K_d = V_H/2\pi$. Equilibrium tracking is ordinarily centered around 180°, so the DC offset in V_d must be canceled out with an appropriate bias circuit. Ripple is a square wave at the signal frequency and has a duty ratio that depends on phase error. Duty ratio is 50% if tracking equilibrates at $\theta_d = 180^\circ$.

Digital ICs are manufactured without regard to small noise voltages that might appear on the high or low logic levels of individual devices. If low noise is needed, it is advisable to use the digital circuit to drive a low-noise analog gate—a *charge pump*—to produce the actual DC output. Chapter 12 is devoted entirely to PLLs with charge pumps.

The flip-flop need not be operated at the actual input frequency; digital counters can divide the input frequency by a factor N. The linear range of the PD, referred to the input signal, becomes $2\pi N$ radians, which is in strong contrast to the much smaller range achievable with a sinusoidal, multiplier PD.

Suppose that the input signal to the RS flip-flop fails. Then the next VCO negative transition will reset the flip-flop, and it stays reset until the signal returns. The loop interprets the steady reset condition as a large phase error and attempts to correct it by lowering the VCO frequency. Eventually, the loop filter or VCO is pushed against a saturation limit and remains in this condition. The problem caused by input signal failure is easily remedied in the simple flip-flop: Just arrange the circuit so that the VCO transitions toggle the flip-flop rather than reset

it [10.11]. Then, if the input fails, the flip-flop toggles back and forth between the two logic levels with a 50% duty ratio, which the loop interprets as zero phase error, so the loop tends to remember its existing state and is prepared to resume tracking quickly upon the return of input.

Signal failure illustrates a general problem of sequential phase detectors: The circuit tends to be intolerant of missing or extra transitions. This behavior should be contrasted to that of multipliers, in which the transition, as such, has little influence; the total waveform determines the DC output. This transition-sensitive property has a major adverse effect on sequential-PD operation in the presence of noise, as discussed in Section 10.4.

10.3 PHASE/FREQUENCY DETECTOR

The most important and best-known sequential PD is the *phase/frequency detector* (PFD). Because it is so widely described and employed, it is here given its own section. Brown [10.12] appears to have been the first to disclose the principle of PFDs, and his article was followed shortly thereafter by commercial products [10.14, 10.15].

10.3.1 PFD Configuration

A basic PFD, illustrated in Fig. 10.8, consists of a pair of D flip-flops (*D-flops*) plus an AND gate and a delay (shown as a buffer in the figure) in a feedback connection. The data terminals of the D-flops are held permanently true. Transitions from the input signal (labeled *R* for *reference*) and from the feedback signal (labeled *V* for *VCO*) are applied to clock terminals of the D-flops. Output from one of the D-flops is labeled UP and the other, DN (for *down*). A clock

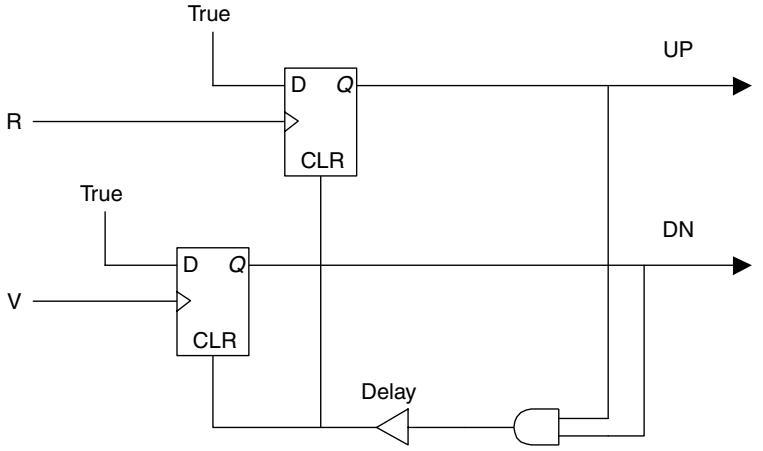


Figure 10.8 Phase-frequency detector (PFD).

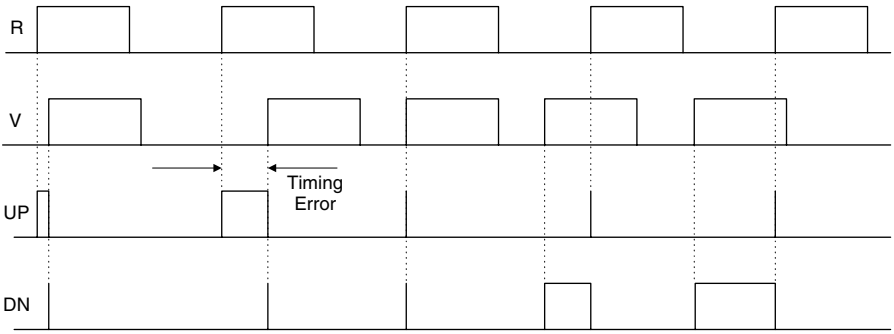


Figure 10.9 Waveforms in a phase-frequency detector.

transition of the correct polarity turns on its associated D-flop. If UP and DN are true simultaneously, as detected by the AND gate, feedback resets both D-flops.

The waveforms in Fig. 10.9 depict idealized operation of the PFD. The signals R and V are shown as rectangular pulses; positive transitions on these pulses actuate the D-flops. The R-pulses in the example are equally spaced, and the V-pulses are assigned various positions to illustrate the operation. (Presumably, V-pulse timing would never be as irregular as shown in the figure.) If R leads V (as at the left side of the figure), the UP D-flop turns on for awhile until the V-pulse turns on the DN D-flop, whereupon both D-flops shut off. If V leads R (as at the right side of the figure), the opposite occurs.

An active UP output tells the PLL to raise the frequency of the VCO since the VCO is lagging behind the input signal. An active DN output tells the opposite. Therefore, UP or DN active outputs give a direction of phase error. The magnitude of the phase error is indicated in the width of the UP or DN pulse, whichever applies.

It is useful to introduce the concept of *net duty ratio*. The duty ratio of an UP or DN pulse is the ratio of the pulse duration to the period of the signal. Denote the duty ratios from the two D-flops as d_{UP} and d_{DN}; the net duty ratio is $d = d_{\mathrm{UP}} - d_{\mathrm{DN}}$. Phase error, in cycles, is given exactly by duty ratio d. If the R and V pulses are aligned exactly, the two D-flops turn on and then off together, very quickly, as illustrated at the center of Fig. 10.9. Net output for that condition is no more than a fleeting glitch resulting from imbalances between UP and DN (or in the succeeding charge pump). That glitch constitutes the phase-detector ripple waveform when the PLL is in equilibrium. Clearly, the ripple energy is very small and its spectral content is widely spread, particularly in comparison to the ripple waveforms for most multiplier PDs shown previously. Both features of the PFD ripple are highly favorable to ripple suppression.

Many authors have described the PFD as a "digital" phase detector because it employs digital logic circuits. That is wrong nomenclature; PFD output information is contained in the widths of the UP and DN pulses, which are continuously variable analog quantities. Phaselock loops that employ PFDs are almost always analog PLLs, not digital.

10.3.2 Delay in PFD

The necessary role of the delay in the feedback path may be seen from inspection of Fig. 10.10. Rather than instantaneous transitions, this figure shows finite rise times instead. A D-flop does not turn on fully until some delay time after the actuating R or V transition begins, and the CLR (reset) pulse does not turn on fully until some delay time after both UP and DN are fully on. The CLR pulse has to be on long enough to assure, with extremely high probability, that both D-flops shut off reliably. That necessary CLR pulse width is assured by inserting delay in the CLR path. Feedback delay is a critical feature of the PFD even though early literature never mentioned it.

Actually, delay has to be rather longer than demanded by reliable switching of the PFD. In most instances, the PFD drives a charge pump, an electronic switch that dispenses charge proportional to phase error into the loop filter on each cycle of phase comparison. (Charge-pump PLLs are the subject of Chapter 12.) Like the D-flops of the PFD, those switches require a finite time to turn on and off. If the UP and DN on-intervals are too short—they are shortest when phase error is small—the charge-pump switches never turn on at all. That inserts a dead zone into the s-curve of the PFD–charge pump combination.

A feedback loop with a dead zone in its s-curve is never able to settle to a firm equilibrium. Instead, it wanders aimlessly around in the dead zone. Wandering shows up as noise, usually of comparatively low frequency within the bandwidth of the PLL, causing unwanted, unfilterable phase-noise modulation on the VCO. Furthermore, the dead zone is a nonlinearity that causes intermodulation among noise components that might be present at the PFD. Intermodulation reshapes the spectrum of that noise, transforming filterable high-frequency noise into unfilterable low-frequency noise. Noise intermodulation is revisited in Chapter 15.

To obviate the worst effects of a dead zone, it is common practice to design sufficient delay into the PFD to make both D-flops turn on long enough so that

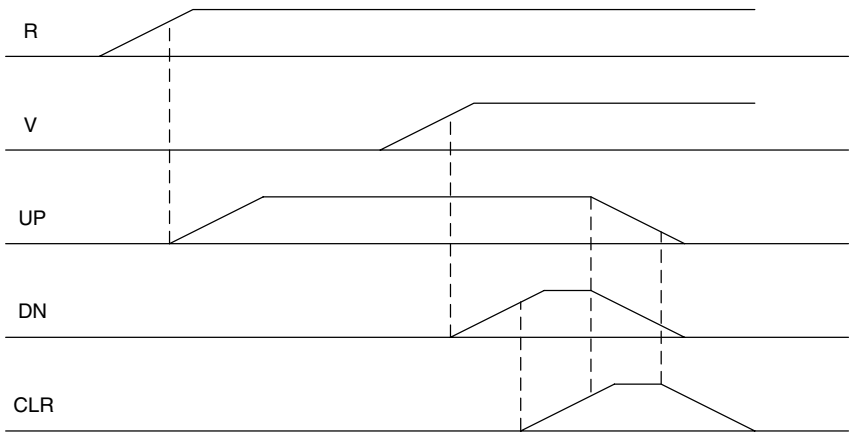


Figure 10.10 Waveforms in a PFD; expanded time scale to show propagation delays.

both charge-pump switches are forced to be all the way on simultaneously during each cycle. Charge pumps are supposed to be designed such that the UP and DN currents are equal, and thus the net charge transferred to the loop filter should be zero while both switches are on simultaneously. A nonzero phase error will cause one charge switch to be turned on for longer than the other, even for very small phase errors, thus eliminating (or at least ameliorating) the dead zone.

The waveforms of Fig. 10.9 are idealized; they assume negligible delay compared to the period of the signals. Delay interferes with the operation of the PFD, particularly at large phase errors. Satisfactory operation requires that the delay be small compared to the signal period. Thus, necessary delay implies an upper limit on the operating frequency of the PFD. This limitation appears in some form in all switching circuits, not just in phase detectors.

10.3.3 PFD State Diagram

Understanding of the PFD is enhanced with the aid of a state diagram [10.13, p. 25; 10.16, 10.17]. The two D-flops, the two memory elements of the PFD, can each be in one of two states, on or off, so there are four distinct possible states among the two elements. The state with both D-flops shut off is known as the *zero* or *null state* (here designated the *N state*), another state in which only the UP D-flop is turned on is called the *UP state*, and a state in which only the DN D-flop is turned on is the *DN state*. One state, that in which both D-flops are turned on simultaneously, is transient because of the feedback that quickly shuts off both D-flops; it is called the *CLR state*.

Figure 10.11 is a diagram of those states (in labeled circles) along with the allowed transitions between states (directed arcs). Each arc has a label that indicates the event (R or V clock edge) that causes a state transition. For example, starting from the null state, an R edge causes a transition to the UP state and a succeeding V edge causes a transition back to null but first passing through the transient CLR state; if a second R edge should occur before arrival of the

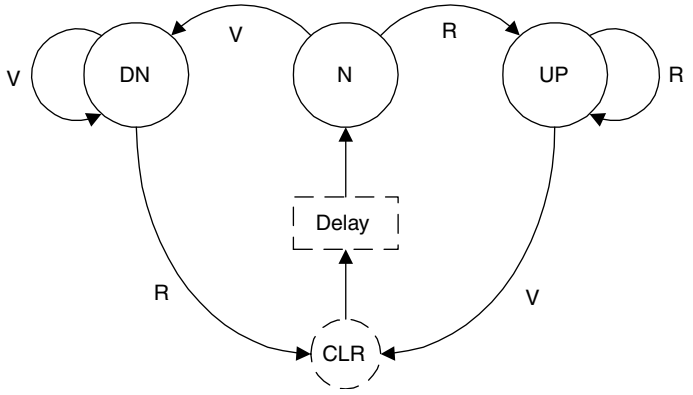


Figure 10.11 State diagram for a PFD; dashed blocks indicate a transient state.

next V edge, the PFD remains in the UP state; and so on for the other possible conditions. [**Comment**: The transient CLR state is customarily omitted from these state diagrams. It is included here because a number of otherwise obscure issues, treated subsequently, are clarified by its inclusion.]

10.3.4 PFD s-Curve

Outcome of a thought experiment helps introduce the *s*-curve of the PFD. Consider the hypothetical test setup shown in Fig. 10.12. It consists of a clock source at frequency f_c, a variable delay τ, a PFD, and a pair of averaging filters on the two outputs of the PFD. Output of the clock generator is split into two paths; one path goes directly to the R terminal of the PFD while the other path goes through the variable delay to the V terminal of the PFD. Averaged outputs of the PFD, labeled d_{UP} and d_{DN}, are to be explained as a function of the delay τ. To that end, refer to Fig. 10.13, consisting of plots of d_{UP} and d_{DN} vs. τ. Suppose that the PFD starts in its null state, that the variable delay at the starting instant is τ_0 (i.e., V edges initially lag the R edges by τ_0), and that the first clock pulse

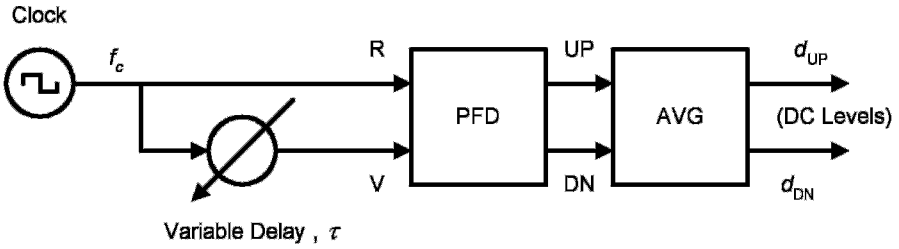


Figure 10.12 Hypothetical test setup to determine the *s*-curve of a PFD.

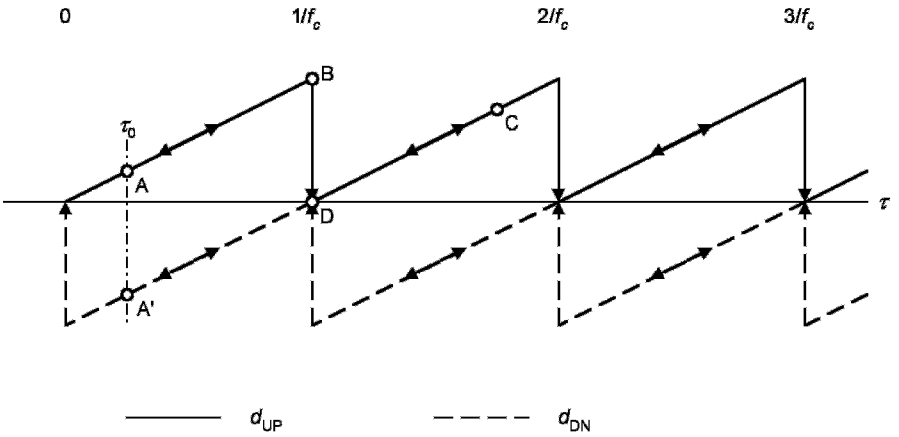


Figure 10.13 *s*-Curve of an ideal PFD.

to reach the PFD is from the R path. Those initial conditions are marked by point A in Fig. 10.13.

Because the R and V signals have the same frequency, every R edge is always followed by a V edge before another R edge can arrive. From the state diagram of Fig. 10.11, the PFD states will cycle from N to UP to CLR to N repeatedly, never entering the DN state as long as τ remains unchanged. For now, assume that the dwell time in the CLR state is negligible compared to $1/f_c$. Consequently, the duty ratio d_{UP} is $\tau_0 f_c$ and the duty ratio d_{DN} is zero. Now gradually increase the variable delay so that V lags R by an increasing amount. Duty ratios will be $d_{\mathrm{UP}} = \tau f_c$ and $d_{\mathrm{DN}} = 0$, up to point B at which $\tau f_c = 1$: that is, that point at which R and V edges coincide and the phase shift between R and V is $360^\circ = 0^\circ$ modulo-360°. Duty ratios of both UP and DN fall to zero at that boundary. Further increases of τ leave $d_{\mathrm{DN}} = 0$ and d_{UP} linearly increasing from zero, following the sawtooth shown in the figure.

Let the delay τ keep increasing until the point C is reached, then reverse the direction of delay variation so that τ decreases thereafter. At first, d_{DN} remains equal to zero and d_{UP} simply retraces the sawtooth that it traversed on its way forward. But retrace ends at point D. Discontinuities in the sawtooth are unidirectional, as indicated by the arrowheads. To the left of D the PFD will now interpret timing as V edges leading R edges, so the DN state will become active, and as long as τ continues to decrease, the PFD will never enter the UP state again. For decreasing τ, the PFD output follows the dashed sawtooth for d_{DN}.

Now the s-curve can be discerned; the thought experiment reveals the existence of two intermeshed s-curves. Which one is active depends entirely on chance in the starting conditions: whether an R or a V edge happens to be the first to arrive, as exemplified by points A and A'. Once one s-curve has been selected, the PFD stays on it, in the absence of disruptive events that might cause it to jump to the other track. Each phase-detection s-curve has a linear shape over a range of $\pm 2\pi$ about its $d = (d_{\mathrm{UP}} - d_{\mathrm{DN}}) = 0$ position. The duty ratio amplitude varies from -1 at one extreme to $+1$ at the other extreme. The slope (related to PD gain) is $(d_{\max} - d_{\min})/4\pi = 1/2\pi$ rad^{-1}. The two s-curves are shifted by 2π from one another.

10.3.5 Frequency Detection in a PFD

The state diagram of Fig. 10.11 and the duty-ratio trajectories of Fig. 10.13 taken together provide an explanation of the frequency-detection capability of a PFD. Assume that the frequency f_{R} of the R input is slightly larger than the frequency f_{V} of the V input and that the frequencies are unvarying. The state diagram shows that the PFD state will circulate among N, UP, and CLR, but never enter DN. The duty-ratio trajectory will uniformly trace out the upper (solid line) portion of Fig. 10.13 and never enter the lower (dashed line) portion. The average duty ratio of a sawtooth trajectory with unity maximum amplitude is 0.5; that is the average indication applied to the loop filter instructing the VCO to increase its frequency. Similarly, if f_{R} is slightly less than f_{V}, the average duty ratio from the PFD will be -0.5.

Numerous instances will arise in which two or more R edges will follow one another before a V edge occurs if f_R is substantially larger than f_V; the UP state persists for one or more entire cycles of f_R. The presence of these events increases the average duty ratio from the value of 0.5 delivered when the two frequencies are nearly equal. As the ratio f_R/f_V becomes large, the average duty ratio thus approaches 1. Similarly, as the ratio f_R/f_V approaches 0, the average duty ratio approaches -1. Goyuer and Meyer [10.18] performed an analysis of the idealized frequency-detection indications and arrived at

$$d = \begin{cases} 1 - \dfrac{0.5 f_V}{f_R} & \text{if } f_R > f_V \\ \dfrac{0.5 f_R}{f_V} - 1 & \text{if } f_V > f_R \end{cases} \tag{10.6}$$

10.3.6 Effects of Delay in a PFD

The foregoing descriptions of phase-detector *s*-curves and frequency-detector characteristics are simplifications based on the negligible influence of feedback delay in the CLR path of the PFD. Feedback delay cannot be neglected if the signal frequencies are large enough; the following explores a couple of effects from that delay.

Figure 10.13 shows discontinuities in the sawtooth *s*-curves at phase errors of $\pm 2\pi$. That idealization is not attainable in real-life PFDs because of the feedback delay. Visualize a PFD operating at a phase error close to $+2\pi$. That is, the V edge comes almost one full cycle later than the R edge. Duty ratio d_{UP} should be nearly $+1$ and d_{DN} should be zero. But those duty ratios will obtain only if all edges are tallied properly by the PFD. And that is where the feedback delay gets in the way. If the delay is long enough that the duration of the CLR state encompasses the next R pulse (closely following the V pulse that initiated the CLR state), that R pulse will have no effect; it will be lost in the refractory interval of the CLR state. The next V pulse is the next pulse that is able to trigger a D-flop and it will activate the DN state, which is soon terminated by the closely following R pulse.

Loss of the one R pulse causes the PFD to shift from one s-curve to the other. The PFD now indicates that R is lagging V by a small amount instead of leading it by almost one full cycle. Depending on the reason why R should have been leading V by such a large amount, this behavior suggests that loss of the pulse might initiate a cycle slip. In any event, the perfect sawtooth *s*-curve of Fig. 10.13 deteriorates in the vicinity of the discontinuities in some fashion that, as far as I know, has not yet been published. Clearly though, the idealized phase-error range of $\pm 2\pi$ will be reduced by the existence of the feedback delay.

Frequency detection also suffers from the presence of feedback delay. Goyuer and Meyer [10.18] conclude that frequency detection fails completely if the feedback delay exceeds half the period of the reference source. Lesser feedback delays reduce the magnitude of the average duty ratio that constitutes the useful

frequency-error indication. The adverse mechanism is the loss of otherwise valid R or V edges in the refractory intervals when the PFD is dwelling in the CLR state. Suppose that $f_R > f_V$; an ideal PFD without feedback delay will cycle among N, UP, and CLR as described above and never enter the DN state. But in a realistic circuit, some R edges will arrive while the PFD is in the CLR state and thus be lost. The next V edge sends the PFD into the DN state, around which it may cycle for several more reference periods. Eventually, the PFD recovers—enters a cycle around the UP state—but frequency error is indicated in the wrong direction as long as the DN-state cycle survives. The frequency-detection predictions of (10.6) are optimistic in the presence of nontrivial feedback delay. The presence of delay-caused reversals were displayed in [10.19].

10.3.7 Extra or Missed Transitions

Some signals or signal conditions can introduce too few or too many transitions at the R input. If a signal transition is missing, or if an extra one appears, the PFD interprets this event as a loss of lock and tries to reacquire lock. Since it has its own memory, the effects of an extra (missing) transition propagate for more than one cycle. If the loop is tracking with small error, a missing transition will cause a very large error indication to appear for at least one cycle. Accordingly, the PFD is intolerant of missing or extra transitions.

As one example, binary data signals predominantly are transmitted in the *non-return-to-zero* (NRZ) *format*, in which the signal value changes from one bit to the next only if there is a change in the data value. The probability of data transitions is 50% for random binary data. A PFD of the kind described above would interpret such an input stream as a signal at a much lower frequency than the actual bit rate and attempt to lock improperly. Modified PFDs have been devised for specialized NRZ streams and have been employed in large numbers in the floppy disk drives of personal computers. Additional elements—an enable (EN) latch and a delay—are added to the PFD elements of Fig. 10.8. The R input is applied to both the EN clock terminal and the delay. When EN is false, the D terminals of the D-flops of the core PFD are held false, so the PFD output is clamped in its null state. A data transition (of either polarity) asserts EN, whose output enables the UP and DN D-flops (i.e., sets their D-inputs true). After a delay of typically half the bit period, the data transition is applied as clock to the UP D-flop and the DN D-flop is clocked by the V signal as in a regular PFD. Feedback from the AND operation on UP and DN is applied to clear EN, which clears the UP and DN D-flops in turn.

As another example, large noise can induce extra or missing zero crossings into a signal waveform; the number of crossings depends on the noise spectrum and the signal-to-noise ratio [10.20]. If the number of crossings per second departs from the signal frequency, the phase-frequency detector acts as though the loop is out of lock and the PLL tries to alter the VCO frequency to bring the loop back to "lock." At the very least, the wrong number of crossings will cause a bias of the PD output; tracking will fail entirely if the number of crossings is sufficiently

wrong. A sequential PD should be used in a noisy environment only with great caution and for well-justified reasons. This problem is mitigated somewhat if the noise spectrum is shaped so that the rate of noise crossings (of the polarity for the clock active edge for the D-flops) is equal to the signal frequency. A noise spectrum with arithmetic symmetry about the signal frequency has the desired property.

10.3.8 Lock Indicator for a PFD

Section 8.4.1 explained the use of a auxiliary phase detector in quadrature with a main phase detector of the multiplier class as a widely used indicator of phase lock. That scheme does not work with a PFD or with the RS flip-flop phase detector; the phase relations are wrong. The PFD, though, lends itself to an effective and simple scheme, as described below.

A two-input OR gate takes as its inputs the UP and DN outputs of the PFD. When the PLL is locked with small phase error, neither UP nor DN is true for any but very short intervals during each comparison cycle. When the PLL is out of lock, either UP or DN will be true, on average over many cycles, for 50% or more of the time. The basis of lock detection is to pass the output of the OR gate through a smoothing filter to extract its average dwell time in the true state and to compare that average against a suitable threshold (say, 25% average true dwell time). The PLL is deemed to be locked if the average true time is below the threshold and unlocked if the average true time is above the threshold.

Lock detectors (all kinds, not just for the PFD) also frequently include a timer that requires the lock indication to persist for a specified time interval before phase lock is declared. The timer is started when the average dwell time falls below threshold and reset to zero whenever the threshold is exceeded before the timer reaches its specified interval.

10.4 BEHAVIOR OF PHASE DETECTORS IN NOISE

Phaselock loops are sometimes required to operate with a very poor signal-to-noise ratio at the signal input to the phase detector. Properly designed PDs of the multiplier class are operable with signals deeply immersed in the noise but sequential PDs are much less robust. This section therefore deals only with PDs of the multiplier class. Although all of the results presented here were developed for analog PDs, they should apply to comparable digital PDs as well.

10.4.1 Bandpass Limiters

An introduction to limiters is needed as a preliminary to examining the effects of noise on phase-detector operation. Attention is restricted to an ideal bandpass hard limiter. It is *bandpass* because a narrowband filter, centered at the signal frequency, precedes the limiter proper. A *hard* limiter has an input voltage v_i and output $v_L = V_L \operatorname{sgn}(v_i)$, a rectangular waveform that preserves the locations of the zero crossings of the filtered input. A zonal filter may (or may not) follow

the limiter to remove all harmonics and pass only the fundamental band. Limiter action has been analyzed for an input consisting of a sinusoidal signal plus Gaussian noise [10.21–10.23]. Various interesting properties are revealed by the analyses and summarized in succeeding paragraphs.

Output power from the limiter is constant, irrespective of input signal-to-noise ratio. Since the output waveform is a square wave of constant amplitude, this result is hardly surprising; the only effect of noise is to cause jitter of the zero crossings of the square wave. Moreover, the output power in each zone (i.e., each harmonic band—fundamental, third harmonic, fifth harmonic, etc.) is constant irrespective of input SNR. A symmetric limiter does not generate even harmonics. In the absence of noise, the fundamental component of the rectangular output of the limiter is a sine wave with amplitude $4V_L/\pi$. When noise is added to the input, the signal component of the output must decrease because the total output signal plus noise is held constant; noise suppresses the signal in a limiter. Signal suppression is given the symbol α and is a function of the input signal-to-noise power ratio ρ_i as measured in the passband of the input filter. Interpret α as the ratio of the fundamental signal amplitude at input SNR ρ_i to the amplitude $4V_L/\pi$ that obtains in the absence of noise. Signal suppression is given by

$$\begin{aligned}\alpha &= \sqrt{\frac{\pi\rho_i}{4}}\left[I_0\left(\frac{\rho_i}{2}\right) + I_1\left(\frac{\rho_i}{2}\right)\right]e^{-\rho_i/2} \\ &\approx \sqrt{\frac{\rho_i}{\rho_i + 4/\pi}}\end{aligned} \tag{10.7}$$

where I_0 and I_1 are modified Bessel functions. The ratio α is plotted in Fig. 10.14. The approximate formula is more than accurate enough for engineering calculations.

The gain K_d of a multiplier-type phase detector is proportional to the signal voltage applied. If the signal voltage is suppressed by a factor α, the PD gain is

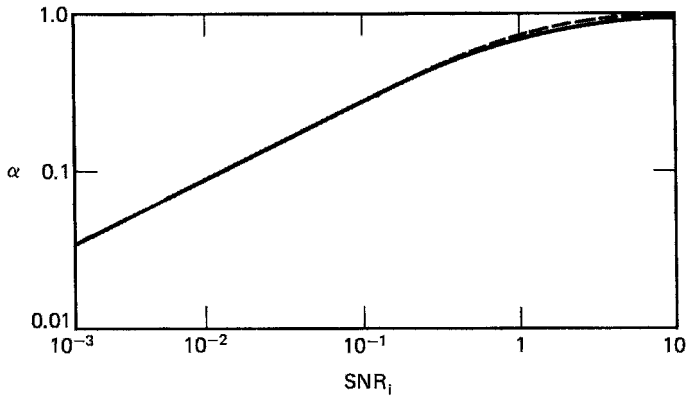


Figure 10.14 Signal suppression factor α of a limiter. The solid curve is an approximation; the dashed curve is exact.

also reduced by a factor α. Consequently, loop gain, damping, and bandwidth are a function of the input signal-to-noise ratio if a limiter precedes the phase detector. Signal suppression is a major effect of a limiter and must be taken into account in the loop calculations. Output signal-to-noise ratio SNR_o in the fundamental zone is also interesting [10.21]. Analyses show that the output SNR_o, is degraded by no more than 1.05 dB for very low input SNR values ($\rho_i \ll 1$) and shows an *improvement* of 3 dB at very large input SNR values.

These SNR_o results are correct but cannot be applied uncritically to the analysis of PLLs, contrary to early thinking. A 3-dB improvement was the first feature to be recognized as irrelevant. Even though a limiter does indeed improve SNR_o by 3 dB for large values of ρ_i, that improvement does not accrue in any way as reduced phase jitter in the PLL. High-SNR improvement reflects the suppression of the AM component of noise; the limiter has no influence on the PM component. Since PLL jitter depends on phase, not amplitude, the suppression of AM noise does not improve tracking performance, certainly not by 3 dB.

Also, jitter degradation at low input SNR ($\rho_i \ll 1$) is not as bad as 1.05 dB. The limiter spreads the noise spectrum of the input [10.23] such that the output spectrum of the fundamental zone has relatively increased density in the tails and decreased density at the center of the spectrum. A narrowband PLL passes mainly the central portion of the spectrum, so noise degradation is less than 1.05 dB. The true degradation depends on the input filter shape, the post filter, and the PD configuration [10.22]. Further discussion is deferred to Section 10.4.4.

10.4.2 Phase-Detector Noise Threshold

Noise has many adverse effects on the operation of phase detectors. One arises from the unavoidable DC offsets in the loop, particularly in the PD itself. Offsets arise from uncompensated biases, unbalanced circuits, rectified noise, incidental frequency discrimination, and a host of even more esoteric sources. Offset is usually dependent on temperature, signal frequency, SNR, and time.

If a limiter is used, the signal amplitude at the PD is suppressed for low SNR at the input. If limiting is not used, the signal amplitude must be small so that signal plus worst-case noise does not overload the PD. In either case, the relative (to the signal) noise-caused offset increases with worsening SNR, partly because of noise effects on offset and partly because of the necessity of restricting the signal to a low amplitude. If the useful output of the PD is so small that it cannot overcome the offset, tracking fails and the loop loses lock. This occurrence is dubbed *phase-detector threshold* and is caused by unavoidable defects in the circuits rather than any inherent property of a PLL. Nonetheless, any real phase-detector circuit has such defects and they must be taken into account in the design.

A poorly balanced phase detector might exhibit a PD threshold for input SNR of about -20 to -15 dB or higher, whereas a well-designed circuit might tolerate -30 dB. Painstaking design efforts are needed to obtain satisfactory operation below about -25 dB. Input SNR is controlled by means of bandpass noise-rejection filters prior to the phase detector.

10.4.3 *s*-Curve Shape in Noise

Another effect is the degradation of the shape of the PD's *s*-curve in the presence of large input noise. Pouzet [10.24] has shown that any periodic *s*-curve loses its noise-free shape and tends toward sinusoidal as the input SNR becomes small. Figure 10.15 shows an example for a rectangular *s*-curve, but similar changes [10.2, Chap. 7; 10.24–10.26] occur for any of the other common shapes. The shape of an *s*-curve at arbitrary SNR can be calculated by Pouzet's analysis. Physical insight is gained from realization that the phase of signal plus noise fluctuates randomly about the mean phase, which is that of the signal alone. Useful DC output may be regarded as the fluctuating input phase averaged over the noise-free *s*-curve, weighted by the probability density of the phase fluctuations.

Represent the mean phase error by θ_e and the noise-free PD characteristic by $g(\theta_e)$. The phase fluctuation caused by the noise is designated θ_n and has probability density $p(\theta_n)$, a function of ρ_i. The resultant phase of signal plus noise is $\theta_e - \theta_n$. The average DC output of the phase detector is

$$V_d(\theta_e, \rho_i) = \int_{-\pi}^{\pi} g(\theta_e - \theta_n)\, p(\theta_n)\, d\theta_n \tag{10.8}$$

where θ_n is taken modulo 2π. Expressed in this manner, the DC output V_d is seen to be the convolution of the noise-free characteristic $g(\theta_e)$ by the input phase probability density $p(\theta_n)$. In the absence of noise, the phase density is a delta function $\delta(\theta_n)$ and the DC output reduces to $V_d(\theta_e, \infty)$. When noise is present, the convolution causes the DC output to be a smeared and diminished version. If $p(\theta_n) = p(-\theta_n)$, as is true if the input noise is Gaussian, and if

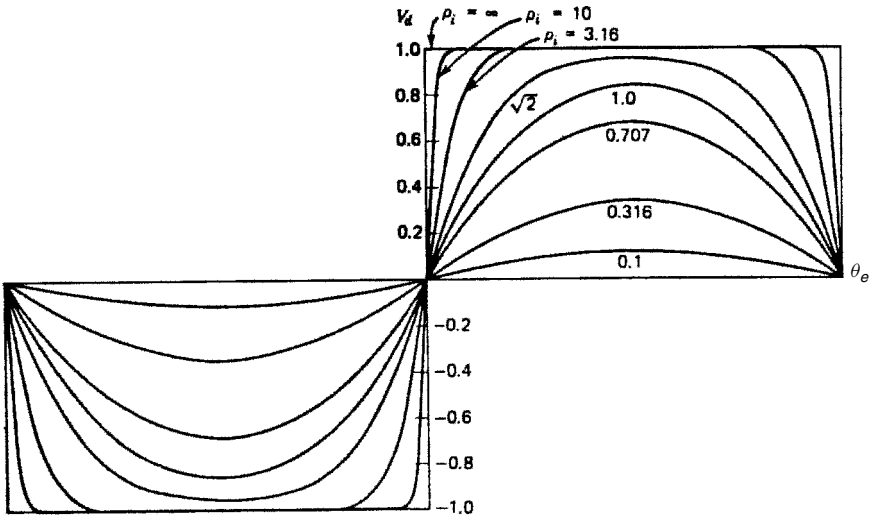


Figure 10.15 Noise degradation of a rectangular *s*-curve.

TABLE 10.1 Increase of PLL Phase Jitter (dB) Due to Limiter (for $\rho_i \ll 1$)

PD s-Curve	Single-Tuned BPF	Rectangular BPF
Sinusoidal		
No limiter	0	0
With limiter	0.25	0.65
Triangular	0.3	0.7
Rectangular	0.36	0.97
Sawtooth	2.9	2.9

$g(-\theta_e) = -g(\theta_e)$, the null at $\theta_e = 0$ will not shift with varying input SNR. If $g(\theta_e)$ is not odd-symmetric, the null can shift—a highly unsatisfactory occurrence.

Not only does noise cause the PD characteristic to degenerate toward sinusoidal, but the slope at the null is reduced; this is signal suppression and has been described in Section 10.4.2 for a sinusoidal PD preceded by a limiter. To find suppression for other s-curves, differentiate (10.8) inside the integral with respect to θ_e and evaluate the differentiated integral at $\theta_e = 0$. From Pouzet's paper one can infer that suppression in any ordinary PD will not deviate radically from that found for the sinusoidal PD.

Piecewise-linear s-curves all require that a limiter precede the phase detector. A triangular s-curve is obtained if the square-wave limiter output is unfiltered and is used to drive a switching phase detector. A rectangular s-curve results if the unfiltered limiter output is sampled. The s-curve is sinusoidal if either input to a PD is sinusoidal. This can occur if the input bandpass signal is not limited, if limiter output is filtered to remove harmonics, or if VCO drive to the PD is sinusoidal. All three alternatives yield identical shapes of the s-curve.

10.4.4 Jitter Dependence on s-Curve Shape

A limiter before the PLL causes an increase in phase jitter at small ρ_i compared to jitter without a limiter. The increase depends on the shapes of the noise-free s-curve and the input bandpass filter. Pouzet has calculated the increase for various conditions; his results are summarized in Table 10.1. The numbers shown represent the asymptotic increase of jitter at very low input SNR for two different extremes of shapes for the prelimiter bandpass filter.

Very little loss is incurred with a sinusoidal s-curve or even a triangular or rectangular s-curve, especially if the input filter is single tuned. However, there is a severe loss with a sawtooth characteristic. Since the actual characteristic degenerates to sinusoidal anyhow, it is difficult to justify the use of a sawtooth PD if the input signal is normally immersed in the noise. Similar results ought to be anticipated from any other extended PD characteristic.

10.5 TWO-PHASE (COMPLEX) PHASE DETECTORS

The need to filter out phase-detector ripple imposes a loop bandwidth that is small compared to the frequency of phase comparison. This section describes a

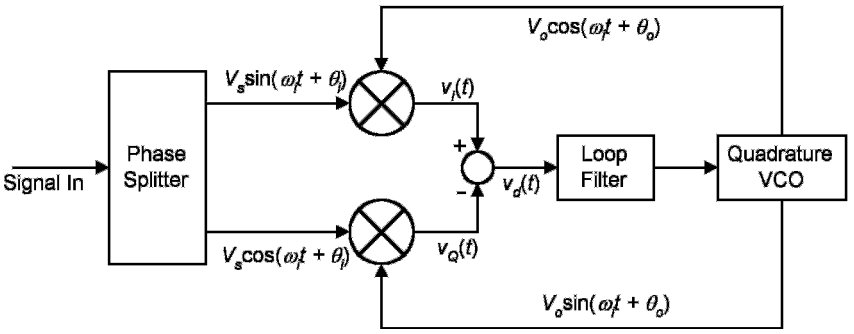


Figure 10.16 Complex (two-phase) phase detector.

method of canceling ripple instead, thereby permitting a comparatively larger loop bandwidth. Cancellation works best in digital PLLs but it has been applied successfully (although rarely) to analog PLLs as well. A technique for cancellation is shown in Fig. 10.16. The input signal is split into two quadrature components, each applied to its own phase detector. Output of the VCO also has two quadrature components, each applied to one of the two phase detectors. Assume that the phase detectors are ideal multipliers and that each pair of quadrature signals is perfectly balanced in amplitude and exact phase quadrature.

The individual phase detectors deliver $v_{\mathrm{I}}(t)$ and $v_{\mathrm{Q}}(t)$, respectively. Just as in Section 6.1.1, each individual PD output contains a DC term proportional to the sine of the phase error plus a double-frequency ripple component. But subtracting one PD output from the other doubles the DC term while canceling the ripple term, so

$$v_d(t) = K_m V_s V_o \sin(\theta_i - \theta_o) \tag{10.9}$$

where the notation (same as in Section 6.1.1) denotes K_m as the multiplier scaling factor, V_s as the amplitude of each PD-input signal, V_o as the amplitude of each VCO signal, θ_i as the input phase, and θ_o as the VCO phase.

To the extent that balancing is perfect, the ripple is suppressed completely. Essentially perfect balancing is feasible in digital PLLs. Ripple suppression of only about 30 dB may be expected in analog PLLs because of the difficulty of attaining closer balances between individual paths in a quadrature pair. Ripple cancellation is perfect (ideally) only for sinusoidal signals and ideal-multiplier PDs. Cancellation fails, for example, if the PDs are full-wave switches, as illustrated in Figs. 10.1 and 10.2*b*. For that case, the ripple waveform on v_d (combined from the two PDs) has a fundamental component of four times the comparison frequency (which relieves the ripple filtering burden somewhat), but its peak-to-peak amplitude is unchanged from that shown in Fig. 10.1.

The arrangement of Fig. 10.16 may be recognized as a single-sideband demodulator in which the lower sideband (at zero frequency) is the one selected and the upper sideband rejected. Refer to the copious single-sideband literature for methods of phase splitting; refer to Section 9.7 for references to quadrature VCOs. Variations on the basic two-phase PD have appeared in [10.27] and [10.28].

Another useful representation is in complex exponential format wherein

$$z_s(t) = V_s e^{j(\omega_i t + \theta_i)}, \qquad z_o(t) = V_o e^{j(\omega_i t + \theta_o)} \tag{10.10}$$

so that the configuration of Fig. 10.16 generates

$$v_d(t) = K_m \mathrm{Im}[z_s z_o^*] = K_m V_s V_o \sin(\theta_i - \theta_o) \tag{10.11}$$

where $\mathrm{Im}[x]$ indicates the imaginary part of x and the asterisk * indicates a complex conjugate.

APPENDIX 10A: PHASE MODULATION DUE TO PHASE-DETECTOR RIPPLE

Phase-detector ripple is a disturbance accompanying normal operation of a phase detector. Ripple is processed by the loop filter and applied to the VCO as part of the control voltage. Ripple in the control voltage generates phase modulation on the VCO output. Those modulation effects are undesirable and have to be minimized. This appendix describes the character of the ripple produced by several common phase detectors and shows how to calculate ripple amplitude. Additional information on ripple is contained in Chapters 11 and 12.

10A.1 Ripple Model

Designate the comparison frequency in the phase detector as f_c. All examples in this appendix assume that the ripple is periodic in $1/f_c$. Phase modulation due to the periodic ripple therefore is also periodic in $1/f_c$. If the unmodulated oscillator frequency f_o is equal to f_c (or is only a low harmonic of f_c), the ripple causes f_c-synchronous distortion within the individual cycles of the VCO waveform; such distortion generates integer harmonics of f_o. Study of this problem involves nonlinear differential equations of the PLL [10.29, 10.30], a study not undertaken here. Instead, assume that $f_o \gg f_c$, either because a frequency divider is in the feedback path between the VCO and the PD (Chapter 15) or because frequency conversions take place in the feedback path (as in phaselock receivers). The phase modulation caused by the ripple then generates sidebands at frequencies $f_o \pm nf_c$, where n takes on positive integer values. Amplitude of a sideband depends on the amplitude and waveshape of the ripple.

Since the ripple is periodic, it can be expanded in a Fourier series and the modulation index of each term in the series determined by linear calculations. Sideband amplitudes and phases can be calculated for each term individually through the well-known Bessel function expansion for sinusoidal phase modulation. Overall amplitudes and phases are the vector sums of the contributions of the sidebands generated by each of the harmonic terms of the Fourier series. In many instances, the lowest-frequency term of the series will dominate the modulation.

This appendix derives the terms of the Fourier series: the peak amplitude $\Delta\theta_n$ of the nth phase-deviation term. Each such term generates a single line, a Dirac delta function with area $\Delta\theta_n^2/2$ rad^2, in the phase-noise spectrum $W_\phi(f)$ (see Section 7.2 for definition of the spectrum) at frequency nf_c. The results presented here deal with ripple in a second-order type 2 PLL (Section 2.2) with no additional filtering at high frequencies. That is unrealistic; adequate suppression of ripple almost always demands additional filtering. These results (i.e., the amplitudes of phase deviations) together with requirements imposed on the allowable strengths of discrete components disclose the additional ripple attenuation needed, through additional filtering or reduction of loop bandwidth. Linear operation of the PLL is assumed throughout. Interpretation of $\Delta\theta_n$ gives warning if the ripple is so large as to drive the PLL into nonlinear behavior but does not tell the consequences of such overload. Nonlinear overload is addressed further in Chapter 11.

10A.2 Basis of Analysis

Denote the ripple component of PD output as $v_{dr}(t)$, which is added to the DC output v_d of the PD. For purposes of analysis, assume that the PLL tracks without any static phase error. The ripple waveforms considered below are based on zero static phase error at the PD. Be aware, though, that most of the waveforms will be altered if the PLL is working at a nonzero phase error and that the Fourier components will be changed from those presented here.

Ripple passes through the loop filter into the control voltage for the VCO. Denote the ripple component of control voltage as $v_{cr}(t)$. Assume that ripple frequency is high enough that ripple transmission through the proportional path of the loop filter greatly exceeds transmission through the integral path so that the latter can be neglected. Therefore, ripple contribution to the control voltage is approximated by $v_{cr}(t) \approx v_{dr}(t)\tau_2/\tau_1$, where τ_2 and τ_1 are as defined in Chapter 2. Frequency modulation imposed on the VCO by the ripple is $\omega_{or}(t) = K_o v_{cr}(t)$ rad/sec, and the corresponding ripple phase modulation is

$$\theta_{or}(t) = \frac{K_o \tau_2}{\tau_1} \int v_{dr}(t)\, dt \tag{10A.1}$$

This equation is applied to each term of the Fourier series expansion of a ripple waveform $v_{dr}(t)$ to find the phase deviation caused by each term.

10A.3 Ripple Examples

Figure 10A.1 shows several ripple waveforms generated by familiar phase detectors. They are periodic in $1/f_c$; the figure is drawn for one period of duration $1/f_c$. The waveforms, as drawn, all have skew symmetry so their Fourier expansions contain only sine terms of the form $V_r a_n \sin(2\pi n f_c t)$, where V_r is the peak amplitude of the ripple and a_n is the nth coefficient of the Fourier series expansion of unit-amplitude ripple.

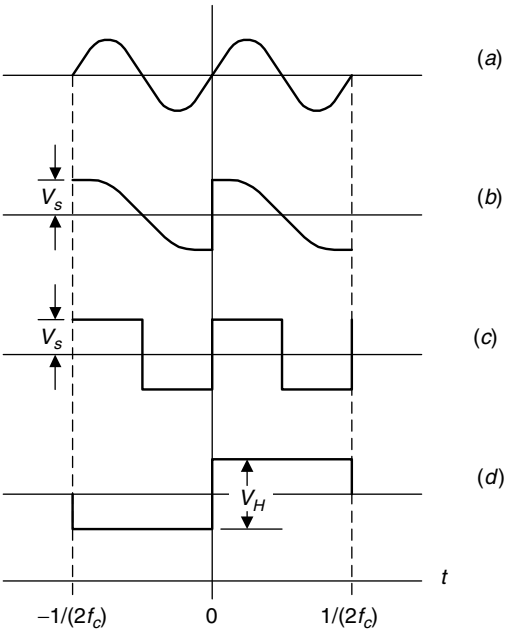


Figure 10A.1 Ripple waveforms in example phase detectors.

In each example phase detector, the PD gain K_d is proportional to V_r. Therefore, ripple amplitude from any particular PD can be described as $K_d = cV_r$, where c is a constant pertaining to the specific PD. From (2.18), the loop gain is defined as $K = K_d K_o \tau_2/\tau_1$ rad/sec. Combining the foregoing expressions, the peak phase-modulation amplitude for the nth term is found to be

$$\Delta\theta_n = \frac{a_n K}{2\pi cnf_c} \qquad \text{rad} \tag{10A.2}$$

The various ripple-waveform examples differ only in c and a_n in their consequences for VCO phase modulation. Table 10A.1 summarizes the ripple characteristics for each of the examples of Fig. 10A.1.

10A.4 Ripple Filters

Additional filtering for suppression of ripple is essential in nearly all PLLs. The simplest filter is a single-pole lowpass network with corner frequency at $f = f_p$. If $nf_c \gg f_p$, the attenuation of the ripple component at $f = nf_c$ is nearly 20 $\log(f_p/nf_c)$ dB. Even more lowpass filters are often cascaded within a loop to attain greater attenuation than that of a single pole. Insertion of extra poles involves a trade-off between ripple suppression on one hand and loop stability and phase margin on the other. See Chapters 2 and 3 for discussions of stability and phase margin.

TABLE 10A.1 Ripple Properties of Example Phase Detectors

PD Name	Waveform	s-Curve	Loop Gain	Equations[a]	Voltage[b]	Deviation, $\Delta\theta_n$	Valid n[c]
Multiplier	Fig. 10A.1*a*	Sine	$K_d = V_r$	(6.4), (6.5)	—	$\dfrac{K}{2\pi n f_c}$	$n = 2$
Full-wave switcher							
Sine input	Fig. 10A.1*b*	Sine	$K_d = 2V_r/\pi$	(10.3)	$V_r = V_s$	$\dfrac{K}{\pi(n^2-1)f_c}$	n = even > 0
Square input	Fig. 10A.1*c*	Triangle	$K_d = 2V_r/\pi$	—	$V_r = V_s$	$\dfrac{2K}{\pi n^2 f_c}$	$n/2$ = odd > 0
RS flip-flop	Fig. 10A.1*d*	Sawtooth	$K_d = V_r/4\pi$	(10.5)	$V_r = V_H/2$	$\dfrac{8K}{\pi n^2 f_c}$	n = odd > 0

[a] Defining equation for K_d.
[b] Relation between the peak signal voltage and peak ripple voltage V_r in the K_d definition.
[c] Applicable values of n for $\Delta\theta_n$ in the preceding column; otherwise, $\Delta\theta_n = 0$ for all other n.

If one spectral component of ripple is dominant (usually, the lowest-frequency component of the Fourier series) and if comparison frequency f_c is confined to a narrow range, a notch network such as a twin-T [10.31] is effective in ripple suppression. Active filters with transmission notches would also be effective, but the literature does not provide examples of their use in PLLs.

REFERENCES

10.1 B. Gilbert, "A Precise Four-Quadrant Multiplier with Subnanosecond Response," *IEEE J. Solid-State Circuits* ***SC-3***, 365–373, Dec. 1968.

10.2 A. Blanchard, *Phase-Locked Loops*, Wiley, New York, 1976.

10.3 W. J. Gruen, "Theory of AFC Synchronization," *Proc. IRE* ***41***, 1043–1048, Aug. 1953.

10.4 C.-S. Yen, "Phase-Locked Sampling Instruments," *IEEE Trans. Instrum. Meas.* ***IM-14***, 64–68, Mar.–June 1965.

10.5 B. R. Eisenberg, "Gated Phase-Locked Loop Study," *IEEE Trans. Aerosp. Electron. Syst.* ***AES-7***, 469–477, May 1971.

10.6 S. Barab and A. L. McBride, "Uniform Sampling Analysis of a Hybrid Phase-Locked Loop with a Sample-and-Hold Phase Detector," *IEEE Trans. Aerosp. Electron. Syst.* ***AES-11***, 210–216, Mar. 1975.

10.7 L. M. Robinson, "TANLOCK: A Phase-Lock Loop of Extended Tracking Capability," *Proc. IRE Conv. Mil. Electron.*, Los Angeles, Feb. 1962, pp. 396–421.

10.8 M. Balodis, "Laboratory Comparision of TANLOCK and Phaselock Receivers," Paper 5-4, *Conf. Rec. Natl. Telem. Conf.*, 1964.

10.9 A. Acampora and A. Newton, "Use of Phase Subtraction to Extend the Range of a Phase-Locked Demodulator," *RCA Rev.* ***27***, 577–599, Dec. 1966.

10.10 J. Klapper and J. T. Frankle, *Phase-Locked and Frequency Feedback Systems*, Academic Press, New York; 1972, Chap. 8.

10.11 C. J. Byrne, "Properties and Design of the Phase-Controlled Oscillator with a Sawtooth Comparator," *Bell Syst. Tech. J.* ***41***, 559–602, Mar. 1962.

10.12 J. I. Brown, "A Digital Phase and Frequency-Sensitive Detector," *Proc. IEEE* ***59***, 717–718, Apr. 1971. Reprinted in [10.13].

10.13 B. Razavi, ed., *Monolithic Phase-Locked Loops and Clock Recovery Circuits*, Reprint Volume, IEEE Press, New York, 1996.

10.14 D. K. Morgan and G. Steudel, *The RCA COS/MOS Phase-Locked-Loop*, Appl. Note ICAN-6101, RCA, Somerville, NJ, Oct. 1972.

10.15 *Phase-Locked Loop Data Book*, 2nd ed., Motorola, Schaumburg, IL, Aug. 1973.

10.16 C. A. Sharpe, "A 3-State Phase Detector Can Improve Your Next PLL Design," *EDN*, Sept. 20, 1976. Reprinted in [10.13].

10.17 J. Tal and R. K. Whitaker, "Eliminating False Lock in Phase-Locked Loops," *IEEE Trans. Aerosp. Electron. Syst.* ***AES-15***, 275–281, Mar. 1979.

10.18 M. Soyuer and R. G. Meyer, "Frequency Limitations of a Conventional Phase-Frequency Detector," *IEEE J. Solid-State Circuits* ***25***, 1019–1022, Aug. 1990.

10.19 A. M. Fahim and M. I. Elmasry, "A Fast Lock Digital Phase-Locked Loop Architecture for Wireless Applications," *IEEE Trans. Circuits Syst. II* ***50***, 63–72, Feb. 2003, Fig. 8.

10.20 S. O. Rice, "Mathematical Analysis of Random Noise," *Bell Syst. Tech. J.* ***23***, 282–332, 1944; ***24***, 46–156, 1945.

10.21 W. B. Davenport, Jr., "Signal-to-Noise Ratios in Band-Pass Limiters," *J. Appl. Phys.* ***24***, 720–727, June 1953.

10.22 J. C. Springett and M. K. Simon, "An Analysis of the Phase Coherent–Incoherent Output of the Bandpass Limiter," *IEEE Trans. Commun.* ***COM-19***, 42–49, Feb. 1971.

10.23 J. H. van Vleck and D. Middleton, "The Spectrum of Clipped Noise," *Proc. IEEE* ***54***, 2–19, Jan. 1966.

10.24 A. H. Pouzet, "Characteristics of Phase Detectors in Presence of Noise," *Proc. Int. Telem. Conf.* ***8***, Los Angeles, 1972, pp. 818–828.

10.25 B. N. Biswas, S. K. Ray, A. K. Bhattacharya, B. C. Sarkar, and P. Banerjee, "Phase Detector Response to Noise and Noisy Fading Signals," *IEEE Trans. Aerosp. Electron. Syst.* ***AES-16***, 150–158, Mar. 1980.

10.26 E. H. Sheftelman, "The Transfer Function Characteristic of a Linear Phase Detector When Its Input Signal–Noise Ratio Is Small," *Proc. IEEE* ***55***, 694, May 1967.

10.27 G. L. Baldwin and W. G. Howard, "A Wideband Phaselocked Loop Using Harmonic Cancellation," *Proc. IEEE* ***57***, 1464, Aug. 1969.

10.28 R. E. Scott and C. A. Halijak, "The SCEM-Phase-Lock Loop and Ideal FM Discrimination," *IEEE Trans. Commun.* ***COM-25***, 390–392, Mar. 1977.

10.29 J. L. Stensby, "On the PLL Spectral Purity Problem," *IEEE Trans. Circuits Syst.* ***CAS-30***, 248–251, April 1983.

10.30 J. L. Stensby, *Phase-Locked Loops*, CRC Press, New York, 1997.

10.31 V. F. Kroupa, *Phase Lock Loops and Frequency Synthesis*, Wiley, Chichester, West Sussex, England, 2003, Sec. 3.1.4.

CHAPTER 11

LOOP FILTERS

Two classes of loop filters are common: those used directly with phase detectors and those used with charge pumps. This chapter deals with the loop filters used directly with the PDs; those used with charge pumps are considered in Chapter 12. A loop filter is a comparatively simple circuit whose linear analysis is well covered in Chapters 2 to 4. This chapter sets forth miscellaneous features of active loop filters, features that are not covered in the circuit analysis but which have considerable practical importance. The benefits of type 2 PLLs have been laid out in Chapter 5 and implementations of suitable loop filters described in Chapters 2 to 4. The discussion to follow assumes that type 2 is the preferred PLL design and that any appreciable shortfall from type 2 is unfortunate.

11.1 ACTIVE VS. PASSIVE LOOP FILTERS

The DC amplifiers of the early days of PLLs had large, drift-prone offset voltages and were generally unreliable. Phaselock loops were built with passive loop filters to avoid DC amplifiers. Early literature on PLLs concentrated on passive filters, even to the extent of implying that a passive filter is the natural method of building a PLL. Only a type 1 PLL can be realized with a passive filter, but performance of a type 1 PLL is impaired by static phase error (Chapter 5). Also, monstrously large capacitors tend to be needed in passive loop filters for narrow-bandwidth PLLs. [**Comment**: Charge pumps often work with passive filters, but their operation is different from that of PLLs without charge pumps. Type 2

Phaselock Techniques, Third Edition, by Floyd M. Gardner

operation can be achieved from a passive loop filter driven by an ideal charge pump; see Chapter 12.]

With the advent of well-behaved low-cost operational amplifiers (*op-amps*), the early reasons for passive loop filters have evaporated. The rest of this chapter deals primarily with issues raised by op-amps in active loop filters.

11.2 DC OFFSET

Any input-referred DC offset in the op-amp of a loop filter will be canceled through PLL feedback by an opposite offset in the output of the phase detector. Such offset in the PD is generated by a steady phase error, usually an undesirable feature. Care is needed to minimize DC offsets in the op-amp and any other DC circuits, including the PD. The literature on operational amplifiers is replete with techniques for reduction of offsets. When an input signal is absent and the loop is out of lock, the only input to the loop filter will be DC offset and random noise. With very high gain in the op-amp (as is usual in op-amps without DC feedback), DC offset accumulates in the integrator function of the loop filter. Eventually, the integrated output of the op-amp is driven to a saturation level. A saturated op-amp might prevent acquisition of phase lock when a signal does appear.

Acquisition problems due to DC offset are eliminated if an acquisition aid such as frequency sweep or frequency detection is incorporated into the PLL. A sweep pattern that avoids the saturation neighborhoods will acquire lock shortly after a signal appears. A frequency detector does not prevent saturation, but properly designed, it will overcome any small offset voltages when the signal appears and drive the integrator out of saturation toward the correct lock frequency.

Acquisition aids might not be permissible under some unusual circumstances: for example, if frequency search is implemented in the transmitter rather than the receiver of a radio link. Saturation has to be avoided in those circumstances, too. Even more restrictively, the resting frequency of the VCO (in the absence of input signal) often has to be held within restrictive limits to assure that a signal can be acquired promptly once it appears. One approach is to lock the PLL to a local reference signal when the information signal is absent and to switch over to the information signal when its presence is detected. For example, the timing-recovery PLL in a disk drive might be locked to the Write clock when the PLL is idling and switched to the recovered data stream when data are to be read from the disk. Two different kinds of phase detector may be used for the two different kinds of input signals applied to the PLL; the PLL switches between the PDs upon command.

Another approach is to spoil the integrator action by applying local DC feedback around the op-amp of the loop filter (a resistor between output and summing-junction terminals). Enough feedback is used to ensure that the maximum amplified offset does not drive the op-amp to saturation or, more stringently, ensures that the maximum amplified offset is held within narrower limits to restrict the range of resting frequency of the VCO. This technique reduces the PLL to type 1 operation, with its inevitable static phase error. Nonetheless,

careful engineering can yield substantial improvement over performance with passive filters.

Circuit diagrams of active filters in PLLs often show the inclusion of a DC-feedback resistor at the op-amp. A folk legend holds that such feedback is somehow necessary, but there is no basis to the legend other than described in the paragraph above. The DC feedback is not needed if suitable acquisition methods are employed to counteract saturation.

11.3 TRANSIENT OVERLOAD

Linear theory of operation is inapplicable in the presence of overload. All components within a PLL have to be protected against overload. Component behavior may be difficult to predict in overload, but operation of the PLL is usually impaired if a component is overloaded. Op-amps, because of their large gains, are particularly susceptible to overload. This section points up two potential sources of overload to guard against.

11.3.1 Overload from PD Ripple

Chapter 10 gave examples of large fast excursions of ripple voltage at the outputs of phase detectors. Ripple filtering is essential to minimize ripple sidebands in the VCO output, but also to prevent overload in the loop filter and—possibly—in the VCO, too. This section concentrates on overload in the op-amp, partly because it is typically the more susceptible element but also because effective protection of the op-amp usually protects the VCO, too.

There are two aspects to op-amp protection; one is that the amplified ripple should not be so large as to exceed the linear output-voltage range of the op-amp. That is also a problem during acquisition and is deferred to the next section. A more subtle aspect is the high-frequency nature of ripple. In particular, consider a PD where the ripple waveform has large, fast transitions. Op-amps usually cannot tolerate step transitions at their inputs; they go into slew limiting, with unpredictable effects on PLL operation. Step transitions or other high-frequency artifacts have to be kept out of the op-amp. Sufficient ripple filtering has to precede the op-amp to ensure linear operation; ripple postfiltering does not protect the op-amp from transient overload.

An extra pole, at a frequency appropriate for ripple suppression, is often provided by connecting a capacitor from the op-amp output back to the summing junction. That avoids a step-voltage output from the op-amp (which the op-amp cannot supply) but still does not protect the op-amp from overload on step inputs because the op-amp cannot supply the step-current output required by the feedback capacitor either. Do not subject an op-amp to step inputs or other large high-frequency inputs.

11.3.2 Overload During Acquisition

During acquisition, a phase detector delivers a beat note at the difference frequency between the incoming signal and the VCO. The beat-note waveform is a

replica of the PD s-curve with a peak amplitude equal to the peak of the s-curve. Unless close attention is paid to this phenomenon, it is easy to design a PLL in which the beat note overloads the loop-filter op-amp. A lowpass filter between the PD and active loop filter may be of some help in combating this problem. However, the potential overload is worst when the difference frequency is near the lock-in frequency ($\approx K$ rad/sec, per Chapter 8), but significant filtering at such a low frequency leads to instability of the locked loop (see Chapters 2 and 3).

Another approach is to keep the product of peak PD voltage and loop-filter gain within the linear bounds of the op-amp output range. That is a desirable goal but might be difficult to arrange in the presence of any or all of the following: need for a large loop gain K (therefore, a large phase-detector gain or large gain through the loop filter); need for a wide tuning range of the VCO or low VCO gain K_o (necessitating a potentially wide range of control voltage).

Yet another approach is to accept the inevitability of overload during acquisition but provide for predictable overload behavior with fast recovery. This can be accomplished either by careful choice of an op-amp or by external limiter circuitry. Overload behavior of op-amps is not often well described in data sheets; experimental determination may be necessary. Watch out especially for such things as gain-polarity reversal on overload or recovery from overload that is far slower than would be expected from the bandwidth of the linear amplifier.

Op-amps can go into slew limiting on a beat note. The best countermeasure is an op-amp fast enough that it will not slew limit on any beat note of expected frequency and amplitude. If slew limiting cannot be avoided, the speed of slewing should be the same in both directions so as to avoid rectification that introduces a DC component (harmful to acquisition) to the distorted output. The existence of millions of PLLs that work perfectly well demonstrates that overload pitfalls can be overcome. The purpose of these warnings is to alert the reader to take care of the problems early in a design cycle and not be surprised by them later.

CHAPTER 12

CHARGE-PUMP PHASELOCK LOOPS*

A charge pump in a PLL is an electronic switch that dispenses charge into the loop filter under control of the phase detector. Charge pumps can be used advantageously with any phase detector that delivers bilevel outputs in which the phase-error information is contained in the duty ratios of the output waveforms. Despite their wider applicability, charge pumps have been associated primarily with the phase/frequency detectors (PFDs) of Section 10.3. Accordingly, this chapter assumes a PFD phase detector. Nonetheless, various charge-pump properties brought out here can be applied to charge pumps working with other kinds of phase detectors.

Many early PFDs operated without charge pumps. The two output terminals of the PFD were regarded, inaccurately, as a differential pair. Waveforms of the pair were converted into a single-ended drive for the loop filter. These PLLs could lock and track, but they did not have the advantages provided by charge pumps, advantages laid out in this chapter.

12.1 MODEL OF A CHARGE PUMP

A typical charge pump comprises two current switches, labeled UP and DN, controlled from the UP and DN terminals of a PFD. The UP switch delivers a pump current I_p into the loop filter when the UP terminal of the PFD is active

*Portions of this chapter have been reprinted, with permission, from [12.1]: F. M. Gardner, "Charge-Pump Phase-Lock Loops," *IEEE Trans. Commun.* ***COM-28***, 1849–1858, Nov. 1980; © 1980 IEEE.

Phaselock Techniques, Third Edition, by Floyd M. Gardner

and the DN switch extracts a pump current I_p from the loop filter when the DN terminal of the PFD is active. A current switch is ideally an open circuit while it is shut OFF. The third (null) state of the PFD, during which both switches are shut OFF, imparts a vital property to charge pumps, a property that does not exist in conventional PLLs. [**Comment**: Voltage switches, rather than current switches, also have been used. Current switches are preferable, as discussed in Section 12.5. Otherwise, only current switches are considered here.]

Assume that the PLL is locked and denote the comparison frequency of the PD as ω_c rad/sec. Let the phase error be $\theta_i - \theta_o = \theta_e$ radians. The ON time of either UP or DN, as applicable, is

$$t_p = \frac{|\theta_e|}{\omega_c} \tag{12.1}$$

for each period $2\pi/\omega_c$, of the input signal. (The subscript p connotes "pump.") These two features—the three-state description and (12.1)—completely characterize the PFD for the purposes of this section.

The loop filter can be either passive or active. A passive loop filter is represented by its two-terminal impedance $Z_F(s)$; an active loop filter is characterized by its transfer impedance (with current in and voltage out). Most attention here is given to passive filters, partly because analysis is simplified thereby but also because the configuration is eminently practical and widely employed.

Because of the switching, the charge-pump PLL is a time-varying network; an exact analysis must take account of the time variations of the circuit topology, and that is a more involved effort than is needed for time-invariant networks. In particular, simple transfer-function analysis is not immediately applicable to time-varying networks. Nevertheless, in many applications the state of the PLL changes by only a very small amount on each cycle of the input signal; loop bandwidth is small compared to the signal frequency. In these cases the detailed behavior within a single cycle is of less concern than the average behavior over many cycles. By applying an averaged analysis, the time-varying operation can be bypassed and the powerful tool of time-invariant transfer functions retained. The remainder of this section is devoted to the derivation of average-operation transfer functions. Be aware, though, that the per-cycle behavior can be important even for quite narrow bandwidths, as will be shown later.

A pump current $I_p \operatorname{sgn}(\theta_e)$ is delivered to the filter impedance Z_F for a time t_p on each cycle. Each cycle has a duration $2\pi/\omega_c$ seconds, so, utilizing (12.1), the error current averaged over one cycle is

$$i_d = \frac{I_p \theta_e}{2\pi} \qquad \text{amperes} \tag{12.2}$$

Equation (12.2) is also the error current averaged over many cycles, provided that both inputs are periodic—that no input cycles are missing. Therefore, phase-detector gain is

$$K_d = \frac{I_p}{2\pi} \qquad \text{A/rad} \tag{12.3}$$

Oscillator control voltage is given by

$$V_c(s) = I_d(s)Z_F(s) = \frac{I_p Z_F(s)\theta_e(s)}{2\pi} \tag{12.4}$$

where $I_d(s)$ is the Laplace transform of $i_d(t)$, and similarly for the other symbols. The phase of the VCO is given by

$$\theta_o(s) = \frac{K_o V_c(s)}{s} \qquad \text{rad} \tag{12.5}$$

where K_o is the VCO gain in rad/sec·V. These expressions lead to the loop transfer functions

$$\begin{aligned} G(s) &= \frac{\theta_o(s)}{\theta_e(s)} = \frac{K_o I_p Z_F(s)}{2\pi s} \\ H(s) &= \frac{\theta_o(s)}{\theta_i(s)} = \frac{K_o I_p Z_F(s)}{2\pi s + K_o I_p Z_F(s)} \\ E(s) &= \frac{\theta_e(s)}{\theta_i(s)} = 1 - H(s) = \frac{2\pi s}{2\pi s + K_o I_p Z_F(s)} \end{aligned} \tag{12.6}$$

12.2 LOOP FILTER

A typical loop filter for use with a charge pump is illustrated in Fig. 12.1; component labels have been assigned to be consistent with earlier notation in Chapters 2 and 3. Defining $b = 1 + C/C_3$ and $\tau_2 = R_2 C$ gives an expression for the filter impedance

$$Z_F(s) = \frac{b-1}{b} \frac{s\tau_2 + 1}{sC\left(\frac{s\tau_2}{b} + 1\right)} \qquad \text{ohms} \tag{12.7}$$

Next, define loop gain as

$$K = \frac{b-1}{b} \frac{K_o I_p R_2}{2\pi} \qquad \text{rad/sec} \tag{12.8}$$

Substituting (12.7) and (12.8) into (12.6) leads to the transfer functions (2.38) to (2.41) previously obtained for a third-order type 2 PLL. The averaged dynamics of a charge-pump PLL, as evidenced by its transfer functions, are the same as for a conventional PLL as laid out in Chapters 2 and 3. [**Comment**: The factor

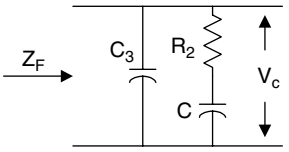


Figure 12.1 Passive loop filter for a charge-pump PLL. (From [12.1]; © 1980 IEEE.)

$(b-1)/b$ appearing in (12.8) was erroneously omitted from the definition of K in [12.1]. That error is corrected throughout the paragraphs that follow.]

Capacitor C_3 is sometimes omitted from the loop filter, leading to a second-order type 2 PLL whose dynamic properties are delineated in Section 2.2. Omission of C_3 is usually unwise; that capacitor provides ripple filtering essential in nearly all practical charge-pump PLLs. In the absence of C_3, the ripple voltage $I_p R_2$ generated across the filter impedance is likely to overload the VCO and the active current switch itself. In practice, many charge-pump PLLs incorporate even one more lowpass pole for enhanced ripple filtering, producing a fourth-order type 2 loop. Ripple in a charge-pump PLL is examined further in Section 12.6.

12.3 STATIC PHASE ERROR

Static phase error (see Section 5.1.1) of the charge-pump PLL is

$$\theta_v = \frac{2\pi\,\Delta\omega}{K_o I_p Z_F(0)} \qquad \text{rad} \tag{12.9}$$

Note from (12.7) that $Z_F(0) = \infty$ in the loop filter of Fig. 12.1, so that the static phase error from (12.9) is zero. This desirable feature is achieved with a passive filter. Achievement of zero static phase error in a conventional PLL requires an active filter with infinite DC gain. Therefore, the charge pump permits zero static phase error (type 2 response) without the need for DC amplification, a valuable property of charge pumps. This effect arises because of the open-circuit switches during the null state of the PFD and does not depend on use of a current switch; the same behavior is found with voltage switches.

Practical circuits may impose shunt loading across the passive filter impedance. Denote the load as a resistor R_s. The resulting static phase error, from (12.9), will be

$$\theta_v = \frac{2\pi\,\Delta\omega}{K_o I_p R_s} = \frac{\Delta\omega R_2}{K R_s}\frac{b-1}{b} \qquad \text{rad} \tag{12.10}$$

Shunt loading is most likely to come from input impedance of the VCO control terminal or from the charge switch itself. Both impedances can be made extremely large. The VCO may be varactor-tuned, which implies nearly-infinite resistance, and the switch is typically a reverse-biased bipolar transistor or an MOS device. Other varieties of VCO could utilize a high-impedance buffer if necessary to isolate a low-impedance input.

Leakage current may be more significant in producing phase error if R_s is very large. The phase error θ_v resulting from a bias current I_b injected continuously into the filter node can be calculated as

$$\theta_b = \frac{I_b}{K_d} = \frac{2\pi\,I_b}{I_p} \qquad \text{rad} \tag{12.11}$$

12.4 STABILITY ISSUES

Stability and stability margins for conventional third-order type 2 PLLs have been addressed at length in Chapters 2 and 3. The same criteria apply to charge-pump PLLs. However, all of the earlier chapters are based on averaged-response, time-continuous, constant-element operation of the loop. Additional critical issues arising from the actual discontinuous operation also need attention. Design efforts based on time-averaged operation should be deferred until an adequate stability margin has been assured for discontinuous operation.

In some sense, the charge-pump PLL operates on a sampled basis and not as a straightforward time-continuous circuit. An approximate analysis of time-discrete stability was outlined in [12.1]. Although the charge-pump switching makes the PLL time varying, the operation inside each switching interval can be represented as time invariant and well approximated as linear. State variables of the PLL—voltages on the capacitors and phase error between input signal and VCO, along with time instants of switching—can be calculated by standard linear circuit analysis during each switching interval. Final conditions in one switching interval serve as initial conditions of the next. Gardner [12.1] shows the details for a second-order type 2 charge-pump PLL.

Conditions for a third-order PLL at the instants of turn-on of the charge pump are described in algebraic and transcendental difference equations. These can be combined into a single difference equation for phase error at those instants. On the assumption of small phase error, the transcendental terms are approximated algebraically and the algebraic difference equation is z-transformed into a rational transfer function in z. The z-plane poles of the PLL are the three roots of the denominator of the rational transfer function.

It turns out that the instability boundary for increasing K corresponds to the real pole intersecting the unit circle at $z = -1$ for all positive values of ω_c, τ_2, and b. That boundary value of the normalized gain is

$$K\tau_2 = \frac{(\omega_c\tau_2)^2}{\pi^2\left(1 + \dfrac{\omega_c\tau_2}{\pi}\dfrac{1-a}{1+a}\dfrac{b-1}{b}\right)} \tag{12.12}$$

where $a = \exp(-2\pi b/\omega_c\tau_2)$.

As an approximation of the stability limit, (12.12) is charted in Fig. 12.2 for several different values of b. A PLL is stable for normalized gain values $K\tau_2$ below the curve for the appropriate b and unstable for gain values above the curve. At first glance it would seem as though small values of b permit higher values of loop gain. That impression is correct, but you will find that regions of smaller b and higher stable gain mostly are regions where the damping of the complex pole pair is too small (see Figs. 3.3 and 3.4). These are regions to be avoided. The stability boundary is crucial but is not the only criterion of good performance.

Good design requires a margin against instability, a choice of gain that is smaller than the gain at the instability boundary. The dashed curves in Fig. 12.2

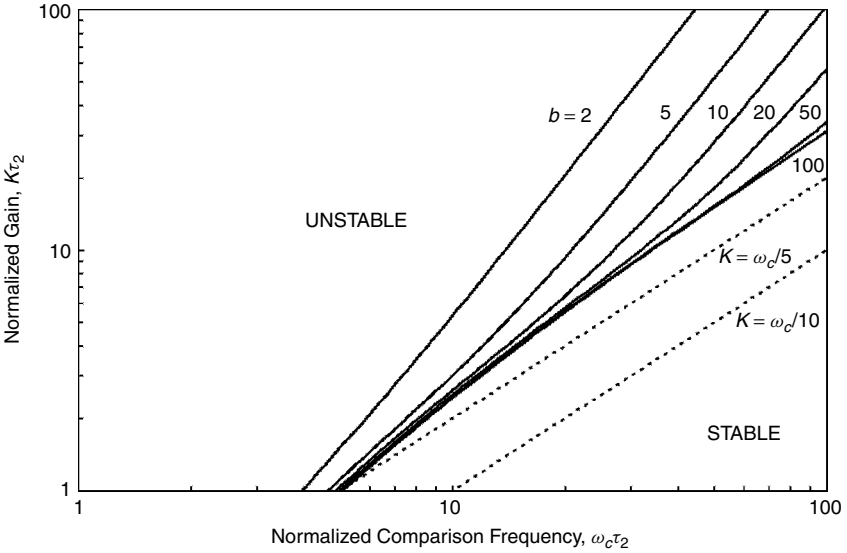


Figure 12.2 Stability boundaries for a third-order charge-pump PLL.

show the locations in the stable region for two choices of gain relative to comparison frequency. A choice of $K = \omega_c/10$ offers a gain margin of about 10 dB, whereas $K = \omega_c/5$ offers 6 dB less margin. Experience indicates that time-continuous analysis via transfer functions (12.6), and the tools of Chapters 2 and 3 furnish good approximations to the behavior of the charge-pump PLL, provided that the stability margin is adequate (e.g., $K \leq \omega_c/10$).

This section has examined instability due to the sampled character of a charge-pump PLL. Many other analog PLLs also have sampling features and would have related instability issues except that their loop gains (loop bandwidths) typically are selected to be far smaller than the comparison frequency. Charge-pump PLLs might bump up against time-discrete instability limits because they often are designed with large loop gains to better track out large phase noise in the VCO. This topic is pursued further in Chapter 15.

Instability due to sampling is the only issue that has been treated in this section. Once sampling stability is assured, the time-continuous instabilities, stability margins, and damping issues described in Chapters 2 and 3 still have to be considered, using Bode plots, root-locus plots, or Nichols plots as may be convenient.

12.5 NONLINEARITIES

A dead zone is one example of nonlinearity in a PFD. Section 10.3.2 tells how delay in the reset feedback of the PFD helps eliminate the dead zone. Although delay is helpful and necessary, it is not sufficient; some residual crossover distortion remains, as described in the paragraphs below and in Chapter 15.

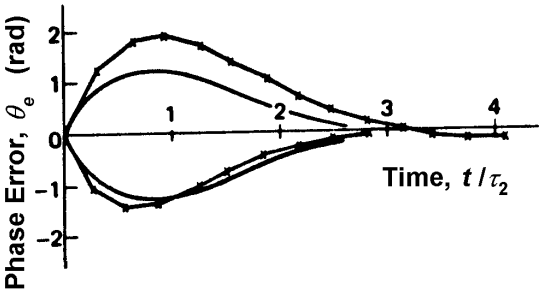


Figure 12.3 Response of a second-order PLL to frequency steps $\Delta\omega = \pm 2K$. Smooth curves, response of time-continuous PLL; marked points, time-discrete response of a charge-pump PLL with $K\tau_2 = 2$ and $K = \omega_c/10$. (From [12.1]; © 1980 IEEE.)

A more subtle nonlinearity exists even if the PFD and charge pump are ideal and no dead zone exists. Influence of this nonlinearity is best introduced by an example illustrated in Fig. 12.3, which shows the simulated transient phase-error response of a PFD/charge-pump PLL to frequency steps of $\Delta\omega = \pm 2K$ rad/sec. Loop parameters for the example are $K\tau_2 = 2$, $K = \omega_c/10$, and $b = \infty$ (second-order PLL with $\zeta = 0.707$). Bandwidth K is rather wide relative to comparison frequency, probably as wide as is prudent. Smooth curves in the figure show the error response for a conventional phase detector whose operation is assumed to be time-continuous. Observe that responses to positive and negative excursions are mirror images of one another, as would be expected from a linear system. Curves with marked data points are simulation results for the PFD–charge pump PLL. Each data point represents the phase error at the instant of turn-on of a new PFD cycle. Time-sequential data points are connected by straight-line segments for clarity of display, but the phase-error trajectories between the data points do not necessarily follow straight lines.

The salient feature of Fig. 12.3 is that the error responses of the PFD–charge pump PLL are not mirror images for positive and negative excitations. The negative response is very close to the theoretical response for the time-continuous PLL, but the positive response is appreciably larger and slower to settle toward equilibrium. A linear system should have the same response (except for sign) for positive and negative excitations. The existence of different responses depending on sign means that the PFD–charge pump system is nonlinear, even when the components are perfect. This nonlinear behavior is inherent in the PFD.

Refer to Section 10.3.1 and Fig. 10.9 on operation of the PFD as an aid to the following explanation of the nonlinearity. If the V pulse leads the R pulse (i.e., phase error θ_e is negative), the pump interval t_p on any one cycle is predetermined by the phase error at the instant of the V pulse. But if the R pulse leads the V pulse, the VCO speeds up while the charge pump is turned on so that the pump interval t_p is shortened from what it would have been in the absence of feedback. The charge transferred into the loop filter on a positive phase error will be less than the charge transferred out on a negative phase error of the same magnitude. Therefore, response to a positive phase error is

slower than response to a negative phase error, as evident in Fig. 12.3. The effect of the nonlinearity is inconsequential under most circumstances, but it can be troublesome in special cases.

A different nonlinearity arises from mismatch between the UP and DN charge pumps. Equal pump currents and instantaneous, time-aligned switching of the current sources have been assumed so far. Real current switches are never matched perfectly. Nonlinearity due to mismatch appears as different gain for pump-UP compared to pump-DN.

Additionally, since the net charge delivered to the loop filter must be zero in steady state, the current switch with the lesser current has to be turned on for a longer interval than the switch with the greater current. Mismatched active intervals can only be generated by a corresponding static phase error. Careful matching of pump currents and switching speeds are needed to minimize the phase error and the nonlinearity.

Yet another nonlinearity exists, most prominently in voltage-switch charge pumps. Suppose that a pair of voltages $\pm V_p$ to be switched are perfectly matched, that the switches are perfect, and that the switches drive the loop filter through a resistor R_1. If the loop filter is at zero voltage, the pump current $\pm V_p/R_1$ is the same for either polarity of phase error. However, if the loop filter has voltage V_F stored on its capacitors, the source pump current will be $(V_p - V_F)/R_1$ and the sink pump current will be $-(V_p + V_F)/R_1$. Here again is another nonlinearity that shows up as gain for one polarity of phase error different from that of the other polarity. Moreover, the severity of nonlinearity depends on the voltage V_F stored on the loop filter.

A similar nonlinearity of lesser severity occurs also with current switches. A practical current source has finite, not infinite, Norton shunt conductance. The actual current delivered therefore depends on the voltage across the load. Even if a pair of current sources are perfectly matched if equal voltages are across them, they depart from match when a nonzero load voltage causes the voltages across the sources to be unequal.

Nonlinearity due to mismatch caused by load voltage can be alleviated by employing an op-amp active filter; the summing junction of the op-amp is always at the same potential, so the charge switches do not see the same range of load voltage as generated in a passive filter.

12.6 RIPPLE SUPPRESSION

Suppose that $C_3 = 0$, so that the PLL is second order. A voltage jump of magnitude $I_p R_2$ is generated on each cycle in which one charge switch or the other (but not both) is turned ON. That voltage jump can exceed the voltage headroom of the charge switches or of the VCO control terminal. Unacceptable overload then occurs and the PLL behavior is unsatisfactory. For this reason, designers rarely omit C_3; it is necessary for restraining the voltage jump. As a point of reference, though, assume temporarily that C_3 has been omitted and that the net UP

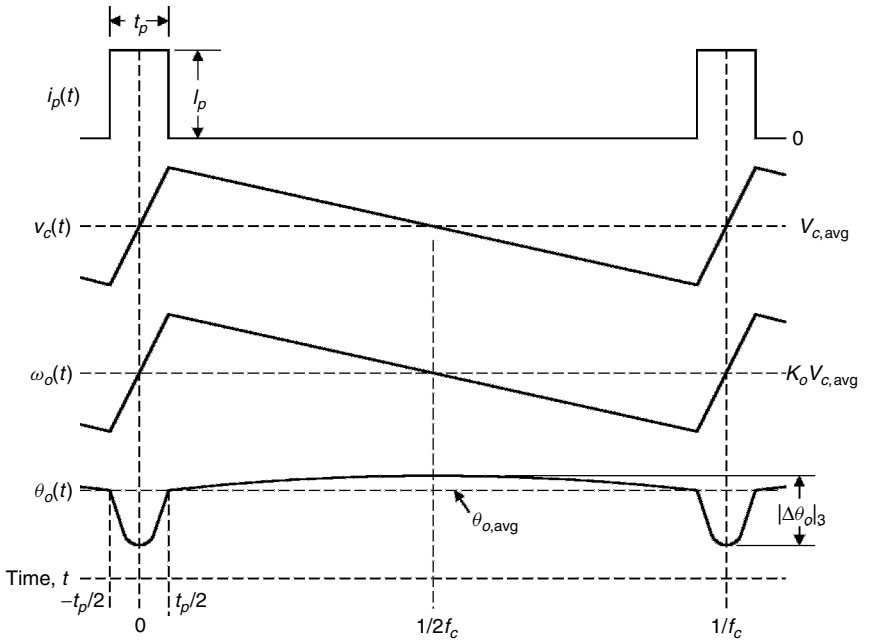


Figure 12.4 Ripple waveforms in a third-order charge-pump PLL.

(or DN) active time is t_p. Voltage jump $I_p R_2$ causes a corresponding frequency jump $K_o I_p R_2$, which when integrated over a time interval t_p, gives rise to a phase ramp with peak-to-peak excursion of $|\Delta\theta_o|_2 = K_o I_p R_2 t_p$ radians.

Now consider a third-order PLL whose charge pump drives into capacitor C_3. Refer to Fig. 12.4 for the ensuing waveforms. As an approximation to simplify analysis, assume that the admittance $\omega_c C_3$ greatly exceeds the admittance $\omega_c C/(\omega_c C R_2 + 1)$ of the rest of the loop filter. If that approximation is valid, the rectangular current pulse of amplitude I_p and duration t_p generates a control-voltage ramp of amplitude $\Delta v_c = I_p t_p / C_3$ peak to peak. Compare that excursion to $I_p R_2$ volts, the voltage jump in the second-order PLL. The ratio of voltage excursions is $t_p/R_2 C_3 = (b-1)t_p/\tau_2$. Substantial reduction obtains for small t_p (a condition existing when the PLL is locked and tracking well), but there is less improvement for larger t_p, such as occurs during acquisition. If problems arise during acquisition, overload from control-voltage excursions should be investigated.

In actuality, the elements C and R_2 present additional admittance to the charge switches, so the voltage swing is not quite as large as indicated. Moreover, the waveform is a segment of an exponential, not a true linear ramp. For these reasons, results of the simplified analysis are slightly on the pessimistic side. As a further approximation, the control voltage is assumed to ramp down linearly from its maximum excursion. The ramp occurs over a time interval $1/f_c - t_p$ and returns exactly to the starting level of the up ramp. [**Comments**: (1) $f_c = \omega_c/2\pi$

hertz is the comparison frequency at the PFD. (2) If the PLL is in steady-state lock, all ripple waveforms have to repeat exactly, with period $1/f_c$. (3) A nonzero value for t_p in steady state comes about from a source of static phase error uncovered in the preceding sections, a source not identified further in this discussion.]

The waveform of $v_c(t)$ is reproduced in the VCO frequency $\omega_o(t) = K_o v_c(t)$ rad/sec, except scaled by the VCO gain K_o. Average values of control voltage and VCO frequency are indicated in Fig. 12.4 by the dashed lines labeled $V_{c,\text{avg}}$ and $K_o V_{c,\text{avg}}$. The ramps have zero mean excursion about the average values. Phase excursions—the integrals of the frequency ramps—are parabolic segments. Peak amplitude of the negative parabola is found to be $K_o I_p t_p^2/8C_3$ radians and that of the positive parabola to be $K_o I_p t_p(1 - t_p f_c)/8 f_c C_3$ radians. Their sum is

$$|\Delta\theta_o|_3 = \frac{K_o I_p t_p}{8 f_c C_3} \qquad \text{radians, peak to peak} \tag{12.13}$$

The ratio of peak to peak phase excursions of the third- and second-order PLLs is

$$\frac{|\Delta\theta|_3}{|\Delta\theta|_2} = \frac{1}{8 f_c R_2 C_3} = \frac{\pi(b-1)}{4\omega_c \tau_2} \tag{12.14}$$

For a numerical example, let $\omega_c\tau_2 = 10$ and $b = 10$, yielding a ratio of $9\pi/40 = 0.7$ (-3 dB). For another example, let $\omega_c\tau_2 = 100$ and $b = 51$, yielding a ratio of $50\pi/400 = 0.39$ (-8 dB). These are not vast improvements in phase excursions; the main benefit of C_3 appears to be in reduction of voltage excursions for small t_p.

12.7 LATE DEVELOPMENTS

Two significant papers [12.4], [12.5] on charge-pump PLLs appeared after the manuscript for this book was completed. They each contain valuable information that is not included in the preceding pages of this chapter. Taking account of the sampled nature of a PFD, [12.4] derives the z-transform transfer functions of the type 2 third-order charge-pump PLL in much the same manner that I did but did not include in this book or in [12.1]. Consult [12.4] for the details of that derivation. In particular, we have found the same characteristic polynomial of the closed-loop transfer functions, the denominator of their equation (40). [**Caution**: Our notations are defined differently. Most particularly, their K is not the same as the K of this book. Nonetheless, our results agree once the differing notations are harmonized.]

Several workers have observed that the z-domain stability limit for this PLL is reached at a gain K for which a real pole passes through the unit circle at $z = -1$. The gain associated with that limit is found by substituting $z = -1$ into the characteristic equation and solving for K; the result is (12.12), as plotted in Fig. 12.2.

The paper also provides information on designing for good phase margin. From simulation results, Hanumolu et al. [12.4] suggest that the stability boundary is well approximated by a restriction on unity-gain crossover frequency of $\omega_{gc} \approx \omega_c/3.5$. Applying the approximation $K \approx \omega_{gc}$ and sketching $K = \omega_c/3.5$ into Fig. 12.2 (sketch not included in the printed figure) suggests that the latter approximation is reasonable for some choices of loop parameters.

Levantino et al. [12.5] explore a PLL incorporating a PFD and a charge pump. Its objective is to achieve fast acquisition of frequency by maximizing the loop bandwidth and judicious placement of closed-loop poles in the z-domain. The authors recommend setting the stabilizing zero at $z = 0.5$ and all three closed-loop poles close to $z = 0$. Theoretically, and unrealizably, such a loop could settle in just three cycles of the PD comparison frequency. (A similar behavior for a first-order digital PLL is described in Section 4.2.6.) The article reports a simulated PLL that settles in just seven cycles, much faster than achieved in normal designs.

A fast-settling design has to be checked carefully for the effects of parameter tolerances on stability margin, transient response, and frequency response. The closed-loop system response $|H(f)|$ of such a PLL provides little filtering of input disturbances. The most plausible application of such a wideband PLL might be for stabilizing of noisy locked oscillators, as in a frequency synthesizer (Chapter 15).

REFERENCES

12.1 F. M. Gardner, "Charge-Pump Phase-Lock Loops," *IEEE Trans. Commun.* ***COM-28***, 1849–1858, Nov. 1980. Reprinted in [12.2] and [12.3].

12.2 W. C. Lindsey and C. M. Chie, *Phase-Locked Loops*, Reprint Volume, IEEE Press, New York, 1986.

12.3 B. Razavi, *Monolithic Phase-Locked Loops and Clock Recovery Circuits*, Reprint Volume, IEEE Press, New York, 1996.

12.4 P. K. Hanumolu, M. Brownlee, K. Mayaram, and U.-K. Moon, "Analysis of Charge-Pump Phase-Locked Loops," *IEEE Trans. Circuits Syst. I* ***51***, 1665–1674, Sept. 2004.

12.5 S. Levantino, M. Milani, C. Samori, and A. L. Lacaita, "Fast-Switching Analog PLL with Finite-Impulse Response," *IEEE Trans. Circuits Syst. I* ***51***, 1697–1701, Sept. 2004.

CHAPTER 13

DIGITAL (SAMPLED) PHASELOCK LOOPS

As will become apparent, many PLLs that have been called "digital" really are hybrids of analog and digital. A truly digital PLL works solely by processing discrete sequences of numbers. A hybrid PLL has a mixture of analog and digital operations. Examples of each appear in this chapter. The terminology *sampled PLLs* is employed here to encompass both kinds. Sampled (time-discrete) phaselock loops can be categorized in a variety of ways. For immediate purposes they are divided into two distinct classes: (1) quasilinear and (2) inescapably nonlinear. All PLLs—digital or analog—contain nonlinearities, as has been noted repeatedly in earlier chapters. Nonlinearities in a quasilinear sampled PLL can be neglected to arrive at useful analysis techniques based on z-domain transfer functions, as in Chapter 4. Two different types of nonlinearities can be neglected: those inherent in the digital phase detector and VCO, nonlinearities that are essentially the same as (or less prominent than) in analog PLLs and those arising from numerical quantization. Quantization nonlinearity is unique to digital operations. Other sampled PLLs contain drastic nonlinearities that entirely prevent any linear approximation. In a severely nonlinear network, the powerful tools of transfer functions, frequency responses, gain, and bandwidth no longer have any definable meaning. Behavior of nonlinear PLLs is much more complicated than that of linear PLLs.

This chapter is divided into three sections on (1) quasilinear PLLs, (2) quantization effects, and (3) nonlinear PLLs. Work on sampled PLLs with digital elements has been in progress at least since the 1960s. Lindsey and Chie [13.1] published a survey article that gives a good overview of the early work and a long

Phaselock Techniques, Third Edition, by Floyd M. Gardner

list of references. They divide sampled PLLs into classes based on operation of the phase detector, a subclass in this chapter. They also denominate many hybrid PLLs as digital, a pervasive usage not continued in this book.

Most systems containing a sampled PLL accept an analog signal as input, even if the PLL itself is purely digital. The analog input is sampled and digitized within the system. In all that follows, it is assumed tacitly that suitable presampling filters suppress harmful spectral folding.

13.1 QUASILINEAR SAMPLED PLLs

In-lock operations of just about all quasilinear PLLs can be approximated by transfer functions similar to those introduced in Chapter 4. This section is devoted primarily to implementations of the constituent elements and to variant configurations of the overall PLL. All nonlinearities, including quantization, are ignored.

13.1.1 Digital-Controlled Oscillators

A number-controlled oscillator (NCO) was examined in Chapters 4 and 9 and a recursive digital sinusoidal oscillator (RDSO) was mentioned in Chapter 9. The generic term *digital-controlled oscillator* (DCO) has been used to mean any oscillator whose frequency is controlled by a digital number. This section examines a couple of DCOs of interest.

Period DCO Figure 13.1 shows an arrangement that was considered extensively in early sampled PLLs [13.3, 13.4]. It consists of a fixed oscillator at frequency f_{ck} followed by a frequency divider with selectable division ratio Q. The output frequency of the divider $f_o = f_{\text{ck}}/Q$ is inversely proportional to the divider ratio. Rather than cope with the nonlinear inverse relation, analysis is simplified by considering the output period $T_o = 1/f_o = Q/f_{\text{ck}}$ instead. Now T_o

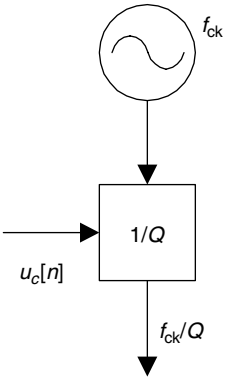


Figure 13.1 Period DCO based on a frequency divider.

is directly proportional to Q and is easily incorporated into difference equations and corresponding transfer functions.

To formulate a difference equation for operation of the period DCO, denote the time of occurrence of the nth leading edge of the divider output stream as $t[n]$. That time of occurrence is determined by

$$t[n] = t[n-1] + Q[n-1]t_{ck} = t[n-1] + u_c[n-1]t_{ck} \tag{13.1}$$

where the control number $u_c[n]$ is substituted for Q and $t_{ck} = 1/f_{ck}$. That is, the divider modulus Q is updated after each output cycle according to the value of the control word $u_c[n]$. Regard the control word as an integer with unit increments. Take the z-transform of (13.1) to obtain

$$T(z) = \frac{z^{-1}}{1 - z^{-1}} t_{ck} U_c(z) \tag{13.2}$$

Equation (13.2) represents a digital integrator with the same form as (4.9) for an NCO. That expression can be incorporated into the linear description of a DPLL in the manner developed in Chapter 4, and the same forms of DPLL transfer functions will be obtained; no further development of transfer functions is needed here. Additional analysis of sampled PLLs containing period oscillators may be found in [13.4] and [13.5].

Observe from (13.1) that the time of the leading edge can be adjusted only in increments of t_{ck}. To achieve fine adjustments requires a large frequency f_{ck} of the fixed oscillator and fast circuits in the frequency divider. This granularity arises because the period DCO is not fully digital—it is a hybrid device. Its output is not a digital number, it is an analog signal whose useful information is carried in the time instants of its leading edges. Those leading edges typically are used to actuate a switch that samples an analog input signal.

The NCO of Chapter 4 does not suffer from this particular timing granularity. Output of the NCO is a sequence of discrete digital numbers representing phase (closely related to time, as explained in Section 2.1.4). Granularity of phase depends solely on the word length of the digital numbers, a length ordinarily much larger than feasible in the divider modulus Q, especially at high frequencies for the period DCO output. The output of an NCO might be processed further to generate samples of the sine or cosine of the phase angle. The output of the RDSO of Chapter 9 directly delivers sine and cosine samples without further processing. Phase relations are implicit in the sines and cosines; the phase resolution depends primarily on the word length.

Phase-Selector DCO Figure 13.2 shows another technique that seems to be used extensively, but it does not seem to have a well-accepted name. The only published account I could find was [13.42] and its references. Like the period DCO of Fig. 13.1, it too is really a hybrid device with analog output and also has a granularity problem. It offers improved granularity without incurring the unreasonably high frequency f_{ck} demanded by the period DCO. The phase-selector

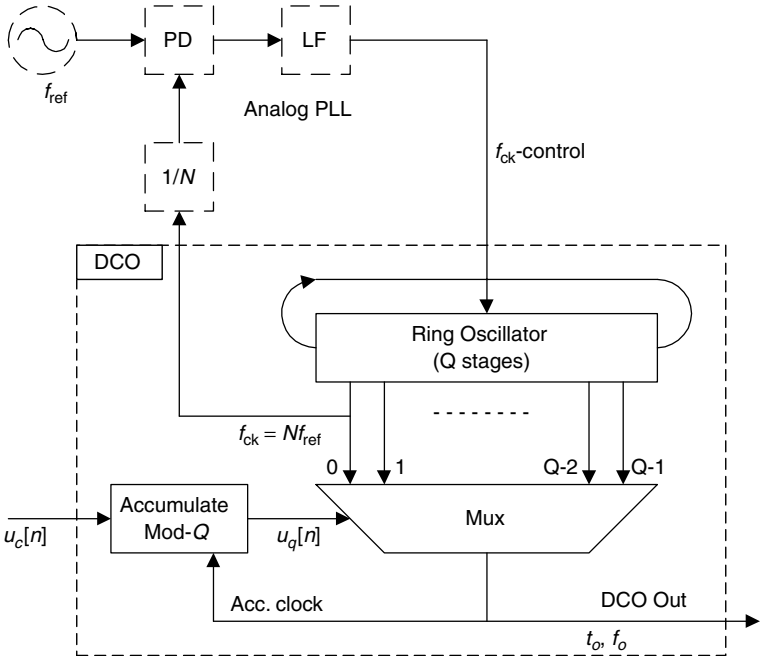


Figure 13.2 Phase-selector DCO based on a phaselocked ring oscillator and a multiplexer.

DCO is based on selecting discrete phases from a tapped delay line with Q taps. To avoid phase unwrapping problems that arise in an open-loop delay line, the delay line is closed in a ring, thus forming a ring oscillator. To establish an accurate oscillation frequency in the ring, the oscillator is phaselocked to a harmonic of a stable fixed oscillator at frequency f_{ref}.

Oscillation frequency around the ring is $f_{\text{ck}} = 1/t_{\text{ck}} = Nf_{\text{ref}}$. With Q equally spaced taps around the ring, the time increment between adjacent taps is $\delta t = t_{\text{ck}}/Q$. Rather than relying on a high clock frequency and a fast divider as in the period DCO, this scheme only requires short delays in each cell of the ring oscillator to achieve a small time increment.

The waveforms at each tap are nominally square waves, each at the same frequency f_{ck} but time-shifted from one another by intervals δt. The phase selector is part of a PLL that generates a control signal $u_c[n]$ which selects, via an accumulator and a multiplexer, the delay-line tap to be used as the DCO output. Feedback around the PLL causes the tap-selection phase to keep up with the phase of the input signal.

Accumulator operation is represented by

$$u_q[n] = \{u_q[n-1] + u_c[n-1]\} \mod Q \tag{13.3}$$

where $u_q \in \{0, 1, \ldots, Q-1\}$ selects the active tap on the ring oscillator. A new tap can be (but usually is not) selected for each output cycle of the DCO. Control

signal u_c may be regarded as an integer in its effect on the DCO; it specifies the increment in tap position from one output cycle to the next. The increment is restricted to the range

$$-\frac{Q}{2} \leq u_c < \frac{Q}{2} \tag{13.4}$$

(A circular increment in excess of halfway around a circle in one direction is indistinguishable from a complementary smaller increment in the other direction. Values of u_c outside the restricted range are aliased.)

In conformance to (13.3), duration of the nth output cycle is

$$t_o[n] = t_{\text{ck}} \left(1 + \frac{u_c[n-1]}{Q}\right) \tag{13.5}$$

and $f_o = 1/t_o$. From the restrictions (13.4) on u_c, the extreme bounds on period are $t_{\text{ck}}/2 < t_o < 3t_{\text{ck}}/2$, so that $2f_{\text{ck}}/3 < f_o < 2f_{\text{ck}}$. Equation (13.3) has the same format as (13.1); they differ only in the coefficient of $u_c[n]$ and in the meaning of Q. Their z-transforms have the same format. Therefore, the phase-selection DCO is really a period DCO, just implemented in a different manner that may be more advantageous.

Although the waveform at any one tap of the ring oscillator nominally is a square wave with 50% duty ratio, the DCO waveform out of the multiplexer might have any duty ratio from almost 100% (for u_c nearly equal to $Q/2$) down to ~33% (for u_c nearly equal to $3Q/2$). The desired timing information is contained in the time instants of the leading edges of the multiplexer output, not in the waveform itself. Great care must be taken in the design of the multiplexer to avoid switching faults—extra or missing edges in the waveform. The design problem can be especially challenging if δt is very small.

The DCO can be operated advantageously at a frequency much higher than that of the input signal to the DPLL; just place a frequency divider in the feedback path out of the DCO to reproduce the input frequency. A divider preserves the timing resolution δt while reducing the phase granularity relative to the input signal.

13.1.2 Hybrid Phase Detectors

Several categorizations of phase detectors may be recognized: (1) true digital vs. hybrid, (2) multiplier vs. sequential, and (3) sample the signal vs. sample the DCO. The two inputs to a true digital phase detector consist of sequences of time-discrete digital numbers, and the PD output is another sequence of time-discrete digital numbers computed numerically from the input samples. One or both inputs to a hybrid phase detector are analog signals; its output is a sequence of digital numbers. A hybrid phase detector incorporates an analog-to-digital converter (ADC) in some, possibly obscure form. This section is devoted to hybrid PDs.

Multiplier vs. Sequential Most hybrid and all true-digital PDs are in the multiplier (combinatorial) category. Recollect from Chapter 10 that a sequential PD measures the time differences between specific edges on the input waveforms. Time differences are not meaningful in sample sequences, so a fully digital PD cannot be sequential. One way to implement a hybrid sequential PD is to use an edge on the input signal to start a high-speed counter and an edge on the feedback signal from the DCO to stop the counter. The resulting count is an indication of phase (or timing) error, the counter serving as the ADC. Such PDs are workable at frequencies low enough for an adequate number of counts to be accumulated, but they grow progressively less satisfactory as the signal frequency increases. For this reason, most hybrid PDs belong to the multiplier category.

Sample the Signal More restrictively, many hybrid PDs and all true-digital PDs are strictly sampling phase detectors. All deliver samples at their outputs (otherwise the outputs are not digital; the PD would then be analog, not hybrid). Many hybrid PDs perform their sampling at their inputs. A common arrangement is illustrated in Fig. 13.3. The analog input signal $r(t)$ is sampled and held at time instants (not equispaced) determined by the DCO. Analog-voltage samples $r[n]$ are applied to the ADC which delivers digital samples $u_d[n]$. The s-curve of this PD, that is, the average of u_d as a function of the phase error θ_e, has the same shape as the waveform of $r(t)$. If $r(t)$ is sinusoidal, then u_d vs. θ_e is also sinusoidal. Other shapes of $r(t)$ yield corresponding different shapes of the s-curve. Replication of the signal waveform in the s-curve is a general property of sampling PDs, at least in the absence of additive noise at the input.

Sample the DCO Figure 13.4 turns around the usual concept of a sampling PD; a trigger edge of the input signal $r(t)$ is used to sample the feedback information. In Fig. 13.4, the DCO or VCO signal is assumed to be at a much higher frequency than the input signal. Output of the DCO or VCO is counted down to

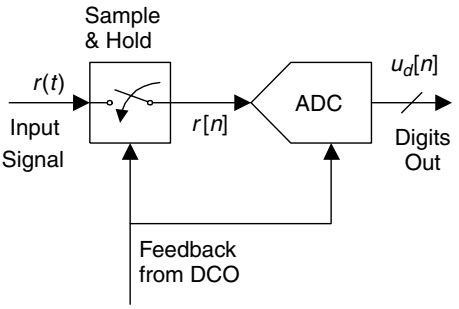


Figure 13.3 Hybrid phase detector consisting of a sampler and an analog-to-digital converter.

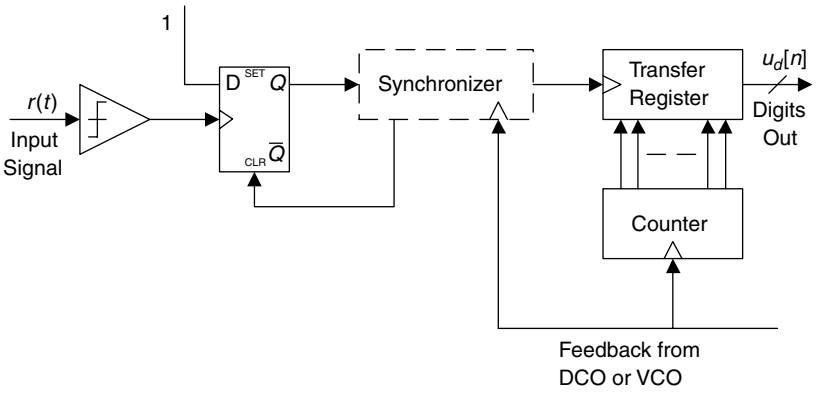


Figure 13.4 DCO sampled at signal edges for phase detection.

the frequency of $r(t)$. One state of the counter is established as a zero state. If the PLL is phaselocked to the input signal, the counter (in the absence of static phase error and of phase jitter) should be at the zero state for each trigger edge of the input signal. The figure shows a count-down arrangement, but the same principle applies if the trigger edge samples the phase register of an NCO. Analog-to-digital conversion is provided by the counter or NCO, as applicable, obviating the conventional ADC ordinarily needed in a PD that samples the signal. If the countdown ratio is Q, the s-curve is quantized in Q equal steps over one full cycle of the oscillator signal. Apart from the quantization, the s-curve is a linear sawtooth over the cycle.

As described so far, the s-curve will have a dead zone of $1/Q$ cycle in its zero state. Dead zones in s-curves should be avoided whenever possible. A simple expedient is to append $\frac{1}{2}$LSB to $u_d[n]$ for each n. That eliminates a state of zero output from the PD; the smallest u_d sample values are $\pm 1/2Q$ instead of zero. At equilibrium, the PLL jumps back and forth between these two minimal phase errors and can never settle to zero. The fast jumping is far preferable to the slow wandering encountered in a dead zone.

Trigger edges of $r(t)$ are asynchronous to the feedback edges. An essential synchronizer block shown in Fig. 13.4 operates as described in the next paragraphs. Asynchronous interfaces can appear in several different locations in a sampled PLL; be alert to the potential need for synchronization. If the transfer register were to be actuated at uncontrolled time instants, transfers could occur as the counter was changing its state. Because of this uncertainly of state, the count applied to the transfer register would contain unacceptable random errors. A synchronizer is armed by the input-trigger edge and then actuates the transfer at an instant when the counter is stable.

The time between arrival of the input trigger and the next feedback edge is variable, so the synchronization in a feedback-sampling PD introduces a source of phase jitter that is not present if the time-continuous input signal is sampled by the time-discrete feedback. If the extra jitter is tolerable, sampling of the

feedback may be far preferable to including the ADC that would be needed in conjunction with sampling of the input signal.

Properties of Sampling PDs The two configurations of sampling PDs shown above (one samples the input signal; the other samples the feedback) generate usable phase-error sequences if the frequency of the sampling stream is any integer subharmonic of the frequency of the sampled signal. This property can be beneficial (it is the basis for subsampling) or it can be hazardous (the PLL could lock up at the wrong frequency if the wrong subharmonic were selected inadvertently). The s-curve is always periodic at the frequency of the sampling stream and has the shape of the waveform of the sampled signal.

Another valuable property of a sampling PD: Ripple is absent from the PD output if the sampling stream and sampled signal are synchronous when the PLL is locked and if there is no more than one sample per cycle of the sampled signal.

13.1.3 Complex-Signal Digital Phase Detector

True digital PLLs are perhaps most widely used today for signal synchronization (recovery of carrier and clock) in data-signal receivers. Many data signals are generated and transmitted in two-dimensional (complex) format. Most radio receivers for data signals are arranged in a complex format, even if the signal itself is one-dimensional (e.g., BPSK signals). This section points out salient properties of a digital PD for complex signals.

Represent the complex input signal after sampling as $v_i[n] = A\exp[j(2\pi f_i t_s + \theta_i)]$ and the complex (two-phase) output of a phaselocked DCO as $v_o[n] = \exp[-j(2\pi f_i t_s + \theta_o)]$, where A is the amplitude of the input signal, f_i is its frequency, and t_s is the uniform sampling interval. There is no necessary relation between f_i and t_s. A truly digital phase detector can be based on the complex product $v_i v_o = A\exp[j(\theta_i - \theta_o)] = A\exp[j\theta_e]$. More specifically, a phase-detector algorithm,

$$u_d[n] = \mathrm{Im}[v_i v_o] = A\sin\theta_e[n] \tag{13.6}$$

illustrates the potentialities of complex signal processing. [**Notation**: Im$[x]$ indicates the imaginary part of x.] Observe that (13.6) contains no ripple component and does not depend on either f_i or t_s. The sampling frequency $1/t_s$ need not be synchronous with the signal frequency f_i. The sampling frequency can be much smaller than the carrier frequency; aliasing of the carrier is permissible if the location of the desired alias is known and adequately separated from other aliases. This is the basis of subsampling.

A glimpse inside the PD process is informative. Recollect that $\exp(jx) = \cos(x) + j\sin x$, so that the complex product $v_i v_o$ consists of four real multiplications and a pair of addition/subtractions. The imaginary part of the product

(the only part that need be computed) is

$$\begin{aligned}
&\mathrm{Im}\{A[\cos(2\pi f_i t_s + \theta_i) + j\sin(2\pi f_i t_s + \theta_i)] \\
&\quad \times [\cos(2\pi f_i t_s + \theta_o) - j\sin(2\pi f_i t_s + \theta_o)]\} \\
&= A\{-[\cos(2\pi f_i t_s + \theta_i)\sin(2\pi f_i t_s + \theta_o)] \\
&\quad + [\cos(2\pi f_i t_s + \theta_o)\sin(2\pi f_i t_s + \theta_i)]\} \\
&= \frac{A}{2}\{[\sin(\theta_i - \theta_o) - \sin(4\pi f_i t_s + \theta_i + \theta_o)] \\
&\quad + [\sin(\theta_i - \theta_o) + \sin(4\pi f_i t_s + \theta_i + \theta_o)]\} \\
&= A\sin(\theta_i - \theta_o) \qquad (13.7)
\end{aligned}$$

Only two real multiplications [enclosed in brackets in (13.7), lines 3 and 4] and one subtraction are needed for computing (13.6). Products of the individual multiplications contain a double-frequency ripple component (as well as the desired difference-frequency component), but those ripple components cancel in the subtraction. Digital implementation can achieve essentially perfect balance between the products and consequent near-perfect cancellation of ripple. Imbalance, and therefore uncanceled ripple, comes about solely from finite-word-length effects. Observe that $\mathrm{Re}[v_i v_o] = A\cos(\theta_i - \theta_o)$, irrespective of any relation between f_i and t_s. This property is equivalent to that of the analog auxiliary phase detector of Section 8.4.1.

13.1.4 DPLLs in Digital Data Receivers

Example DPLL configurations shown in Figs. 13.5 to 13.7 illustrate several techniques employed for carrier recovery in data receivers. All elements and all connections in these figures are digital. Connections with double lines represent complex signals (almost universally generated in digital processing of passband signals), whereas connections with single lines are real signals. These figures are greatly simplified for purposes of illustration; configurations of actual receivers are more complicated.

Basic DPLL Configuration Figure 13.5 is a base for further discussion. It shows only a carrier-recovery DPLL embedded within a digital receiver. A phase detector (PD), loop filter, and NCO are digital elements introduced in Chapter 4. Two new elements appear in the figure: a sine/cosine processor and a phase rotator. The sine/cosine processor accepts the real NCO phase samples $\varepsilon_o[n]$ (in fractional cycles) as input and delivers sine and cosine samples of those phases to produce a complex local oscillator signal $\exp(-j\theta_o[n])$, where $\theta_o[n] = 2\pi\varepsilon_o[n]$. Scaling by 2π is implicit in the sine/cosine process; it has to be included as a factor in the loop gain when developing the transfer functions

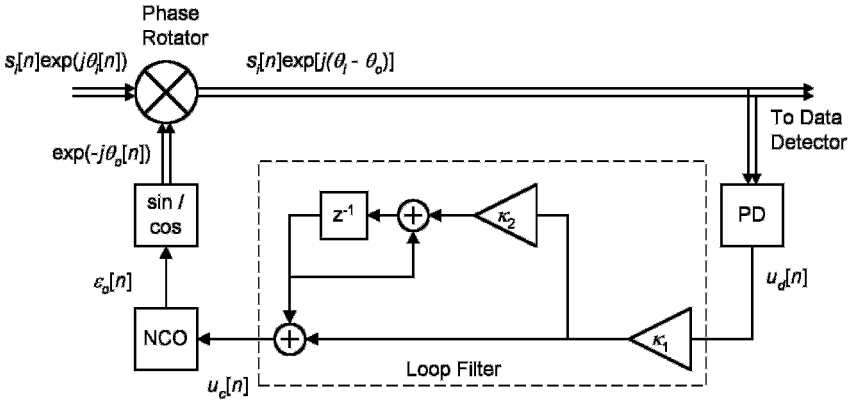


Figure 13.5 Carrier-recovery digital PLL (double lines for complex signals; single lines for real signals).

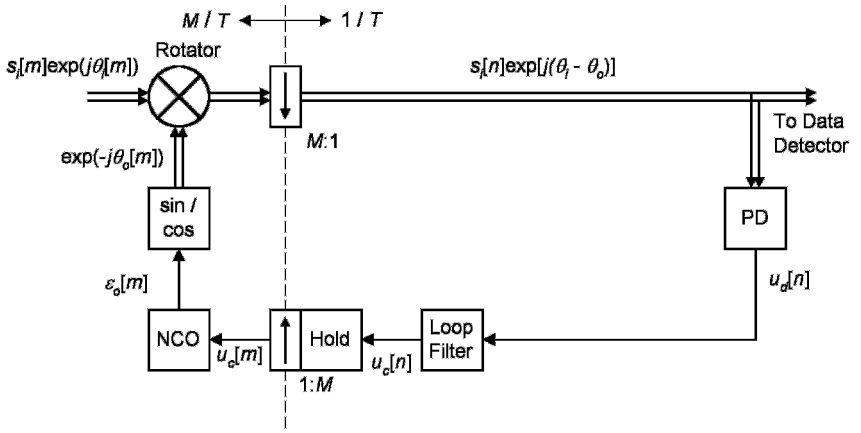


Figure 13.6 Digital PLL running at two sample rates M/T and $1/T$.

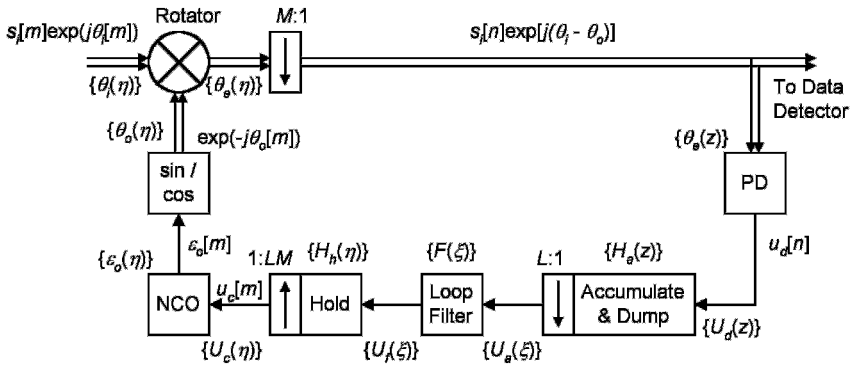


Figure 13.7 Digital PLL running at three rates, showing a hold element and an accumulate & dump element.

if signal phases are measured in radians and NCO phases in cycles. The phase rotator performs a complex multiplication between the incoming complex data signal $s_i[n]\exp(j\theta_i[n])$ and the local oscillator signal to produce the complex, ripple-free, difference-frequency signal $s_i[n]\exp[j(\theta_i - \theta_o)]$.

In general, the data signal and the local-oscillator signal are at nonzero frequencies, whose existence is incorporated into linearly varying values of θ_i and θ_o. The average frequencies in θ_i and θ_o are equal when the PLL is locked, so the linearly varying components cancel out from the difference $\theta_e = \theta_i - \theta_o$. Samples with index n have been taken at uniform time increments corresponding to data-symbol interval T. Assume that timing of the samples has been synchronized to the received data symbols; timing recovery—a rich subject—is not examined further in this section.

The phase detector provides an indication of phase error once per symbol interval. (One sample per symbol is the optimum design for digital phase detection.) A PD algorithm often used for quadrature-amplitude modulation (QAM) signals is $u_d[n] = \mathrm{Im}\{c^*[n]s_i[n]\exp[j(\theta_i - \theta_o)]\}$, where $c[n]$ is an estimate of the data value of the nth symbol and * indicates a complex conjugate. But no matter what PD algorithm is used, operation in a quasilinear mode yields an approximation $u_d[n] \approx \kappa_d(\theta_i[n] - \theta_o[n])$.

In analog PLLs, the action of the phase detector is twofold: (1) it provides an indication of phase error, but (2) it also provides a frequency translation from passband to baseband. The digital PLL of Fig. 13.5 separates those two actions in a manner typical of complex-signal receivers: The phase rotator provides the passband to baseband frequency translation while the PD extracts phase error information from the complex baseband signal. Despite these novelties, the configuration of Fig. 13.5 is described by the same difference equations and transfer functions developed in Chapter 4.

Multirate Sampling Figure 13.6 introduces *multirate processing* [13.6, 13.7]. The front portions of a digital receiver customarily need to be sampled at a rate M/T that is higher than the symbol rate $1/T$. Values of M in the range 2 to 4 are widespread. (The sampling ratio M need not be an integer and not even a rational number; see [13.8–13.11].) Data recovery and phase detection are performed at the symbol rate, so the higher rate has to be down-sampled to $1/T$. Any feedback from symbol-rate portions of the receiver into higher-rate portions have to be up-sampled to the higher rate.

For illustrative purposes, down-sampling by the ratio $M:1$ in Fig. 13.6 is shown as following the phase rotator, an unlikely location for the simple configuration of the figure but realistic in more complicated arrangements. That necessary down-sampling requires subsequent up-sampling at some location between the PD and the local-oscillator signal applied to the phase rotator. An appropriate location for up-sampling is between the loop filter and the NCO.

Hold Process Up-sampling often is accomplished by a *zero-order hold* process which accepts samples $u_c[n]$ from the loop filter at rate $1/T$ and delivers M

identical samples $u_c[m] = u_c[n]$ at rate M/T. If u_c is regarded as the NCO phase increment (in fractional cycles) per sample interval, the NCO phase advances by $Mu_c[n]$ fractional cycles in the time interval T. The hold function inserts a dimensionless gain factor M into the loop equations.

Why should the phase increment be characterized in the time interval T rather than in the time interval T/M? Since the PD is sampled at $1/T$, it can only sense changes at time intervals T. It is common practice in multirate PLLs to relate all phases and times to the PD, the location where the loop error is sensed.

Accumulate & Dump Process Loop bandwidths of synchronizer PLLs in data receivers ordinarily are much smaller than the symbol rate $1/T$. Practical bandwidths range from ~3 to 5% of the symbol rate at the upper extreme to as small as 0.1% of the symbol rate, or even less. Recognizing the small bandwidth, many workers have asked themselves why it is necessary to update the loop so often. Why not down-sample after the PD and update at a slower rate, thus reducing computing burden? Several good arguments against this approach arise in the sequel, but the technique has been applied in some instances; the next few paragraphs relate its principles.

Figure 13.7 places one more element—an accumulate & dump process—into the digital PLL. An accumulate & dump constitutes one of the simplest techniques for down-sampling. It adds up L succeeding input samples at rate $1/T$ and delivers their sum in a single sample at the output. Therefore, the process comprises a filter (in the accumulation) and a down-sampler (in the dump). The loop filter that follows the accumulate & dump runs at a sample rate of $1/LT$. The hold process that follows the loop filter must up-sample by $1:LM$ to achieve a rate of M/T.

How has the loop gain been affected by the insertion of the accumulate & dump? For comparison, refer to Fig. 13.5 for the configuration with single-rate sampling. Replace the loop filter with a straight-through connection by setting $\kappa_1 = 1$ and $\kappa_2 = 0$, equivalent to the loop-filter transfer function $F(z) = 1$. In response to L successive output samples from the PD of equal value u_d, the phase of the VCO will be advanced by Lu_d fractional cycles.

Now refer to Fig. 13.7, which has the accumulate & dump with its $L:1$ down-sampling and the zero-order hold with its $1:LM$ up-sampling. (Ignore labels in braces $\{\cdot\}$ in the figure until Appendix 13A.) Once again, assume that the loop filter is a straight-through connection with gain = 1. In response to L successive equal input samples u_d, the accumulate & dump outputs a single sample of amplitude Lu_d. Up-sampling in the hold process generates LM samples of amplitude Lu_d, so the phase of the NCO advances by L^2Mu_d fractional cycles. That is a factor of LM larger than the phase advance due to L equal PD samples in the configuration of Fig. 13.5. Since the hold contributed a gain factor of M, the net contribution of the accumulate & dump to loop gain is a factor L.

These gain factors for the accumulate & dump and for the hold apply only for a constant input signal, that is, a signal at zero frequency. Signals at nonzero frequencies will be filtered. Understanding of the filter properties is needed for

informed design of the feedback loop. Section 2.3 of [13.6] lays out the principles of resampling in multirate systems. Appendix 13A applies those principles to develop transfer functions for the DPLL of Fig. 13.7.

[**Caution**: Down-sampling induces aliasing of the input signals. If spectral content (such as noise or interference) exists in the PD output at frequencies in excess of $1/2LT$, those frequencies will be aliased into frequencies in the range 0 to $1/2LT$ and will be more difficult to reject by filtering in the PLL. If such disturbances exist in significant amplitudes, down-sampling with an accumulate & dump should be avoided.]

13.1.5 Loop Stability

Linearized analysis of quasilinear digital and hybrid PLLs leads to a simple stability criterion: All poles of the system transfer function must be within the unit circle for the loop to be stable. But all PLLs are nonlinear, even apart from quantization effects; the linearized analysis does not necessarily reveal all facts relating to stability. Several authors [13.12–13.15] have investigated hybrid PLLs for the effects of nonlinearities (excluding quantization issues) on stability and acquisition. They have uncovered behavior that would not be expected from experience with analog PLLs. A prudent designer would (1) become familiar with these analyses to guard against the perils identified therein and (2) incorporate comfortable stability margins into sampled PLLs as protection against unexpected surprises.

13.2 QUANTIZATION

Digital numbers necessarily have finite precision—they are quantized. This section explores the effects of quantization upon PLL phase jitter. As will become apparent, numerous questions remain open for further investigation.

13.2.1 Lessons from Related Studies

Quantization has received close attention in digital signal processing [13.16–13.19], in delta–sigma ($\Delta\Sigma$) converters [13.20, 13.21], and articles on quantization per se [13.22–13.28]. Several well-known results are directly applicable to PLLs, as described in the following paragraphs.

Quantization as Additive Noise Effects of quantization are often modeled as white noise of uniform probability density added to an otherwise linear system. Most treatments of this approach warn emphatically that the model is valid only on the essential condition that the signal or external additive noise is large enough to make the quantization error be uncorrelated with the signal. That condition is not usually met in a PLL, because additive noise is often small or absent in PLLs of interest and phase error is small and the control signal to the NCO is largely quiescent when the loop is locked. Simulations (Gardner, unpublished

results) have demonstrated that the additive-noise model of quantization error is very poor in a PLL with otherwise-small noise at its input.

Effects of Additive Noise Numerous studies have shown that additive noise "linearizes" the stair-step character of a quantizer. That is, the average value of a signal plus zero-mean noise over many samples approaches the true signal value alone, despite the quantization of individual samples. With sufficient additive noise, the quantizing error can be treated as additive white noise with uniform probability density. Surprisingly small additive noise—standard deviation on the order of one quantizing increment—may be sufficient. This effect applies as well to digital PLLs, as appears in the sequel.

Limit Cycles A digital feedback network may be subject to limit cycles—unwanted periodic oscillations due to quantization—not predicted by linear analysis. Limit cycles occur in recursive digital filters and in $\Delta\Sigma$ converters. They also occur in digital PLLs and their properties are a main topic of this section. Rigorously speaking, a true limit cycle is strictly periodic—the same sequence of sample values repeats exactly in every period. That can occur only if the frequency of the incoming signal is in an exact rational ratio to the frequency of the sampling clock. If, as is common, the incoming signal and sampling clock are derived from independent oscillators, the frequency ratio will be irrational; the two frequencies are incommensurate. In that case, the same sequence of sample values (specifically, of phase error) never repeats exactly and the oscillations are not periodic. Experts on nonlinear dynamics use the name *quasiperiodic orbit* instead. Besides the requirement on periodicity, a limit cycle is supposed to be the only orbit in its neighborhood, whereas many similar "limit cycles" with varying details may develop in the PLLs to be described, depending on initial conditions. The term *limit cycle* is improper when applied to quantized PLLs. The term is used here nonetheless, despite its impropriety, because of its greater familiarity in the wider engineering community.

13.2.2 Quantization Considerations in Hybrid PLLs

Many of the early articles on hybrid PLLs neglected quantization entirely; these dealt with enough then-new concepts without raising the complications of quantization as well. Other articles dealt with extremely coarse quantization. The latter class of systems is regarded here as inherently nonlinear, not quasilinear, and its discussion is deferred to Section 13.3. A search of the early literature turned up only two papers [13.29, 13.30] on hybrid PLLs with multilevel quantization. Both used a model similar to Fig. 13.8, consisting of analog sinusoidal signal plus noise as input, a sampler as a phase detector, an analog-to-digital converter providing quantization, a loop filter, and a period DCO. The D'Andrea and Russo paper [13.29] dealt solely with a first-order loop; the Pomalaza-Raez and McGillem paper [13.30] treated a second-order type 2 loop in addition. All quantization in [13.29] is in the ADC; [13.30] also makes provision for separate

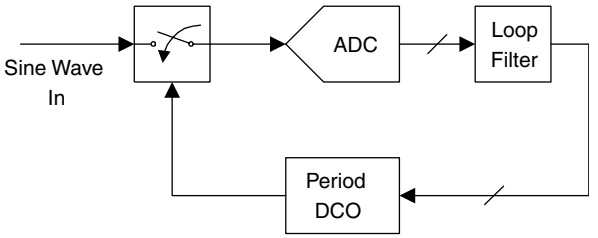


Figure 13.8 Hybrid PLL.

quantization in the DCO. The quantization in [13.29] is uniform in phase error, not in signal amplitude; this kind of s-curve might most easily be implemented by the arrangement of Fig. 13.4. Both quantizers have a midtread characteristic, implying a dead zone around zero phase error. Both papers assume low to modest signal-to-noise ratio and are concerned primarily with the issues classically associated with analog PLLs operating in noise, such as probabilities of steady-state phase error, time to loss of lock (due to noise), and speed of acquisition. Both make use of Markov chains for analysis of performance.

13.2.3 Effects of Frequency (NCO) Quantization

The results reported in [13.29] and [13.30] are valuable for evaluating the performance of hybrid or digital PLLs in the presence of external noise but furnish only limited insight into behavior under noise-free conditions where quantization effects predominate. A different approach is needed for analysis of noise-free PLLs. Quantization has to be examined in each individual element of a DPLL and then quantization of all elements together. This section concentrates on frequency quantization in the NCO. Results can readily be extended to frequency quantization in any digit-controlled oscillator in any digital or hybrid PLL.

Frequency quantization has been studied by Gardner [13.31], who performed simulations and inferred generalized properties therefrom and by Teplinsky, Feely, and Rogers [13.32] and Teplinsky and Feely [13.33], who performed elaborate nonlinear analyses that corroborated applicable portions of [13.31]. At the time of this writing, very little seems to have been published on quantization elsewhere in the loop, with one exception: Da Dalt [13.43] has investigated a digital PLL with a binary-quantized ("bang-bang") phase detector.

Study Model Figure 13.9 is a simplified model of the DPLL considered in [13.31–13.33]. Notation has been changed somewhat to conform to that of Appendix 13A. It closely resembles the configuration of Fig. 13.7 with the minor exceptions that the $M:1$ signal-path down-sampling of Fig. 13.7 is absent (equivalently, $M = 1$) and the integrator in the loop filter is delay-free. Its quasilinear transfer functions are nearly the same as those derived in Appendix 13A; compare eq. (19) of [13.31] to (13A.20) and (13A.21).

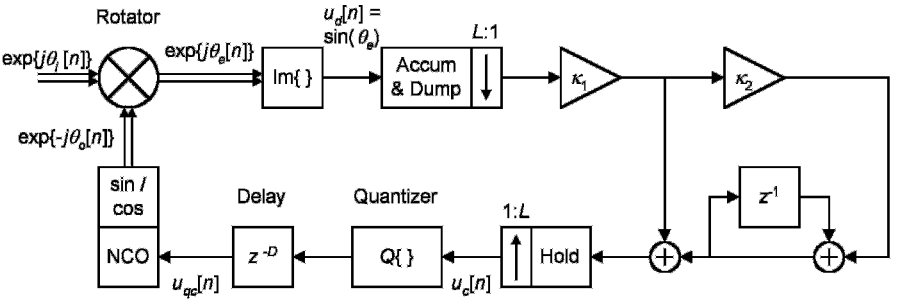


Figure 13.9 Simulation model of digital PLL for studies of frequency quantization.

Input Signal The input signal is modeled as a unit-amplitude unmodulated complex exponential $\exp\{j\theta_i[n]\}$, where the input phase is defined by

$$\theta_i[n] = 2\pi n f_i t_s + \theta_i[0] \tag{13.8}$$

where t_s is the sample interval and f_i is the frequency of the sampled signal. Define a normalized frequency $f_i t_s$ with units of cycles per sample interval. The analysis assumes that $|f_i t_s| < 0.5$, implying that θ_i cannot change by more than π from one sample to the next. Negative frequencies are just as valid as positive frequencies in a complex signal. Complex white zero-mean Gaussian noise was added to the signal for some simulation trials as reported below. The in-phase and quadrature components of the noise each had variance σ_v^2.

Phase Detector The phase rotator of Fig. 13.9 plus the box labeled Im{·} (designating the imaginary part) constitute a phase detector whose s-curve is sinusoidal and whose gain is $\kappa_d = 1$ rad^{-1}. As explained in Section 13.1.3, a complex-signal phase detector is free of ripple in its output.

Loop Filter and Delay The accumulate & dump, the proportional-path coefficient κ_1, the integral path with coefficient κ_2 and an integrator, and the hold operation are substantially the same as in Fig. 13.7. All delay in the loop has been lumped into a single box with delay $D \geq 1$.

Quantizer The quantizer was modeled with 2^b uniform increments, $b =$ positive integer, subject to the quantizing rule

$$Q(u_c) = u_{qc} = \text{IP}[2^b u_c], \qquad |u_c| \leq 0.5 \tag{13.9}$$

where u_c is the output from the hold element. The ends of the range are irrelevant to this discussion since no simulated signals u_c approached them at all closely.

These definitions provide a quantizer with a riser at $u_c = 0$ and a zero-valued output in the interval $u_c \in [0, 2^{-b})$. In hindsight, it might have been preferable to have avoided a zero-valued region, but the simulation difficulties it raised

were circumvented by judicious selection of signal frequency $f_i t_s$. Since the zero value is in the NCO tuning characteristic and not the PD s-curve, its presence has no essential effect on the nature of the limit cycles described below; it does not constitute a dead zone in the error detector. It was convenient to further normalize the signal frequency to the quantization interval according to

$$\mu_i = 2^b f_i t_s \tag{13.10}$$

NCO and the Sine/Cosine Process The operations of these elements are the same as described previously, the sole difference being the quantization of the frequency-control word u_{qc} delivered to the NCO so that the NCO can take on only 2^b discrete frequencies. All other quantities in the simulation were represented in floating-point numbers, whose granularity was minuscule compared to 2^{-b}. On the same normalized frequency scale (13.10) that defines μ_i, the NCO can run at any integer frequency μ_o from -2^{b-1} to $2^{b-1} - 1$. By contrast, μ_i takes on values from a continuum $(-0.5, 0.5)$. The frequency increment from the quantized NCO is $\delta f_o = (2^b t_s)^{-1}$, a relation useful for evaluating the performance of PLLs whose frequency is quantized by oscillators other than NCOs.

Noise-Free Limit Cycles Unless the normalized signal frequency μ_i is an integer (an event of zero probability if the received signal and the PLL clock are generated by independent oscillators), the NCO frequency μ_o can never be the same as μ_i. Therefore, the NCO cannot settle to a single frequency and still maintain phase lock. Simulations revealed the existence of limit cycles (more correctly, quasiperiodic orbits) in the phase error θ_e, the loop-filter output u_c, the quantizer output u_{qc}, and the NCO frequency μ_o.

Two example plots of limit-cycle waveforms (out of many hundreds generated in simulations) are shown in Figs. 13.10 and 13.11. Abscissas are labeled with sample numbers. Denoting the output of the integrator in the loop filter of a type 2 PLL by $u_I[n]$, the quantity X2 in the plots is defined by $\mathrm{X2} = 2^b u_I[n]$. In all waveform plots, be aware that the simulation program draws straight lines between adjacent samples for visual separation of closely spaced points, but all points really are discrete. Furthermore, jagged connecting lines are an artifact of the pixel raster on the computer screen and not a feature of the limit cycles.

The limit cycles of a quantized PLL exhibit an astonishing variety of features that do not appear in operation of a nonquantized PLL. In the absence of additive noise, prominent attributes of the steady-state limit cycles include the following:

- Waveforms are strongly dependent on FP$[\mu_i]$, the fractional part of the normalized signal frequency.
- For a type 2 PLL (which exhibits zero static phase error), the same limit cycle waveforms appear for all frequencies with the same FP$[\mu_i]$, irrespective of IP$[\mu_i]$.
- If FP$[\mu_i]$ is a ratio p/q (p and q relatively prime integers), the limit cycles are periodic in q; the same sequence of sample values repeats exactly. For

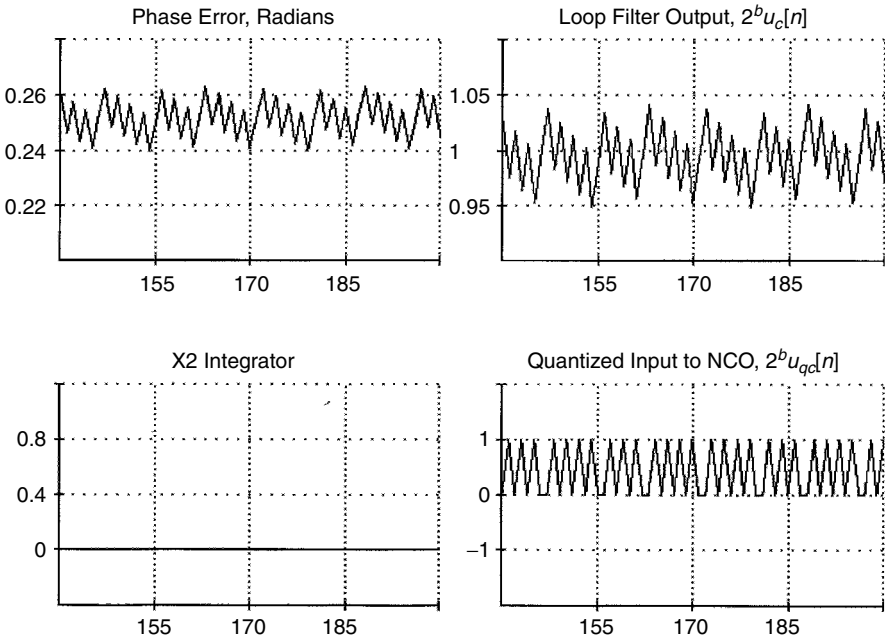


Figure 13.10 Limit-cycle waveforms: first-order DPLL, $\mu_i = 0.44, b = 8, D = 1, L = 1$.

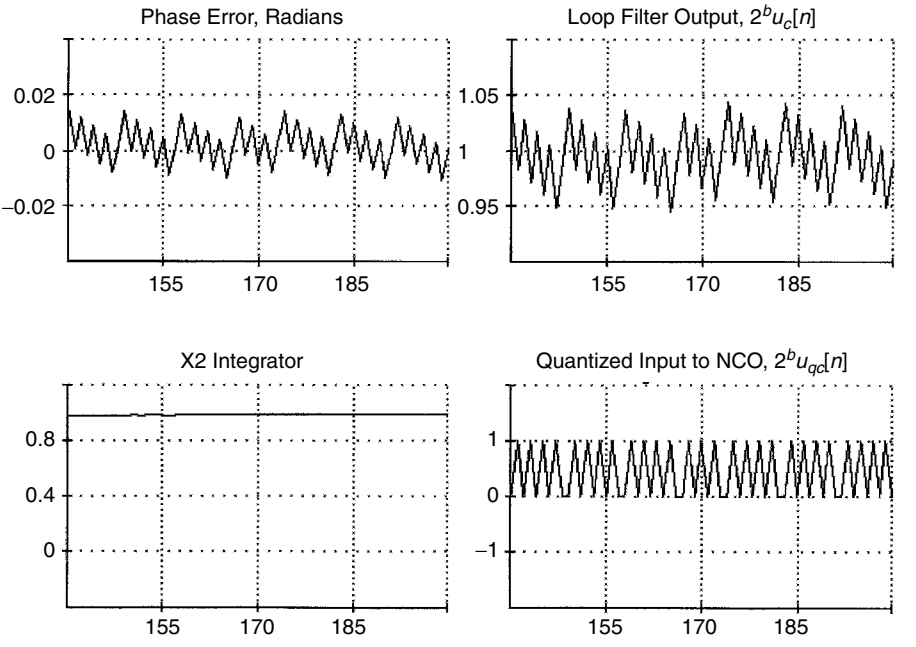


Figure 13.11 Limit-cycle waveforms: type 2 DPLL, $\mu_i = 0.44$, $b = 8$, $D = 1$, $L = 1$.

example, in Figs. 13.10 and 13.11, where $\mu_i = 0.44 = 11/25$, a period 25 can be discerned.

- If FP[μ_i] is irrational, the limit cycles cannot be periodic; they never repeat exactly.
- If a limit cycle is periodic, its spectrum cannot be white. The spectrum must consist of discrete harmonics of the fundamental frequency of the limit cycle. This fact vitiates the common assumption that quantizing noise has a flat spectrum.
- For loop parameters of practical interest (adequate damping—meaning small-enough κ_2—and comfortable stability margin) and $\mu_i \neq$ integer, the quantized NCO frequency jumps only between the two values IP[μ_i] and 1 + IP[μ_i]. Feedback action in the PLL adjusts the relative dwell times at these two NCO frequencies, so that the average NCO frequency is exactly μ_i, thereby permitting phaselock despite the fact that the NCO cannot run at a frequency $\mu_i \neq$ integer.
- In Figs. 13.10 and 13.11, the output of the loop filter remains close to $2^b u_c = 1 + \text{IP}[0.44] = 1$ throughout the equilibrium limit cycle. That level is provided either from nonzero PD output from static phase error ($\approx$0.25 rad in the example) propagating through the proportional path of the type 1 PLL of Fig. 13.10 or from the integrator in the loop filter of the type 2 PLL of Fig. 13.11. In the absence of quantization, the average filter output would 0.44 exactly, not $\sim$1.
- The phase-error limit cycles in Figs. 13.10 and 13.11 have the same waveshape except for the offset due to static phase error in the first-order PLL of Fig. 13.10. Moreover, changing κ_1 and b has no influence on the waveshape of the limit cycle (other than amplitudes of the excursions and altered static phase error in the first-order PLL), even for quantization as coarse as $b = 1$. Changing κ_2 has no effect either, provided that κ_2 is small enough.
- If gain κ_2 of the integral path is small enough, response speed in the integrator path is very slow, so slow that it is unable to follow the limit cycle in the proportional path. As far as the steady-state limit cycle is concerned, the type 2 PLL appears nearly indistinguishable from a type 1 PLL, with the correct offset applied externally as a bias to the quantizer.
- Under most practical conditions, the peak-to-peak excursions of the phase error were found to be approximated by

$$\frac{2\pi(D + L - 1)}{2^b} \quad \text{rad} \tag{13.11}$$

- Notice that (13.11) omits the gain coefficients κ_1 and κ_2; the phase excursions due to NCO quantization are independent of loop gain. Independence of κ_1 and of small-enough values of κ_2 were verified in simulations. Large values of κ_2 have an adverse effect, as discussed subsequently.
- Observation of phase-error limit cycles revealed waveforms composed of sawtooth sections (e.g., as visible in Figs. 13.10 and 13.11), which suggested

uniform distribution of amplitude. If the distribution is indeed uniform within the bounds of (13.11), the phase-error variance due to quantization of the NCO would be

$$\sigma_{\theta e}^2 = \frac{1}{12}\left[\frac{2\pi(D+L-1)}{2^b}\right]^2 \qquad \text{rad}^2 \tag{13.12}$$

- Evaluation over many trials found that simulated $\sigma_{\theta e}^2$ agreed well with (13.12) for $D = 1$ and $L = 1$, but varied from (13.12) by 2 : 1, or even somewhat more, in either direction for larger D or L. A heuristic argument justifying (13.11) is proposed further below.
- Evaluation of the phase excursions by experiment is statistically dubious because of sampling inadequacies. Teplinsky and Feely [13.33] develop a more sophisticated analysis via nonlinear mathematics (but only for $L = 1$ and $D = 1$) and conclude that the peak-to-peak excursion of phase error never exceeds $4\pi/2^b$ rad, and that only for integer μ_i. They further conclude that (13.11) is correct for all irrational μ_i and approaches $2\pi(1 + 1/q)/2^b$ radians if $\mu_i = p/q$, provided that κ_2 is small enough.
- If μ_i is rational (and noise is absent), the steady-state limit cycle in a first-order PLL repeats its sample values exactly over every period. In consequence, the properties of the limit cycle—such as phase excursions and static phase error—depend on the initial conditions; different limit cycles arise for different initial conditions.
- If μ_i is irrational, the equilibrium limit cycle never repeats exactly and the properties of the limit cycle do not depend on initial conditions.

Integer Frequencies The limit cycles (if any) for $\mu_i =$ integer are quite different from those described so far. Simulations have shown that PLL behavior is atypical and degenerate with an integer frequency. Comments below are for noise-free conditions. In a first-order PLL, if μ_i is an integer and within the lock-in range of the loop, the phase error ultimately freezes at a value that sets $\mu_o = \mu_i$ and no further changes occur—there is no limit cycle. Notice that because of the quantization, there is an entire range of phase errors that will set $\mu_o = \mu_i$; the loop freezes as soon as the phase error enters that range.

In a type 2 PLL, the phase error cannot freeze at a nonzero value because the integrator in the loop filter can hold a constant output only for a zero input. An illustrative limit cycle for $\mu_i = 0$ in a type 2 PLL is shown in Fig. 13.12. The illustrated behavior is instructive for any integer μ_i. If the PLL were started with initial conditions $\theta_e[0] = 0$ and $0 \le 2^b u_I[0] = X2 < 1$, it would simply freeze in its initial state, with no limit cycle. However, a nonzero initial phase error starts the PLL off into vigorous activity, as exemplified in Fig. 13.12. Phase error during the limit cycle takes on only two distinct values, neither one of which is zero. [**Exception**: If $\theta_e[0]$ were an exact integer multiple of $2\pi/2^b$, the phase error would eventually settle to zero in a transient that terminated in finite time.] Because phase error is never zero, the integrator output must always be

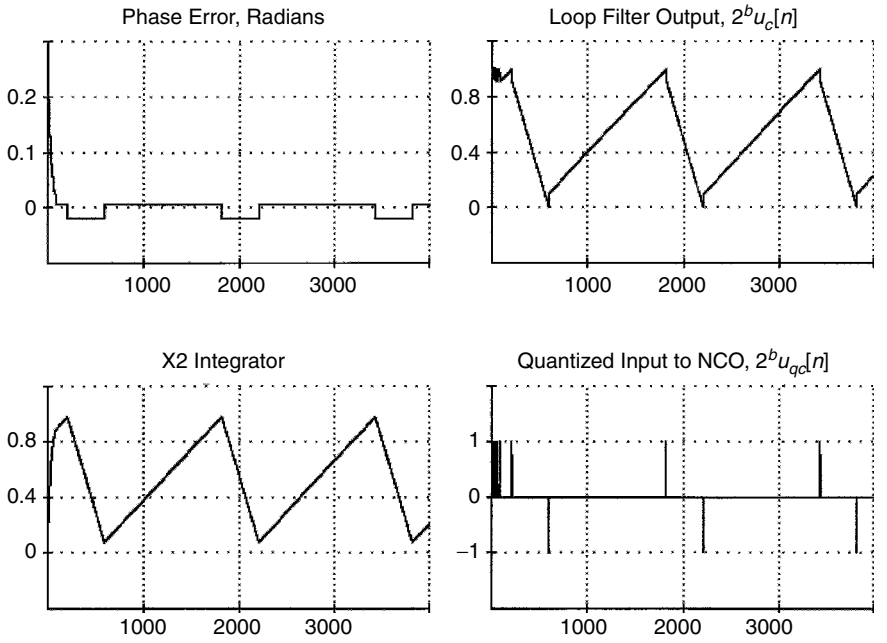


Figure 13.12 Limit-cycle waveforms: type 2 DPLL, $\mu_i = 0$, $b = 8$, $D = 1$, $L = 1$.

slewing, never standing still. Different initial conditions would lead to different details in the limit cycle.

Despite the extreme activity of the integrator, the frequency of the NCO $\mu_o = 0 = \mu_i$ for all but brief instants. NCO frequency jumps to $+1$ or -1 for those brief instants as output of the loop filter passes over the corresponding boundaries of the quantizer. The jumps would be to $\mu_o = \mu_i \pm 1$ for integer $\mu_i \neq 0$. Contrary to the limit cycles for noninteger μ_i, where μ_o jumped only between the two frequencies IP(μ_i) and $1 + \text{IP}(\mu_i)$, μ_o jumps among three frequencies for integer μ_i. The simulated PLL for Fig. 13.12 is identical to the one for Fig. 13.11; only the signal frequency has changed.

The limit cycles in Figs. 13.10 and 13.11 are maintained primarily through the proportional path of the PLL, whereas the limit cycles in Fig. 13.12 are dominated by the integral path. It is useful to think of two intertwined loops in a type 2 PLL, each subject to its own limit cycle because of quantization. The notion of intertwined loops is exploited subsequently.

Effect of Accumulation and Delay The foregoing applies only to $L = 1$ (no accumulation and down-sampling) and $D = 1$ (no excess delay in loop). Gardner [13.31] reports simulation results for $L > 1$ and $D > 1$; analyses in [13.32] and [13.33] were restricted to $L = 1$ and $D = 1$. This section describes some of the simulation results as shown in Figs. 13.13 and 13.14. In both figures, results are for a first-order PLL on the premise that the limit cycle will be similar

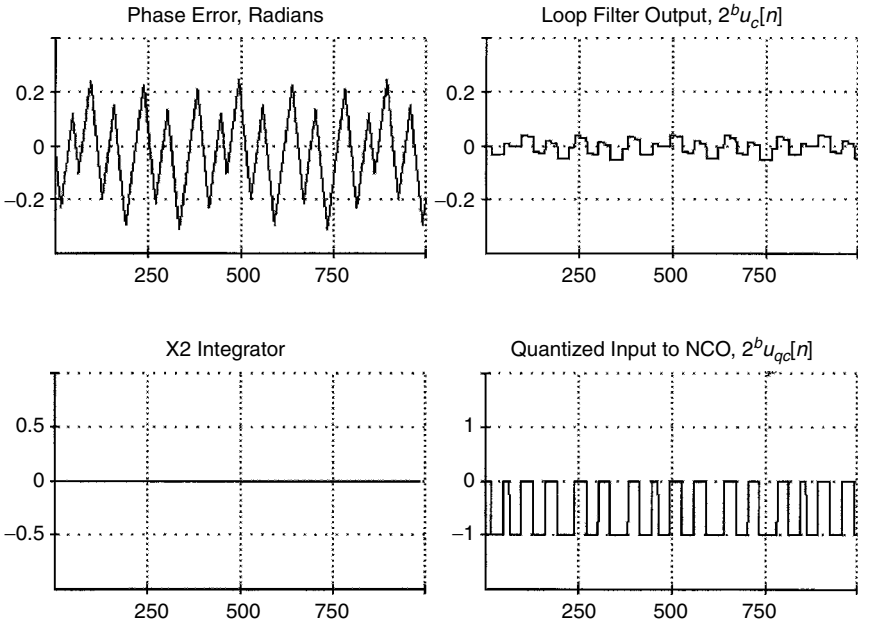


Figure 13.13 Limit-cycle waveforms: first-order DPLL: $\mu_i = -0.56$, $b = 8$, $D = 1$, $L = 16$.

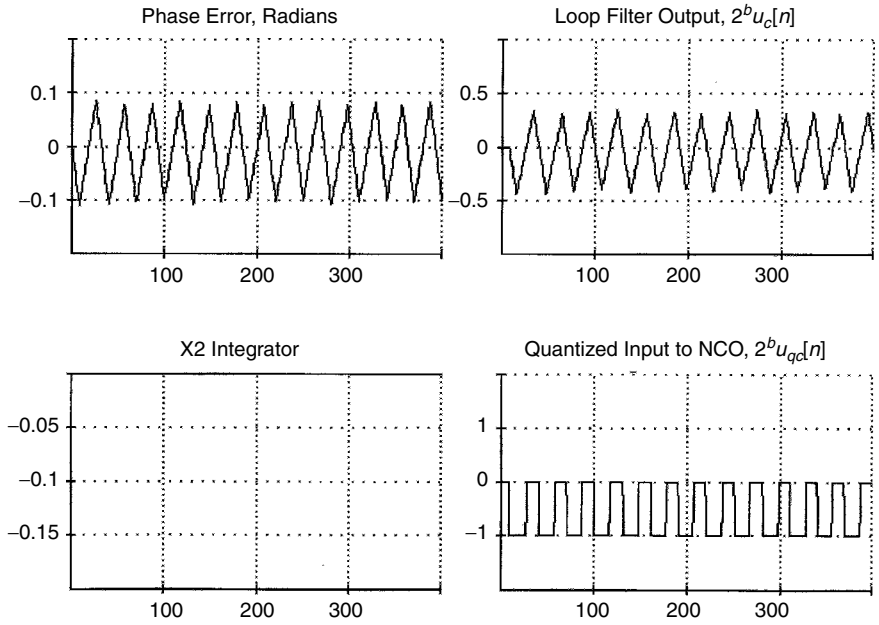


Figure 13.14 Limit-cycle waveforms: first-order DPLL, $\mu_i = -0.56$, $b = 8$, $D = 8$, $L = 1$.

for type 1 and type 2 PLLs, except for integer input frequencies. A negative frequency $\mu_i = -0.56 = -1 + 0.44 = -14/25$ was employed to avoid problems in the simulator with static phase error. Except for static phase error, $\mu_i = -0.56$ is equivalent to $\mu_i = +0.44$ as employed in Figs. 13.10, to which Figs. 13.13 and 13.14 should be compared.

Figure 13.13, $L > 1$ Since accumulation ratio L is a factor in the loop gain κ [see (13A.18)], and since the loop update rate is $1/Lt_s$, the setting for simulated gain κ_1 in the proportional path was reduced by a factor $1/L^2$ compared to its value in Fig. 13.10. This expedient first compensates for the factor L in the loop gain, and second, reduces the loop bandwidth by a factor $1/L$, as is appropriate for the reduced update rate. It also maintains the same stability margin in both instances ($\sim$26 dB). Delay $D = 1$ is used for this example.

The plots of Fig. 13.13 for $L = 16$ show a large increase in phase-error excursion compared to Fig. 13.10, a much smaller static phase error (due to the choice of signal frequency), and substantial changes in all waveforms. Careful inspection shows the period of the phase-error limit cycle to have increased to $25 \times 16 = 400$ samples and its peak-to-peak excursion to have increased approximately 16-fold as well. Normalized frequency of the NCO now jumps between 0 and -1, as is mandated by the choice of a negative input frequency, but the jumps are still only between the two levels, despite the large increase of L. NCO frequency remains steady for 16 sample intervals because of the down-sampling and subsequent up-sampling with the hold operation. Therefore, each phase-error excursion necessarily is 16 times larger than for $L = 1$. For the same reason, the period of the limit cycle is 16 times larger.

Figure 13.14, $D > 1$ Because delay D does not enter into the loop gain or the update rate, the same gain κ_1 in the proportional path was chosen for the simulated PLL of Fig. 13.14 as was used for the PLL of Fig. 13.10. With $D = 8$, delay reduced the stability margin to 6.6 dB, (from 26 dB), probably rather lower than desirable. Accumulation ratio was set at $L = 1$ for this example. The plots of Fig. 13.14 show greatly changed waveforms from Fig. 13.10, a peak-to-peak phase error excursion of about eightfold larger, and an NCO frequency still jumping between only two levels: 0 and -1 for the conditions in Fig. 13.14.

Derivation of Phase Excursion Consider a first-order PLL with $D = 1$ and $L = 1$. Postulate a phase-error threshold Θ_E defined by the property

$$2^b u_{qc} = \begin{cases} \mathrm{IP}(\mu_i), & \theta_e < \Theta_E \\ 1 + \mathrm{IP}(\mu_i), & \theta_e > \Theta_E \end{cases} \tag{13.13}$$

If $2^b u_{qc} = \mathrm{IP}(\mu_i)$, the phase error will advance by $\Delta\theta_+ = 2\pi \mathrm{FP}(\mu_i)/2^b$ radians over the next sample interval, whereas the phase error will fall back by

$\Delta\theta_- = 2\pi[\mathrm{FP}(\mu_i) - 1]/2^b$ radians if $2^b u_{qc} = 1 + \mathrm{IP}(\mu_i)$. Starting from $\theta_e = \Theta_E - \varepsilon(\varepsilon \to 0)$, the greatest possible positive excursion is to $\Theta_E + \Delta\theta_+$, since only negative excursions are possible if $\theta_e > \Theta_E$. Similarly, starting from $\theta_e = \Theta_E + \varepsilon$, the greatest possible negative excursion is $\Theta_E + \Delta\theta_-$. Therefore, the maximum possible peak-to-peak excursion of θ_e is

$$\begin{aligned}\Delta\theta_e &\le (\Theta_E + \Delta\theta_+) - (\Theta_E + \Delta\theta_-) = \Delta\theta_+ - \Delta\theta_- \\ &= \frac{2\pi}{2^b} \qquad \text{rad} \end{aligned} \tag{13.14}$$

irrespective of Θ_E.

From the analyses of [13.32] and [13.33], it is apparent that (13.14) is not a true upper bound on the peak-to-peak phase excursions for all signal frequencies. Defective as it may be, its derivation still offers a description of the physical process that generates the excursion. Now incorporate the simulation findings for $L > 1$ and $D > 1$. The examples presented in Figs. 13.13 and 13.14 demonstrated peak-to-peak phase-error excursions that were proportional to L and to D. As a rough hypothesis, assume that delay is inserted into the loop by either delay or by accumulation and subsequent hold. Inserted delay is the sum from both sources. A delay of D obviously contributes D sample intervals of loop delay, but the accumulation and hold contribute only $L - 1$ intervals since there is no delay for $L = 1$: that is, no accumulate and hold operations. The phase of the NCO continues to increase (decrease) by $2\pi/2^b$ rad for $(D + L - 1)$ sample intervals after the Θ_E boundary has been passed, thereby increasing the excursion by that amount. Applying this heuristic reasoning to (13.14) gives (13.11).

Effect of Gain Coefficients κ_1 and κ_2 Teplinsky et al. [13.32] and [13.33] state that the peak-to-peak excursion of the phase-error limit cycle is independent of κ_1 (provided that the loop is stable) but does depend on κ_2. The simulations of [13.31] found the same behavior; examples from the simulation study provide some insight. Figures 13.15 and 13.16 show limit cycles for two type 2 PLLs that are identical except for the values of the integral-path gain coefficients κ_2. These simulations were run for $L = 16$, but that is incidental and irrelevant. They were selected because they showed the κ_2 effect most clearly of all trial runs. Other simulations with $L = 1$ also showed the effect, but not nearly as forcefully.

Prior to these illustrated simulations with a type 2 PLL, another simulation was run with a first-order PLL (i.e., $\kappa_2 = 0$) with the same $L = 16$, the same μ_i, and $\kappa_1 = 2^{-11}$ instead of 2^{-14} as in Figs. 13.15 and 13.16. Its phase-error limit cycle was identical (except for a small static phase error) to that of Fig. 13.15. Several conclusions can be drawn from these results: (1) loop type has no influence on the waveform of the phase-error limit cycle (except for static phase error), (2) proportional-path gain (in a stable loop) has no influence on the waveform of the phase-error limit cycle, and (3) the value of $\kappa_2 = 2^{-5}$ in Fig. 13.15 is small enough to have no discernible effect on the waveform or excursion of the phase-error limit cycle.

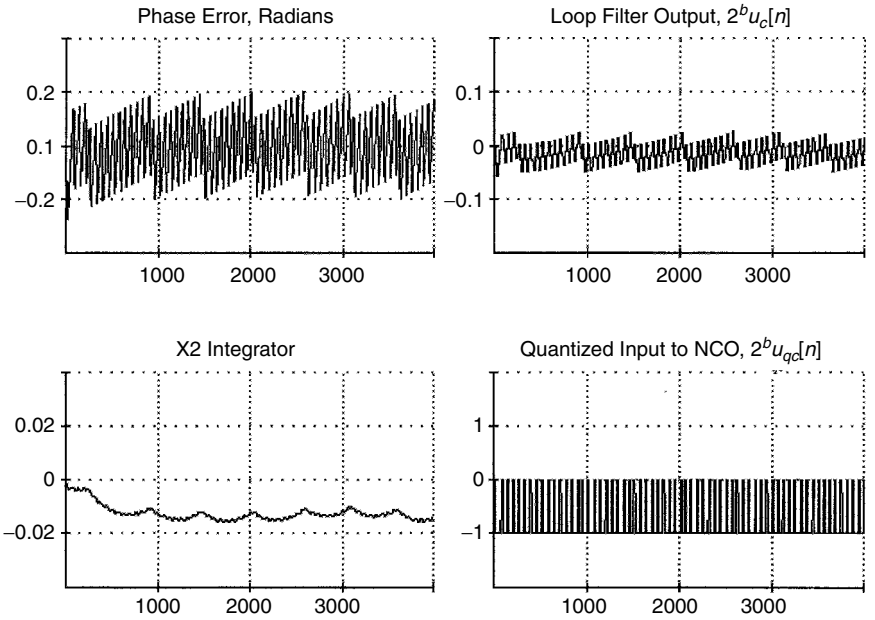


Figure 13.15 Limit-cycle waveforms: type 2 DPLL: $\mu_i = -0.7425$, $b = 8$, $D = 1$, $L = 16$, $\kappa_1 = 2^{-14}$, $\kappa_2 = 2^{-5}$.

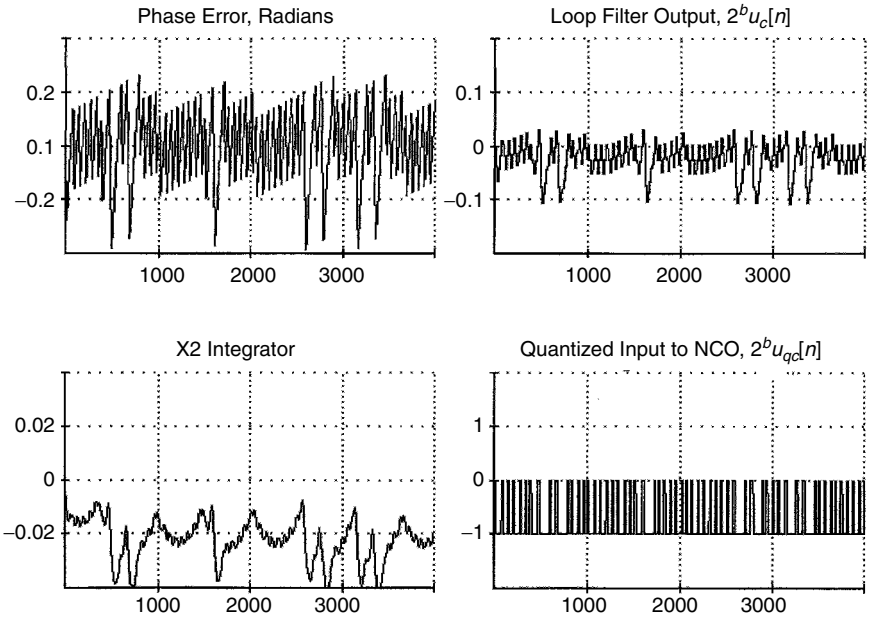


Figure 13.16 Limit-cycle waveforms: type 2 DPLL: $\mu_i = -0.7425$, $b = 8$, $D = 1$, $L = 16$, $\kappa_1 = 2^{-14}$, $\kappa_2 = 2^{-3}$.

Several features may be noted from examining Fig. 13.15:

- The peak-to-peak phase-error excursion conforms closely to (13.11).
- The peaks lie equally above and below zero (once the initial transient dies out), indicating the success of the integrator in removing static phase error.
- Quantized input to the NCO jumps between only two levels.
- The output of the integrator is always slightly negative from zero.
- The waveform of the integrator shows small-amplitude scalloping that corresponds to the lowest-frequency (period = 512?) variations in the phase-error waveform.

Figure 13.16 shows the limit cycles for the same PLL under the same conditions as in Fig. 13.15 except that κ_2 has been increased from 2^{-5} to 2^{-3}. Remnants of the limit cycles from Fig. 13.15 can still be seen. The phase-error remnants appear to be just about the same as for the smaller κ_2 and the larger amplitude of the scalloping remnants in the integrator is explained entirely by the larger κ_2.

But the most striking feature of Fig. 13.16 is the emergence of large departures from the comparatively regular limit cycles of Fig. 13.15. It is as though another, more vigorous limit cycle associated with the integral path of the loop filter is breaking through and overcoming the more sedate limit cycle associated with the proportional path. This explanation may be overly dramatic—microscopic examination of the waveforms undoubtedly would reveal a mundane explanation in terms of switching levels. Nonetheless, the notion of contending limit cycles provides a useful model for thinking about several other phenomena besides the effect of κ_2. The idea has already arisen in explaining the large limit cycles in type 2 PLLs for integer μ_i; a variation on the idea will appear in explaining the effect of additive noise on limit cycles.

Comparison of Figs. 13.15 and 13.16 suggests the existence of a threshold value (or threshold range) of κ_2. For values well below the threshold, the phase-error limit cycle is dominated by the proportional path, to the near (maybe complete) exclusion of influence from the integral path. For κ_2 values well above the threshold, the integral path dominates. Theory is not yet developed sufficiently to specify the threshold; simulations still are necessary.

Effect of Additive Noise In the absence of quantization, the phase variance at the output of the NCO in a DPLL in response to an input consisting of the sum of a unit-amplitude complex-exponential signal plus complex white Gaussian noise of variance $2\sigma_v^2$ (σ_v^2 in each rectangular component of the complex noise) is

$$\sigma_{\theta no}^2 = 2B_L t_s \sigma_v^2 \qquad \text{rad}^2 \tag{13.15}$$

where B_L is the noise bandwidth as defined in Chapter 6 and t_s is the sample interval.

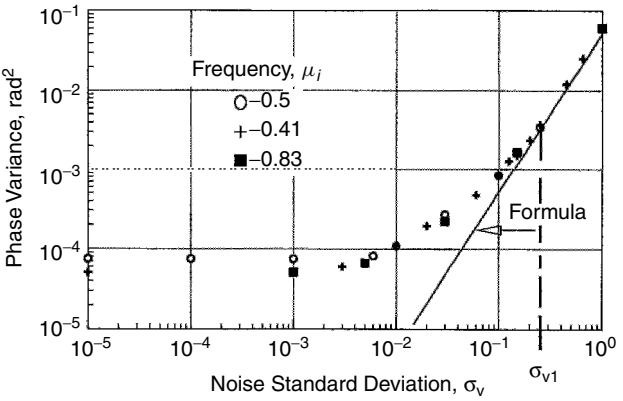


Figure 13.17 Phase error variance vs. noise: $b = 8$, $D = 1$, $L = 1$, $\kappa_1 = 2^{-6}$, $\kappa_2 = 0$. (From [13.31; © 1996 IEEE.)

Phase fluctuations due to frequency quantization join the fluctuations due to additive noise. Figure 13.17 shows a few examples of simulated phase variance resulting from combined quantization and additive noise. For small σ_v, the quantization component dominates, as is evident from the flatness of the plots. For large σ_v, the additive noise dominates, closely approaching the theoretical formula (13.15) (solid straight line labeled "formula"). These results are typical of many trials with a variety of conditions of noise, signal frequency, and PLL parameters.

An attempt was made to deduce an empirical formula for the total variance in the transition region where noise and quantization contributions are roughly equal. Simple addition of the individual variances (of noise alone and of quantization alone) or of the individual standard deviations gave a very poor match to the observed total variances. After the fortuitous discovery of Fig. 13.18, in which the total variance *decreases* with increasing σ_v in the transition region (a behavior induced by near-instability in the particular PLL), it was realized that the additive noise and the quantization limit cycle combine in nonlinear fashion and that no simple addition rule could be correct.

Observation of limit-cycle waveforms revealed that additive noise tends to disrupt the limit cycle rather than add to it. A typical waveform might have segments in which the noise-free limit cycle appears unimpaired, followed by segments in which the limit cycle has been destroyed and replaced by noise. Two processes are contending with one another in the nonlinear system, and no simple characterization is evident.

As a substitute for a combined rule for variance, a simple rule of thumb was devised for an input noise level $\sigma_v = \sigma_{v1}$ such that the contribution of additive noise would predominate over that of quantization. The rule of thumb was based on the observation that in almost all practical cases, a noise-free limit cycle out of the quantizer jumps back and forth between only two adjacent quantization

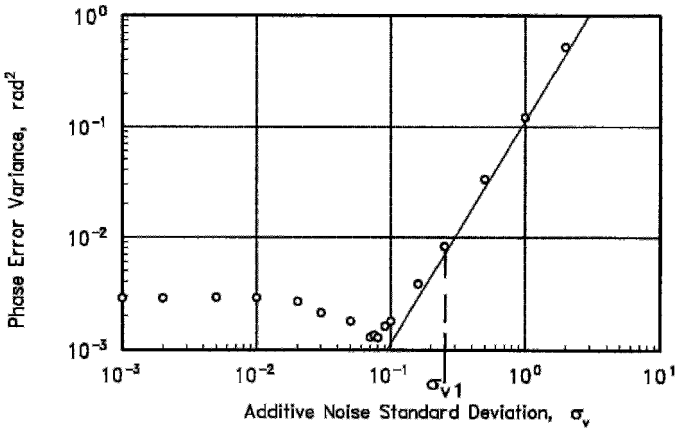


Figure 13.18 Phase error variance vs. noise: $\mu_i = -0.41$, $b = 8$, $D = 8$, $L = 1$, $\kappa_1 = 2^{-6}$, $\kappa_2 = 0$. (From [13.31]; © 1996 IEEE.)

levels. It was further observed that the input $u_c[n]$ to the quantizer had a limit cycle with peak-to-peak excursions much smaller than one quantizing interval in most instances. The rule of thumb arbitrarily defines σ_{v1} as the input-noise level that produces a noise component at the input to the quantizer with standard deviation equal to one quantizing interval.

A formula for σ_{v1} is obtained as follows. Only the quadrature component of input noise, with variance σ_v^2, need be considered since the in-phase component does not appear in the PD output (see Section 6.1.1). Assume, reasonably, that transmission of the noise through the loop filter is mostly through the proportional path so that transmission through the integral path can be neglected entirely for the rule. Thus, the loop filter scales the input noise by a factor κ_1.

An accumulate and dump—that is, $L > 1$—complicates the derivation. The transfer functions of Appendix 13A were derived with the approximation that the input-signal bandwidth is small compared to the filter update rate $1/Lt_s$. But that approximation is not valid for input noise that is white over a two-sided bandwidth $1/t_s$. To take account of the accumulator, recognize that the accumulator just adds L samples of the noise. If these L samples are independent and identically distributed with standard deviation σ_v, the standard deviation out of the accumulator is simply $\sigma_v L^{1/2}$ and the standard deviation at the input to the quantizer is $\kappa_1 \sigma_v L^{1/2}$. Setting that product equal to 2^{-b} (one quantizer increment) gives the rule of thumb

$$\sigma_{v1} = \frac{1}{2^b \kappa_1 \sqrt{L}} \tag{13.16}$$

Markers for σ_{v1} in Figs. 13.17 and 13.18 show that the total phase variance is well past the transition region between quantization dominance and noise dominance and is very close to the formula of (13.15) at $\sigma_v = \sigma_{v1}$. Observations of waveforms with $\sigma_v = \sigma_{v1}$ reveal that the majority of noisy samples of the quantizer

output occur at the two levels that are activated in the absence of noise, a substantial portion of the samples reach one further level in each direction, fewer extend two levels further, and only a very few reach three levels further. This is exactly the sort of quantizer output to be expected from noise standard deviation of one quantizing interval.

Static Phase Error A first-order digital PLL suffers from a static phase error from the necessity to have a nonzero output from the loop filter to tune the NCO to its correct average frequency. Excessive phase error causes the loop to lose lock; there is a hold-in limit, just as in an analog PLL. An example of static phase error is shown in Fig. 13.19 for noise-free and noisy conditions. Formulas for the phase error were inferred from observation of numerous simulation trials. In the simulation model of Fig. 13.9, the formula in the absence of noise is

$$\theta_v = \sin^{-1}\frac{\mathrm{IP}(\mu_i)+1}{2^b L\kappa_1} + \frac{2\pi(D+L-1)}{2^b}[\mathrm{FP}(\mu_i)-0.5] \qquad \text{rad} \qquad (13.17)$$

which vanishes only at $\mu_i = 0$. The inverse sine term in (13.17) is responsible for the stepwise character apparent in the noise-free portion of Fig. 13.19, and the second term imparts a tilt that is just barely discernible on each step top. Also visible in expanded plots, but invisible in Fig. 13.19, are relatively small, frequency-dependent deviations from the straight-line step tops indicated by (13.17).

With sufficient added noise, the steps are wiped out and the static phase error approaches

$$\theta_v = \sin^{-1}\frac{\mu_i + 0.5}{2^b L\kappa_1} \qquad \text{rad} \qquad (13.18)$$

a result nearly the same as that in an analog PLL with sinusoidal PD.

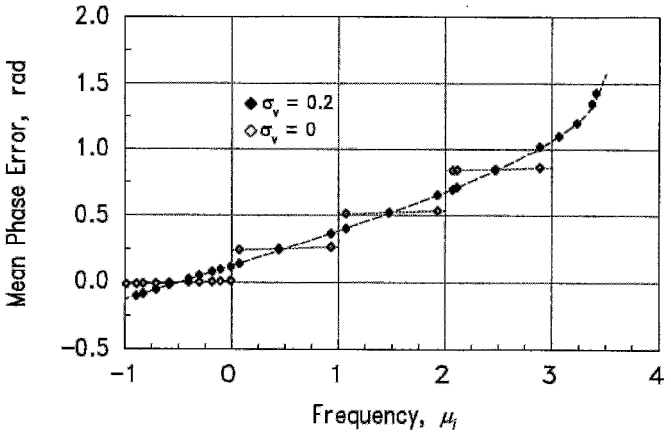


Figure 13.19 Static phase error: $b = 8$, $D = 1$, $L = 1$, $\kappa_1 = 2^{-6}$, $\kappa_2 = 0$. (From [13.31]; © 1996 IEEE.)

Design Rules Several guidelines to minimize phase-error excursions can be extracted from all of the foregoing:

- Use small quantizing increments (large b) in the frequency quantizer (an obvious expedient).
- Use a type 2 PLL to avoid static phase error (an expedient well known from analog PLLs).
- Avoid accumulation and down-sampling in the loop filter.
- Minimize delay D in the loop.

13.2.4 Quantization in a Phase Detector and an Integrator

Despite the rich collection of strange behaviors recounted in the preceding section, much remains to be learned about the nonlinear effects of quantization of NCO frequency. Operation of every other element in a digital PLL is also quantized, and the effects of those quantizations also need to be understood. Nonetheless, excepting Da Dalt [13.43], frequency quantization appears to be the only one that has been studied in any depth so far. This section raises preliminary ideas and questions on quantization in the phase detector and in the integrator of the loop filter. Definitive treatment lies in the future.

Phase-Detector Quantization Consider a first-order DPLL. Output of its phase detector is $u_d[n]$ and the control signal to the NCO is $u_c[n] = \kappa_1 u_d[n]$. Assume plausibly that the PD is arranged to avoid a dead zone and that multiplication by κ_1 preserves all bits of u_d for u_c. If the LSB of u_c aligns with the LSB of the NCO, the PD quantization matches that of the NCO. With that condition, it seems as though the PLL quantization effectively is that of the NCO, with behavior as described at length in Section 13.2.3. This conclusion may be wrong; intuition insists that quantization of u_d has to have some effect. The issue is open for further investigation.

One or more bits will be wasted if the LSB of u_c falls below the LSB of the NCO. Performance is the same as if the lost bits were truncated from u_c. There is no obvious benefit from carrying extra bits in the PD. The loop fails entirely if the MSB of u_c (the sign bit) falls below the LSB of the NCO. This feature implies that the quantization of NCO frequency in some sense establishes a minimum bandwidth of its PLL, at least for the DCO configuration that has been considered so far. If the NCO indeed is arranged with a sign bit as part of its input, the sign bit from the control word always must align with the NCO sign bit, separately, if necessary, from alignment of the rest of the control word.

Integrator Quantization Now consider a type 2 PLL that includes an integral path in the loop filter. To reduce the number of conditions to be examined, assume that the LSB of u_c aligns with the LSB of the NCO input word. The output of

the integrator is

$$u_I[n] = \sum_{k=0}^{n-1} \kappa_1 \kappa_2 u_d[k] + u_I[0]$$

so its smallest increment, if all input bits are retained, would be a factor κ_2 smaller than the smallest increment accepted by the NCO. That raises the question: Should all of those bits be retained, no matter how small κ_2 may be?

Argument for Retaining All Bits Although the LSBs of individual samples of $u_I[n]$ are not included in $u_c[n]$, those LSBs add up over many samples and their sum eventually does enter into $u_c[n]$ and thus the NCO frequency. That sum is useful information. Frequency information stored in the integrator is remembered more accurately in the event of signal dropout if all bits are retained.

Argument Against Retaining All Bits It is difficult to see how bits for frequency increments much smaller than the frequency resolution of the NCO can have much effect on operation of the PLL, particularly for small values of κ_2. Retention of all bits requires a long word in the integrator, imposing a computing and hardware burden.

Compromise Retain some of the extra bits but not all. Rules for retention or discard would be a good subject for investigation.

13.3 IRREMEDIABLY NONLINEAR PLLs

This section deals with a particular family of sampled PLLs whose essential nonlinearities thwart any attempt at approximate linearization. Such concepts as gain, bandwidth, and transfer functions are properties of linear systems, concepts with no meaning in a severely nonlinear system. Analysis of a nonlinear system is much more difficult than that of a linear system and the results of analysis are less comprehensive, as forcefully illustrated by the seemingly uncomplicated PLLs examined in the sections that follow. The example PLLs can be built almost entirely from elements that are readily implemented as standard digital integrated circuits, making them attractive from a hardware standpoint. Characterization gaps due to inadequate analysis have to be mitigated by simulations.

13.3.1 Configuration of a Nonlinear PLL

Attention here is limited to a particular configuration of hybrid PLL, one that performs two-level quantization of its input signal and that alters the phase of its DCO by only one small fixed increment upon each adjustment cycle. In the past, the configuration sometimes has been called an *incremental phase modulator* (IPM); the origin of the name has been lost.

Prior Literature Similar, related configurations have appeared in the earlier literature. All of the authors give the name *digital PLL* (DPLL) to their models, but all models accepted analog signals at their inputs and delivered analog signals from their DCOs, just like the hybrid configuration examined in this section. Cessna and Levy [13.34] found that respectable performance was possible from a PLL with two-level quantization. Like most of the other publications, this one was concerned largely with performance in the presence of additive noise. Holmes [13.35] analyzed a related model in the presence of noise with the help of Markov chains, deriving phase-jitter statistics and mean time to first cycle slip. Ransom and Gupta [13.36] describe a bit-synchronization loop that works on similar principles. D'Andrea and Russo [13.37] describe graphical methods for representing the state trajectories of the nonlinear PLL. Reprint volume [13.38] contains [13.35]. Walker [13.40] examines a highly nonlinear analog PLL that has characteristics partially resembling those examined in this section.

Block Diagram Figure 13.20 depicts one version of an IPM. This version resembles a first-order quasilinear PLL in some properties. A type 2 IPM is introduced after the first-order model has been explored. In the simple model considered here, the input signal (noise-free for now) is a sinusoid with phase $\psi_i(t)$ that is sampled once per cycle by the output of a DCO. Each sample goes through a slicer (also called a limiter or clipper or comparator or one-bit quantizer). Figure 13.20 shows the slicer following the sampler, but it could equally well precede the sampler; operation of the IPM is the same either way. If the sampler is preceded by the slicer, the sampler can be implemented as a D flip-flop, potentially leaving the slicer as the only mixed-signal circuit in the entire PLL.

Binary samples are used to clock an up/down counter whose output $c[n]$ is $+1$ upon overflow, -1 upon underflow, and zero otherwise. Earlier authors have described the counter as a *sequential filter*, but it is argued here that the counter is only a means of establishing an equivalent of loop gain in the PLL. A ring oscillator arrangement similar to Fig. 13.2 is shown for the DCO in Fig. 13.20, but a counter arrangement, as shown in Fig. 13.1, has also been used. Both arrangements operate in the same fashion: a control input $c[n] = +1$ from the

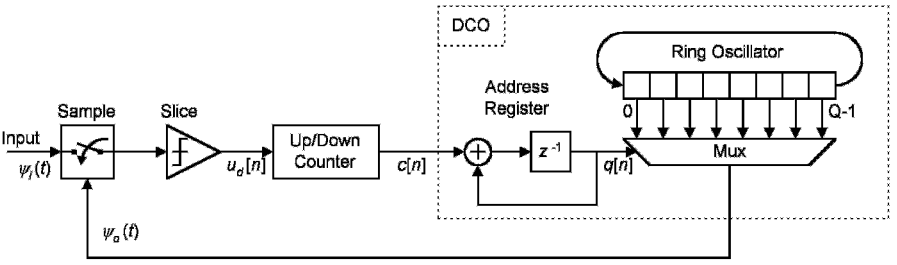


Figure 13.20 Incremental phase modulator (IPM) as an example of a severely nonlinear PLL.

up/down counter causes the phase of the DCO output (the phase of the drive to the sampler) to be advanced by an increment $2\pi/Q$, a control input $c[n] = -1$ causes the phase to be retarded by $2\pi/Q$, and $c[n] = 0$ (often the most prevalent condition) causes the DCO phase to remain unchanged. These statements are amplified subsequently.

Variations on the configuration are possible. A frequency division in the ratio $R:1$ could be inserted after the DCO so that sampling was performed at a frequency $1/R$ of the DCO frequency. Phase increments at the divider output are $1/R$ that of the DCO. Another modification might be *subsampling*: that is, a sampling rate (i.e., DCO frequency) that is a submultiple of the signal frequency. This expedient could relieve problems of circuit speed. The s-curve of a subsampled PD has multiple cycles in one period of the sampling rate.

13.3.2 Operation of the PLL Elements

The sinusoidal time-continuous analog signal at the input to the PLL has a phase expressed as

$$\psi_i(t) = \omega_i t + \theta_i \tag{13.19}$$

where ω_i (rad/sec), the radian frequency of the signal, and θ_i (rad), the phase of the signal, are to be regarded as fixed or only slowly varying.

Phase Detector The nth sample is taken at a time t_n so it has a value $\sin(\omega_i t_n + \theta_i)$. Sampling instants are not uniformly spaced. The output of the slicer is

$$u_d[n] = \text{sgn}\{\sin[\psi_i(t_n)]\} \tag{13.20}$$

which takes on only the two values $+1$ and -1. The output of the slicer is applied to the up/down counter.

Up-Down Counter Operations of the counter are depicted in the state diagrams of Fig. 13.21. The counter counts up by one increment for each $u_d = +1$ and counts down by one increment for each $u_d = -1$. Two different counter arrangements are examined here. One is a single-loop counter and the other is a double-loop counter. They each impart roughly the same gross characteristics to the PLL but have greatly different influences on some of the fine details, especially the phase limit cycles.

The single-loop counter of Fig. 13.21*a* is a simple up-down counter connected in a ring. There are no end states in this counter; the count goes endlessly from one state to the next in either direction. A $+1$ from the PD increases the count by 1 unit to the next-higher-numbered state; a -1 from the PD decreases the count by 1 unit to the next-lower-numbered state; a 0 from the PD causes the counter state to remain unchanged. State numbers are an arbitrary formalism for this model. Since the states exist in a closed ring, no state serves as a natural origin or end. In the figure, states (0) and (6) have been selected as "end" states

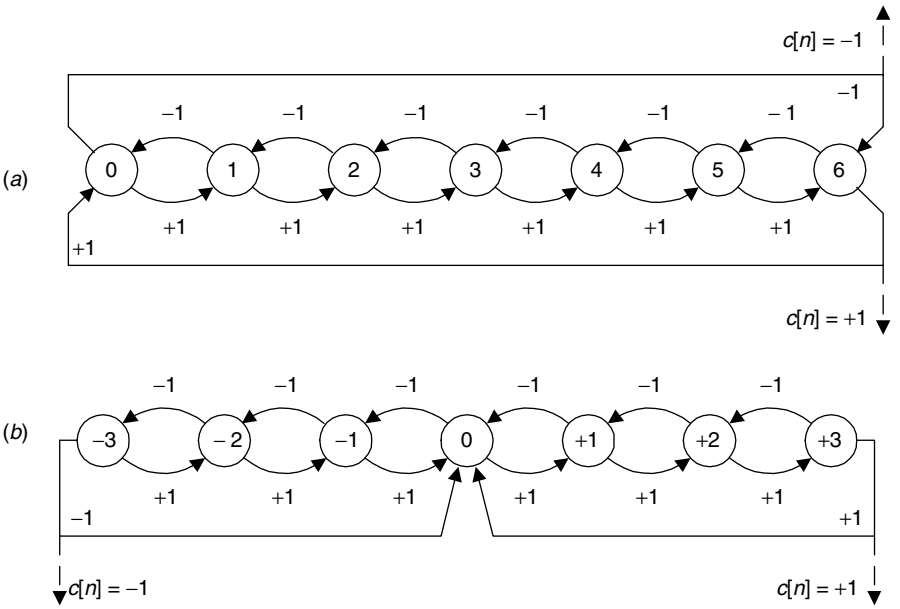


Figure 13.21 State diagrams of two up/down counters: (*a*) single loop; (*b*) double loop.

in the sense that if a $+1$ is delivered from the PD when the counter is in the (6) state, the state will be driven to the (0) state, whereas if a -1 is delivered from the PD when the counter is in the (0) state, the counter will be driven to the (6) state. Any other pair of adjacent states could have been selected instead to achieve identical behavior.

Transitions between the two "end" states generate the counter output $c[n]$. A transition from (6) to (0) generates $c[n] = +1$, a transition from (0) to (6) generates $c[n] = -1$, and any transition between any other pair of adjacent states generates $c[n] = 0$. Notice particularly that $c[n] = \pm 1$ is emitted by the *transitions* between states (0) and (6) and not by occupation of any particular state. A state within the counter is denoted by $s_c[n]$.

The size of the counter is designated by C_i ($i = 1$ for the single-loop counter): the net number of input samples of value $+1$ needed to drive the count around its loop exactly once. The size of the counter in Fig. 13.21*a* is $C_1 = 7$, which happens to be equal to the number of states. It is not generally necessary that C_1 be a power of 2; any positive integer is acceptable.

The more complicated counter of Fig. 13.21*b* has two loops that share a central state denoted (0). As a consequence of establishing a central state, a pair of end states $C_2 - 1$ and $-(C_2 - 1)$ are also defined. Counter *size* C_2 is the net number of input samples needed to drive the count once around a loop. In Fig. 13.21*b* the counter size is $C_2 = 4$, which happens to be the number of states in each loop. Notice that since the (0) state is shared between the two loops, the counter has only seven states total. Notice also that C_2 must be an even positive integer for

a central state to exist. Counter size C_i is one of the key parameters of the PLL. State transitions exist from either end state to (0), but not in the reverse direction. A transition from $C_2 - 1$ to (0) emits $c[n] = +1$, a transition from $-(C_2 - 1)$ to (0) emits $c[n] = -1$, and any other transition emits $c[n] = 0$.

Digit-Controlled Oscillator The DCO of Fig. 13.20 contains a ring oscillator with Q equally spaced taps running at a fixed radian frequency ω_{ck}. In practice, the ring oscillator would be phaselocked to a stable, accurate reference but that is not shown in the figure. A multiplexer selects among the Q taps to adjust the phase of the DCO output. An address register with content $q[n]$ controls the multiplexer to select the tap with number designation q.

[**Comments**: (1) Tap numbering in the figure is such that an increase of q advances the phase of the DCO. (2) Designation of the origin of the ring—that is, the tap assigned as tap-0—is arbitrary; the ring is in a closed circle and has no distinguishable beginning or end. (3) The phase increment from one tap to the next is $2\pi/Q$ radians.]

The address register is part of a recycling accumulator with difference equation

$$q[n] = \{q[n-1] + c[n-1]\} \quad \text{mod-}Q \tag{13.21}$$

where $q \in (0, 1, \ldots, Q-1)$. This accumulator resembles an NCO except that its control inputs $c[n]$ only take on values 0 and ± 1. Instead of an accumulator and address register, the multiplexer tap could be selected by a Q-stage shift register connected in a ring, with one and only one tap true at any instant. The shift register would be shifted forward or backward or held steady by the sequence $\{c[n]\}$. As another alternative, the entire DCO of Fig. 13.20 could be replaced by a count-down arrangement similar to that of Fig. 13.1 with the requisite DCO phase shifts implemented by pulse-swallowing means. The material to follow does not depend on the particular configuration of the DCO as long as the external properties are the same.

In any of these DCO configurations, be sure that the phase shifting is done without switching faults of any kind. Since sampling in the PD is triggered by a waveform edge from the DCO, it is crucial that there be no missing or extra edges to disrupt operations. The phase at the output of the DCO at the time t_n of the nth sample is

$$\psi_o(t_n) = \omega_{ck} t_n + \theta_o(0) + \frac{2\pi q[n]}{Q} \qquad \text{rad} \tag{13.22}$$

Origin of the time variable can be chosen such that $\theta_o(0) = 0$, eliminating it from further consideration. The phase error $\psi_e(t_n) = \psi_i(t_n) - \psi_o(t_n)$ then becomes

$$\psi_e(t_n) = \Delta\omega t_n + \theta_i - \frac{2\pi q[n]}{Q} \qquad \text{rad} \tag{13.23}$$

where $\Delta\omega = \omega_i - \omega_{ck}$ and ψ_e is evaluated modulo-2π. The PLL is deemed to be locked if $|\psi_e|$ is kept small.

13.3.3 PLL State Diagrams

Understanding of the behavior of the PLL is aided by a state diagram as exemplified in Fig. 13.22 (for the double-loop counter) and Fig. 13.23 (for the single-loop counter). The diagrams incorporate the finite states of the digital portions of the PLL along with the time-continuous nature of the input phase. In addition to showing the states, the diagrams also display the pertinent phases relative to the $q = 0$ tap of the DCO.

Input phase ψ_i can take any position around the continuous outer circle. The phase ψ_o of the DCO can only take on the discrete locations marked by short radial lines intersecting the circle and labeled with a q value. Arrows in the diagrams indicate example locations of these phases, and the arc between them indicates the phase error $\psi_e = \psi_i - \psi_o$. Each q-labeled location for ψ_o represents one of Q states for the DCO. Inside the circle are eight chains, one for each q position. These chains represent the up/down counters of Fig. 13.21. Although only one such counter exists in a system, the counter's states are replicated for each q position to display all of the digital states of the PLL.

The chains are positioned radially, with the low-numbered states toward the center of the circle and the positive states toward the periphery. A positive u_d

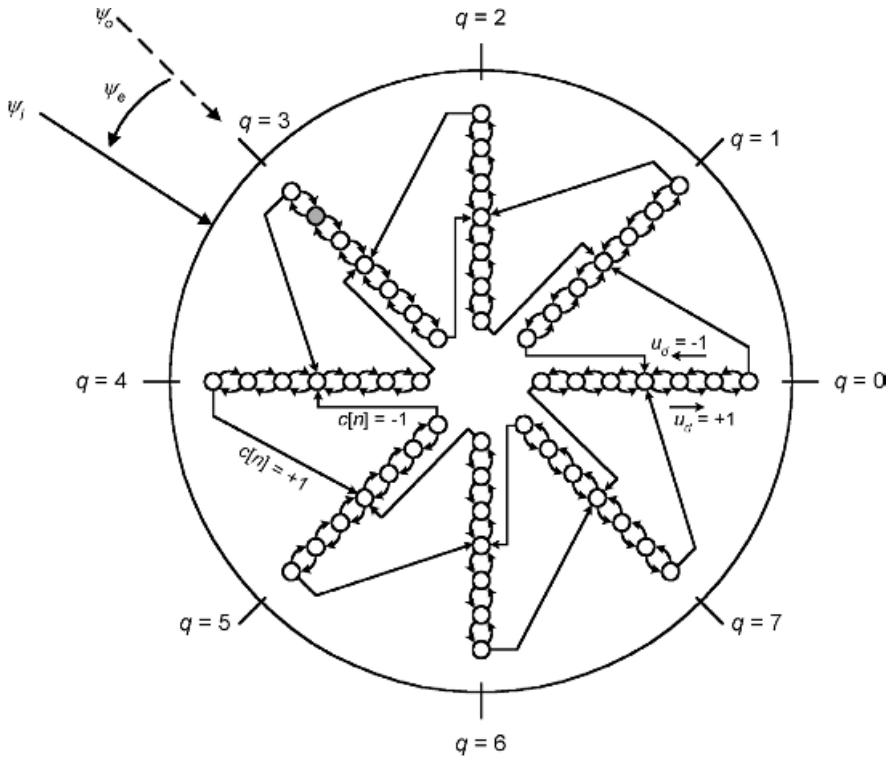


Figure 13.22 Phase and state diagram of a first-order IPM with a double-loop counter.

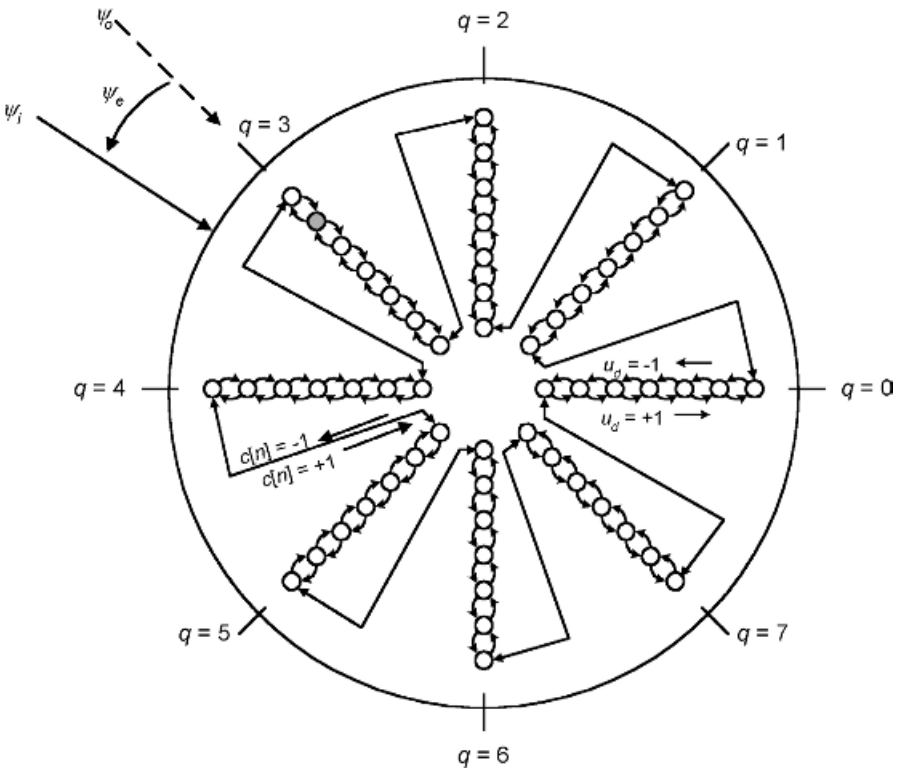


Figure 13.23 Phase and state diagram of a first-order IPM with a single-loop counter.

drives the counter state outward; a negative u_d drives the state inward. Counter output $c[n] = \pm 1$ is emitted whenever u_d drives a counter beyond one of its end states and the DCO q state is advanced or retarded by 1. Recycling of the counter appears in the diagrams as a transfer from one chain to the next higher or lower, depending on the sign of $c[n]$.

Transfer between chains in Fig. 13.22 goes from an end state of the source chain to the (0) state of the target chain; separate lines are shown from the two ends to the two different targets. Transfer from one end of a chain in Fig. 13.23 goes to the opposite end of the receiving chain, and transfer in the other direction goes between the same ends. For that reason, and because the drawing becomes impossibly cluttered otherwise, only a single bidirectional path is shown for each interchain transfer path in Fig. 13.23.

Each state of the PLL is defined by its coordinates $\{q[n], s_c[n]\}$: one coordinate for the DCO and one for the counter. Regarding the diagrams as a snapshot, the state in the example of Fig. 13.22 is shown as $q = 3, s_c = 2$; the q coordinate is identified by the ψ_o arrow pointing to $q = 3$, and the s_c coordinate is identified by shading of that state in the chain. In Fig. 13.23, the coordinates are $q = 3, s_c = 5$ instead, because of the differences in the counters.

13.3.4 Operation of the Nonlinear PLL

Now that the individual elements have been characterized, the behavior of the closed loop can be determined with the help of the state diagrams. Initially, the input is assumed to be noise-free; later sections address performance in jitter and in additive noise.

Limit Cycles Phase error is shown as positive in the snapshot of Fig. 13.22 (for the PLL with double-loop counter), so the next sample of u_d will drive the counter state to $s_c = 3$ and the next one thereafter recycles to $s_c = 0$ but also drives the DCO state to $q = 4$. Assuming that ψ_i is constant (i.e., $\omega_i = \omega_{\text{ck}}$), the phase error now is negative, so succeeding u_d samples drive the counter consistently negative. Upon the fourth negative u_d, the counter recycles to (0) and the DCO phase returns to $q = 3$. Here is a limit cycle. It has a peak-to-peak amplitude of phase of $2\pi/Q$ radians and a period of $2C_2$ sample intervals. Phase quantization has to be fine enough that the limit-cycle amplitude meets any requirements imposed on the system. As a practical matter, phase quantization in this kind of DCO rarely can be made as fine as that attainable in an NCO.

Next consider the behavior of the PLL of Fig. 13.23 (for the PLL with a single-loop counter). From the illustrated state of $q = 3$ and $s_c = 5$, the next sample of $u_d = +1$ advances the counter state to $s_c = 6$. The next one thereafter advances it to $s_c = 0$ and $q = 4$. Now the phase error is negative, so the next sample is $u_d = -1$ and the state of the PLL is immediately driven back to $q = 3$ and $s_c = 6$. Whereupon the next $u_d = +1$ and the state is driven to $q = 4$ and $s_c = 0$. Here is a different limit cycle, brought about by the different counter arrangement. It has the same peak-to-peak amplitude of phase of $2\pi/Q$ radians but has a period of only two sample intervals, irrespective of counter size C_1. Which counter arrangement is better? Jitter of the short-period limit cycle is easier to filter in any following PLL (if one exists), but the high activity from $c[n]$ may impede control access to the DCO by added features, such as an integrator for type 2 operation (described in Section 13.3.5). Other yet-unrecognized considerations might favor one counter over the other.

Phase Acquisition How much time, in sample intervals N_A, is required for the noise-free PLL to bring itself into its steady-state limit cycle from an initial phase error $|\psi_e[0]| < \pi$? Assume that $\Delta\omega = 0$ and consider only positive values of $\psi_e[0]$ (since acquisition behavior should be symmetric for either direction of phase error). The edge of the steady-state limit cycle is reached once phase error is reduced below $2\pi/Q$, which corresponds to

$$q[N_A] = \text{IP}\left\{\frac{Q\psi_i(0)}{2\pi}\right\}$$

and the initial phase error is defined by

$$\psi_e(0) = \psi_i(0) - \frac{2\pi q[0]}{Q}$$

from which is found the total number of DCO phase increments traversed to acquisition as

$$q[N_A] - q[0] = \text{IP}\left\{\frac{Q\psi_e(0)}{2\pi}\right\} \tag{13.24}$$

Except for the first phase increment, the counter requires C_i sample intervals ($i = 1$ or 2) per phase increment traversed. In the first phase increment, the number of sample intervals needed depends on the initial state $s_c[0]$ of the counter and the arrangement of the counter. Denote the number of sample intervals needed to traverse the first phase increment by $\text{IS}_i\{s_c[0]\}$, so

$$N_A = \text{IP}\left\{\frac{Q|\psi_e(0)|}{2\pi}\right\} C_i + \text{IS}_i\{s_c[0]\} \qquad \text{sample intervals} \tag{13.25}$$

In a linear PLL, a large magnitude of phase error causes a large output from the PD (at least up to the phase error corresponding to the peak of the s-curve), but the two-level PD in the nonlinear PLL under consideration has the same PD output irrespective of the magnitude of phase error. Phase acquisition proceeds at an average rate $2\pi/C_i Q$ radians per sample interval, independent of phase error.

Frequency Tracking Limits How far can the DCO be adjusted from its reference frequency ω_{ck}? If ψ_o is advanced by one increment $2\pi/Q$ on each sample interval, the upper tunable frequency is $\omega_{\text{ck}}(1 + 1/Q)$ and the minimum is $\omega_{\text{ck}}(1 - 1/Q)$ if the phase is retarded instead. Clearly, fine quantization and large tuning range are incompatible in an IPM. This kind of PLL can be used only if frequencies are constrained within a narrow range: for example, in data synchronizers with accurately known bit rates.

Now let frequency difference $\Delta\omega$ be nonzero but constant. How large can $|\Delta\omega|$ become before phaselock fails? If the value of every PD sample is $u_d = +1$, the phase ψ_o of the DCO can advance by one full cycle of ω_{ck} in QC_i sample intervals. If the signal frequency causes input phase ψ_i to advance any faster, the DCO phase cannot keep up and lock will be impossible. Therefore, the hold-in limit of the first-order IPM is approximately

$$\Delta\omega_H \approx \pm\frac{\omega_{\text{ck}}}{QC_i} \qquad \text{rad/sec} \tag{13.26}$$

where the approximation assumes that $QC_i \gg 1$. Compare (13.26) to $\Delta\omega_H = \pm K$, the hold-in range for a first-order analog PLL with a sinusoidal s-curve in its PD. By analogy, the ratio ω_{ck}/QC_i for the nonlinear PLL might be construed to resemble loop gain K for the analog PLL. It is this property that identifies the up/down counter as a means of setting gain rather than serving as a filter.

Now follow the PLL state diagrams to visualize the behavior of ψ_o in the presence of a frequency offset $\Delta\omega < \Delta\omega_H$. To that end, define sectors $sct[q, q + 1]$ of the circle by the q values of the bounding ψ_o phases. In Figs. 13.22 and 13.23, the ψ_i marker is located in $sct[3,4]$. A constant frequency offset appears as a

constant-rate rotation of the ψ_i marker around the circle. The limit cycle (when it exists) of ψ_o jumps back and forth between the two sector boundaries. Once the limit cycle has started in any one sector, it continues uninterrupted until ψ_i rotates out of the sector. Then the limit cycle stops (ψ_o holds constant at the boundary value crossed by ψ_i) and the counter state slews to follow ψ_i until an end of the counter is passed and ψ_o is bumped to the next sector boundary. At that point, the limit cycle resumes and continues until ψ_i again rotates into the next sector.

In Section 13.2 it was observed that quantization of NCO frequency placed a lower limit on PLL loop gain; the feedback loop is blocked if the MSB of the scaled phase-detector output is smaller than the LSB of the NCO control word. No such lower limit occurs in the IPM; ω_{ck}/QC_i, the equivalent of loop gain, can be made as small as desired without severing the feedback loop (but at the cost of vanishing hold-in range).

Response to Input Jitter Suppose that phase ψ_i of the input signal is modulated deliberately or has unwanted jitter. Collectively, lump both kinds of phase variation under the term *jitter*. How does the nonlinear PLL respond to input jitter, assuming $\Delta\omega$ to be negligibly small? The linear analyses of Chapter 5 do not apply to a nonlinear PLL; only qualitative piecemeal explorations are offered here, rather than a comprehensive nonlinear analysis.

Consider one possible model for jitter in the state diagrams of Figs. 13.22 and 13.23. In each of those diagrams, intended as single-sample snapshots, the input phase is shown as a single marker arrow at angle ψ_i on the circle. That is a starting point that can be expanded by overlaying identical displays for multiple samples. If the input were jitter-free (and $\Delta\omega = 0$), all displays would be identical and the overlaid conglomeration would still show a single marker for ψ_i. The positions of the multiple ψ_i markers would not coincide if the input were jittered. Instead, a cloud of markers would form around an average position. Quantitative analysis of the response to the cloud would require, at the least, knowledge of the statistical properties of the jitter, knowledge that is often unavailable. In a linear system, it is sufficient to know the transfer function of the PLL and the spectrum of the jitter. No such simple tool is available for a nonlinear PLL. Only qualitative observations on special cases can be offered here.

As a first case, suppose that the cloud is contained entirely within one of the quantization sectors on the phase circle of the state diagram. That is certainly possible if the jitter amplitude is small enough and the cloud is sufficiently far away from a sector boundary. All input phases inside the sector appear the same to the PLL, each generating identical PD outputs. Under those conditions, the nonlinear PLL completely suppresses in-sector jitter; none of the input jitter on ψ_i appears in the output phase ψ_o. Small-enough input jitter gets lost in the phase quantization.

Next, consider an isolated sample whose phase lands outside the sector containing the bulk of the cloud. There are two possibilities in this case: (1) the sign of the phase error for the isolated sample is the same as for the cloud, so

that there is no effect, or (2) the sign of the phase error for the isolated sample is opposite from that of the cloud. The second condition causes the state of the up/down counter to back up one increment from its previous position rather than proceeding one increment in the opposite direction as it would in response to any sample in the bulk of the cloud. This backup lengthens the period of the one isolated limit cycle and increases the phase dwell time in the direction of the disturbance but does not change the peak-to-peak ψ_o phase excursion.

Response to single isolated samples helps explain behavior if phases of many samples are outside the central sector. Those samples whose phase error is the same as that of the main cloud have no perceptible effect; each sample whose phase error is opposite that of the main cloud causes the state of the PLL to back up by one increment of the up/down counter, increasing the phase dwell time in that direction. If enough backups accumulate, the DCO phase will jump by one increment in the direction of the jittered samples. Probability of such a jump depends on the statistics of the jitter and on the size C_i of the counter.

Intuition suggests that a large C_i reduces the probability of a jump for any given statistical properties of the jitter, but closer inspection raises doubts. Although a larger C_i requires more out-of-sector samples to be accumulated before a jump is initiated, a larger C_i also accumulates the effects of a larger number of samples. Further study is needed to determine which action has the greater influence. If jitter is not so bad as to cause loss of lock, the DCO phase eventually returns to the central phase sector. If the cloud extends over both boundaries of the central sector, jumps can occur in either direction.

13.3.5 Type 2 Nonlinear PLL

As explained above, the DCO is tunable over a frequency range of $\pm\omega_{ck}/Q$, but the first-order IPM can hold lock over a range of only $\pm\omega_{ck}/QC_i$. A larger hold-in range is often needed; the prevalence of type 2 analog PLLs makes a type 2 IPM an obvious candidate for achieving the larger range, as laid out in this section. An alternative approach is proposed in [13.39].

Figure 13.24 shows one way to implement a type 2 IPM. Elements in the proportional path—sampler, slicer, and up/down counter—are the same as in the

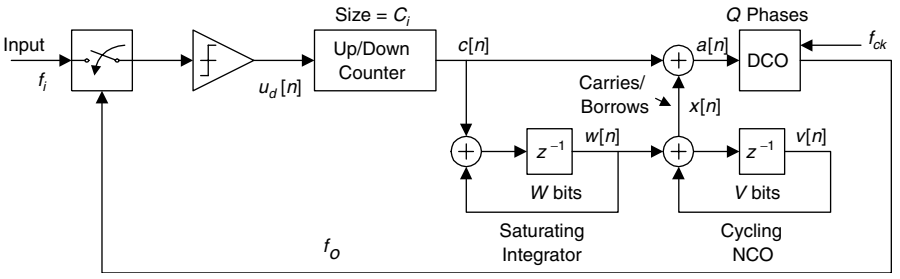


Figure 13.24 Type 2 IPM.

first-order IPM of Fig. 13.20. The DCO can have any configuration equivalent to those discussed in earlier sections. The frequency of the input signal is f_i and average frequency f_o delivered by the DCO has to equal f_i when the PLL is locked. All operations within the PLL, including the sampling, run at f_o, not f_{ck}.

Frequency offset is defined as $\Delta f = f_i - f_{ck}$. The objective of an integral path in the PLL is to deliver additional unit-magnitude control updates to the DCO at a rate close enough to Δf so that the proportional path is not stressed excessively. The integral path in Fig. 13.24 contains a saturating integrator and a cycling NCO. The integrator sums the sequence $\{c[n]\}$ from the up/down counter and feeds that sum $w[n]$ to the NCO as a frequency control. Since $c[n] \in (0, \pm 1)$, the integrator can be implemented as another up/down counter, except that it must saturate at its ends, not recycle. The integrator register has a word length of W bits.

The NCO is much the same as those explored earlier; its difference equation is

$$v[n] = \{v[n-1] + w[n-1]\} \mod 2^V$$

where $v[n]$ is the content of the NCO register and V is the number of bits in the register. The NCO recycles if the sum in braces exceeds $2^V - 1$ or is less than zero; that is, $v[n] \in (0, \ldots, 2^V - 1)$. Useful output of the NCO $x[n] \in (0, \pm 1)$ comes from the 1-bit over- and underflow carries and borrows that the adder generates upon recycling.

The control signal to the DCO is $a[n] = c[n] + x[n]$, $a[n] \in (0, \pm 1, \pm 2)$. Instead of just switching by only one DCO phase increment at a time, as in the first-order IPM, a type 2 PLL also has to be able to switch by two increments at a time. That increased switching requirement, along with the multibit operations in the integrator and NCO, makes for greater complexity in the type 2 IPM.

How many bits W are needed in the integrator and V in the NCO? Consider the NCO first. The average frequency of the NCO for fixed $w[n]$ is $f_{NCO} = wf_o/2^V$, and since w is an integer, the frequency increment is $\delta f_{NCO} = f_o/2^V$. To reduce stress on the proportional path, the frequency increment should be small compared to the hold-in limit of the first-order IPM as given by (13.26); the NCO word length determines its frequency quantization. Define λ such that $\delta f_{NCO} = \lambda\ \Delta f_H$, where $\Delta f_H = \Delta\omega_H/2\pi = f_{ck}/QC_i$ and $0 < \lambda < 1$. As an approximation, $f_o \approx f_{ck}$, so the NCO word length is determined by

$$2^V \approx \frac{QC_i}{\lambda} \tag{13.27}$$

As a numerical example, if $\lambda = 0.5$, $Q = 32$, and $C_i = 16$, then $V = 10$.

A rough equivalence may be deduced between gain κ_2 in the integral path of a quasilinear PLL and the NCO word-length-related ratio λ/QC_i. A large value for λ entails high activity in the integral path, inducing consequent high activity in the proportional path and an implication of inadequate damping (whatever "damping" might mean in a system that is so nonlinear). A small-enough value

for λ provides slow response in the integral path and ample damping. One should suspect that the loop may become unstable or otherwise fail to lock if λ is too large. As of this writing, the stability of highly nonlinear PLLs does not appear to have been pursued extensively in the PLL literature.

Word length W in the integrator establishes the frequency range that the integrator path can accommodate. The largest range feasible within the context of an IPM (no more than one carry or borrow from the NCO per clock interval) is $\pm f_o/Q$. For the NCO to approach close to that rate requires that $w[n] = \pm(2^V - 1)$ for all n, but that word length requires that $W = V + 1$ bits, an apparent misfit. However, it can be achieved if W is formatted as sign plus magnitude, where the sign determines whether the NCO adds or subtracts $w[n]$ but only the magnitude enters into the NCO word $v[n]$. Each reduction of one bit from W will halve the frequency range of a type 2 IPM. The first halving to $W = V$ permits more conventional arithmetic to be used in the NCO. Not much has been published on the behavior of a type 2 IPM. Such features as the limit cycles, acquisition speeds, stability limits, and so on, remain to be explored.

13.3.6 Effects of Additive Noise

Let the input signal be accompanied by additive noise according to

$$s(t) = A \sin \psi_i(t) + y(t) \tag{13.28}$$

where A is the peak amplitude of the signal and $y(t)$ is zero-mean Gaussian noise with variance σ_y^2. Phase-detector samples are $u_d[n] = \text{sgn}\{s(t_n)\} = \text{sgn}\{A \sin[\psi_i(t_n)] + y(t_n)\}$. The signal-to-noise ratio is defined as

$$\rho^2 = \frac{A^2}{2\sigma_y^2} \tag{13.29}$$

Effect on Phase Detector Operation In the absence of noise, $u_d[n] = +1$ if $\sin[\psi_i(t_n)] > 0$, but the presence of noise could induce $u_d[n] = -1$ instead for some samples. The probability P_+ of $u_d[n] = +1$, given a particular $\psi_i(t_n) = \psi$ and $\text{SNR} = \rho$, is given by

$$\text{P}_+ = \Pr\{s(t_n) > 0 | \psi\} = \int_{-A \sin \psi}^{\infty} p(y)\, dy = \frac{1}{\sqrt{2\pi}} \int_{-(A/\sigma) \sin \psi}^{\infty} e^{-x^2/2}\, dx \tag{13.30}$$

and the probability of a negative sample is $\text{P}_- = 1 - \text{P}_+$.

The average rate (in positive-valued samples per second) is

$$r = f_s(\text{P}_+ - \text{P}_-) = f_s(2\text{P}_+ - 1) \tag{13.31}$$

where f_s is the sample rate. Combining (13.30) and (13.31), the normalized rate r/f_s is

$$\frac{r}{f_s} = 2\mathrm{P}_+ - 1 = \frac{2}{\sqrt{\pi}} \int_0^{\rho \sin \psi} e^{-z^2}\, dz = \mathrm{erf}(\rho \sin \psi) \tag{13.32}$$

the net fraction of increments (positive or negative) relative to total samples. This is the useful output of the PD and may be regarded as its s-curve. Figure 10.15 is a plot of (13.32) with slightly modified notation.

The maximum rate is attained for $\psi = \pi/2$, so the largest average steady-state phase error for which the first-order IPM can hold lock is $\mathrm{Max}(r/f_s) = \mathrm{erf}(\rho)$. For large ρ, $\mathrm{erf}(\rho) \approx 1$, and for small ρ, $\mathrm{erf}(\rho) \approx 2\rho/\pi^{1/2}$. Hold-in limits therefore are $\Delta f_H \approx f_s/QC_i$, as found earlier in (13.26) for large ρ and

$$\Delta f_H \approx \frac{2 f_s \rho}{\sqrt{\pi}\ QC_i} \tag{13.33}$$

for small ρ.

Previously, phase-detector gain has been defined as the slope of the s-curve at its zero crossings. The same concept can be applied to the two-level sampling PD of the IPM by first differentiating (13.32) with respect to ψ to obtain

$$\frac{dr}{d\psi} = \frac{2 f_s \rho \cos \psi}{\sqrt{\pi}} = \frac{2 f_s \rho}{\sqrt{\pi}}$$

at $\psi = 0$. Next, approximate r by its Taylor series expansion for small ψ, yielding

$$\frac{r}{f_s} \approx \frac{2\rho\psi}{\sqrt{\pi}} \tag{13.34}$$

for small-enough ψ and all ρ. Notice that r becomes very large and the region of validity of (13.34) is very small for increasing ρ.

The slope of the s-curve at $\psi = 0$ of an analog phase detector is the PD gain K_d. Equivalently, a PD gain for the IPM in the presence of noise could be defined from (13.34) as $K_d = 2\rho f_s/\sqrt{\pi}$ positive samples per second per radian in an analog formulation or $\kappa_d = K_d/f_s = 2\rho/\sqrt{\pi}$ rad^{-1} in a dimensionless digital formulation. Furthermore, the average "gain" of the counter and DCO is $2\pi/QC_i$ radians per sample, so the "gain" of the first-order IPM is $K = 4\pi^{1/2} f_s \rho/QC_i$ rad/sec in analog formulation or $\kappa = 4\pi^{1/2}\rho/QC_i$ (dimensionless) in digital formulation. The utility of the definitions diminish as ρ becomes large. Equivalence to either an analog PLL (with units for the gain) or a digital PLL (dimensionless) comes about because the PLL in question is a hybrid and partakes of qualities of both implementations.

Effects on DCO Phase Fluctuations Modeling of the first-order IPM in terms of chains of states is conducive to analysis as Markov chains. Every transition from one state to the next has a transition probability that depends on the

relation of the states in the pair to the phases of the signal and DCO and to the statistics of the noise or other disturbances. If the probabilities can be assigned (not always feasible), various summations provide statistics of phase errors. References [13.34–13.36] formulate such analyses and provide results of probability distributions and variance of DCO phase fluctuations as well as cycle-slip statistics (mean time to first slip). Evaluation of the statistics usually requires extensive computer calculation for each case.

13.3.7 Application to Bit Synchronizers

The IPM model so far has been restricted to a sinusoidal input signal. Another signal, one for which the IPM might be more applicable, is a binary NRZ symbol stream: that is, a two-level waveform with uniform symbol duration $T = 1/f_i$. Transitions between levels from one symbol to the next occur with probability $d < 1$. A simplistic model assumes that transitions are instantaneous. In practice, data rates f_i are often specified with small tolerances; a rate uncertainty of $\pm 0.1\%$ of the bit rate would be considered quite large.

A modified IPM has been a candidate for recovering clock from a bit stream of this description. Modifications are needed in the phase detector and in the up-down counter to apply the IPM to a bit stream. Analyses of the earlier parts of this section are applicable once the modifications have been taken into account. A phase detector for a bit stream has to have a three-level output $u_d[n]$ with possible values 0 in addition to ± 1, where n now refers to symbol index. A value 0 is delivered if there is no transition between two symbols; $+1$ or -1 indicates whether the signal leads or lags the DCO phase. Ransom and Gupta, [13.36], Walker [13.40], and Gardner [13.41] describe examples of timing error detectors that can be adapted to deliver ternary outputs.

A two-level signal requires at least two samples per symbol to extract the needed timing error information, but the PD delivers only one sample per symbol. That is, the average rate of sampling the signal must be $2/T$, whereas PD samples are generated at $1/T$. Moreover, since some PD samples are zeros, the average rate of $+1$ or -1 is lessened by the probability $d < 1$ of transitions in the symbol stream. A zero applied to the up-down counter neither advances nor retards the count. Each state diagram of the counter should have a reentrant path to show that the state remains unchanged for $u_d[n] = 0$. Operation of the DCO depends on the PD and up-down counter only through the sequence $\{c[n]\}$ but is otherwise unchanged from the explanations given in Sections 13.3.2 and 13.3.3.

The existence of zero-valued PD samples lengthens, by a factor $1/d$, the average number of symbol-rate samples needed to shift the DCO phase. Therefore, the average period of a limit cycle or the average number of symbol-rate samples needed to acquire the signal phase is increased by a factor $1/d$. Because of the fewer useful PD samples, frequency hold-in range is decreased by a factor d.

Random variations of limit-cycle period, phase-acquisition time, and hold-in range occur because $u_d = 0$ occurs randomly. [**Comment**: Since a limit cycle is strictly periodic by definition, a closed trajectory with random variations in

duration no longer is a limit cycle.] Another issue comes about because a bit stream can have long strings of the same symbol value—long strings without a transition. Since frequency offset $\Delta\omega$ is almost never zero, the signal phase continues to rotate relative to ω_o, so a large phase error can build up during a transition-free string. Frequency tolerances have to be tight enough to constrain that buildup of phase error.

APPENDIX 13A: TRANSFER FUNCTION OF A MULTIRATE DPLL

This appendix develops the transfer functions for the multirate DPLL of Fig. 13.7, based on the multirate theory established in [13.6, Sec. 2.3]. Applicable notation is shown in Fig. 13.7 in braces. Also described are useful engineering techniques for analyzing multirate processes.

13A.1 Nomenclature

Three different sampling rates appear in Fig. 13.7: $1/T$ at the phase detector input and output, M/T at the signal input, the hold output, the NCO, the sine/cosine process, and the phase rotator, and $1/LT$ at the output of the accumulate & dump and in the loop filter. The symbol rate is $1/T$; L and M are integers. All elements in the DPLL are assumed to be quasilinear processes that can be represented by z-transform transfer functions. To enhance clarity of explanation, three different transform variables are employed in the development: z in the $1/T$ sampling region, $\xi = z^L$ in the $1/LT$ sampling region, and $\eta = z^{1/M}$ in the M/T sampling region. The final results are expressed solely in terms of z.

13A.2 Phase-Detector Operation

The PD equation is

$$U_d(z) = \kappa_d \theta_e(z) \tag{13A.1}$$

where κ_d is the PD gain in rad^{-1} and $\theta_e(z) = \theta_i(z) - \theta_o(z)$ is the z-transform of the phase error at the PD input.

13A.3 Accumulate & Dump and the Loop Filter

An accumulate & dump can be modeled as a finite impulse response (FIR) filter with L equal-weight taps, operating entirely at sampling rate $1/T$, followed by an $L:1$ down-sampler. The transfer function of the filter is

$$H_a(z) = 1 + z^{-1} + z^{-2} + \cdots + z^{-(L-1)} = \frac{1 - z^{-L}}{1 - z^{-1}} \tag{13A.2}$$

Notice that its DC response is $H_a(1) = L$. The frequency response of the filter on the unit circle is

$$H_a(e^{j\omega T}) = e^{-j(L-1)\omega T/2} \frac{\sin(L\omega T/2)}{\sin(\omega T/2)}, \qquad |\omega T| \le \pi \tag{13A.3}$$

so the filter has a delay of $(L-1)T/2$ and an amplitude response

$$|H_a(e^{j\omega T})| = \left|\frac{\sin(L\omega T/2)}{\sin(\omega T/2)}\right| \tag{13A.4}$$

The amplitude response has nulls at $f = \omega/2\pi = k/LT$ for $|k| < L/2, k \neq 0$. The Nyquist folding frequencies incident to down-sampling to $1/LT$ are at odd multiples of $1/2LT$, midway between the nulls, so all nulls alias to $f = 0$.

The output of the accumulate & dump after down-sampling by $L:1$ is represented [13.6] by the z-transform

$$U_a(\xi) = \frac{1}{L} H_a(\xi^{1/L}) U_d(\xi^{1/L}) \tag{13A.5}$$

This result neglects all aliasing; see [13.6, eq. (2.64)] for the effects of aliases. Down-sampling usually should be avoided if aliasing is not negligible. Transmission through the loop filter $F(\xi)$ is [13.6]

$$U_f(\xi) = U_a(\xi)F(\xi) = \frac{1}{L} U_d(\xi^{1/L}) H_a(\xi^{1/L}) F(\xi) \tag{13A.6}$$

For the proportional-plus-integral loop filter in Fig. 13.5, the loop-filter transfer function is

$$F(\xi) = \kappa_1 \left(1 + \frac{\kappa_2 \xi^{-1}}{1 - \xi^{-1}}\right) \tag{13A.7}$$

13A.4 Hold Process

Operation of the hold is equivalent to an up-sampling by $1:LM$ in which $LM - 1$ zero-valued samples at spacing T/M are inserted between input samples at spacing LT, followed by an FIR filter $H_f(\eta)$ with LM taps, each with unit weight. The up-sampled filtered signal is represented by

$$U_c(\eta) = U_f(\eta^{LM}) H_h(\eta) \tag{13A.8}$$

and the transfer function of the equivalent FIR filter is

$$H_h(\eta) = \frac{1 - \eta^{-LM}}{1 - \eta^{-1}} \tag{13A.9}$$

which is the same form as the FIR filter in the accumulate & dump, except for the number of taps.

[**Comment**: Crochiere and Rabiner [13.6] show that the interpolation filter following an up-sampler generally is time varying and cannot be represented by a simple transfer function. The equivalent filter in a zero-order hold is an exception, perhaps unique, that can be represented with a time-invariant transfer function, as demonstrated in Section 13A.7.]

13A.5 NCO, Phase Rotator, and $M:1$ Down-Sampling

Phase $\theta_o(\eta) = 2\pi\varepsilon_o(\eta)$ generated from the NCO and sine/cosine process is

$$\theta_o(\eta) = 2\pi U_c(\eta)\frac{\eta^{-1}}{1-\eta^{-1}} \tag{13A.10}$$

assuming that the NCO gain defined in (4.4) and (4.9) is $\kappa_v = 1$. The phase error is

$$\theta_e(\eta) = \theta_i(\eta) - \theta_o(\eta) \tag{13A.11}$$

Equations (13A.10) and (13A.11) are correct and would be satisfactory in a PLL with only one sampling rate, but they are not useful in the particular circumstances of the DPLL of Fig. 13.7. In particular, it is not clear how the z-transform of $\theta_e[n]$ at sample rate $1/T$ is to be determined from knowledge of $\theta_i(\eta)$ and $\theta_o(\eta)$ at sample rate M/T.

The artificial model of Fig. 13A.1 is proposed as a substitute to furnish a simplifying approximation. Figure 13A.1 introduces two significant improvisations: (1) a fictitious up-sampler in the signal path and (2) a split of the $1:LM$ hold process into two fictitious hold processes. First consider the fictitious up-sampler. Pretend that the input sequence $\{\theta_i[m]\}$ at rate M/T is generated by up-sampling a sequence $\{\theta_i[n]\}$ at rate $1/T$ by a ratio $1:M$. No receiver is ever likely to be configured in that manner; the fictitious up-sampler is merely an analytical expedient. The sequence..., $\theta_i[n]$, $\theta_i[n+1]$,... becomes $\theta_i[n]$, $\theta_i[n+1/M]$, $\theta_i[n+2/M]$,..., $\theta_i[n+(M-1)/M]$, $\theta_i[n+1]$,.... This model implies that $\theta_i[n]$ is preserved unaltered in $\theta_i[n+k/M]$ for $k=0$—a property not possessed by sampling-rate expanders in general. That does not matter for dealing with θ_i since the receiver has full possession of all M samples in the interval T; the value of the fictitious $\theta_i[n]$ prior to up-sampling is irrelevant.

Next, split the $1:LM$ hold process into a $1:L$ hold followed by a $1:M$ hold. Output samples issue forth from the $1:L$ hold element at a rate $1/T$, so they can be given index n (the same as the PD I/O samples) and z can be the associated

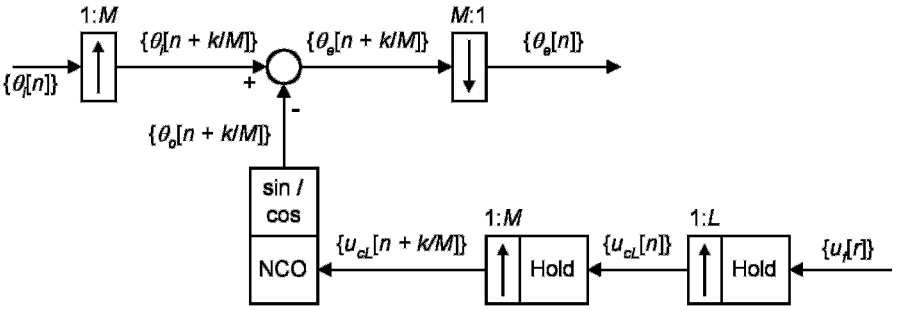


Figure 13A.1 Model for the operation of a hold element.

transform variable. Denote those fictitious samples as $u_{cL}[n]$; their z-transform is

$$U_{cL}(z) = U_f(z^L)\frac{1 - z^{-L}}{1 - z^{-1}} = U_f(z^L)H_a(z) \tag{13A.12}$$

where the transfer function of the equivalent interpolation filter, derived in Section 13A.7, is identical to that of the accumulate & dump filter of (13A.2).

Now $u_c[m]$ becomes a $1:M$ up-sampling and M-fold repetition of $u_{cL}[n]$. It is true that $\theta_o[m+1] = \theta_o[m] + 2\pi u_c[m]$, but it is also true that $\theta_o[n+1] = \theta_o[n] + 2\pi M u_{cL}[n]$ because of the M-fold repetition of $u_{cL}[n]$, even though $u_{cL}[n]$ is fictitious. This simple relation is a consequence of the zero-order hold and probably does not apply to any other interpolation filter. The related z-transform is

$$\theta_o(z) = \frac{2\pi M z^{-1} U_{cL}(z)}{1 - z^{-1}} \tag{13A.13}$$

(This formulation neglects a delay of approximately $T/2$ in the equivalent filter of the $1:M$ hold operation, a delay that can be lumped into all the other delays that eventually have to be taken into account.)

The effect of the memoryless phase rotator is simply to subtract $\theta_o[m]$ from $\theta_i[m]$ to produce $\theta_e[m]$; it contains no filter effects. Rotated output after $M:1$ down-sampling is

$$\theta_e[n] = \theta_i[n] - \theta_o[n] \tag{13A.14}$$

which has the z-transform

$$\theta_e(z) = \theta_i(z) - \theta_o(z) \tag{13A.15}$$

Several issues related to the $M:1$ down-sampler deserve comment. First, how does the down-sampler select one good sample out of M and reject all others? That selection is the duty of the timing-recovery operation, a subject itself worthy of chapters or entire books for adequate coverage, but not otherwise treated here. Next, no antialias filtering is shown prior to the $M:1$ signal down-sampler in Fig. 13.7 or 13.8. An actual receiver will have filters before or after the phase rotator, but they have little effect on the signal phase, which has tacitly been assumed to be changing slowly compared to the symbol rate. Finally, any filters placed after the rotator are inside the feedback loop, and their delay has to be lumped into the overall loop delay.

13A.6 Transfer Functions

Figure 13A.2 gathers together the information developed in the preceding sections of this appendix and presents it as an open loop with $\theta_e(z)$ as input and $\theta_o(z)$ as output. Parentheses associated with each block or signal indicate applicable equation numbers. The only departure from the preceding explanation is in the delay D lumped with the NCO and sine/cosine process; (13A.13) is written with $D = 1$.

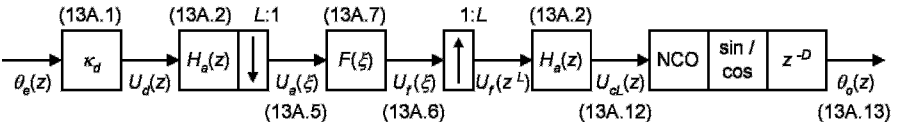


Figure 13A.2 Open-loop model of a multirate DPLL.

Substituting z^L for ξ in the blocks sampled at rate $1/LT$ and combining all the equations yields the open-loop transfer function

$$G(z) = \frac{\theta_o(z)}{\theta_e(z)} = \frac{2\pi M \kappa_d z^{-D}}{L(1-z^{-1})} F(z^L)[H_a(z)]^2$$
$$= \frac{2\pi M \kappa_d z^{-D}}{L(1-z^{-1})} F(z^L) \left(\frac{1-z^{-L}}{1-z^{-1}}\right)^2 \tag{13A.16}$$

From (4.11), the transfer function of a proportional-plus-multiple-integral loop filter can be written as

$$F(z^L) = \kappa_1 \left\{1 + \frac{\kappa_2 z^{-L}}{1-z^{-L}} \left[1 + \frac{\kappa_3 z^{-L}}{1-z^{-L}}(1+\cdots)\right]\right\} \tag{13A.17}$$

with gain κ_1 in the proportional path. Following Section 3B.2, the filter $H_a(z)$ from (13A.2) may be interpreted as a lowpass high-frequency filter with DC gain $H_a(1) = L$. Incorporating these factors, loop gain κ is defined as

$$\kappa = \frac{2\pi M[H_a(1)]^2 \kappa_d \kappa_1}{L} = 2\pi M L \kappa_d \kappa_1 \tag{13A.18}$$

whereupon the open-loop transfer function takes the form

$$G(z) = \frac{\kappa z^{-D}}{1-z^{-1}} \frac{F(z^{-L})}{\kappa_1} \frac{[H_a(z)]^2}{L^2}$$
$$= \frac{\kappa z^{-D} F(z^L)}{\kappa_1 L^2 (1-z^{-1})} \left(\frac{1-z^{-L}}{1-z^{-1}}\right)^2 \tag{13A.19}$$

For a type 2 DPLL, the transfer function for $F(z^L)$ is given by (13A.7), yielding an open-loop transfer function

$$G(z) = \frac{\kappa z^{-D}}{L^2(1-z^{-1})} \left(1 + \frac{\kappa_2 z^{-L}}{1-z^{-L}}\right) \left(\frac{1-z^{-L}}{1-z^{-1}}\right)^2$$
$$= \frac{\kappa z^{-D}[1 - z^{-L}(1-\kappa_2)](1-z^{-L})}{L^2(1-z^{-1})^3}$$
$$= \frac{\kappa z^{-D}[1 - z^{-L}(1-\kappa_2)](1+z^{-1}+z^{-2}+\cdots+z^{-(L-1)})}{L^2(1-z^{-1})^2} \tag{13A.20}$$

This PLL has a real zero at $z = (1-\kappa_2)^{1/L} \approx 1 - \kappa_2/L$ and $L-1$ complex zeros on the unit circle at $\omega = 2\pi q/LT$, $|\omega T| \le \pi$, $q =$ integer $\neq 0$.

The closed-loop system transfer function is

$$H(z) = \frac{G(z)}{1+G(z)} = \frac{\kappa z^{-D}[1 - z^{-L}(1-\kappa_2)](1 + z^{-1} + z^{-2} + \cdots + z^{-(L-1)})/L^2}{(1-z^{-1})^2 + \kappa z^{-D}[1 - z^{-L}(1-\kappa_2)](1 + z^{-1} + z^{-2} + \cdots + z^{-(L-1)})/L^2} \tag{13A.21}$$

and the closed-loop error transfer function is

$$E(z) = \frac{1}{1+G(z)} = \frac{(1-z^{-1})^2}{(1-z^{-1})^2 + \kappa z^{-D}[1 - z^{-L}(1-\kappa_2)](1 + z^{-1} + z^{-2} + \cdots + z^{-(L-1)})/L^2} \tag{13A.22}$$

The order of this PLL is $(2L + D - 1)$ instead of second order. It is second order only for minimum $L = 1$ (i.e., no accumulate & dump) and minimum delay $D = 1$.

13A.7 Transfer Function of a Hold Filter

A zero-order hold with L-fold repetition accepts an input sequence $\{x[r]\}$ at rate $1/LT$ and delivers an output sequence $\{y[n]\}$ at rate $1/T$. An up-sampler intersperses $L-1$ zero-valued, equally spaced output samples, between adjacent input samples, and a filter with impulse response $h[k]$ delivers L output samples for each r, samples that are T spaced and $x[r]$ valued. The impulse response is defined as

$$h[k] = \begin{cases} 1, & k = 0 \text{ to } L-1 \\ 0, & k \neq 0 \text{ to } L-1 \end{cases} \tag{13A.23}$$

From [13.6, eq. (2.78)]

$$y[n] = \sum_{r=-\infty}^{\infty} h[n - rL]x[r] \tag{13A.24}$$

That is, $y[n] = x[r]$ for $n = rL, rL+1, \ldots, rL+L-1$. The z-transform of $y[n]$ is

$$\begin{aligned} Z\{y[n]\} &= \sum_{r=-\infty}^{\infty} \sum_{n=rL}^{(r+1)L-1} x[r]z^{-n} = \sum_{r=-\infty}^{\infty} x[r] \sum_{n=rL}^{(r+1)L-1} z^{-n} \\ &= \sum_{r=-\infty}^{\infty} x[r] \sum_{k=0}^{L-1} z^{-(k+rL)} = \sum_{r=-\infty}^{\infty} x[r]z^{-rL} \sum_{k=0}^{L-1} z^{-k} \\ &= \frac{1-z^{-L}}{1-z^{-1}} \sum_{r=-\infty}^{\infty} x[r]z^{-rL} = \frac{1-z^{-L}}{1-z^{-1}} X(z^L) \end{aligned} \tag{13A.25}$$

REFERENCES

13.1 W. C. Lindsey and C. M. Chie, "A Survey of Digital Phase-Locked Loops," *Proc. IEEE* ***69***, 410–431, Apr. 1981. Reprinted in [13.2].

13.2 W. C. Lindsey and C. M. Chie, eds., *Phase-Locked Loops*, Reprint Volume, IEEE Press, New York, 1986.

13.3 F. D. Natali, "Accurate Digital Detection of Angle-Modulated Signals," *IEEE EASCON Conv. Rec.*, 407–413, 1968. Reprinted in [13.2].

13.4 G. S. Gill and S. C. Gupta, "First-Order Discrete Phase-Locked Loop with Applications to Demodulation of Angle-Modulated Carrier," *IEEE Trans. Commun.* ***COM-25***, 454–462, June 1972. Reprinted in [13.2].

13.5 A. Weinberg and B. Liu, "Discrete Time Analyses of Nonuniform Sampling First- and Second-Order Digital Phase Lock Loops," *IEEE Trans. Commun.* ***COM-22***, 123–137, Feb. 1974.

13.6 R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, 1983.

13.7 P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice Hall, Englewood Cliffs, NJ, 1993.

13.8 F. M. Gardner, "Interpolation in Digital Modems, Part I: Fundamentals," *IEEE Trans. Commun.* ***41***, 501–507, Mar. 1993.

13.9 L. Erup, F. M. Gardner, and R. A. Harris, "Interpolation in Digital Modems, Part II: Implementation and Performance," *IEEE Trans. Commun.* ***41***, 998–1008, June 1993.

13.10 U. Mengali and A. N. D'Andrea, *Synchronization Techniques for Digital Receivers*, Plenum Press, New York, 1997, Sec. 7.3.

13.11 H. Meyr, M. Moeneclaey, and S. Fechtel, *Digital Communication Receivers*, Wiley, New York, 1998, Chap. 9.

13.12 W. C. Lindsey and C. M. Chie, "Acquisition Behavior of a First-Order Digital Phase-Locked Loop," *IEEE Trans. Commun.* ***COM-26***, 1364–1370, Sept. 1978.

13.13 H. C. Osborne, "Stability Analysis of an *N*th Power Digital Phase-Locked Loop, Part I: First-Order PLL," *IEEE Trans. Commun.* ***COM-28***, 1343–1354, Aug. 1980. Reprinted in [13.2].

13.14 H. C. Osborne, "Stability Analysis of an *N*th Power Digital Phase-Locked Loop, Part II: Second- and Third-Order DPLLs," *IEEE Trans. Commun.* ***COM-28***, 1355–1364, Aug. 1980. Reprinted in [13.2].

13.15 G. M. Bernstein, M. A. Lieberman, and A. J. Lichtenberg, "Nonlinear Dynamics of a Digital Phase Locked Loop," *IEEE Trans. Commun.* ***37***, 1062–1070, Oct. 1989.

13.16 A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, 1975, Chap. 9.

13.17 L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1975, Sec. 5.3.2.

13.18 *Digital Signal Processing, II*, Part 4C, "Limit Cycles," Reprint Volume, IEEE Press, New York, 1976.

13.19 J. G. Proakis and D. G. Manolakis, *Introduction to Digital Signal Processing*, Macmillan, New York, 1988, Sec. 10.4.1.

13.20 J. C. Candy and G. C. Temes, eds., *Oversampling Delta-Sigma Data Converters*, Reprint Volume, IEEE Press, New York, 1992.

13.21 S. R. Norsworthy, R. Schreier, and G. C. Temes, eds., *Delta-Sigma Data Converters*, Reprint Volume, IEEE Press, New York, 1997.

13.22 M. Bertocco, C. Narduzzi, P. Paglierani, and D. Petri, "A Noise Model for Digitized Data," *IEEE Trans. Instrum. Meas.* ***IM-49***, 83–86, Feb. 2000.

13.23 N. M. Blachman, "The Intermodulation and Distortion Due to Quantization of Sinusoids," *IEEE Trans. Acoust. Speech Signal Process.* ***33***, 1417–1426, Dec. 1985.

13.24 R. M. Gray, "Quantization Noise Spectra," *IEEE Trans. Inf. Theory* ***IT-36***, 1220–1244, Nov. 1990.

13.25 R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory* ***IT-44***, 2325–2383, Oct. 1998.

13.26 M. F. Wagdy, "Effect of Various Dither Forms on Quantization Errors of Ideal A/D Converters," *IEEE Trans. Instrum. Meas.* ***IM-38***, 850–855, Aug. 1989.

13.27 M. F. Wagdy and M. Goff, "Linearizing Average Transfer Characteristics of Ideal ADC's via Analog and Digital Dither," *IEEE Trans. Instrum. Meas.* ***IM-43***, 146–150, Apr. 1994.

13.28 B. Widrow, I. Kollar, and M. C. Liu, "Statistical Theory of Quantization," *IEEE Trans. Instrum. Meas.* ***IM-45***, 353–361, Apr. 1996.

13.29 N. A. D'Andrea and F. Russo, "Multilevel Quantized DPLL Behavior with Phase- and Frequency Step Plus Noise Input," *IEEE Trans. Commun.* ***COM-28***, 1373–1382, Aug. 1980.

13.30 C. A. Pomalaza-Raez and C. D. McGillem, "Digital Phase-Locked Loop Behavior with Clock and Sampler Quantization," *IEEE Trans. Commun.* ***COM-33***, 753–759, Aug. 1985.

13.31 F. M. Gardner, "Frequency Granularity in Digital Phaselock Loops," *IEEE Trans. Commun.* ***44***, 749–758, June 1996.

13.32 A. Teplinsky, O. Feely, and A. Rogers, "Phase-Jitter Dynamics of Digital Phase-Locked Loops," *IEEE Trans. Circuits Syst. I* ***46***, 545–558, May 1999.

13.33 A. Teplinsky and O. Feely, "Phase-Jitter Dynamics of Digital Phase-Locked Loops, Part II," *IEEE Trans. Circuits Syst. I* ***47***, 458–473, Apr. 2000.

13.34 J. R. Cessna and D. M. Levy, "Phase Noise and Transient Times for a Binary Quantized Digital Phase-Locked Loop in White Gaussian Noise," *IEEE Trans. Commun.* ***COM-20***, 94–104, Apr. 1972.

13.35 J. K. Holmes, "Performance of a First-Order Transition Sampling Digital Phase-Locked Loop Using Random-Walk Models," *IEEE Trans. Commun.* ***COM-20***, 119–131, Apr. 1972. Reprinted in [13.38].

13.36 J. J. Ransom and S. C. Gupta, "Performance of a Finite Phase State Bit-Synchronization Loop with and Without Sequential Filters," *IEEE Trans. Commun.* ***COM-23***, 1198–1206, Nov. 1975.

13.37 N. A. D'Andrea and F. Russo, "A Binary Quantized Digital Phase Locked Loop: A Graphical Analysis," *IEEE Trans. Commun.* ***COM-26***, 1355–1363, Sept. 1978.

13.38 W. C. Lindsey and M. K. Simon, eds., *Phase-Locked Loops and Their Application*, Reprint Volume, IEEE Press, New York, 1978.

13.39 H. Yamamoto and S. Mori, "Performance of a Binary Quantized All Digital Phase-Locked Loop with a New Class of Sequential Filter," *IEEE Trans. Commun.* ***COM-26***, 35–44, Jan. 1978.

13.40 R. C. Walker, "Designing Bang-Bang PLLs for Clock and Data Recovery in Serial Data Transmission Systems," in B. Razavi, ed., *Phase-Locking in High-Performance Systems*, Reprint Volume, IEEE Press, New York, and Wiley, New York, 2003, pp. 34–45.

13.41 F. M. Gardner, "A BPSK/QPSK Timing-Error Detector for Sampled Receivers," *IEEE Trans. Commun.* ***COM-34***, 423–429, May 1986.

13.42 D. E. Calbaza and Y. Savaria, "A Direct Digital Period Synthesis Circuit," *IEEE J. Solid-State Circuits* ***SC-37***, 1039–1045, Aug. 2002.

13.43 N. Da Dalt, "A Design-Oriented Study of the Nonlinear Dynamics of Digital Bang–Bang PLLs," *IEEE Trans. Circuits Syst. I* ***52***, 21–31, Jan. 2005.

CHAPTER 14

ANOMALOUS LOCKING

Preceding chapters described various lock failures caused by such things as loop instability, too much noise, or excessive rates of change in signal phase or frequency. This chapter concentrates on several ways in which a PLL might lock to the wrong phases or frequencies. Also, another mechanism of lock failure is identified. Techniques are suggested for avoiding these problems.

14.1 SIDELOCKS

Various signals either include sidebands with discrete spectral lines caused by periodic modulation components, or else necessary nonlinear operations on the signal generate such lines. Modulation on the latter signals is said to be *cyclostationary* (statistics of the modulation are periodic). Data streams modulated onto a carrier are a prime example of cyclostationary signals. A PLL of sufficiently narrow bandwidth can lock onto any discrete spectral line of adequate amplitude that it encounters, whether that line is the carrier, for which lock is wanted, or is one of the periodic sideband components, an unwanted and problematic lock. Lock to a sideband component is known as *sidelock*, a condition to be avoided.

The known techniques for avoiding sidelock are not plentiful. One technique is to restrict tuning of the PLL to the near vicinity of the closely specified carrier frequency of a modulated signal. The restricted tuning range must not approach close enough to the nearest periodic sideband to allow the PLL to fall into a

Phaselock Techniques, Third Edition, by Floyd M. Gardner

sidelock. That method is not workable if the frequency uncertainty exceeds about half the frequency spacing from the carrier to the closest lockable sideband.

Another technique is to employ a frequency-lock loop (FLL) for frequency acquisition prior to phase locking. The FLL is required to reduce initial frequency error to a small region in the close vicinity of the correct carrier frequency. Typical frequency-difference discriminators indicate zero frequency error when the VCO frequency is at the center of gravity of the power spectrum of the input signal plus noise plus interference. If the desired signal is appreciably stronger than any accompanying noise and interference, if the spectrum of the desired signal is symmetric about the carrier frequency, and if the signal spectrum has not been distorted asymmetrically by filters or multipath, the FLL can settle to near the carrier frequency. Otherwise, the equilibrium FLL frequency will be biased away from the carrier.

As a rather more complicated technique, a spectrum analysis can be performed on a segment of the received signal. The carrier would be identified from prior knowledge of spectral characteristics and the PLL retuned as necessary to align to the desired carrier. Operations of this sort are strongly dependent on the specifics of the individual application.

Yet another technique would be to examine the demodulated signal after phaselock had been acquired and decide (by means highly specific to the particular signal) whether correct lock had been achieved. If not, search would be instituted for another lockable component at a higher (or lower) frequency. Search ends when a correct lock is identified.

I am unaware of any successful examples of either of the latter two methods.

14.1.1 Periodic Modulations

Some examples of periodic modulations include:

- Periodic tones amplitude- or angle-modulated onto a carrier
- Color bursts in a color television signal occurring once per horizontal line [14.1]
- RF pulses from a coherent radar

Each of these sources has a spectrum consisting of a carrier and pairs of discrete-frequency sidebands. Frequency spacing is equal to the modulation frequency. For the burst or pulse signals, the amplitudes of the sidebands depend on the *duty ratio*: the ratio of pulse (or burst) width to the repetition period. For a small duty ratio, the amplitudes of the close-in sidebands are only slightly less than that of the carrier.

[**Comment**: In a pulsed (or burst) signal, the signal phase has to be coherent from one pulse to the next for meaningful phaselock to hold over multiple pulses. Coherence requires that the source oscillator in the transmitter must run continuously and that the transmitter has to be switched ON and OFF in a stage following the oscillator. If the oscillator is switched instead, the signal phase is not coherent

from one pulse to the next and a receiver PLL has to reacquire phase on each individual pulse. That is an issue entirely different from the sidelock problem.]

In his classic paper on color reference in color TV, Richman [14.1], employs both a restricted tuning range for the PLL and also employs frequency-aided acquisition with a quadricorrelator (Section 8.3.4). Eisenberg [14.3], Mengali [14.4], and Schiff [14.5] explore various aspects of a gated PLL that is turned on only in the presence of the burst or pulsed signal. If duty ratio is small, the PLL is a sampled system and has to be analyzed as such. Eisenberg [14.3] concentrates on the sampled nature of the gated PLL. Mengali [14.4] explores the noise analysis. Schiff [14.5] sets formal design conditions for avoidance of sidelock in a second-order type 1 PLL.

A gated PLL has to hold the signal properties accurately during the OFF interval and switch between ON and OFF with negligible disturbances. Accurate holding requires that the frequency memory be set accurately during the ON interval and have small drift during the OFF interval. A type 2 PLL is advisable since it has zero steady-state phase error during the ON interval (meaning that all frequency information has been stored in the integrator) and a good integrator holds its charge well when given a zero-valued input during the OFF interval. It is important to isolate the frequency memory—the integrator—from DC offsets and from noise or interference during the OFF interval to preserve the memory. A small duty ratio makes stringent demands on the quality of the hold operation.

14.1.2 Cyclostationary Modulations

Numerous papers [14.6–14.10] have been published on sidelocks that occur with passband data signals. In essence, the signal consists of a data stream at a uniform symbol rate $1/T$ that is translated by suppressed-carrier modulation to a carrier frequency f_c. Almost all of these papers have denominated the mis-locking phenomenon as *false lock*, but they really deal with sidelocks. The term *false lock* belongs more properly to an entirely different phenomenon, explained in Section 14.4.

A suppressed-carrier signal with random zero-mean data modulation typically has no discrete line components in its spectrum. The data modulation and the RF cycles are cyclostationary, a property that allows carrier information and symbol timing to be recovered from a signal that contains neither a carrier component nor any discrete sideband component at the symbol rate.

For an MPSK signal, a signal in which the modulation takes on M uniformly spaced phases, the carrier-recovery circuits are able to sidelock to frequencies at integer multiples of $1/MT$ to either side of the desired carrier frequency f_c instead of the carrier itself. All of the papers cited try to explain this behavior by analyses of detailed operations within the carrier-recovery circuits. The analyses are not wrong, but they obscure an alternative viewpoint that explains more simply why the sidelocks occur. The papers cited analyze a *Costas loop* (a popular technique for carrier recovery) or other similar PLL that incorporates nonlinear operations in its phase detector. An appropriate nonlinear operation is an essential ingredient for carrier recovery from a suppressed-carrier signal.

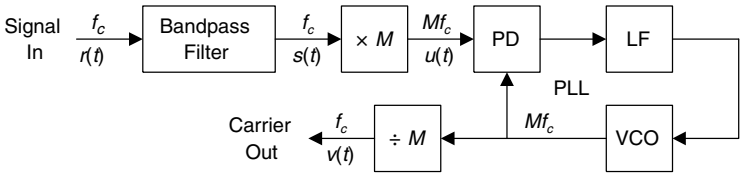


Figure 14.1 Carrier regenerator.

Rather than delve into the details of the carrier-recovery phase detectors, consider instead a *carrier regenerator* arrangement as in Fig. 14.1. A suppressed-carrier data signal $r(t)$ with carrier frequency f_c is received and applied to a bandpass filter that suppresses out-of-band noise and interference. Filter output $s(t)$ at carrier frequency f_c is applied to a $\times M$ frequency multiplier. The heart of a $\times M$ frequency multiplier is a memoryless nonlinear device of at least Mth order. (More severe nonlinearities, such as absolute value, are often beneficial.) The output $u(t)$ of the frequency multiplier is at frequency Mf_c. Furthermore, the modulation phases have also been multiplied by M, which is equivalent to rotating all M input phases into coincidence at a single phase. It is this nonlinear operation that regenerates a discrete carrier component from a signal that has no carrier component. Frequency-multiplier output $u(t)$ is delivered to an ordinary PLL acting as a narrowband filter that suppresses additive and self-noise from the regenerator and produces a clean signal at frequency Mf_c at the VCO. The output of the VCO is applied to a $1/M$ frequency divider to produce a signal $v(t)$ at frequency f_c to be used for coherent demodulation of the data signal.

Besides regenerating a carrier component from a suppressed-carrier signal, intermodulation in the nonlinearity also regenerates discrete spectral components at multiples of the symbol rate $1/T$ from Mf_c. After frequency division by M, the discrete spectral components are spaced by multiples of $1/MT$ from the carrier frequency f_c. Figure 14.2 shows the simulated spectra of $s(t)$ and $u(t)$ for a 4PSK signal passed through a fourth-power nonlinearity. The absence of all discrete components is evident in the spectrum of $s(t)$, whereas spectral spreading, a discrete carrier component at $4f_c$, and symbol-rate lines at $4f_c \pm 1/T$ are evident in the spectrum of $u(t)$. Only one pair of symbol-rate lines appear in Fig. 14.2*b* because the example signal is strictly bandlimited to an RF bandwidth of less than $2/T$. Other close-in components are too weak to be discerned, and those further out are nonexistent because of the strict bandlimiting. Additional symbol-rate-spaced components would appear in the spectrum of $u(t)$ if the input signal were not bandlimited so stringently.

Almost invariably, the bandwidth of a PLL is small compared to $1/MT$; the PLL will lock to any of these spectral components that have sufficient amplitude if the PLL tunes to its frequency. These are sidelocks, even though the input signal $r(t)$ has no discrete-frequency sidebands. How does the regenerator behavior relate to PLLs with nonlinear phase detectors, such as a Costas loop, as treated in the references cited? It can be shown that performance of a PLL with nonlinear PD is mathematically the same as that for regenerator with an

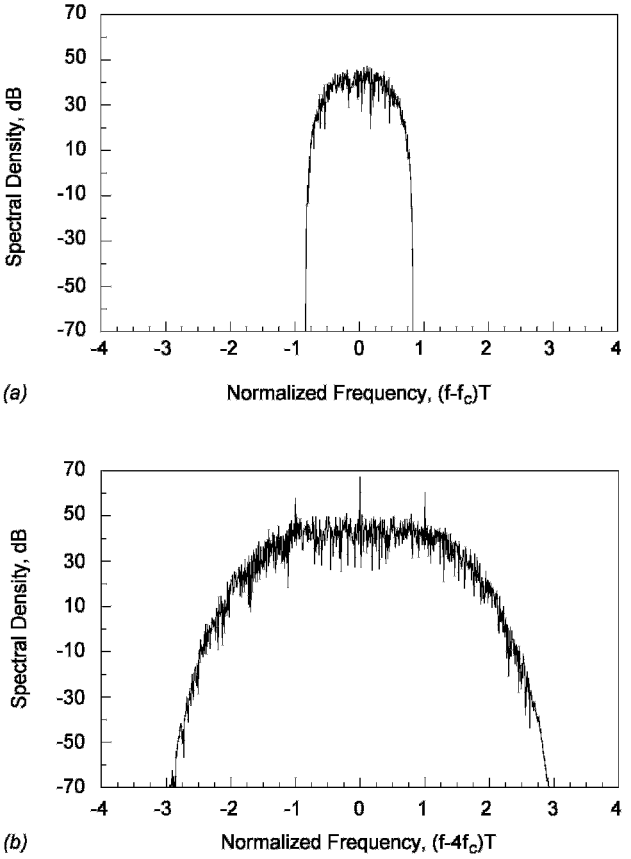


Figure 14.2 Simulated spectra of a 4PSK signal: (*a*) filtered received signal $s(t)$; (*b*) signal $u(t)$ after $\times 4$ frequency multiplication. Note the regenerated lines in (*b*): carrier at zero and symbol-rate sidelines at ± 1. (Adapted from [14.11].)

equivalent nonlinearity. For instance, the classical Costas loop (for 2PSK signals), with a nonlinearity furnished by multiplication of the I-channel by the Q-channel, is readily shown to perform in the same way as a regenerator with square-law nonlinearity: the *squaring loop* [14.12]. Visualize a nonlinear PD as a nonseparable combination of the $\times M$ nonlinearity, the ordinary phase detector, and the $1/M$ frequency divider of Fig. 14.1. Although this model is exaggerated, the clock lines appearing in Fig. 14.2 furnish a much quicker explanation of the sidelocks observed than can ever be rendered by detailed analyses of the microscopic behavior of PLLs with nonlinear PDs.

14.1.3 Alias Locks

Many nonlinear phase detectors employed in carrier synchronizers operate in a sampled manner, typically at the symbol rate $1/T$. That sampling leads to aliasing

of the incoming signal, especially significant if the signal bandwidth is substantially larger than $1/T$. Anomalous locks to aliases are analyzed in [14.13–14.15]. Alias locks intermixed with sidelocks produce a rich collection of possible wrong-lock frequencies.

14.2 HARMONIC LOCKS

Under suitable conditions, related in this section, a PLL can lock to the harmonics of the input signal frequency f_i. The lock frequency f_o of the VCO might be a subharmonic of the signal ($f_o = f_i/M$), a superharmonic ($f_o = Nf_i$), or a fractional harmonic ($f_o = Nf_i/M$), where N and M are relatively prime integers. The ability to lock (if you want harmonic locks) or the vulnerability to lock (if you do not want harmonic locks) depends on properties of the phase detector and the signals applied to it.

Consider an ideal multiplier used as a phase detector. Suppose that the two signals delivered to a PD are periodic but not necessarily sinusoidal. Since they are periodic, each can be resolved into a Fourier series of sinusoids at integer multiples of its base period. The PD multiplies each term in one Fourier series by each term in the other series; the output of the PD is the sum of these term-by-term products. Zero-order products at DC arise if and only if both terms in the product have the same frequency. The amplitudes of individual products depend on the amplitudes of the constituent Fourier terms and the phase differences between them. The amplitude of a product is a sinusoidal function of the phase angle between the two constituent terms and has the same period as that of each of the terms. This sinusoidal function constitutes an s-curve on which the PLL may be able to lock.

As an example, suppose that one input to the multiplier is a square wave and the other is a sine wave. The common switching PD (Section 10.1.1) is equivalent in its behavior. A square wave contains all odd harmonics of its fundamental frequency, so the PLL of this example potentially could lock to all odd subharmonics of the frequency of the input signal. As another example, suppose that both inputs to the PD are square waves. (An exclusive-OR PD is a practical realization of this condition.) Then fractional harmonics $f_o = Nf_i/M$ are potentially lockable for all odd N and M. As a third example, a sampling phase detector contains all harmonics, even and odd, in the impulse stream driving the sampler. The sampling PD can generate s-curves for any of these harmonics.

Harmonic operation of a multiplier-class PD is readily understood from examining harmonics common to both inputs. The same cannot be said for sequential phase detectors. Experience suggests that the popular phase/frequency detector (PFD) of Section 10.3 is free of harmonic locks, but other sequential PDs are not. No simple rule on harmonic locking has been devised for sequential PDs; each case has to be painfully worked out for itself.

The need to cope with missing transitions makes phase detectors for bit synchronizers particularly vulnerable to unwanted fractional-harmonic locks. This statement applies to both sequential-class PDs and multiplier-class PDs.

14.3 SPURIOUS LOCKS

Phase detectors explored in depth in previous chapters all have had well-behaved s-curves. Each period of each s-curve had one and only one stable point of equilibrium—only one zero crossing with proper slope. That good behavior is not always found in all s-curves for all signals. Some s-curves for particular PDs and signal formats can have more than one stable crossing, raising the possibility of *spurious locks* to the wrong phase.

An example shown in Fig. 14.3 is for a carrier phase-error detector for 16QAM data signals using a decision-directed algorithm $u_d[n] = \text{Im}\{c_n^* s[n]\}$, where $s[n]$ is the complex value of the sample of the nth symbol in a data stream and c_n^* is the complex conjugate of the estimate of the nth symbol. This algorithm is widely used for phase detection for QAM signals. Although the example algorithm has a digital formulation, similar behavior occurs with analog implementation [14.16].

The s-curve in Fig. 14.3 is drawn for just one octant of the circle. The entire s-curve is skew symmetric in the octant to the left and repeats periodically in each quadrant. (s-Curve periodicity in quadrants is inherent for any signal with quadrantal symmetry.) The desired stable crossing of the s-curve is located at zero phase error. The curve is well behaved to about 16° and then deteriorates sharply for larger errors because of mistakes in decisions on the symbol value c_n. Mistakes occur because the phase error has rotated some of the symbol samples into the wrong decision cells. Because of these mistakes, the resulting s-curve has two spurious positive-slope zero crossings, one at about 31° and another at

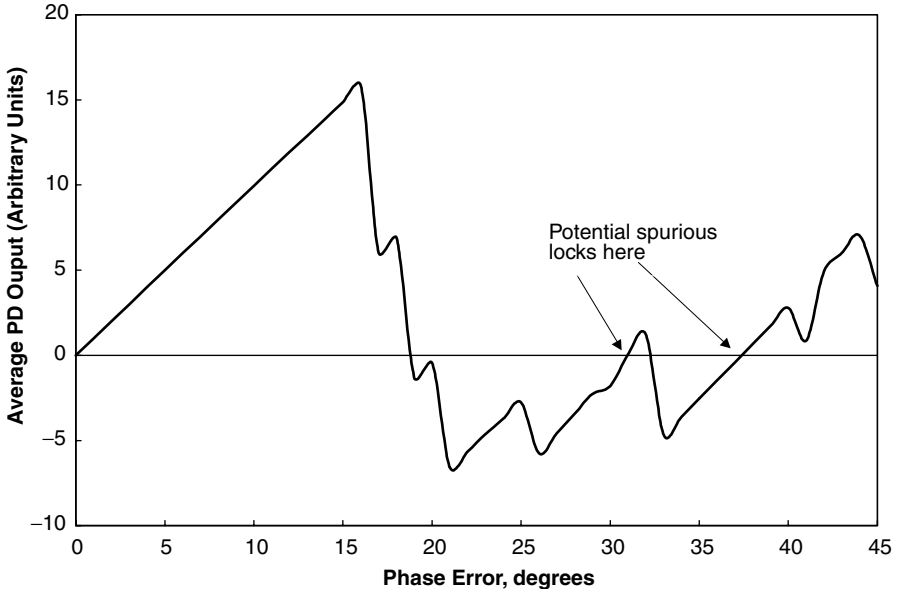


Figure 14.3 One octant of the s-curve of a phase detector for a 16QAM signal.

about 38°. The PLL might lock at either of these crossings if one of them were encountered before phase of the PLL reached the correct crossing at 0°.

The lockable phase ranges about these spurious crossings obviously are not as extensive as about the desired crossing at zero error, so the spurious locks will be comparatively weak in the presence of disturbances. An adequate rate of phase sweep (frequency offset) during acquisition of lock would drive the phase error right through these potential spurious locks and not stop before reaching the correct lock point.

Another expedient is to employ only a subset of the QAM constellation (such as only the four innermost points) for phase detection, thereby obtaining an *s*-curve without spurious crossings. Another technique is to employ a simpler constellation for acquisition (e.g., 4QAM), thereby avoiding spurious locks, and then switch to a larger constellation after correct lock has been acquired. Spurious locks are not confined to phase detectors for data signals; they might arise for other kinds of phase detectors and signals. A design engineer should always know the *s*-curve of a PD to avoid unwelcome surprises.

14.4 FALSE LOCKS

The preceding anomalous locks have all been genuine phase locks, even though the lock was to the wrong frequency or phase. This section deals with *false locks* that are not phase locks at all, but aberrations of the pull-in mechanism. False lock might prevent phase lock entirely. The explanation for false locks historically has been based on phase shifts in passband filters within the feedback loop, located in the intermediate frequency (IF) stages of phaselock receivers. This account follows the established approach, but keep in mind that excessive phase shifts in baseband circuits within the PLL also lead to false locks.

A simplified block diagram of a typical superheterodyne phaselock receiver is shown in Fig. 14.4. The incoming signal at frequency f_1 is mixed down to a

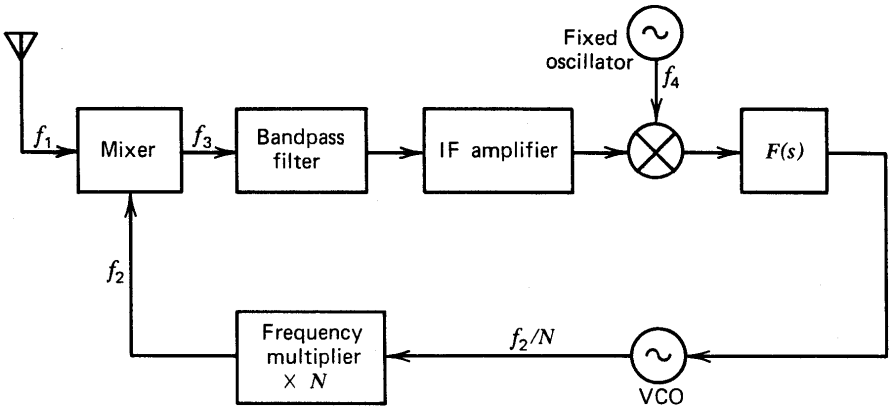


Figure 14.4 Long-loop phaselock receiver.

convenient intermediate frequency labeled f_3. A fixed oscillator at frequency f_3 is compared against IF amplifier output in a phase detector; the loop is closed through the loop filter, VCO, frequency multiplier, and mixer. Simple phaselock loops of the kind treated in earlier chapters are often known as *short loops*; the more complicated loop of Fig. 14.4 is called a *long loop*, for obvious reasons.

14.4.1 IF Filter Analysis

A narrowband IF filter is often employed as a means to provide a satisfactory signal-to-noise ratio at the phase detector (see Section 10.4.2). False locks were observed, to the mystification of early victims, in phaselock receivers using narrowband IF filters with steep skirts. To see how false lock comes about, it is first necessary to devise a method of bringing the IF bandpass filter into the linear analysis of the PLL. To this end, consider the hypothetical test setup of Fig. 14.5. What is the effect of the filter on the modulation of the test signal? Specifically, the amplitude and phase of the modulation output compared with the modulation input are desired as a function of modulation frequency.

The result, expressed as a modulation transfer function, denoted $F_m(s)$, is stated but not proved. If (1) the filter has a narrow, symmetrical passband, (2) the signal generator is tuned to the center frequency of the filter, and (3) the modulation deviation is very small, the approximate one-sided modulation transfer function is obtained by translating the actual filter transfer function to zero frequency and discarding the response at negative frequencies, as shown in Fig. 14.6. An equivalent two-sided response is derived in [14.17].

Now suppose the open-loop response of the PLL of Fig. 14.4 were to be measured by opening the loop in its low-frequency portion and applying a low-frequency sinusoidal test signal. Total open-loop response would consist of the product $F_m(s)G(s)$, whose factors are the normal response $G(s)$ of the loop and the modulation transfer function $F_m(s)$ of the IF filter. This combined open-loop response is substituted into the transfer functions of Chapters 2 and 3 for determination of pole locations, stability, damping, and all the other valuable tools of linear analysis. The effect of the IF filter is the same as if additional lowpass filters were incorporated into the baseband portion of the PLL. In particular, false lock is possible if too many lowpass poles are present in the baseband portion of a PLL, even in a short loop with no bandpass filters at all. The analysis that

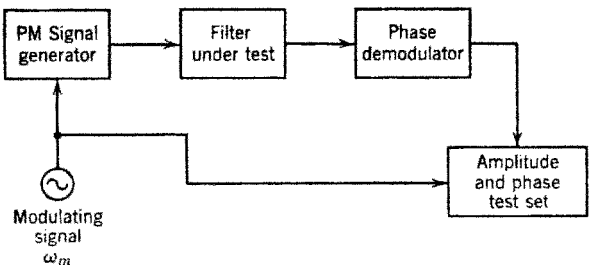


Figure 14.5 Test setup for measurement of a modulation transfer function $F_m(s)$.

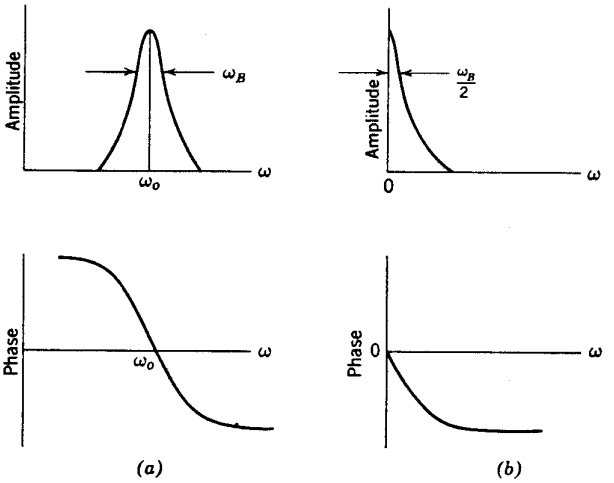


Figure 14.6 Transfer characteristics of a bandpass filter: (*a*) bandpass transfer function; (*b*) equivalent modulation transfer function.

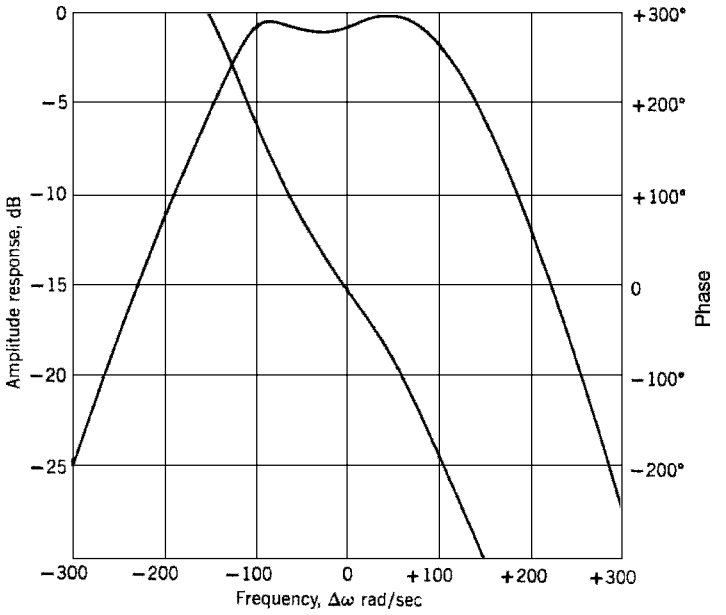


Figure 14.7 Example bandpass frequency response (from measurements on a crystal filter).

follows deals only with analog PLLs, but similar analysis can be applied to digital PLLs, which are equally subject to false locks.

To provide an example of a bandpass filter, Fig. 14.7 shows a response scaled from measurements on an actual crystal filter. The equivalent modulation transfer response is shown in the Bode plot of Fig. 14.8, along with that of a basic

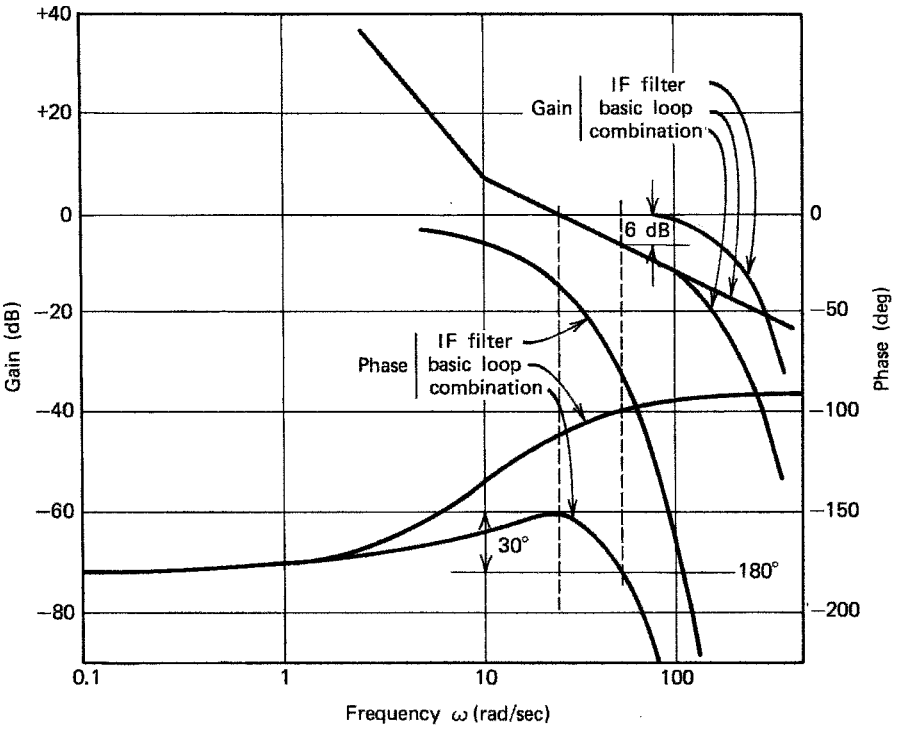


Figure 14.8 Bode plot of a long-loop PLL containing an example crystal IF filter.

second-order type 2 PLL and the combination of filter and basic PLL. The bandwidth of the IF filter (3 dB) is 240 rad/sec, whereas the loop has been chosen arbitrarily with $1/\tau_2 = 10$ rad/sec. Loop gain (ignoring the IF filter) has been selected so that $\zeta = 0.707$, so $\omega_n = 14.1$ rad/sec and $K = 20$ rad/sec. These numbers are reasonable for a phaselocked receiver with a very narrow bandwidth, such as those employed in deep-space applications.

The combined Bode plot shows a phase margin of 30° and a gain margin of 6 dB. Although the loop is stable, its response is very different from that expected in the absence of the IF filter. If loop gain is fixed (by AGC or limiter) so that it cannot exceed the value used for the example, the stability margins are barely adequate, not ample. However, if the gain of the example is a threshold gain and increases of gain are to be expected with improved signals, the gain margin is completely inadequate. If the gain doubles, the loop will oscillate. A more conservative design would use a substantially wider IF filter bandwidth.

14.4.2 Origin of False Locks

Even if the loop transfer function is stable, a narrow IF filter can cause false locks in which frequency acquisition halts and the PLL appears to lock at a frequency

that bears no obvious relation to the input frequency. Until the source of false lock is recognized, the phenomenon can be a disturbing and mystifying experience. Subsequent pages describe how false lock, or the related problem of frequency pushing, is a disorder of the pull-in mechanism described in Section 8.3.1 and is almost inevitable in some degree in a PLL that includes extra filtering or delay. Existence of false lock is yet another reason not to rely on pull-in for the frequency-acquisition method.

Investigations of false lock have been reported in [14.18–14.21]. The approximate analysis presented here follows a slightly different approach. Consider an unlocked loop with input $V_s \sin \omega_i t$ and VCO output $V_o \cos \omega_o t$. Phase detector output is a beat note at a frequency $\Delta\omega_i = \omega_i - \omega_o$. If $\Delta\omega_i$ is sufficiently larger than the loop gain K, the beat note will be nearly sinusoidal and take the form $K_d \sin(\Delta\omega_i t)$. In passing through the loop, the beat note is attenuated by a factor $\eta(\Delta\omega_i)$ and phase shifted by an angle $\psi(\Delta\omega_i)$. Frequency-modulating voltage applied to the VCO is $\eta K_d \sin(\Delta\omega_i t + \psi)$, so the VCO output is (approximately)

$$v_o(t) = V_o \cos\left[\omega_o t - \frac{\eta K_o K_d}{\Delta\omega_i} \cos(\Delta\omega_i\, t + \psi)\right] \tag{14.1}$$

The spectrum (Fig. 8.6) of $v_o(t)$ consists of a carrier line at ω_o and an infinite series of sideband lines at frequencies $\omega_o + k\Delta\omega_i$. The line for $k = 1$ is at a frequency of $\omega_o + (\omega_i - \omega_o) = \omega_i$, which is exactly the input frequency. Using a Fourier series analysis, the VCO component at ω_i is found to be

$$V_o J_1\left(\frac{\eta K_o K_d}{\Delta\omega_i}\right) \sin(\omega_i t + \psi) \tag{14.2}$$

where $J_1(\cdot)$ is the first-order Bessel function of the first kind.

When this line is multiplied in the phase detector against the input signal $V_s \sin(\omega_i t)$, the resulting DC component is

$$V_d = \frac{1}{2} V_s V_o K_m J_1\left(\frac{\eta K_o K_d}{\Delta\omega_i}\right) \cos\psi = K_d J_1\left(\frac{\eta K_o K_d}{\Delta\omega_i}\right) \cos\psi \tag{14.3}$$

where K_m is the multiplier gain coefficient, as defined in Section 6.1.1.

For an example: In a standard, second-order type 2 loop, for large enough $\Delta\omega_i$ and in the absence of an IF filter, the parameters η and ψ are $\eta = \tau_2/\tau_1$ and $\psi = 0$. Since $K_o K_d \tau_2/\tau_1 = K$ for this special case, equation (14.3) becomes

$$V_d \approx K_d J_1\left(\frac{K}{\Delta\omega_i}\right) \tag{14.4}$$

Equation (14.4) is an approximation to the pull-in voltage v_p of (8.6); the two expressions agree asymptotically for a large frequency difference and disagree by less than 10% if $|\Delta\omega_i| > 2K$.

Now suppose that additional filtering is added into the standard loop. It is very difficult to avoid adding at least one extra pole for ripple filtering, the operational amplifier in an active filter contributes at least one more pole, and a third pole in the VCO control line is nearly inescapable. If a long loop is used, the filters in the IF amplifier contribute additional equivalent lowpass poles. Up to a dozen extra poles are not at all unusual. Define a relative attenuation coefficient

$$\eta' = \frac{\eta K_o K_d}{K} \tag{14.5}$$

In the standard loop, $\eta' = 1$. Departure of $\eta'(\Delta\omega_i)$ from unity describes the magnitude response of any additional filtering within a physical loop. Accordingly, (14.4) is modified to

$$V_d \approx K_d J_1\left(\frac{\eta' K}{\Delta\omega_i}\right)\cos\psi \tag{14.6}$$

The pull-in voltage (14.4) of the standard loop is multiplied by the cosine of the added phase shift. For $K/\Delta\omega_i \ll 1$ (the only region of validity for the approximations of this analysis) the Bessel function is approximated by

$$J_1\left(\frac{\eta' K}{\Delta\omega_i}\right) \approx \frac{1}{2}\frac{\eta' K}{\Delta\omega_i} \tag{14.7}$$

so the pull-in voltage is further reduced by a factor η'. A suitable approximation of the pull-in voltage, including the effects of additional filtering, is

$$V_d \approx \frac{\eta' K_d K}{2\Delta\omega_i}\cos\psi \tag{14.8}$$

If $\eta'(\Delta\omega_i)$ and $\psi(\Delta\omega_i)$ are known, the pull-in and false-lock properties of the loop may be calculated from (14.8).

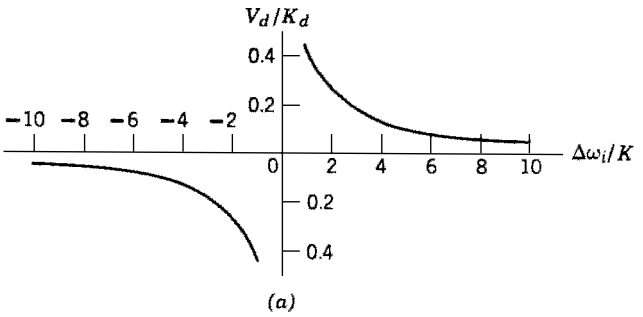
Strictly speaking, the abbreviated analysis presented above applies directly only to a short loop. When the analysis is modified to take account of a long loop, the DC output of the PD can be estimated simply by cascading the equivalent modulation transfer function $F_m(s)$ with the actual loop filter $F(s)$ and calculating a new η' and ψ, provided that the bandpass amplifier is linear. If the bandpass circuit contains a limiter, the bandpass contribution to ψ is unaffected by the nonlinearity, but the contribution to η is more complicated. At large SNR the limiter tends to wipe off any influence on η contributed by bandpass networks preceding the limiter.

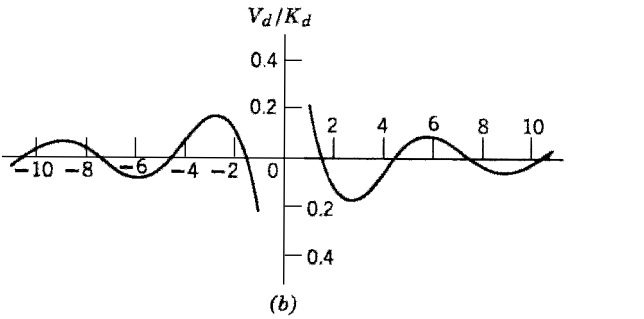
14.4.3 False-Lock Properties

As an example, let excess phase be $\psi = -(\pi/3)(\Delta\omega_i/K)$ and $\eta' = 1$, a fair approximation to the IF filter and PLL shown in Figs. 14.7 and 14.8. (Take note

that this approximate phase is equivalent to a simple delay of $\tau = \pi/3K$; the analysis works correctly for pure delays as well as for more general filters with nonconstant delay.) Using this expression for ψ, the DC phase-detector output is as plotted in Fig. 14.9b. Immediately evident in the plot are nulls of the pull-in voltage corresponding to the zeros of cos ψ, nulls that do not occur in the standard loop (Fig. 14.9*a*). The polarity of V_d is unchanged from that of the standard loop for small $\Delta\omega_i$, so pull-in occurs correctly, albeit more weakly, because of the reduced amplitude of V_d.

However, if the frequency difference is somewhat outside the first null, the polarity of V_d is reversed from standard, and pull-in no longer proceeds normally. Instead, the reversed polarity causes the loop to push out away from the correct lock frequency. Pushing continues until the frequency difference increases to coincide with the second null, which is a stable tracking point of false lock. A true phaselock is not achieved at the false-lock null—a frequency error still exists—but the loop is unable to move itself away from the null.





$$V_d \simeq \frac{\eta' K_d K}{2\Delta\omega_i} \cos\psi$$

Figure 14.9 PLL pull-in characteristics showing the effect of excess phase shift: (*a*) standard loop only; (*b*) loop with excess phase shift from an example crystal filter.

A false lock can be very confusing to an operator. Output from the loop phase detector will have zero DC component, whereas the quadrature PD (correlation detector) will show a DC output, indicating that lock has been achieved. If coherent AGC is used, the magnitude of the quad PD output might even be correct for indicating lock. An oscilloscope connected to the PD output will show the presence of a beat note, but only if noise is small enough. In fact, it is possible that a false lock may go completely unrecognized—until ridiculous data become apparent.

Obviously, false locks must be avoided. One method of avoidance is to use an IF filter of sufficient bandwidth. Another is to recognize that phase shift, for a given bandwidth, increases with the number of equivalent lowpass poles in the filter. If only a single-tuned circuit is used, maximum phase shift in the filter is 90° and there is no finite false-lock null.

With two tanks (two poles in the equivalent low pass modulation transfer function) the maximum phase shift is 180° and the only finite spurious nulls are unstable. Frequency pushing beyond these finite nulls drives the PLL frequency to one or the other limits of its tuning range, not to the correct signal frequency.

Rough sketches of pull-in voltage for various numbers of poles are shown in Fig. 14.10. Actual false locks are encountered only if there are four or more poles

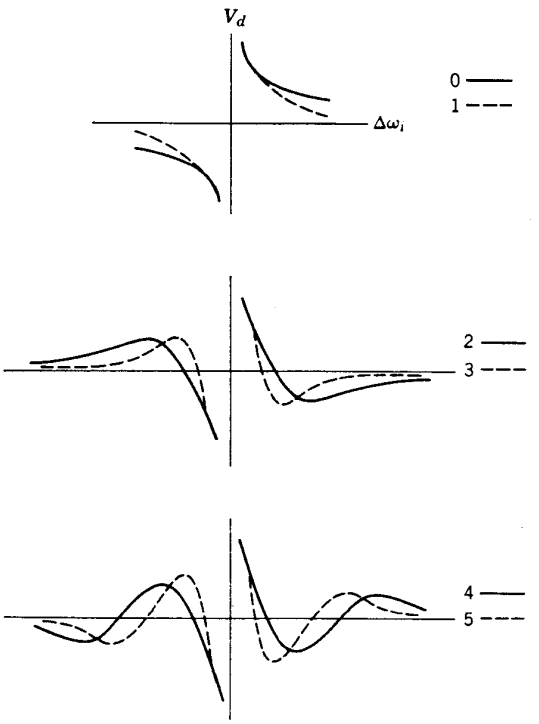


Figure 14.10 PLL pull-in characteristics. Numerals indicate an equivalent number of extra lowpass poles in the loop.

in the lowpass equivalent filter. Numerous poles are found in filters with very steep skirts—so-called rectangular filters. Evidently such filters are not entirely suitable for use in a phaselock receiver. A conservative design for the IF filter would utilize only one or two poles. (A single quartz crystal conveniently provides one equivalent pole.) Actually, there are certain to be other band-restricting elements within the loop, and there will always be more excess phase shift than is provided by the recognizable poles. The main IF filtering should be kept simple to provide some margin against these secondary effects, not all of which are easily predicted.

The foregoing analysis takes into account only the normal signal path through the loop. Unfortunately, bitter experience has shown that insidious paths often contribute more to false lock than does the obvious main path. Beat-note coupling through an inadequately isolated power supply line is a particular offender.

14.4.4 Remedies for False Lock

Chapter 8 demonstrated that the maximum trackable frequency sweep rate depends on bandwidth; a narrowband loop can track only a slowly changing frequency. Therefore, if acquisition is performed by sweep techniques, it may be possible to sweep rapidly enough that the false locks will be unable to hold but slowly enough to succeed in acquiring correct lock. This possibility is complicated by any limiters or AGC that may be used and by IF signal-to-noise ratio. Another alternative, if input SNR is large enough, is to employ a frequency discriminator to aid frequency acquisition. The output of the discriminator has to be large enough to overcome any of the wrong-polarity phase-detector outputs associated with false lock or frequency pushing.

The best remedy, though, is a *split-loop receiver*, devised by McGeehan and Sladen [14.22]. A block diagram is shown in Fig. 14.11. In this arrangement, the two paths of a type 2 PLL are well separated; they have separate baseband filters, separate VCOs, and span different portions of the receiver. Recollect

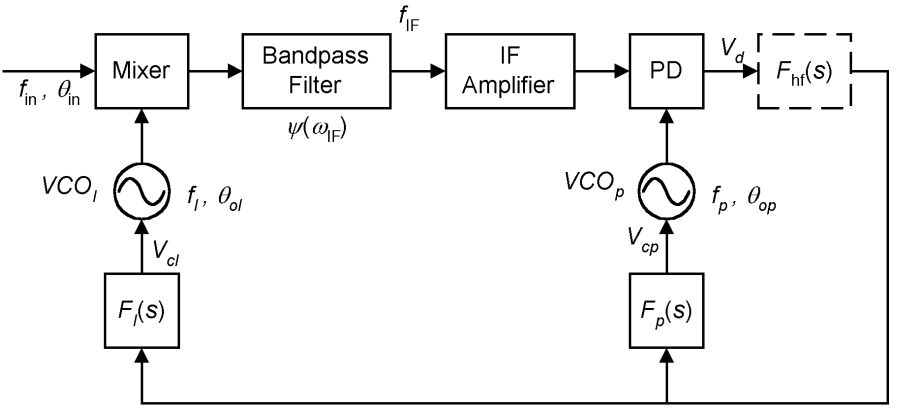


Figure 14.11 Block diagram of a split-loop PLL.

(Section 8.3.1) that pull-in voltage v_p is generated mainly through operations in the proportional path, with a negligible contribution from the integral path. The proportional path in a split-loop receiver is contained in a short loop that does not encompass the phase shifts of the IF filter. The integral path in the split loop is connected in a long loop that does include the IF filter, but the phase shifts in that filter now have a negligible influence on the pull-in voltage. In consequence, a split loop avoids false locks that would arise in a long loop because of phase shifts in the IF filter.

Transfer Functions In Fig. 14.11, let $F_p(s) = K_1$ and $F_I(s) = K_2/s$. Gains of the VCOs are K_{op} and K_{oI}, the PD has gain K_d, and the phases out of the two VCOs are designated θ_{op} and θ_{oI}. Using the methods introduced in Chapter 2, the error response transfer function is

$$E(s) = \frac{\theta_e}{\theta_{in}} = \frac{s^2}{s^2 + K_d F_{\text{hf}}(s)(sK_1K_{op} + K_2K_{oI})} \tag{14.9}$$

from which loop gain is identified as $K = K_d F_{\text{hf}}(0) K_1 K_{op}$ rad/sec. If $F_{\text{hf}}(s) \equiv 1$, the PLL is second-order with $\omega_n^2 = K_d K_2 K_{oI}$ and damping

$$\zeta = \frac{K_1 K_{op}}{2} \sqrt{\frac{K_d}{K_2 K_{oI}}} \tag{14.10}$$

The open-loop gain $G(s)$ is found from $E(s) = 1/[1 + G(s)]$ as

$$G(s) = K_d F_{\text{hf}}(s) \left(\frac{K_1 K_{op}}{s} + \frac{K_2 K_{oI}}{s^2} \right) \tag{14.11}$$

Because the split loop has two VCOs, there is no obvious single definition of closed-loop transfer function of the system. Instead, two system transfer functions can be defined, one for each VCO. For the proportional-path loop the transfer function is

$$H_p(s) = \frac{\theta_{op}}{\theta_{in}} = \frac{sK_d F_{\text{hf}}(s) K_1 K_{op}}{s^2 + K_d F_{\text{hf}}(s)(sK_1K_{op} + K_2K_{oI})} \tag{14.12}$$

and for the integral path

$$H_I(s) = \frac{\theta_{oI}}{\theta_{in}} = \frac{K_d F_{\text{hf}}(s) K_2 K_{oI}}{s^2 + K_d F_{\text{hf}}(s)(sK_1K_{op} + K_2K_{oI})} \tag{14.13}$$

Observe that $H_I(s)$ is an all-pole transfer function (if F_{hf} has only poles and no finite zeros), so that it will exhibit gain peaking (see Section 2.2.4) only if it has sufficiently underdamped poles. Stated differently: $H_I(s)$ can be designed to have no gain peaking whatsoever if that should be advantageous. Notice also that

$H_p(s)$ has bandpass response with a null at zero frequency. That is, the integral path has full control over the DC steady-state response of the split loop, with no contribution from the proportional path.

Frequencies at Phaselock Assuming low-side injection at the mixer, the receiver frequencies are related by $f_{\text{in}} - (f_I + f_p) = 0$ when the PLL is locked. (High-side injection is accommodated by appropriate changes of signs in this formula.) The formula constrains only the sum $(f_I + f_p)$, not the two VCO frequencies individually. What are the individual frequencies f_I and f_p once the loop has locked? Assuming a perfect integrator in the integral path, a constant frequency f_{in}, a noise-free input and the absence of unwanted DC offsets in the phase detector or in the loop filter, the phase error goes to zero at equilibrium lock. Represent the tuning rule for VCO_p as $\omega_p = \omega_{0p} + K_{op}V_{cp}$, where V_{cp} is the control voltage and ω_{0p} is the "free-running" frequency of VCO_p. If the phase error is zero, the PD output voltage is also zero and so is the control voltage V_{cp}. Therefore, the frequency of VCO_p at lock is ω_{0p}, as is the signal frequency in the IF amplifier.

If the integrator is imperfect (as are all analog-circuit integrators), the proportional path will take on some part of the DC tracking burden and VCO_p will be detuned by some amount from ω_{0p}. Digital PLLs can have integrators with infinite DC gain so that they need not experience this steady-state detuning.

14.5 LOCK FAILURES IN CHAINS OF PLLs

So far, Chapter 14 has dealt with locks to an incorrect frequency or phase. This last section deals with lock failures in chains of PLLs under conditions of negligible additive noise, conditions that appear superficially favorable. Long-distance data communication links operating over wire lines or optical fibers often incorporate numerous repeaters in a chain. A repeater consists of a synchronizer that recovers the timing of its input and a regenerator that detects each data symbol and delivers a cleaned-up, retimed data stream at its output. A typical repeater employs a PLL for its synchronizer, although there are other schemes that employ bandpass filters instead. Experience has revealed that the link fails due to excessive cycle slips if the chain contains too many repeaters.

The failures are analyzed in [14.23] and [14.24] and in references cited in those papers. In essence, the randomness of the data stream induces a certain amount of jitter in the output of each synchronizer. Each synchronizer sees the same data stream, so the same jitter is induced in each repeater and passed to the next repeater on the regenerated data stream. Jitter accumulates along the chain; input to the nth repeater has jitter from each preceding synchronizer, each contribution filtered by the transfer functions of intervening synchronizers. The spectrum of the accumulated jitter has a strong peak at a frequency comparable to the bandwidth of the PLL. Synchronizers too far down the chain, where jitter grows too large, will slip cycles rather than track properly.

The usual second-order type 2 PLL necessarily has a certain amount of gain peaking in its closed-loop response $H(s)$, as described in Section 2.2.4. Gain peaking constitutes jitter amplification, a potential disaster in a chain of synchronizers. Constraint of jitter amplification is the reason for tight limits (commonly, 0.1 dB maximum) on gain peaking specified for synchronizers used in telecommunications networks. Section 2.2.4 showed that a damping factor ζ of not less than 4.5 is required to achieve gain peaking no more than 0.1 dB.

Some designers employing damping factors of 20 to 30 still discovered consistent lock failures in long chains. Indeed, [14.24] analyzes synchronizer chains with first-order PLLs whose transfer function does not gain any exhibit peaking but nonetheless suffer cycle slips from induced jitter. One concludes that gain peaking is an aggravating factor in chain failures but not the basic cause. Jitter induced in a synchronizer in a low-noise environment (as is typical of many landline networks) is largely due to self-noise [14.25] resulting from the randomness of the data stream and restricted bandwidth of the transmission medium. Methods for suppression of self-noise are well known [14.26–14.28] but not often applied in synchronizers for wire or fiber lines because of complexity and cost.

Instead, use of *jitter attenuators* has become commonplace (see Section 17.5.2). A jitter attenuator is a combination of an elastic buffer [a first-in first-out buffer (FIFO)] and a PLL with a small loop bandwidth. A data signal is received in a regular data receiver using a phaselocked synchronizer with a bandwidth sufficiently large to track incoming jitter reliably. Data output from the receiver is clocked into the FIFO by the clock output from the large-bandwidth data synchronizer and clocked out by the clock output from the small-bandwidth PLL of the jitter attenuator. The fill indicator of the FIFO serves as the phase detector for the jitter-attenuator PLL; the idea is to maintain an average fill of 50% of FIFO capacity.

Jitter at frequencies outside the PLL bandwidth are attenuated by the lowpass filtering of phase provided by PLLs. In particular, the large peak in spectrum of the accumulated jitter is well outside the bandwidth of a practical jitter-attenuating PLL and so is strongly attenuated. Jitter of low-enough frequency passes through the attenuator PLL, but its amplitude typically is small enough that the wideband synchronizers of downstream repeaters can cope with it.

Large-amplitude jitter is absorbed by the FIFO, jitter that would require large bandwidth in a data synchronizer to assure reliable lock. The jitter attenuator can tolerate large jitter at its input and still have a small bandwidth in its PLL. Notice that input to the attenuator PLL is a clock signal, not a data signal. No self-noise of the kind that afflicts the synchronizer PLL exists to cause additional jitter in the attenuator PLL.

REFERENCES

14.1 D. Richman, "Color-Carrier Reference Phase Synchronization Accuracy in NTSC Color Television," *Proc. IEEE* ***43***, 106–133, Jan. 1954. Reprinted in [14.2].

14.2 W. C. Lindsey and M. K. Simon, eds., *Phase-Locked Loops & Their Application*, Reprint Volume, IEEE Press, New York, 1978.

14.3 B. R. Eisenberg, "Gated Phase-Locked Loop Study," *IEEE Trans. Aerosp. Electron. Syst.* ***AES-7***, 469–477, May 1971.

14.4 U. Mengali, "Noise Performance of a Gated Phase-Locked Loop," *Trans. IEEE Aerosp. Electron. Syst.* ***AES-9***, 55–59, Jan. 1973.

14.5 L. Schiff, "Burst Synchronization of Phase-Locked Loops," *IEEE Trans. Commun.* ***COM-21***, 1091–1099, Oct. 1973.

14.6 G. L. Hedin, J. K. Holmes, W. C. Lindsey, and K. T. Woo, "Theory of False Lock in Costas Loops," *IEEE Trans. Commun.* ***COM-26***, 1–12, Jan. 1978. Reprinted in [14.9].

14.7 K. T. Woo, G. K. Huth, W. C. Lindsey, and J. K. Holmes, "False Lock Performances of Shuttle Costas Loop Receivers," *IEEE Trans. Commun.* ***COM-26***, 1703–1712, Nov. 1978. Reprinted in [14.9].

14.8 M. K. Simon, "The False Lock Performance of Costas Loops with Hard-Limited In-Phase Channel," *IEEE Trans. Commun.* ***COM-26***, 23–34, Jan. 1978. Reprinted in [14.9].

14.9 W. C. Lindsey and C. M. Chie, eds., *Phase-Locked Loops*, Reprint Volume, IEEE Press, New York, 1986.

14.10 S. T. Kleinberg and H. Chang, "Sideband False-Lock Performance of Squaring, Fourth-Power, and Quadriphase Costas Loops for NRZ Data Signals," *IEEE Trans. Commun.* ***COM-28***, 1335–1342, Aug. 1980.

14.11 F. M. Gardner and J. D. Baker, *Simulation Techniques*, Wiley, New York, 1997, p. 316.

14.12 W. C. Lindsey and M. K. Simon, *Telecommunication Systems Engineering*, Prentice Hall, Englewood Cliffs, NJ, 1973, Secs. 2–4 and 2–5.

14.13 K. Kiasaleh, "On False Lock in Suppressed Carrier MPSK Tracking Loops," *IEEE Trans. Commun.* ***COM-39***, 1683–1697, Nov. 1991.

14.14 M. K. Simon and K. T. Woo, "Alias Lock Behavior of Sampled-Data Costas Loops," *IEEE Trans. Commun.* ***COM-28***, 1315–1325, Aug. 1980.

14.15 T. Shimamura, "On The False-Lock Phenomena in Carrier Tracking Loops," *IEEE Trans. Commun.* ***COM-28***, 1326–1334, Aug. 1980.

14.16 M. K. Simon and J. G. Smith, "Carrier Synchronization and Detection of QASK Signal Sets," *IEEE Trans. Commun.* ***COM-22***, 98–105, Feb. 1974.

14.17 R. Lawhorn and C. S. Weaver, "The Linearized Transfer Function of a Phase Locked Loop Containing an IF Amplifier," *Proc. IRE* ***49***, 1704, Nov. 1961.

14.18 J. A. Develet, Jr., "The Influence of Time Delay on Second-Order Phase Lock Loop Acquisition Range," *Int. Telem. Conf.*, London, 1963, pp. 432–437. Reprinted in [14.2].

14.19 W. A. Johnson, *A General Analysis of the False-Lock Problem Associated with the Phase-Lock Loop*, Rep. TDR-269 (4250-45)-1, Aerospace Corp., Los angeles, CA, Oct. 2, 1963 (NASA Accession N64-13776).

14.20 R. C. Tausworthe, *Acquisition and False-Lock Behavior of Phase-Locked Loops with Noisy Inputs*, JPL SPS 37–46, Vol. IV, pp. 226–234, Jet Propulsion Laboratory, Pasadena, CA, Aug. 31, 1967.

14.21 B. N. Biswas, P. Banerjee, and A. K. Bhattacharya, "Heterodyne Phase Locked Loops—Revisited," *IEEE Trans. Commun.* ***COM-25***, 1164–1170, Oct. 1977.

14.22 J. P. McGeehan and J. P. H. Sladen, "Elimination of False-Locking in Long Loop Phase-Locked Receivers," *IEEE Trans. Commun.* ***COM-30***, 2391–2397, Oct. 1982.

14.23 H. Meyr, L. Popken, and H. R. Mueller, "Synchronization Failures in a Chain of PLL Synchronizers," *IEEE Trans. Commun.* ***COM-34***, 436–445, May 1986. Reprinted in [14.9].

14.24 M. Moeneclaey, S. Starzak, and H. Meyr, "Cycle Slips in Synchronizers Subject to Smooth Narrow-Band Loop Noise," *IEEE Trans. Commun.* ***36***, 867–874, July 1988. Comments and discussion: *IEEE Trans. Commun.* ***45***, 19–22, Jan. 1997.

14.25 F. M. Gardner, "Self-Noise in Synchronizers," *IEEE Trans. Commun.* ***28***, 1159–1163, Aug. 1980.

14.26 L. E. Franks and J. P. Bubrouski, "Statistical Properties of Timing Jitter in a PAM Timing Recovery Scheme," *IEEE Trans. Circuits Syst.* ***CAS-21***, 489–496, July 1974.

14.27 A. N. D'Andrea and M. Luise, "Design and Analysis of a Jitter-Free Clock Recovery Scheme for QAM Systems," *IEEE Trans. Commun.* ***41***, 1296–1299, Sept. 1993.

14.28 A. N. D'Andrea and M. Luise, "Optimization of Symbol Timing Recovery for QAM Data Demodulators," *IEEE Trans. Commun.* ***44***, 399–406, Mar. 1996.

CHAPTER 15

PLL FREQUENCY SYNTHESIZERS

Synthesizers are employed in an ever-wider variety of electronic products to generate any one of a number of operating frequencies. Synthesizers based on PLLs are popular because of their potential excellent performance, relative simplicity, and low cost. Phaselock synthesizers have received abundant attention in books such as [15.1–15.7], and the journal literature. Synthesizers were a subject of intense current research and copious innovation at the time this book was written. Important new results undoubtedly have emerged subsequently. This chapter is an abbreviated summary of the basic principles of phaselock synthesizers. It is intended as a guide to the subject; see the references for more thorough treatments.

15.1 SYNTHESIZER CONFIGURATIONS

Phaselock synthesizers have diverse configurations; several examples are provided in this section and variations are examined in a later section.

15.1.1 Basic Configuration

Figure 15.1 shows the basic configuration of a phaselock synthesizer, a configuration on which all the others are based. The synthesizer contains a reference source at frequency f_r and a VCO at frequency f_o. The reference frequency is divided by an integer R to the comparison frequency $f_c = f_r/R$, and the

Phaselock Techniques, Third Edition, by Floyd M. Gardner

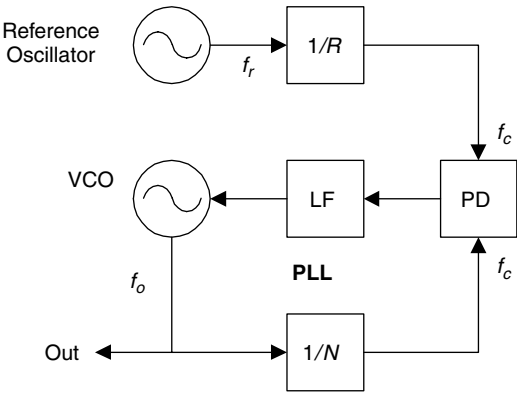


Figure 15.1 Basic PLL synthesizer.

VCO frequency is divided by N. The two divided waves are then compared in a phase detector. Phaselocking imposes the condition of $f_r/R = f_o/N = f_c$, so the output frequency is locked to a rational fraction of the reference according to

$$f_o = \frac{Nf_r}{R} = Nf_c \tag{15.1}$$

This is the basic equation of PLL synthesizers.

Frequency dividers can be programmable. The output frequency f_o is selected by setting the divider ratios R and N. Dividers are mostly implemented with digital counters, although there is a niche for other kinds of dividers, as pursued further in Section 15.2.1. Because frequency is inversely proportional to R (the period is directly proportional to R) and because uniform frequency intervals are needed ordinarily, the R divider is usually held fixed in any one application. For that reason, the sequel concentrates mainly on the comparison frequency f_c and not on $f_r = Rf_c$.

Fractional long-term stability and accuracy of the output frequency are the same as that of the reference. Fractional accuracy is the frequency error as a fraction of the reference or output frequency, as applicable. Output phase noise ideally (but not realizably) is that of the reference times N/R for jitter frequencies within the loop bandwidth, and it is that of the VCO for jitter frequencies outside the loop bandwidth.

The output frequency f_o of the basic synthesizer is selectable in increments f_c, the phase-comparison frequency. Loop bandwidth must be substantially smaller than f_c to suppress ripple adequately and to assure loop stability. If the desired increments are small, the loop bandwidth must be extremely small. On the other hand, a large loop bandwidth is preferred so as to achieve rapid acquisition and to stabilize the short-term jitter of the VCO. A severe conflict exists between these competing goals, a conflict that underlies great efforts that have been expended on phaselock synthesizers.

15.1.2 Alternative Configurations

Two modified configurations are illustrated in Figs. 15.2 and 15.3. They are illustrative of many different configurations that share similar techniques for circumventing the frequency spacing vs. bandwidth conflict of the basic PLL synthesizer configuration.

Output Division In Fig. 15.2, the VCO frequency is P times the output frequency of Nf_c/P. Frequency increments of f_c/P are obtained, even though the phase comparison is performed at a frequency of f_c. The bandwidth conflict is relieved by a factor of P at the cost of operating the VCO and the N and P dividers at P times the output frequency desired. The technique is an economical solution to a serious problem, but limitations on VCO frequency and divider speed inhibit its general application.

Multiple-Loop Synthesizers The multiple loops of Fig. 15.3 combine output division and frequency-translation mixers to avoid the conflicts of the basic loop. As shown, the example uses the same comparison frequency $f_c = f_r/R$ at each phase detector, but this is neither necessary nor particularly desirable. Assuming that filters at all mixers select the difference mixing product (rejecting the sum product) and that $f_1 > f_2/P_2$ and $f_2 > f_3/P_3$, the output frequency is

$$f_1 = \frac{f_r}{R}\left(N_1 + \frac{N_2}{P_2} + \frac{N_3}{P_2P_3}\right) \tag{15.2}$$

The output frequency is selectable in increments of $f_r/RP_2P_3 = f_c/P_2P_3$. More loops can be added to achieve even smaller increments without reducing the comparison frequency. The lower loops might all be constructed as identical modules for ease of manufacturing.

Mixers produce many unwanted output products that accompany the one component desired. Existence of unwanted products raises the possibility of spurious components in the synthesizer output or even locks to the wrong frequency [15.3].

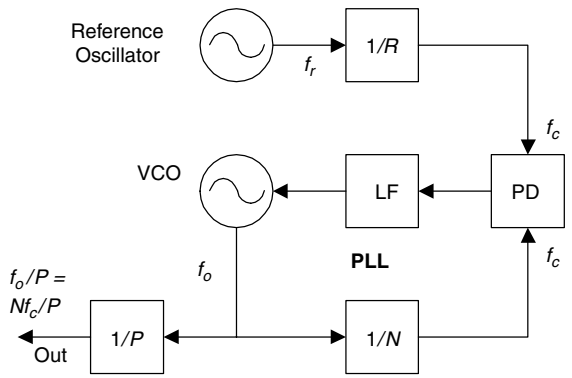


Figure 15.2 Synthesizer with a separate output divider.

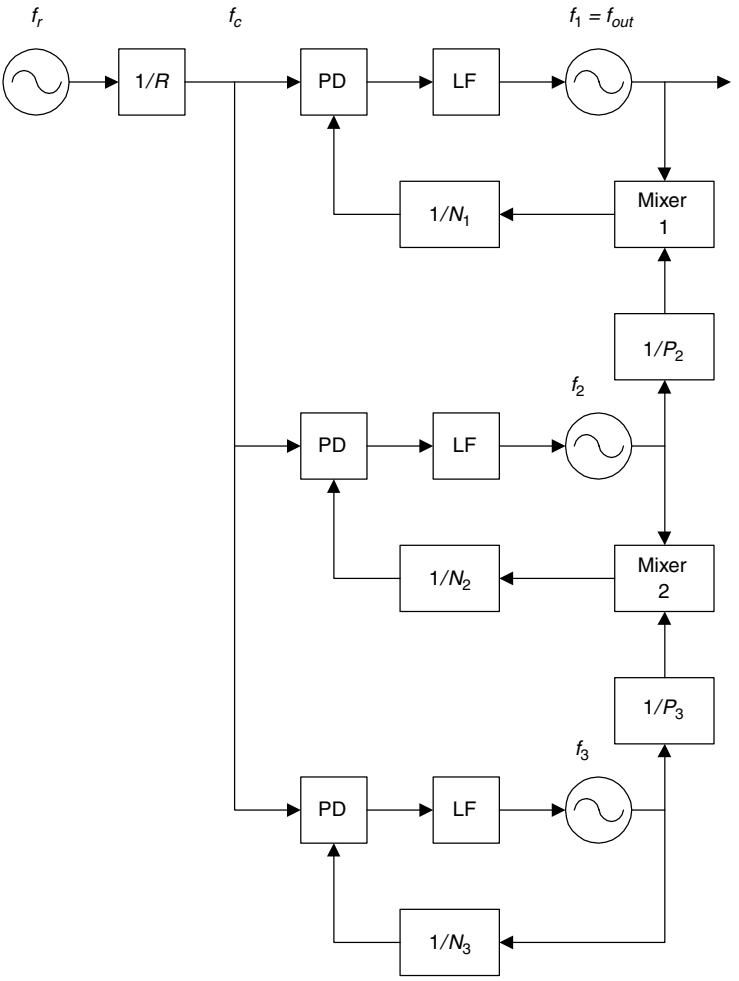


Figure 15.3 Multi-PLL synthesizer with mixers.

Good filters and careful frequency plans are essential for adequate attenuation of spurious products. Although not obvious from the diagram, the presence of mixers also reduces the allowable range of output frequencies.

Historically, some of the lowest phase-noise synthesizers have been built with multiple PLLs and discrete components. They are very costly compared to single-loop IC synthesizers, but so far, the latter have not achieved nearly the same phase-noise performance.

15.2 FREQUENCY DIVIDERS

A frequency divider is an essential component in a phaselock synthesizer. Two kinds of dividers are known: digital counters and analog dividers. Digital counters

are by far the more versatile and widely used, but analog dividers also have a niche in which they are sometimes worth considering.

15.2.1 Analog Frequency Dividers

Analog frequency dividers (also called *subharmonic generators*) have been studied for many years. The literature has papers on *regenerative dividers* [15.8–15.12], *injection-locked oscillators* [15.13–15.20], and *parametric* (nonlinear-capacitor) dividers [15.21–15.23]. Parametric dividers are inherently inefficient; they require much more fundamental input power than they can deliver at the subharmonic output. Furthermore, tuning and power-level adjustment are both critical, so they are not much used in practice. As for the other two categories, Verma, Rategh, and Lee [15.20] contend that regenerative dividers and injection-locked oscillators are simply different applications of the same principle and that they belong to the same family. They present analyses to characterize both.

Analog frequency dividers have been useful only in narrow frequency ranges because of the tuned circuits or other filters necessary for proper operation. For the same reason, they have only a fixed division ratio and are not programmable. Large division ratios have not been feasible. On the other hand, the narrow bandwidths are favorable from a noise standpoint, they are operable at frequencies higher than possible with digital counters, and they can be designed to consume less power than do digital counters of the same division ratio.

15.2.2 Digital Counters as Frequency Dividers

Digital counters are by far the most popular frequency divider in PLL synthesizers. They are easily programmable, allowing output at many different frequencies; they can provide very large division ratios; they are built with digital circuits, thereby avoiding most of the problems of analog circuits; they are readily built as integrated circuits along with other digital devices; they are wideband devices, allowing a wide range of synthesized frequencies; they have no cumbersome tuned circuits or other filters that are difficult to incorporate onto an IC chip; they do not require adjustment to work properly; and they typically are low-cost devices. A digital counter cannot work to frequencies as high as those possible with an analog frequency divider; it consumes more power than an analog divider, and its large bandwidth makes it noisier than an analog divider of the same frequency and divider ratio.

Use of a digital counter usually constrains the choice of associated phase detector. The typical output of a counter is nominally a rectangular waveform, often of short duration compared to the cycle period $1/f_c$. Unless restrictive special measures are taken, the information in the output resides solely in transitions (signal edges) of one polarity. If the phase detector has to accept this constraint, sequential PDs or sampling PDs are about the only kinds that are compatible. Because of the valuable properties of phase/frequency detectors (PFDs;—Section 10.3), this type has been prevalent in synthesizers built on integrated circuits.

A multiplier PD such as an exclusive-OR gate becomes feasible if a separate divide-by-2 follows a programmable-counter divider. Divide-by-2 provides a square-wave output with 50% duty ratio. However, since the divide-by-2 is not part of the programmable divider, the comparison frequency f_c is only half of the frequency out of the programmable divider, so the frequency resolution is $2f_c$. Use of a multiplier or sampling PD requires separate provisions for frequency acquisition, a process that is a built-in feature of a PFD.

Counters are of two kinds: either all stages in a counter change state at the same instant (*synchronous counters*) or else the change of state propagates down the stages of the counter (*ripple counter*). More recently, counters for frequency dividers have been *resynchronized*: The output transition is triggered from the input clock after the counter has reached a designated state. Synchronous or resynchronized counters have shorter jitter-contributing paths and should always be used instead of ripple counters for lower phase noise. Levantino et al. [15.24] present analysis and measurements showing that output jitter in a resynchronized counter comes almost entirely from the resynchronizer alone and very little from the much larger accumulated internal jitter of the multiple stages within the counter.

15.3 FRACTIONAL-*N* COUNTERS

So far, this account has explored only division by an integer, a constraint that leads to the underlying conflict between frequency resolution and loop bandwidth of the basic configuration of Fig. 15.1. The basic configuration could have a more favorable trade-off between resolution and bandwidth if the N counter were able to divide by a fractional ratio. This section examines approaches to fractional-N counting.

15.3.1 Dual-Modulus Counters

A forerunner of fractional-N counters is introduced here before attacking true fractional-N operation. The usual programmable counters are limited in their speed; their highest feasible frequency of operation is substantially less than the highest frequency of toggling of individual stages. One way to cope with a VCO frequency too high for a programmable counter is to put a fixed-ratio prescaling counter ahead of it. If the prescaler's division ratio P is small and fixed, the prescaler will have the ability to work at substantially higher frequencies than will a following programmable counter, which, in turn, has a reduced input frequency that it can tolerate. Unfortunately, the frequency resolution of this configuration is now Pf_c instead of f_c of the basic configuration of Fig. 15.1. A fixed prescaler worsens the resolution vs. bandwidth conflict by a factor P.

A *dual-modulus prescaler* [15.25] has long been a routine solution to the resolution impairment caused by a fixed prescaler, as shown in Fig. 15.4. The N and A counters in the figure are both programmable. The P counter, upon

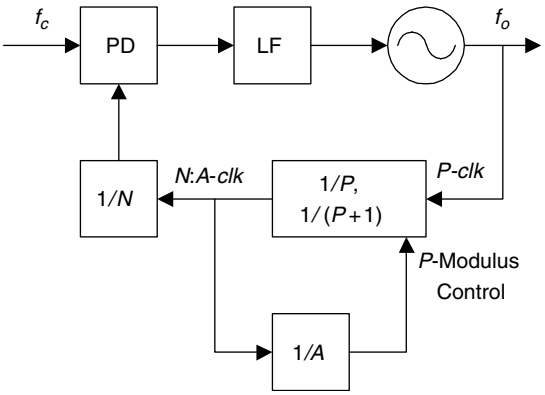


Figure 15.4 Synthesizer with a dual-modulus divider.

command, divides either by P or by $P + 1$; it is a dual-modulus counter. (N, A, and P all are integers.) A total count $N_{fb} = NP + A$ is desired on each comparison cycle, resulting in a VCO frequency of $f_o = N_{fb} f_c = f_c(NP + A)$, with frequency resolution f_c. (The subscript *fb* connotes *feedback*.)

The configuration for Fig. 15.4 is derived from

$$\begin{aligned} N_{fb} &= NP + A = NP + A + AP - AP \\ &= (N - A)P + A(P + 1) \end{aligned} \tag{15.3}$$

which calls for dividing the VCO frequency by P for $N - A$ cycles of the VCO and by $P + 1$ for A cycles. That is exactly what the counters in Fig. 15.4 accomplish. Assume that the N and A counters begin each PD comparison cycle preset to values N and A and that they count down toward zero. Also, assume that the modulus control is set so that the P counter begins a cycle counting the VCO output—the *P-clk*—by $P + 1$. With that modulus setting, the N and A counters are each decremented by 1 on each output cycle of the P counter—on each cycle of *N:A-clk*.

When the state of the A counter reaches zero, it halts its own decrementing and changes the P-modulus control so that the P counter now counts by P instead of $P + 1$. Decrementing of the N counter continues until its state reaches zero, whereupon all three counters are reset to the initial conditions. One complete cycle has included A subcycles with $P + 1$ VCO cycles and $N - A$ subcycles with P VCO cycles, as required by (15.3). Division by $P + 1$ is commonly achieved by preventing the P counter from advancing its count for one *P-clk* cycle after P successive advances, a technique known as *pulse swallowing*.

Although the counter arrangement of Fig. 15.4 is sometimes said to be an example of fractional-N counting, observe that N_{fb} is always a whole number. No fractions are involved; a dual-modulus prescaler is not really a fractional-N counter, no matter what it may have been called. No improvement in resolution is

gained over the basic synthesizer configuration of Fig. 15.1. Observe also that one PD comparison cycle comprises exactly N_{fb} VCO cycles, without any variation that might induce phase fluctuations. Keep this last feature in mind when true fractional-N counters are explained later.

Divider ratios N, P, and A cannot be chosen independently. They are restricted to the limits $N \geq A$ and $A < P$. Egan [15.6, Chap. 4] provides a more extensive discussion of multimodulus counters.

15.3.2 Fractional-*N* PLLs with Analog Compensation

Resolution of a PLL synthesizer could be improved if the feedback frequency divider operated with fractional division ratios, not just integers. An arrangement [15.26] for fractional division is shown in Fig. 15.5; it was long known as *the* fractional-N PLL. To understand its operation, first ignore the DAC and the subtracter. That leaves a PLL consisting of the usual PD, LF, VCO, and a dual-modulus divider capable of dividing by N or by $N + 1$. Division by $N + 1$ is accomplished with a pulse swallower.

Control of pulse swallowing is exercised by an NCO that is clocked by the divider output (designated *C-clk*). Accumulator content in the NCO is incremented by a frequency-control word u_c on each cycle of *C-clk*. Represent the difference equation of the NCO as

$$\varepsilon_o[n] = \{\varepsilon_o[n-1] + u_c\} \quad \text{mod-}Q \tag{15.4}$$

where Q is an integer (commonly an integer power of 2, but not necessarily so) and u_c is any nonnegative integer less than Q. The NCO is arranged so that the addition emits a carry upon each overflow of the accumulator. Like any similar NCO, the average rate of overflows is

$$f_{\text{NCO}} = \frac{f_c u_c}{Q} \tag{15.5}$$

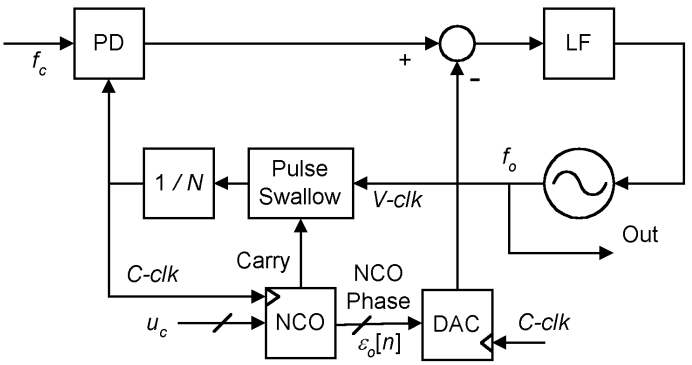


Figure 15.5 Fractional-N synthesizer.

Carries are applied to the pulse swallower. Each carry causes one *V-clk* pulse to be swallowed so that the feedback counter divides the VCO frequency f_o once by $N + 1$ instead of by N. An average fraction u_c/Q cycles of *C-clk* will experience overflows of the NCO accumulator and thus have $N + 1$ counts of *V-clk*, while an average fraction $(1 - u_c/Q)$ cycles of *C-clk* will be free of overflows; these cycles have N counts of *V-clk*. The average count rate therefore is

$$N_{\text{avg}} = (N + 1)\frac{u_c}{Q} + N\left(1 - \frac{u_c}{Q}\right) = N\left(1 + \frac{u_c}{Q}\right) \tag{15.6}$$

In other words, the configuration of Fig. 15.5 permits genuine fractional-N frequency division, on average. Frequency resolution now depends not only on f_c but also on Q, which can be made very large. Fractional-N counting promises a breakout from the resolution vs. bandwidth conflict of the basic PLL synthesizer; fine resolution appears compatible with a high comparison frequency.

But fractional-N counting alone does not provide acceptable performance. Fractional-N counting is attained only on average, not uniformly. Switching between division by N and division by $N + 1$ causes excessive phase jitter, as explained with the help of Fig. 15.6. The figure plots phase vs. time for the reference at frequency f_c and for the divided feedback phase. Both phases are shown unfolded to demonstrate that they increase without bound. Reference phase is a straight line with a constant slope of f_c cycles per second.

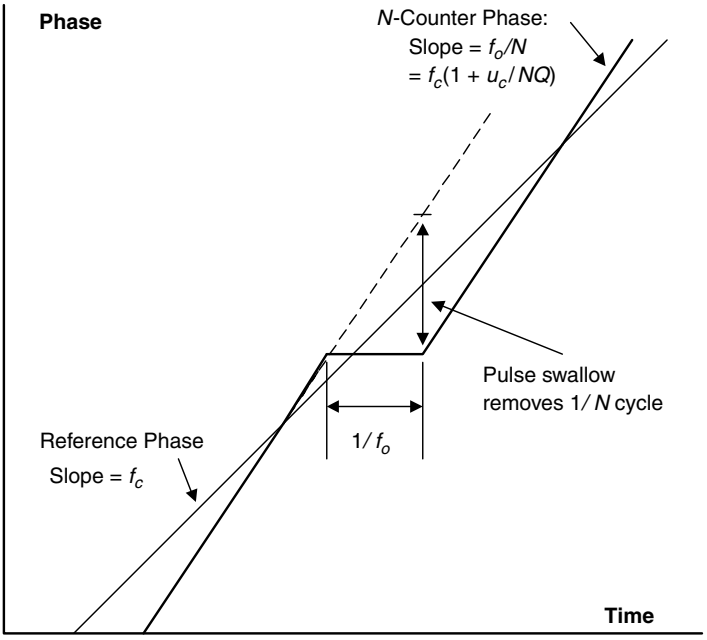


Figure 15.6 Irregular counting in a fractional-N synthesizer.

In the absence of pulse swallowing, the feedback phase (N-counter phase) is a segment of a straight line with a slope of f_o/N cycles per second, a slope that is greater than f_c if $u_c > 0$. Each swallowed pulse arrests the increase of feedback phase for $1/f_o$ seconds, one cycle of the *V-clk*, thenceforth causing the accumulated feedback phase to lag $1/N$ cycles of f_c behind the phase that would have been attained in the absence of the swallow. [**Comment**: Count value in a counter is a time-discrete process. A plot of count vs. time is a staircase. Think of the phase of a counter as a fictitious time-continuous process; the count represents samples of the phase.]

Phase error presented to the PLL is the difference between the reference phase and the N-counter phase. If PLL bandwidth is large (one objective of fractional-N counting), the phase error propagates through to the VCO and causes excessive jitter. The benefit of large bandwidth is lost if the PLL bandwidth is made sufficiently small to filter out the phase error. The spectrum of the phase error has discrete components at the NCO frequency and its harmonics, but also at other frequencies that arise because the time intervals between carries are not uniform unless u_c divides Q. Kroupa's text [15.7, Sec. 12.2] discusses the spurious signals generated in NCOs, as do several articles in his reprint volume [15.27].

The contents $\varepsilon_o[n]$ in the NCO accumulator are digital samples proportional to the phase error at each tick of *C-clk*. Samples are applied to a properly scaled digital-to-analog converter and the analog output is subtracted from the output of the phase detector. The objective is to cancel PD output caused by the phase error and thereby avoid phase jitter in the VCO. Excellent linearity in the PD is necessary, as is critical adjustment of the DAC and the cancellation circuits. One technique is to merge the DAC and the current sources of the charge pump of the PD to assure that they drift in unison. Details of cancellation circuits tend to be proprietary and are not often revealed publicly. Some related patents are listed in [15.28].

It appears that the largest discrete spurious output from the VCO can be suppressed to about 70 dB below the desired carrier. Such performance is not nearly as good as that achievable with a mixer-type synthesizer (e.g., Fig. 15.3), but it is adequate for many purposes and is much less costly.

15.3.3 Fractional-*N* PLLs with Delta–Sigma Modulators

A flood of articles on delta–sigma ($\Delta\Sigma$) PLL synthesizers appeared while this book was being written. These PLLs employ all-digital $\Delta\Sigma$ modulators [15.29, 15.30] to alter the division ratio of a multimodulus frequency divider once per reference cycle. The average division ratio over many cycles is a fractional number. Moreover, the $\Delta\Sigma$ modulator shapes the noise spectrum of its output so as to suppress noise at low frequencies and concentrate it at high frequencies, where the lowpass frequency response of the PLL can attenuate the noise. A $\Delta\Sigma$ PLL is attractive because it achieves its purposes almost entirely by digital means and needs no critical cancellation to yield performance comparable to that of the older fractional-N synthesizers described in Section 15.3.2.

Early papers on the subject include [15.31] and [15.32], with the latter cited extensively in subsequent articles. Also, Egan [15.6, Sec. 8.3] describes $\Delta\Sigma$ techniques without mentioning "delta–sigma." Numerous later papers have been collected in [15.33] and [15.34] and an entire book [15.35] on $\Delta\Sigma$ fractional-N synthesis has been published. Galton's tutorial overview [15.36] is a good introduction to the subject. Dramatic new results were appearing almost monthly while this chapter was being written, a trend that surely continued thereafter. A prudent reader would check subsequent literature for later developments.

A simplified block diagram of a $\Delta\Sigma$ PLL is shown in Fig. 15.7. The loop incorporates all of the usual elements of a PLL synthesizer. It is distinguished from the basic synthesizer by a multimodulus frequency divider capable of dividing by different (necessarily integer) ratios on succeeding cycles of *C-clk*. The multimodulus divider has more than just the two choices of division ratio of a dual-modulus divider.

Control of the division ratio is exercised by a fully digital $\Delta\Sigma$ modulator. Input to the modulator is a digital number u_c that identifies the desired average division ratio in the multimodulus divider, including an integer part and a fractional part. The modulator generates a digital stream of numbers, one per cycle of *C-clk*, each number specifying an integer division ratio that the multimodulus divider can execute. The average of the modulator sequence is the desired fractional division ratio, even though each term in the sequence is an integer. A $\Delta\Sigma$ modulator may be quantized to as few as one bit in each output sample (whereupon the multimodulus divider can have only two different division ratios), or it may have many more bits (allowing more choices of division ratios). [**Comment**: The numbers generated in the modulator may not be the actual division ratio $N[n]$. Instead, they are more likely the necessary instructions to pulse swallowers for attaining the ratios specified.]

The rapidly changing division ratio causes large phase jitter in the feedback signal applied to the phase detector. The system copes with that jitter through the noise-shaping properties of $\Delta\Sigma$ modulators. A modulator of kth order includes k accumulators in its internal digital-signal path. Each accumulator shapes the quantization error spectrum by a factor $1 - z^{-1}$, so a kth-order converter shapes

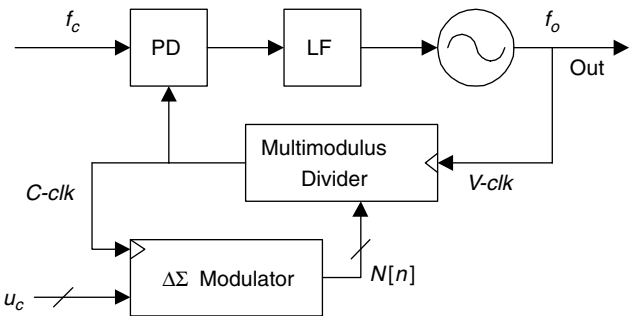


Figure 15.7 Delta–sigma fractional-N synthesizer.

the noise by $(1 - z^{-1})^k$. Shaping inserts k zeros into the noise spectrum at $z = 1$ ($f/f_c = 0$), and the spectral density rises to a peak at $z = -1$ (i.e., $f/f_c = 0.5$). The PLL has a lowpass response [transfer function $H(s)$, introduced in Chapter 2] to phase jitter originating in the feedback path. Adequate lowpass filtering in the PLL removes most of the predominant high-frequency noise of the $\Delta\Sigma$ modulator, leaving residual jitter that is acceptable in many applications.

Effective suppression of the high-frequency noise of a kth-order modulator requires a PLL of at least $(k + 1)$th order. Stable operation of higher-order PLLs requires a narrower loop bandwidth (smaller loop gain K) than is attainable with lower-order PLLs. Narrower bandwidth is contrary to the goal of wider bandwidth that helped inspire fractional-N dividers in the first place. Designers face a trade-off between adequate filtering and wide bandwidth.

Phase jitter caused by the rapid variations of division ratio inherent to $\Delta\Sigma$ modulation can be canceled by DACs [15.37, 15.38], in principle similar to traditional fractional-N synthesizers. That approach is counter to the goal of avoiding critical cancellation schemes, but is effective nonetheless. Another improvement is to apply lowpass digital filtering [15.39] to the $\Delta\Sigma$ control sequence that drives the multimodulus divider, thereby reducing the high-frequency content of the control sequence and alleviating the burden placed on filtering by the PLL itself. A larger loop bandwidth is facilitated thereby.

When supplied with a constant (DC) input, the state trajectory in a $\Delta\Sigma$ modulator is likely to follow a periodic limit cycle [15.40]. (Its behavior is closely related to that of a digital PLL generating an integer frequency, as described in Section 13.2.3.) The periodicity of the limit cycle causes the spectrum of the resulting phase noise to consist of discrete lines rather than a continuous density. Discrete lines are often unacceptable and have to be suppressed. One technique to spread the lines is to apply small-amplitude zero-mean dither to the frequency-control word u_c [15.36]; dither can be generated by a pseudorandom shift-register sequence. Another technique was reported in [15.41]: If a suitable initial condition is established in the first accumulator in a *MASH* modulator [15.30] (also called a *cascade* or *multistage* modulator) of third or higher order, the periodicity is suppressed effectively and the jitter spectrum is continuous. One would suspect that similar measures in higher-order other-than-MASH configurations might also be effective.

Small, previously unimportant nonlinearities in the PLL have major consequences with $\Delta\Sigma$ PLLs. A nonlinear circuit generates intermodulation and harmonics from the components of its input signal, and the sampling inherent in dividers and phase detectors causes spectral foldover. The result is that the high-intensity high-frequency components of the $\Delta\Sigma$ sequence, components that are supposed to be attenuated by the lowpass frequency response of the PLL, are aliased into low frequencies, where they are not filterable. To achieve nearly the full promise of $\Delta\Sigma$ PLLs requires that the designer carefully seek out and counteract all nonlinearities exposed to the large excursions of the $\Delta\Sigma$ sequence.

One such nonlinearity arises in the multimodulus divider itself in the form of differing delays for different moduli [15.36]. Resynchronization of the divider

output is an obvious remedy. Other nonlinearities lurk in a phase-frequency detector (PFD), as described in Section 10.3. Residual dead zone is especially injurious [15.39, 15.42], but charge-pump imbalance also plays a role. Furthermore, the timing of samples in a PFD is not uniform since the charge current is initiated by the reference when reference phase leads the divider feedback but is initiated by the divider output when the reference lags the feedback.

To combat the nonuniform sampling, [15.43] proposed that the PFD and charge pump be followed by a sample-and-hold (S&H) circuit to resample at uniform intervals under control by the reference clock. Rather than resampling after a PFD, one might consider employing an S&H phase detector in place of a PFD to avoid entirely the nonlinearities of a PFD. A sampling PD has no frequency-detection capabilities, so other measures would be required for acquisition. I had found no published account of a $\Delta\Sigma$ PLL with an S&H PD at the time this book was being written.

Due to the nonlinearities, the modeling and analysis of $\Delta\Sigma$ PLLs is more complicated than the straightforward linear methods covered so far in this book. Various approaches to modeling and analysis may be found in [15.4], [15.27], [15.35–15.37], [15.39], and [15.42–15.44].

15.4 NOISE PROPAGATION IN A PLL

Successful development of a high-performance synthesizer (incorporating low phase noise, fast acquisition, close spacing of synthesized frequencies, low power, low cost) is a challenging task. It needs painstaking attention to a multitude of disparate matters. Implementation on an IC requires circuit simulation (especially if nonlinearities are significant) before laying out the chip, and simulation is also often advisable for older packaging methods. Computer programs might keep track of all the details of the design; designers need all the help they can get.

This section takes a different approach, however; it looks at some basic features of PLL synthesizers. The tools employed are linear analysis using transfer functions. Computer help is no more complicated than a spreadsheet. A preliminary design with these simple tools establishes a good baseline for the more intensive efforts needed for the development of a high-performance synthesizer.

15.4.1 Transfer Functions for Oscillator Noise

The simple model in Fig. 15.8 helps establish noise transfer functions for the PLL. Pretend for now that noise is confined to the two oscillators: the reference and the VCO. That is not a realistic assumption—noise arises in every element of a PLL—but oscillator noise should predominate in a high-performance synthesizer. No noise performance can be better than that arising from the oscillators alone.

The PLL of Fig. 15.8 has a $1/N$ frequency divider in its feedback path. All elements—oscillators, phase detector, loop filter, and frequency divider—are assumed to be ideal. Phase noise is treated as if it were applied externally by

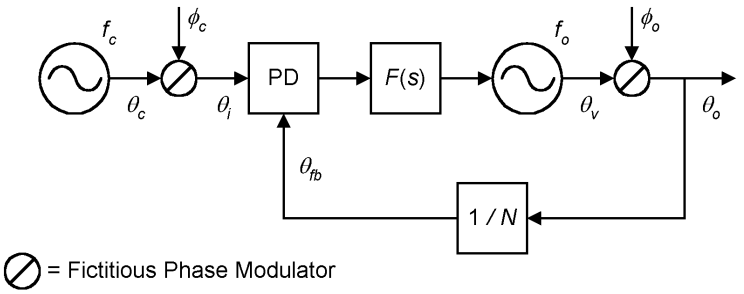


Figure 15.8 Oscillator noise sources in a PLL synthesizer.

means of fictitious phase modulators following each oscillator. A reference source at comparison frequency f_c has a phase θ_c to which is added phase noise ϕ_c radians with spectral density $W_{\phi c}(f)$ rad/sec·Hz. Signal phase applied to the PD is $\theta_i = \theta_c + \phi_c$.

Output from the VCO at frequency f_o has phase $\theta_o = \theta_v + \phi_o$ radians where ϕ_o is phase-noise modulation with spectral density $W_{\phi o}(f)$ rad/sec·Hz and θ_v is established by the control voltage on the VCO. Feedback phase θ_{fb} applied to the PD is simply θ_o/N. Phase error is $\theta_e = \theta_i - \theta_{fb}$, and the phase of the VCO is given by $\theta_v = \theta_e K_d K_o F(s)/s$. (The notation for PLL elements was defined in Chapter 2.)

Upon manipulation of the pertinent equations, the noisy phase out of the VCO is found to be

$$\theta_o(s) = \frac{K_d K_o F(s)\theta_i + s\phi_o}{s + K_d K_o F(s)/N} = H(s)\theta_i + E(s)\phi_o \tag{15.7}$$

where

$$\begin{aligned} H(s) &= \frac{K_d K_o F(s)}{s + K_d K_o F(s)/N} \\ E(s) &= 1 - H(s) = \frac{s}{s + K_d K_o F(s)/N} \end{aligned} \tag{15.8}$$

Phase noise spectral density of θ_o therefore is

$$W_{\theta o}(f) = |H(f)|^2 W_{\phi c}(f) + |E(f)|^2 W_{\phi o}(f) \qquad \text{rad}^2/\text{Hz} \tag{15.9}$$

The transfer function of the loop filter of a second-order type 2 PLL is $F(s) = K_1 + K_2/s$, as in (2.14). Substituting into (15.8) and defining $K = K_d K_o K_1/N$ gives the corresponding transfer functions:

$$\begin{aligned} H(s) &= \frac{NK(s + K_2/K_1)}{s^2 + K(s + K_2/K_1)} = \frac{N(sK + K^2/4\zeta^2)}{s^2 + sK + K^2/4\zeta^2} \\ E(s) &= \frac{s}{s^2 + K(s + K_2/K_1)} = \frac{s^2}{s^2 + sK + K^2/4\zeta^2} \end{aligned} \tag{15.10}$$

Comparing (15.10) against (2.20) and (2.21) reveals that $E(s)$ depends on N only through its contribution to K, but $H(s)$ additionally is proportional to N. A PLL with a frequency divider in its feedback path acts as a frequency multiplier and, like all frequency multipliers, magnifies phase noise at its input by a factor N. A large value for N is detrimental to output phase noise.

15.4.2 Bandwidth Trade-off

This section provides a graphical example, based on Fig. 15.8, that applies the transfer-function analysis of the preceding section. The example demonstrates an important principle: Phase-noise performance of a PLL synthesizer should be optimized by the proper choice of loop bandwidth (loop gain K).

Figure 15.9 illustrates the various constituents of the example. The reference source for the example is a 10-MHz quartz crystal oscillator (XTAL) with specified phase-noise spectrum $W_{\phi c}(f)$ as plotted in the figure. The VCO is a 12-GHz dielectric resonator oscillator (DRO) with a specified phase-noise spectrum $W_{\phi o}(f)$. The DRO is to be phaselocked to the 1200th harmonic of the reference, so the frequency-divider ratio is 1/1200. Phase-noise magnification of the reference is shown on the plot by translating the phase-noise spectrum of the 10-MHz oscillator upward by $20\log(1200) = 61.6$ dB. The oscillator spectra are plotted with solid lightweight lines. Dashed lines show frequency responses for $|H(f)|/N$ and $|E(f)|$ for a particular choice of $K/2\pi = 1$ kHz and $\zeta = 0.707$ for PLL parameters. Response for $|H(f)|$ has been divided by N to keep the plot on-scale. The reference phase noise is filtered by $H(f)$, and the VCO phase

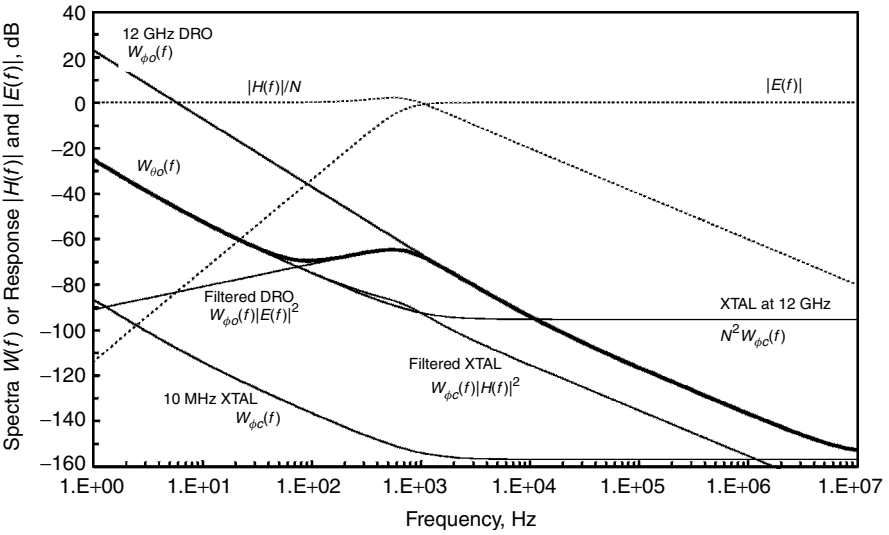


Figure 15.9 Example phase-noise spectra in a PLL synthesizer ($K/2\pi = 1$ kHz and $\zeta = 0.707$).

noise is filtered by $E(f)$, also plotted with lightweight lines. Filtering is calculated simply as the addition of decibel plots of phase-noise spectrum and filter response. Finally, the PLL output phase-noise spectrum $W_{\theta o}(f)$ is plotted with a heavyweight line in accordance with (15.9). The result, characteristic of PLL synthesizers, follows the N-magnified spectrum of the crystal oscillator at low frequencies and the spectrum of the VCO at high frequencies.

It is clear from Fig. 15.9 that a larger bandwidth of the PLL would improve the output phase noise. Figure 15.10 shows output phase-noise spectra for various choices of loop gain K for a PLL with damping held constant at $\zeta = 1$. The minimal phase-noise spectrum $W_{\theta o}(f)$ is attained for $K/2\pi \approx 10$ kHz or perhaps slightly larger. Referring to Fig. 15.9, observe that the VCO phase noise and N-multiplied reference phase-noise curves cross over near $f = 10$ kHz. This observation leads to the following general rule:

If noise in a PLL synthesizer is dominated by the reference oscillator and VCO, the optimum bandwidth for the PLL is close to the frequency of crossover of the phase-noise spectra of the two oscillators.

Many other sources of noise exist in any synthesizer, so the rule is only a starting point for the noise analysis. Moreover, the combination (15.9) is the very best noise performance that can be delivered by a PLL for any given pair of oscillators; other sources of noise inevitably must deteriorate the overall performance. The rule is applicable only if a crossover frequency actually exists and if it is low

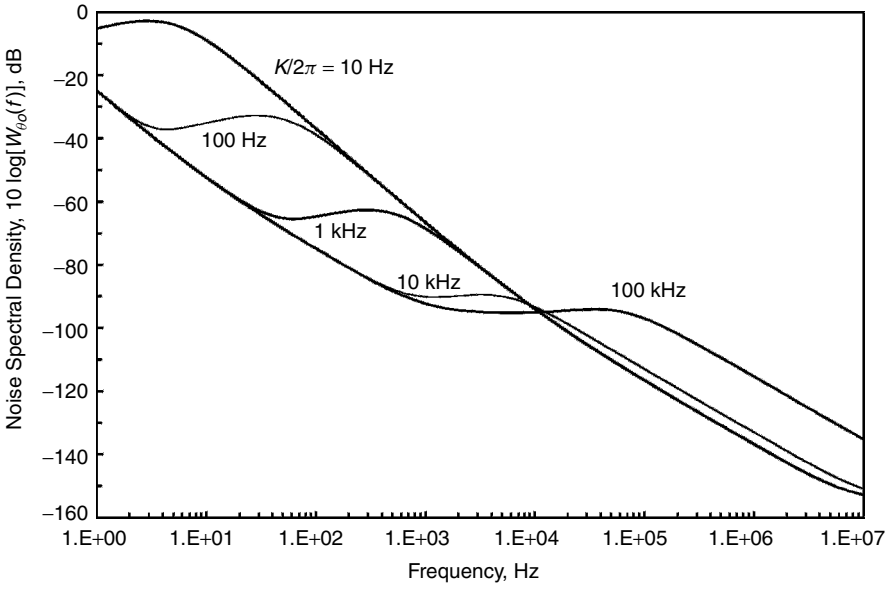


Figure 15.10 Phase-noise spectra for various loop gains K in a PLL synthesizer ($\zeta = 1$; same oscillators as in Fig. 15.9).

enough that a PLL with the optimum bandwidth can have adequate stability margin. In practice, the noise of the VCO in many contemporary synthesizers is so large that the optimum bandwidth is unattainable or even nonexistent. The bandwidth should be made as large as possible in that case, subject to considerations of loop stability and attenuation of PD ripple.

A couple of features often appear in the spectrum of a PLL synthesizer. A transition region in the spectrum $W_{\theta o}(f)$ is more or less flat between the steeply sloping regions of the spectra of the two individual oscillators. That transition flatness imparts a distinctive shelf to the output spectrum. From Fig. 15.10, a wide-extended shelf suggests excessive bandwidth of the PLL; better noise performance may be obtainable with narrower bandwidth.

The transition regions are not entirely flat; a certain amount of peaking can be discerned. Peaking comes about partly from the nature of the transition itself, being somewhat more prominent for narrower than for wide PLL bandwidths. But the peaking also comes about from peaking of $|H(f)|$ (see Section 2.2.4). Peaking is reduced by increases of damping ζ. Peaks are much more apparent when the phase-noise spectrum is plotted to a linear frequency scale; the peaks look like "ears" when viewed on an RF spectrum analyzer and are distinctive features of a PLL synthesizer. Figure 15.10 also suggests that some noise at high frequencies beyond $f = K/2\pi$, the noise due to the floor in the reference-noise spectrum, might be improved by means of additional high-frequency filtering in the PLL, especially if the loop bandwidth exceeds the optimum.

15.4.3 Other Noise Sources

Every element in a synthesizer contributes noise. The assumption in the preceding section, that phase noise of the two oscillators predominates and that all else can be neglected, is an unattainable ideal. Resistors contribute thermal noise, active devices contribute shot noise, many devices contribute flicker noise, phase-detector ripple and mixer spurious products contribute spurious components with discrete spectra, and other nearby circuits contribute ingress interference with assorted spectral properties. Stringent design of a synthesizer entails characterization of the noise from each source and evaluation of its contribution to the phase noise of the output.

That task of evaluation would be overwhelming if all sources had to be analyzed together. Fortunately, if the PLL can be regarded as being linear with respect to the noise sources (not always correct; be careful), linear analysis permits each to be considered separately and the individual effects of all superposed to determine the total phase noise. Superposition is the technique pursued in this section.

The PLL in Fig. 15.11 has the same configuration as that of Fig. 15.8, except for three additional noise sources: additive noises V_{nd} at the output of the phase detector and V_{nc} at the input of the VCO, plus phase noise ϕ_{dv} at the output of the frequency divider. Represent their spectral densities as $W_{vnd}(f)$, $W_{vnc}(f)$, and $W_{\phi dv}(f)$, with units of V^2/Hz for the two voltages and rad^2/Hz for the phase.

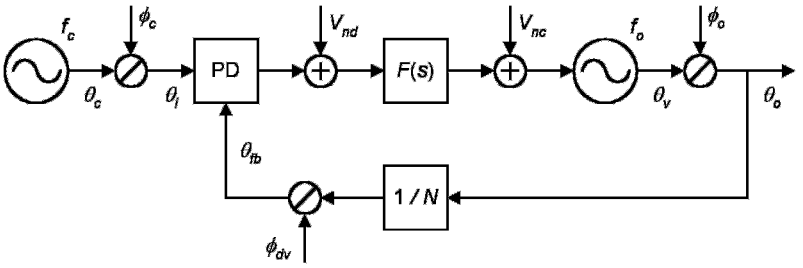


Figure 15.11 Various noise sources in a PLL synthesizer.

(Additive noise out of the PD might be current noise instead of voltage; notation and units are to be changed appropriately.) These noise sources are just examples; they are accompanied by many other sources in every synthesizer.

Through transfer-function analysis, their individual contributions to the output phase-noise spectrum are found to be

$$\begin{aligned} &\frac{|H(f)|^2 W_{vnd}(f)}{K_d^2} \\ &\frac{|E(f)|^2 K_o^2 W_{vnc}(f)}{(2\pi f)^2} \qquad \text{rad}^2/\text{Hz} \\ &|H(f)|^2 W_{\phi dv}(f) \end{aligned} \tag{15.11}$$

Inspection of these contributions reveals:

- Additive noise occurring between the phase detector and the filter is lowpass-filtered by the PLL.
- A large PD gain K_d is favorable for reducing the effects of additive noise arising before the loop filter.
- Additive noise occurring between the loop filter and the VCO is bandpass-filtered if the PLL is of type 2 or higher.
- Small VCO gain K_o is favorable for reducing the effects of additive noise arising after the loop filter.
- Phase noise originating in the frequency divider is filtered by the PLL in exactly the same way as phase noise from the reference oscillator, as in (15.9).

Noise properties of various circuit components have been reported in [15.24] (frequency dividers), [15.45] (frequency multipliers and dividers, mixers—and thus some phase detectors), [15.46] (many components), and [15.49] and [15.50] (frequency dividers). Egan [15.45, 15.49] has long emphasized the sampled character of frequency dividers built from digital counters, a property also brought out in [15.24]. Because the output of a digital divider is sampled at a lower

rate than the input, the input noise is aliased to lower frequencies. Noise at the output of the divider consists of noise that is generated within the divider circuits (analyzed in detail in [15.24]) and aliases of the input phase noise. The PLL is unable to distinguish either type of divider noise from reference noise, as shown in (15.11). References [15.24], [15.45], and [15.49] offer approximations for predicting divider-aliased noise, but further study is needed.

Noise in the digital portion of phase/frequency detectors ought to have effects similar to those in frequency dividers (except that there is no down-sampling); it should be feasible to analyze the effects by methods described in [15.24]. No such analysis seemed to have been published by the time this book was written.

Charge pumps suffer from shot noise and flicker noise. In consequence, the charge delivered to the loop filter fluctuates from one ON interval to the next, even for fixed ON intervals. Those charge fluctuations constitute a noise source in the PLL. Analysis of charge-pump noise was explored briefly in [15.39] and [15.44]. Further analysis would be valuable, taking the sampled character of the ON-interval noise into account. Presumably a charge pump does not contribute significant noise during its OFF intervals.

Noise sources in ordinary time-continuous components such as resistors and operational amplifiers in the loop filter are well understood. Linear analysis consists of evaluating the noise characteristics of the sources and determining the transfer function from each source to the VCO phase-noise spectrum $W_{\theta o}(f)$, as illustrated in the following example. Refer to Fig. 15.12 for a noise model of an active loop filter in a second-order type 2 PLL. Noise is contributed by the two resistors R_1 and R_2 and by the input-referred equivalent voltage noise e_{na} and current noise i_{na} of the operational amplifier. A resistance R generates white thermal noise with one-sided spectral density $W_R = 4kTR$, where k is Boltzmann's constant and T is absolute temperature. The spectra $W_{ena}(f)$ and $W_{ina}(f)$ of the op-amp noises contain flicker $(1/f)$ and white components specific to each device.

Assume that the op-amp is ideal: infinite gain; infinite bandwidth; infinite input impedance. In consequence, the current into the input terminal must be zero and the voltage on the input terminal must also be zero. Consequently, the equation

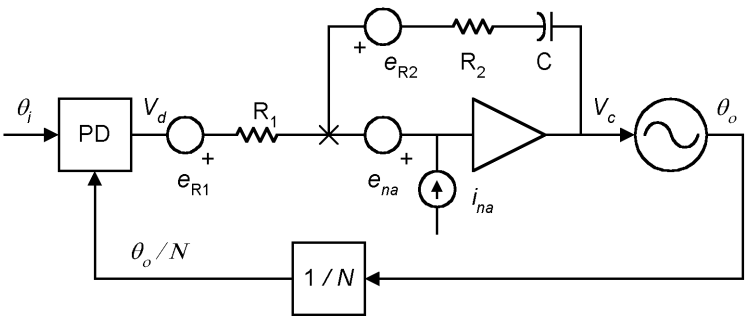


Figure 15.12 Noise sources in the loop filter of a PLL synthesizer.

for the current entering node X in Fig. 15.12 is written as

$$\frac{V_d + e_{R1}}{R_1} + i_{na} + \frac{V_c + e_{R2}}{R_2 + 1/sC} + e_{na}\left(\frac{1}{R_1} + \frac{1}{R_2 + 1/sC}\right) = 0 \qquad (15.12)$$

Let $\theta_e = -K_d(\theta_i - \theta_o)$. (The minus sign compensates the phase reversal in the op-amp.) Setting $\theta_i = 0$, the closed-loop phase noise θ_o due to the loop-filter noise sources is found to be

$$\theta_o(s) = -\frac{K_o}{s^2 + K_o K_d(sCR_2 + 1)/NR_1C} \times \left[\frac{sCR_2 + 1}{CR_1}(e_{R1} + i_{na}R_1 + e_{na}) + s(e_{R2} + e_{na})\right] \qquad (15.13)$$

Define $K = K_o K_d R_2/NR_1$ and apply the definitions of (15.10) to obtain

$$\begin{aligned}\theta_o(s) &= -\frac{1}{s^2 + K(s + 1/R_2C)} \times \left[\frac{NK(s + 1/R_2C)}{K_d}(e_{R1} + i_{na}R_1 + e_{na}) + sK_o(e_{R2} + e_{na})\right] \\ &= -\frac{H(s)}{K_d}(e_{R1} + i_{na}R_1 + e_{na}) - \frac{K_o E(s)}{s}(e_{R2} + e_{na}) \qquad (15.14)\end{aligned}$$

The contribution to spectral density $W_{\theta o}(f)$ of the phase fluctuations of θ_o is

$$\begin{aligned}W_{\theta o}(f) &= [W_{R1} + W_{ina}(f)R_1^2]\frac{|H(f)|^2}{K_d^2} + W_{R2}\frac{K_o^2|E(f)|^2}{4\pi^2 f^2} \\ &\quad + W_{ena}(f)\left|\frac{H(f)}{K_d} + \frac{K_o E(f)}{j2\pi f}\right|^2 \qquad \text{rad}^2/\text{Hz} \qquad (15.15)\end{aligned}$$

(Take heed that $|[H(f)/K_d] + [K_o E(f)/j2\pi f]|^2 \neq |H(f)/K_d|^2 + |K_o E(f)/j2\pi f|^2$; nonzero cross products also exist.)

REFERENCES

15.1 V. F. Kroupa, *Frequency Synthesis*, Wiley, New York, 1973.

15.2 G. Gorski-Popiel, ed., *Frequency Synthesis: Applications and Techniques*, IEEE Press, New York, 1975.

15.3 V. Manassewitsch, *Frequency Synthesizers: Theory and Design*, 3rd ed., Wiley, New York, 1987.

15.4 J. A. Crawford, *Frequency Synthesizer Design Handbook*, Artech House, Norwood, MA, 1994.

15.5 U. L. Rohde, *Microwave and Wireless Synthesizers: Theory and Design*, Wiley, New York, 1997.

15.6 W. F. Egan, *Frequency Synthesis by Phase Lock*, 2nd ed., Wiley, New York, 2000.

15.7 V. F. Kroupa, *Phase Lock Loops and Frequency Synthesis*, Wiley, Chichester, West Sussex, England, 2003.

15.8 J. W. Horton, "Generation and Control of Electric Waves," U.S. patent 1,690,299, Nov. 6, 1928.

15.9 R. L. Miller, "Fractional-Frequency Generators Utilizing Regenerative Modulation," *Proc. IRE* ***27***, 446–457, July 1939.

15.10 S. Plotkin and O. Lumpkin, "Regenerative Fractional Frequency Generators," *Proc. IRE* ***48***, 1988–1997, Dec. 1960.

15.11 C. W. Helstrom, "Transient Analysis of Regenerative Frequency Dividers," *IEEE Trans. Circuit Theory* ***CT-12***, 489–497, Dec. 1965.

15.12 E. Rubiola, M. Olivier, and J. Groslambert, "Phase Noise in the Regenerative Frequency Dividers," *IEEE Trans. Instrum. Meas.* ***IM-41***, 353–360, June 1992.

15.13 R. Adler, "A Study of Locking Phenomena in Oscillators," *Proc. IRE* ***34***, 351–357, June 1946. Reprinted in [15.18].

15.14 K. Kurokawa, "Noise in Synchronized Oscillators," *IEEE Trans. Microwave Theory Tech.* ***MTT-16***, 234–240, Apr. 1968. Reprinted in [15.17].

15.15 R. Adler, "A Study of Locking Phenomena in Oscillators," *Proc. IEEE* ***61***, 1380–1385, Oct. 1973.

15.16 K. Kurokawa, "Injection Locking of Microwave Solid-State Oscillators," *Proc. IEEE* ***61***, 1386–1410, Oct. 1973.

15.17 V. F. Kroupa, ed., *Frequency Stability: Fundamentals and Measurement*, Reprint Volume, IEEE Press, New York, 1983.

15.18 W. C. Lindsey, and M. K. Simon, eds., *Phase-Locked Loops & Their Application*, Reprint Volume, IEEE Press, New York, 1978.

15.19 H. R. Rategh and T. H. Lee, "Superharmonic Injection-Locked Frequency Dividers," *IEEE J. Solid-State Circuits* ***SC-34***, 813–821, June 1999.

15.20 S. Verma, H. R. Rategh, and T. H. Lee, "A Unified Model for Injection-Locked Frequency Dividers," *IEEE J. Solid-State Circuits* ***38***, 813–821, June 2003.

15.21 J. M. Manley and H. E. Rowe, "Some General Properties of Nonlinear Elements, Part I: General Energy Relations," *Proc. IRE* ***44***, 904–913, July 1956.

15.22 D. Leenov and A. Uhlir, "Generation of Harmonics and Subharmonics at Microwave Frequencies with P-N Junction Diodes," *Proc. IRE* ***47***, 1724–1729, Oct. 1959.

15.23 R. A. Mostrom, "The Charge-Storage Diode as a Subharmonic Generator," *Proc. IEEE* ***55***, 735–736, July 1965.

15.24 S. Levantino, L. Romano, S. Pellerano, C. Samori, and A. L. Lacaita, "Phase Noise in Digital Frequency Dividers," *IEEE J. Solid-State Circuits* ***39***, 775–784, May 2004.

15.25 *Phase-Locked Loop Data Book*, 2nd ed., Motorola, Phoenix, AZ, Aug. 1973.

15.26 D. D. Danielson and S. E. Froseth, "A Synthesized Signal Source with Function Generator Capabilities," *Hewlett-Packard J.*, 18–26, Jan. 1979.

15.27 V. F. Kroupa, ed., *Direct Digital Frequency Synthesizers*, Reprint Volume, IEEE Press, New York, 1999.

15.28 D. P. Owen, "Fractional-*N* Synthesizers," *Microwave J.*, 110–121, Oct. 2001.

15.29 J. C. Candy, and G. C. Temes, eds., *Oversampling Delta–Sigma Data Converters*, Reprint Volume, IEEE Press, New York, 1992.

15.30 S. R. Norsworthy, R. Schreier, and G. C. Temes, eds., *Delta–Sigma Data Converters*, Reprint Volume, IEEE Press, New York, 1997.

15.31 B. Miller and R. J. Conley, "A Multiple Modulator Fractional Divider," *IEEE Trans. Instrum. Meas.* ***40***, 578–583, June 1991.

15.32 T. A. Riley, M. A. Copeland, and T. A. Kwasniewski, "Delta–Sigma Modulation in Fractional-*N* Frequency Synthesis," *IEEE J. Solid-State Circuits* ***28***, 553–559, May 1993.

15.33 B. Razavi, ed., *Phase-Locking in High-Performance Systems*, Reprint Volume, IEEE Press, New York, and Wiley, New York, 2003.

15.34 Special Issue on Integrated Phase-Locked Loops, *IEEE Trans. Circuits Syst. II* ***50***, Nov. 2003.

15.35 B. De Muer and M. Steyaert, *CMOS Fractional-N Synthesizers: Design for High Spectral Purity and Monolithic Integration*, Kluwer Academic, Norwell, MA, 2002.

15.36 I. Galton, "Delta–Sigma Fractional-*N* Phase-Locked Loops," original tutorial in [15.33], pp. 23–33.

15.37 S. Pamarti and I. Galton, "Phase-Noise Cancellation Design Tradeoffs in Delta–Sigma Fractional-*N* PLLs," *IEEE Trans. Circuits Syst. II* ***50***, 829–838, Nov. 2003.

15.38 S. E. Meninger and M. H. Perrott, "A Fractional-*N* Frequency Synthesizer Architecture Utilizing a Mismatch Compensated PFD/DAC Structure for Reduced Quantization-Induced Phase Noise," *IEEE Trans. Circuits Syst. II* ***50***, 839–849, Nov. 2003.

15.39 T. A. D. Riley, N. M. Filiol, Q. Du, and J. Kostamovaara, "Techniques for In-Band Phase Noise Reduction in $\Delta\Sigma$ Synthesizers," *IEEE Trans. Circuits Syst. II* ***50***, 794–803, Nov. 2003.

15.40 R. M. Gray, "Quantization Noise Spectra," *IEEE Trans. Inf. Theory* ***IT-36***, 1220–1244, Nov. 1990. Reprinted in [15.29].

15.41 M. Kozak and I. Kale, "Rigorous Analysis of Delta–Sigma Modulators for Fractional-*N* PLL Frequency Synthesis," *IEEE Trans. Circuits Syst. I* ***51***, 1148–1162, June 2004.

15.42 B. De Muer and M. S. J. Steyaert, "On the Analysis of $\Delta\Sigma$ Fractional-*N* Frequency Synthesizers for High-Spectral Purity," *IEEE Trans. Circuits Syst. II* ***50***, 784–793, Nov. 2003.

15.43 M. Cassia, P. Shah, and E. Bruun, "Analytical Model and Behavioral Simulation Approach for a $\Sigma\Delta$ Fractional-*N* Synthesizer Employing a Sample-Hold Element," *IEEE Trans. Circuits Syst. II* ***50***, 850–859, Nov. 2003.

15.44 M. Perrott, M. Trott, and C. Sodini, "A Modeling Approach for $\Sigma\Delta$ Fractional-*N* Frequency Synthesizers Allowing Straightforward Noise Analysis," *IEEE J. Solid-State Circuits* ***37***, 1028–1038, Aug. 2002. Reprinted in [15.33].

15.45 W. F. Egan, "The Effects of Small Contaminating Signals in Nonlinear Elements Used in Frequency Synthesis and Conversion," *Proc. IEEE* ***69***, 797–811, July 1981.

15.46 V. F. Kroupa, "Noise Properties of PLL Systems," *IEEE Trans. Commun.* ***COM-30***, 2244–2252, Oct. 1982. Reprinted in [15.17], [15.47], and [15.48].

15.47 W. C. Lindsey and C. M. Chie, eds., *Phase-Locked Loops*, Reprint Volume, IEEE Press, New York, 1986.

15.48 B. Razavi, *Monolithic Phase-Locked Loops and Clock Recovery Circuits*, Reprint Volume, IEEE Press, New York, 1996.

15.49 W. F. Egan, "Modeling Phase Noise in Frequency Dividers," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* ***UFFC-37***, 307–315, July 1990.

15.50 V. F. Kroupa, "Jitter and Phase Noise in Frequency Dividers," *IEEE Trans. Instrum. Meas.* ***IM-50***, 1241–1243, Oct. 2001.

CHAPTER 16

PHASELOCK MODULATORS AND DEMODULATORS

Phaselock demodulators are widely used for the reception of amplitude modulation (AM), phase modulation (PM), and frequency modulation (FM). Coherent (as opposed to noncoherent) demodulation of AM or PM is almost always accomplished with the help of a phaselock loop. Phaselock FM demodulators can achieve lower thresholds than can conventional FM discriminators. Angle modulators (PM and FM) are sometimes mechanized by means of phaselock loops. This chapter surveys the role of PLLs in modulators and demodulators.

16.1 PHASELOCK MODULATORS

There are numerous methods of producing phase modulation or frequency modulation. In one method the baseband message is inserted into the low-frequency portion of a PLL so as to phase- or frequency-modulate the VCO. [**Comment**: The distinction between PM and FM is artificial; both might be termed *angle modulation* and treated in a unified manner. In this chapter, the term *PM* implies small phase deviation, with a remanent carrier present, whereas *FM* has no such implications. The distinction is more apparent in the modulator and demodulator configurations than in the signals themselves.] Center-frequency stability is established by a fixed oscillator as a reference. Phaselocking forces the average VCO frequency to be equal to the reference frequency. The locked loop tracks out the frequency drift of the VCO.

Phaselock Techniques, Third Edition, by Floyd M. Gardner

16.1.1 Modulator Fundamentals

A block diagram of an angle-modulated PLL is shown in Fig. 16.1. Phase modulation is accomplished by adding the modulating voltage V_p to the output of the phase detector V_d. The loop attempts to null the sum $V_p + V_d$; this is possible only if a phase error generates a V_d that cancels V_p. Phase modulation of the VCO causes phase error to appear at the PD.

Through the transfer-function methods of Chapter 2, the VCO phase modulation generated by the voltage V_p is found to be

$$\theta_o(s) = \frac{V_p(s)}{K_d} H(s) \tag{16.1}$$

Modulation sensitivity of the circuit is $1/K_d$ rad/V. Since $H(s)$ is a lowpass function, the loop bandwidth must be larger than the highest modulation frequency to avoid linear distortion. The phase detector characteristic must be linear to avoid nonlinear distortion of the modulation. Nonlinearity of the VCO is reduced by feedback; VCO nonlinearity is tolerable if loop bandwidth is sufficiently larger than modulating frequency.

Frequency modulation is produced by adding a baseband voltage V_f into the VCO control terminal along with the output of the loop filter. The closed-loop phase modulation of the VCO is readily shown to be

$$\theta_o(s) = \frac{K_o V_f(s)}{s + K_o K_d F(s)} = \frac{K_o E(s) V_f(f)}{s} \tag{16.2}$$

Since the output frequency Ω_o is the derivative of phase, the Laplace transform of the VCO frequency modulation is

$$\Omega_o(s) = s\theta_o(s) = K_o E(s) V_f(s) \tag{16.3}$$

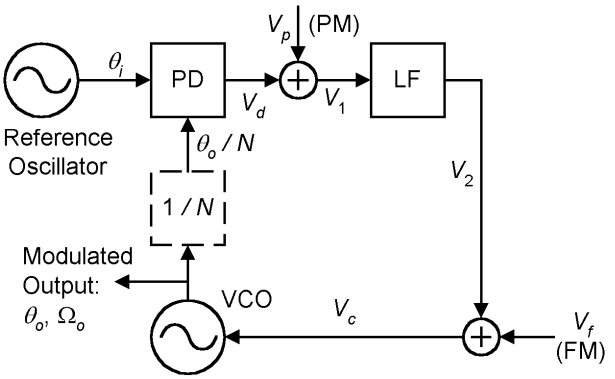


Figure 16.1 Angle modulation by a PLL.

Recollect from Chapter 2 that $E(s)$ is a highpass function. Therefore, the highpass corner frequency (closely related to loop bandwidth) must be smaller than the lowest modulation frequency. The phaselock frequency modulator cannot generate a constant-frequency offset. Also, to avoid nonlinear distortion, the VCO control characteristic must be linear; K_o has to be a constant over the frequency-deviation range. Feedback compensates for nonlinearity of the PD characteristic.

Output phase deviation (for either PM or FM) induces a phase error at the phase detector. Since the phase detector has limited range, it is not possible to obtain a large modulation index (peak phase deviation) if the VCO drives the PD directly. Extended-range PDs (e.g., the PFD of Section 10.3, but beware of crossover distortion) are of some help, but the best of these detectors restricts the phase deviation to less than 2π radians, peak. To achieve a larger modulation index, one could operate the VCO at an integer harmonic N of the input reference frequency and divide the VCO frequency by N (dashed block in Fig. 16.1) before applying feedback to the phase detector. In this manner the peak phase error is $1/N$ times the peak deviation of the VCO. Arbitrarily large indices can be generated by making N sufficiently large.

16.1.2 PLL Measurements via Modulations

Figure 16.1 also shows techniques for measurements of the response of a PLL by taking advantage of its modulation capabilities. The loop is locked to an unmodulated reference signal and a test signal is injected at V_p or V_f, as dictated by the circumstances of the system being tested. Responses are available for measurement as follows:

- For injection at V_p:

$$\begin{aligned} V_d(s) &= -H(s)V_p(s) \\ V_1(s) &= E(s)V_p(s) \end{aligned} \tag{16.4}$$

- For injection at V_f:

$$\begin{aligned} V_2(s) &= -H(s)V_f(s) \\ V_c(s) &= E(s)V_f(s) \end{aligned} \tag{16.5}$$

Such measurements are one means for checking that a PLL actually has the behavior intended for it—that no error has crept into its design or construction.

16.1.3 Delta–Sigma PLL Modulators

The inability of a conventional PLL FM modulator to accept low-frequency modulation, or modulation with a DC component, is a serious limitation in many instances. A $\Delta\Sigma$ approach [16.1–16.4] both circumvents the limitation and fits

in nicely with a frequency synthesizer. In a $\Delta\Sigma$ synthesizer such as that illustrated in Fig. 15.7, let the frequency control u_c comprise not only a fixed digital number representing the desired center frequency of the signal but also digital samples of the required modulation. The synthesizer generates a signal with average frequency and desired modulation determined by the control u_c, with no impediment to low frequencies in the modulation, not even to a DC component. Of course, the bandwidth of the PLL has to be wide enough to accommodate the modulation, the comparison frequency at the PD has to be large enough that the modulation is adequately sampled, and the loop bandwidth still must be narrow enough to suppress the high-frequency quantization noise of the $\Delta\Sigma$ converter. The references cited examine other aspects of importance to design engineers.

16.2 PHASELOCK DEMODULATORS

Phaselock loops are used for the demodulation of many kinds of modulated signals. Applications include coherent amplitude detectors (product detectors), phase demodulators (PM detectors), and frequency demodulators (FM discriminators). Figure 16.2 shows pickoff points in a PLL for the recovery of each type of modulation and establishes nomenclature for the discussion to follow.

16.2.1 PLLs for AM Demodulation

A PLL is not directly responsive to amplitude modulation, as will be shown first. An explanation of the role of PLLs in coherent AM demodulation then follows.

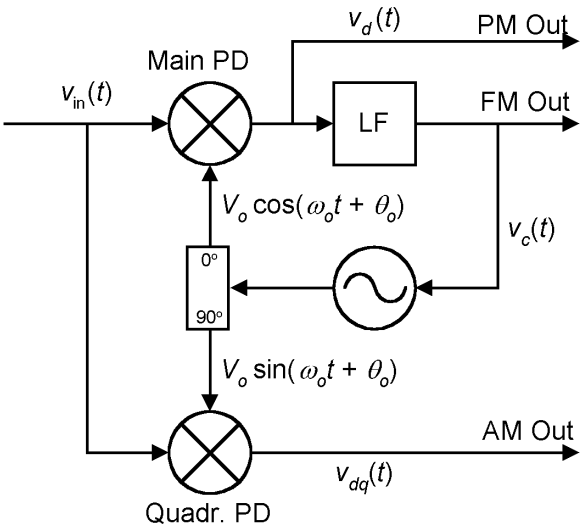


Figure 16.2 Demodulation options for a PLL.

PLL Response to Amplitude Modulation Let the noise-free input signal be amplitude modulated as

$$v_{\text{in}}(t) = V_s x(t) \sin(\omega_i t + \theta_i) \tag{16.6}$$

where $x(t)$ is arbitrary dimensionless amplitude modulation and the other symbols are the same as in Chapters 2 and 6. The phase detector is conveniently modeled as a multiplier that generates the product of v_{in} and the VCO output. Discarding double-frequency terms, the PD output is found to be

$$v_d(t) = K_d x(t) \sin \theta_e \tag{16.7}$$

Only the average (DC) output of the phase detector is useful in establishing phaselock; any fluctuating components, such as ripple, are just potential sources of tracking disturbance that are to be suppressed. The average value of V_d is

$$\text{avg}[v_d(t)] = \overline{x}(t) K_d \sin \theta_e \tag{16.8}$$

where the overbar indicates a time average. There is useful output (i.e., the loop is able to lock) only if $\overline{x} \neq 0$. The modulation must have a DC component for a discrete carrier component to be present; an ordinary PLL is unable to lock to a suppressed-carrier signal. (Phase detectors with special nonlinearities are used for locking to suppressed-carrier signals, but those are not under consideration at this juncture.)

Represent the amplitude modulation as $x(t) = x'(t) + \overline{x}$, where x' has zero mean and $\overline{x} \neq 0$. The phase detector output becomes

$$v_d(t) = [x'(t) + \overline{x}] K_d \sin \theta_e \tag{16.9}$$

But if the loop is phaselocked and is tracking properly, $\theta_e \approx 0$ and there is near-zero output from the phase detector, irrespective of the properties of $x'(t)$. Therefore, to a first approximation, a PLL does not respond to AM that might be present on its input.

For a more concrete example, consider a signal with sinusoidal AM of modulation frequency ω_m and index m to be applied to a perfectly tuned first-order PLL, so the loop equations are, discarding double-frequency terms:

$$\begin{aligned} v_{\text{in}}(t) &= V_s(1 + m \sin \omega_m t) \sin(\omega_i t + \theta_i) \\ v_d(t) &= K_d(1 + m \sin \omega_m t) \sin(\theta_i - \theta_o) \\ \frac{d\theta_o}{dt} &= K_o v_d \end{aligned}$$

Let $\theta_i = 0$ and recollect that $K_o K_d = K$. Combining the equations above yields the differential equation of the loop,

$$\frac{d\theta_o}{dt} = -K(1 + m \sin \omega_m t) \sin \theta_o$$

or rearranging yields

$$\frac{d\theta_o}{\sin\theta_o} = -K(1 + m\sin\omega_m t)\,dt$$

Integrating both sides gives

$$\ln\left(\tan\frac{\theta_o}{2}\right) = -Kt + \frac{mK}{\omega_m}\cos\omega_m t + C$$

where C is a constant of integration. Taking exponentials

$$\tan\frac{\theta_o}{2} = \exp(-Kt)\exp\left(\frac{mK}{\omega_m}\cos\omega_m t\right)\exp(C)$$

which vanishes for large t.

Therefore, if a PLL ultimately tracks with zero average phase error, the presence of amplitude modulation does not alter the equilibrium condition nor does it introduce any phase modulation of the VCO. If steady-state phase error is not zero, there is a complicated, nonlinear interaction between amplitude modulation and phase error, an issue to be revisited when considering FM demodulators.

Coherent Amplitude Detectors Following the development of Chapter 6, consider the input to the PLL of Fig. 16.2 to be an amplitude-modulated signal plus additive narrowband Gaussian noise:

$$v_{\text{in}}(t) = x(t)V_s\sin(\omega_i t + \theta_i) + n_c(t)\cos(\omega_i t + \theta_i) - n_s(t)\sin(\omega_i t + \theta_i)$$

Multiply the input v_{in} by a 90° phase-shifted version $v_q = V_o\sin(\omega_i t + \theta_o)$ of the VCO output as in Fig. 16.2, whereupon the difference-frequency output of the multiplier is

$$v_{dq}(t) = K_d\left[x(t)\cos\theta_e + \frac{n_c(t)}{V_s}\sin\theta_e - \frac{n_s(t)}{V_s}\cos\theta_e\right]$$

If the VCO is tracking properly, θ_e is nearly zero, so the output of the coherent amplitude detector (CAD) is, closely,

$$v_{dq}(t) \approx K_d\left[x(t) - \frac{n_s(t)}{V_s}\right] \tag{16.10}$$

which consists of the linear sum of the desired amplitude modulation plus the component of noise modulation that lies in phase with the carrier. The quadrature noise component and moderate amounts of phase modulation are rejected. The CAD performs amplitude demodulation.

Amplitude demodulation also could have been performed by a simple envelope detector, provided that $x(t)$ is always positive. If $x(t)$ becomes negative ("overmodulation" in radio-engineering parlance), an envelope detector generates severe distortion. The coherent amplitude detector imposes no such restriction; it reproduces $x(t)$ without distortion even if $x(t)$ reverses polarity. Moreover, the CAD will even demodulate a suppressed-carrier signal, provided that there is some means of generating a properly phased local reference: that is, some means of locking the VCO to the proper phase.

Coherent amplitude detectors are also used for low-distortion demodulation of single-sideband (SSB) and vestigial sideband (VSB) signals. A remanent pilot carrier must be transmitted with the signal if the local carrier reference is to be phaselocked (as is essential for coherent demodulation of VSB). The I/Q demodulators so common in the reception of digital data signals are all coherent amplitude detectors, one for the in-phase components of the complex amplitude modulation and one for the quadrature components.

A major advantage of a CAD lies in its linear processing of signal and noise. Its output is the linear superposition of signal and noise; there is no intermodulation between the two, irrespective of the input signal-to-noise ratio. Unlike an envelope detector, a coherent amplitude demodulator does not suffer a threshold degradation (also called *squaring loss*) if the input signal falls below the noise.

16.2.2 Phase Demodulation

Assume that the input signal is phase-modulated according to

$$v_{\text{in}}(t) = V_s \sin[\omega_i t + \theta_i(t)] \tag{16.11}$$

where $\theta_i(t)$ is the phase modulation. If the peak phase excursion is small enough that the PLL remains in its linear domain, the linear transfer-function analysis of Chapter 2 applies and the output of the phase detector can be represented as

$$V_d(s) = K_d E(s) \theta_i(s) \tag{16.12}$$

Since the error transfer function $E(s)$ has a highpass response, phase modulation of sufficiently high frequency appears unaltered at the output of the phase detector. A lower-frequency component is reduced by the feedback factor at that frequency.

Carrier-tracking coherent PM systems must be designed so that the demodulator loop does not suppress the modulation. Subcarriers are often employed to move the information spectrum outside the PLL bandwidth. A subcarrier also moves the signal information away from low-frequency noise and drift disturbances of the VCO. Undistorted demodulation is achieved if the peak phase excursion remains within the linear portion of the phase-detector s-curve. To enhance linearity, an extended-range PD described in Chapter 10, might sometimes be useful. However, since all PD s-curves revert to sinusoidal for low input SNR (Section 10.4.3), the sinusoidal s-curve is of great importance.

Distortion can be tolerated in some applications; demodulation to a subcarrier might be one example. The phaselock loop demands that the signal contain a trackable carrier, but the modulation index is otherwise unrestricted. Assume that the modulation is sinusoidal with a modulating frequency ω_m that is outside the bandwidth of the PLL. The modulated input phase is

$$\theta_i(t) = \Delta\theta \sin \omega_m t \tag{16.13}$$

The loop is unable to track the modulation, so the phase error $\theta_e = \theta_i$ and the output of the phase detector is

$$v_d(t) = K_d \sin(\Delta\theta \sin \omega_m t) \tag{16.14}$$

which is a nonlinear function of the modulation.

Some examples of distorted output waveforms are given in Fig. 16.3 for various choices of peak deviation $\Delta\theta$. Distortion clearly worsens as $\Delta\theta$ becomes larger. The plots all are for a zero-mean value of θ_e; any phase offset would cause asymmetric distortion. It is sometimes useful to select $\Delta\theta$ so as to maximize the fundamental modulation-frequency output from the phase detector. Equation (16.14) is periodic, so it can be expanded in a Fourier series of sinusoids; the Fourier coefficient of the nth harmonic is the nth-order Bessel function $J_n(\Delta\theta)$. The coefficient of the fundamental is $J_1(\Delta\theta)$, which has its maximum at $\Delta\theta = 103^\circ$ (1.8 rad). The remanent carrier amplitude is proportional to $J_0(\Delta\theta)$.

More efficient signal design is achieved if the subcarrier has a square waveform instead of a sinusoidal waveform. The amplitude of the fundamental component of the subcarrier recovered from the demodulator is $(4/\pi) \sin \Delta\theta$ and the amplitude of the remanent carrier is proportional to $\cos \Delta\theta$. The peak of the square-wave phase deviation is $\Delta\theta$. Figure 16.4 shows the modulation and carrier coefficients for sinusoidal and square modulation plotted against $\Delta\theta$. For

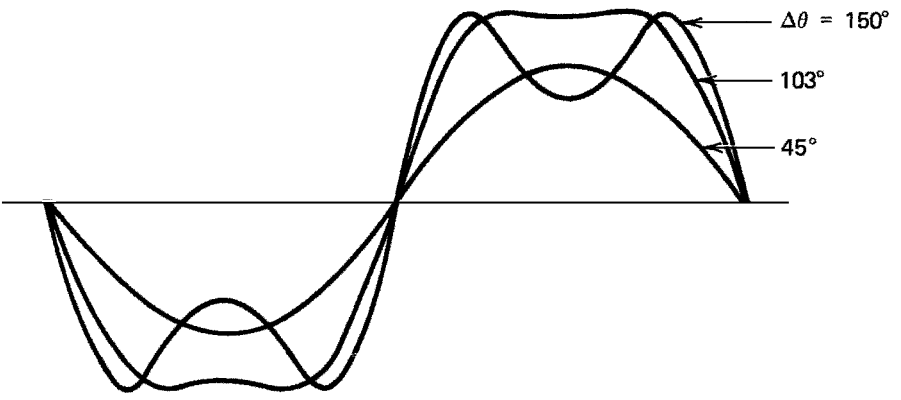


Figure 16.3 Phase demodulator waveforms: sinusoidal s-curve in a PD; sinusoidal modulation $\theta_i(t) = \Delta\theta \sin \omega_m t$.

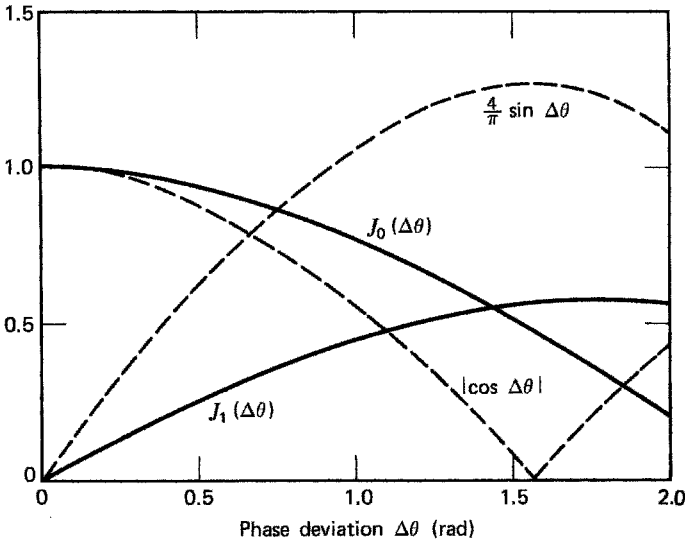


Figure 16.4 Phase modulation parameters.

equal levels of carrier suppression, the square wave provides larger amplitude of recovered subcarrier at the PM demodulator.

If the modulation consists of multiple tones, nonlinearity causes intermodulation [16.5] between the tones, as well as the production of harmonics. It is entirely possible for IM products to be stronger than some of the recovered tones. Design of a PM communications link [16.6] is a complex matter; the foregoing paragraphs give only a few examples of some of the problems that may arise.

A coherent phase demodulator does not generate intermodulation between signal and noise, despite the modulation distortion. Therefore, it is capable of working deeply into noise without a squaring loss penalty. It shares this property with the coherent amplitude detector, a fact that should be no surprise since the two demodulators have identical circuits and differ only in the phase of the local reference.

16.2.3 Frequency Demodulation

Suppose that a frequency-modulated input signal is applied to a PLL. For the loop to remain in lock, it is necessary that the frequency of the VCO track the incoming frequency very closely. The frequency of the VCO is proportional to the control voltage (v_c in Fig. 16.2), so the control voltage must be a close replica of the modulation on the signal. Modulation may therefore be recovered from the VCO control voltage. This is the principle of the phaselock FM demodulator (PLD). The PLD is a modulation-tracking loop, as defined in Section 5.2.4.

A modulation-tracking loop can be used to demodulate large-deviation PM as well as FM. Control voltage v_c is an analog of frequency modulation, but the

original phase modulation can be retrieved by integrating v_c. Integration of v_c is a nontrivial problem, best accomplished in a digital PLL and integrator [16.7].

By use of the linear analysis of Chapter 2, the transfer function relating control voltage $V_c(s)$ to signal phase modulation $\theta_i(s)$ is found to be

$$V_c(s) = \frac{s\theta_i(s)H(s)}{K_o} \tag{16.15}$$

Denote the instantaneous frequency modulation by $m(t)$ in rad/sec. Phase and frequency modulations are related by $m(t) = d\theta_i(t)/dt$, since frequency is simply the derivative of phase. Taking Laplace transforms gives $L\{m(t)\} = M(s) = s\theta_i(s)$ and substituting into (16.15) gives

$$V_c(s) = \frac{M(s)H(s)}{K_o} \tag{16.16}$$

which shows the transfer function between frequency modulation and the resulting VCO control voltage. The message recovered is equivalent to the original message, filtered by the closed-loop transfer function $H(s)$ and divided by the VCO gain factor K_o. If the loop is linear and if its bandwidth is large enough compared to the message bandwidth, $v_c(t)$ is a faithful reproduction of $m(t)$. To avoid distortion, it is evident that the VCO control characteristic must be linear, since K_o appears directly in (16.16); that is, K_o must truly be a constant and not a function of v_c.

Phase-detector gain enters (16.16) only through its influence on $H(s)$, which is significant only at higher modulation frequencies since $H(0) = 1$ irrespective of K_d. For this reason, and because of the reduction of PD distortion by feedback that is noted in Chapter 5, low-distortion operation is possible with a nonlinear phase detector. Avoidance of linear filtering distortion, avoidance of nonlinear PD distortion, and indeed, the very ability to maintain track (Chapter 5) are all enhanced by large bandwidth of the PLL. These interrelated reasons all point to a loop bandwidth that is larger than the message bandwidth. It becomes apparent later that the loop bandwidth should actually be larger than the RF bandwidth of the modulated signal, a conclusion that is not obvious at this point.

16.2.4 FM Noise

Let the frequency modulation be sinusoidal with peak deviation Δf Hz and modulating frequency f_m Hz. Therefore, $m(t) = 2\pi\,\Delta f \sin(2\pi f_m t)$ and the input signal becomes

$$v_{\text{in}}(t) = V_s \sin\left(\omega_i t + \frac{\Delta f}{f_m}\cos 2\pi f_m t\right)$$

so the PLD output signal is

$$v_c(t) = \frac{1}{K_o} 2\pi\ \Delta f\ \sin 2\pi f_m t \tag{16.17}$$

where it has been assumed that $H(j2\pi f_m) \approx 1$; that is, the loop does not filter the modulation appreciably.

Preceding the demodulator is a bandpass filter centered at the signal frequency ω_i and with noise bandwidth B_i Hz. The filter is assumed to have sufficient bandwidth, amplitude flatness, and phase linearity to cause negligible distortion to the signal. A lower-bound constraint is $B_i > 2\Delta f$, and practical bandwidths are usually substantially larger. White Gaussian noise of one-sided density N_0 V^2/Hz is added to the signal at the filter input. Signal-to-noise power ratio at the filter output is

$$\rho_i = \frac{V_s^2}{2B_i N_0} \tag{16.18}$$

the *carrier-to-noise ratio* (CNR) that appears throughout the FM literature.

The effect of noise in a PLL is represented in Chapter 6 as an additive noise generator $n'(t)$, with one-sided spectral density $W_{n'}(f)$ rad^2/Hz, inserted into the linearized phase detector [Fig. 6.2 and equations (6.7) and (6.17)]. If the bandpass filter has a rectangular passband (not a necessary assumption) and $\rho_i \gg 1$ (necessary to assure linearity), $W_{n'}(f) = 2N_0/V_s^2$ for $0 \le f < B_i/2$ and is zero otherwise.

By use of the transfer-function methods of Chapters 2 and 6, the spectral density of noise appearing in the control voltage v_c is found to be

$$W_{vc}(f) = \frac{|(2\pi f)H(f)|^2}{K_o^2} W_{n'}(f) \qquad V^2/\text{Hz} \tag{16.19}$$

If $W_{n'}(f)$ is flat, $W_{vc}(f)$ has the familiar parabolic shape associated with demodulated FM noise, within the passband of $H(f)$.

The signal and noise of the control voltage are processed through an external lowpass post filter. As is customary, assume the passband to be rectangular with cutoff frequency equal to the modulating frequency f_m. The recovered modulation is passed without loss, but all higher-frequency components of noise are suppressed completely. Assume further that $|H(f)|$ is flat from DC to f_m. Noise intensity at the output of the post filter is given by

$$\sigma_{nf}^2 = \int_0^{f_m} W_{vc}(f)\, df \approx \int_0^{f_m} \frac{(2\pi f)^2 W_{n'}(f)}{K_o^2}\, df = \frac{8\pi^2 N_0 f_m^3}{3K_o^2 V_s^2} \qquad V^2 \tag{16.20}$$

The output signal-to-noise ratio is given by the mean square of $v_c(t)$ in (16.17) divided by the mean-squared noise (16.20):

$$\text{SNR}_o = \frac{3\Delta f^2 V_s^2}{4N_0 f_m^3} = \frac{3\Delta f^2 B_i \rho_i}{2f_m^3} \tag{16.21}$$

[Although (16.21) was derived for sinusoidal modulation, similar expressions can be obtained for any other modulation format.]

If $\Delta f = B_i/2$ (the maximum deviation that remains within the input filter) and if the modulation index is defined as $\beta = \Delta f/f_m$, then

$$\text{SNR}_o = 3\beta^3 \rho_i \tag{16.22}$$

which is the classical expression of FM improvement factor [16.8, 16.9]. This result is exactly the same as that obtained for a conventional frequency discriminator. For large CNR a PLD has noise performance identical to that of ordinary discriminator circuits.

To achieve the FM improvement, a conventional discriminator must be preceded by a limiter. Ordinary discriminator circuits are amplitude sensitive and the limiter is essential to suppress the AM component of noise. The PLD furnishes the FM improvement without employing a limiter. In effect, the PLL ignores the component of noise that lies in phase with the signal and is disturbed only by the quadrature component. It is demonstrated in the next section that a limiter worsens threshold performance; ability to deliver FM improvement without incurring limiter loss is one motive for using a phaselocked discriminator in place of a conventional circuit.

16.3 FM THRESHOLD

The ideal performance of (16.22) is achieved at high CNR, but below some minimum CNR, known as the *threshold*, the output SNR_o deteriorates very rapidly with further reduction of CNR. This section is devoted to an exposition of the threshold effect and how a PLD can lower the threshold CNR compared to that of conventional frequency discriminators.

Be forewarned that no good quantitative theory has yet been devised for rigorous explanation of the threshold of a PLD. Operation of the loop is in the nonlinear region, but the nonlinear methods reported in earlier chapters are inadequate to cope with a PLL subjected to bandlimited additive noise together with modulation. This section attempts a heuristic explanation of PLD threshold based on experimental evidence. The explanations are inadequate in that the threshold CNR cannot be predicted nor can the optimum loop configuration be calculated. However, sufficient information is given so that an engineer can optimize design parameters by experiment.

16.3.1 Threshold Characterization

Excessive deterioration of output SNR_o is the best recognized manifestation of the FM threshold, as sketched in Fig. 16.5. At high CNR the output SNR_o is linearly proportional to CNR, as in (16.21). The plot of SNR_o vs. CNR has unit slope on log-log coordinates for large CNR. There is a break in the slope at threshold CNR and the curve is much steeper for low CNR. An exact breakpoint is difficult to recognize since the curve and its slope are continuous. The point

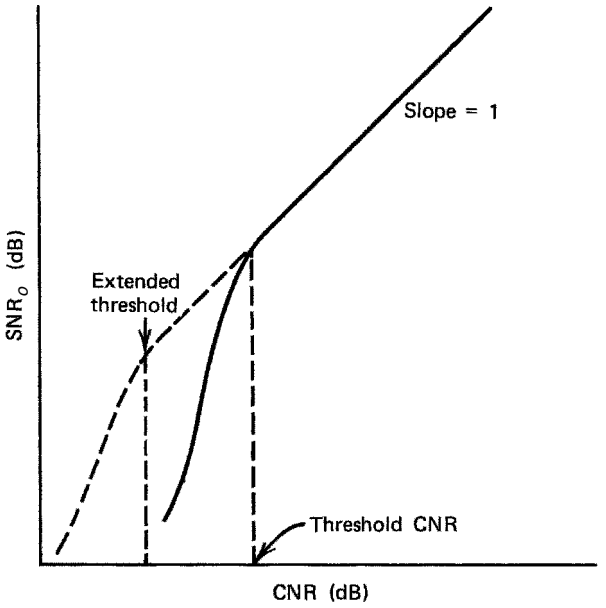


Figure 16.5 FM threshold effect on SNR_o.

of 1-dB deterioration from the extended straight line of SNR_o at high CNR is the customary definition of threshold CNR, but that choice is entirely arbitrary.

The threshold performance of an ideal frequency discriminator is a reference against which to compare other demodulators. A discriminator is *ideal* if it produces an output baseband voltage proportional to the rate of change of phase of the bandpass process applied to its input: the instantaneous frequency of signal plus noise. The phase is that of the resultant of the desired signal plus added noise. An ideal discriminator is insensitive to AM components of signal or noise; performance of well-designed conventional limiter-discriminator circuits is close to ideal.

The term *ideal* in no way implies *optimum* with regard to threshold. All good discriminators have the same performance at large CNR, but the ideal discriminator does not have the lowest threshold. If a discriminator has a lower threshold than that of an ideal discriminator, it is said to be an *extended-threshold demodulator*. An example of threshold extension is sketched in Fig. 16.5.

The subthreshold SNR_o of an ideal discriminator can be calculated from an exact analysis by Shimbo [16.10, 16.11], an analysis produced after a long series of approximate analyses by earlier authors. The difficulty of the problem is perhaps best illustrated by the fact that the exact analysis was not published until some 45 years after the nature of FM was recognized [16.12].

In many applications the SNR_o below threshold is of little interest because normal operation occurs almost exclusively at CNR values above threshold. (The disturbances accompanying below-threshold operation are often much more

disruptive than might be expected from SNR_o considerations alone; their nature will be described shortly.) It is often sufficient, for signal design and link budget purposes, to be able to predict the threshold CNR. A phaselock demodulator is valuable because it offers threshold extension with a relatively simple circuit. The amount of extension is not predictable by any existing theory and depends on signal parameters. Very roughly, a few decibels of improvement are achieved in typical applications.

16.3.2 FM Clicks

The prediction of threshold could, of course, be performed by evaluation of Shimbo's equations, but the approximate method by Rice [16.9] is easier to use and also has concepts that aid understanding of the PLD. Outputs of below-threshold discriminators are observed to contain large-amplitude short-duration spikes or *clicks* (to use Rice's term). These clicks appear only very rarely above threshold. More frequent appearance of clicks is a manifestation of the onset of threshold.

Note the wording of the last sentence. It is not stated that clicks cause threshold or that clicks are the only manifestation of below-threshold operation; neither would be true. Nonetheless, the threshold CNR can be predicted with good accuracy if the average click rate can be calculated.

A click results when noise causes the resultant of signal plus noise to take on (or lose) one complete cycle as compared to the signal alone. A phaser diagram (Fig. 16.6) illustrates the generation of clicks. The phaser reference is chosen so

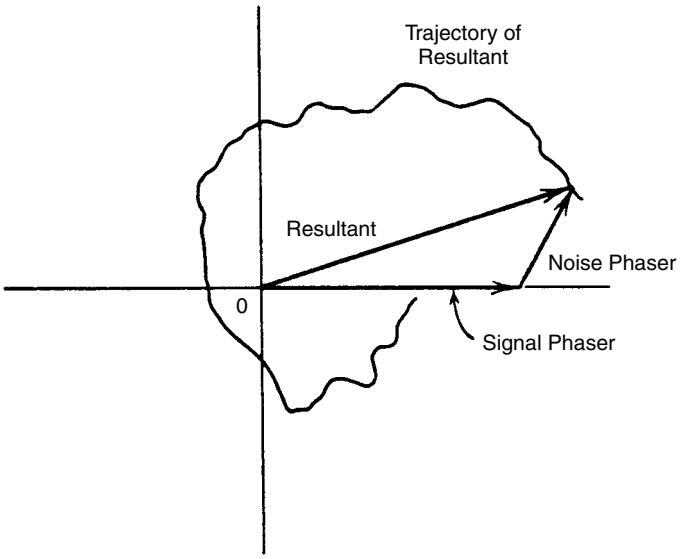


Figure 16.6 Phaser diagram of click generation.

that the signal remains fixed at 0° and constant amplitude while the noise adds to the signal with random fluctuating amplitude and phase. The resultant traces out a continuous trajectory in the complex plane. A cycle is gained or lost (i.e., a click is generated) every time the trajectory encircles the origin.

A click is possible only if instantaneous noise exceeds the signal amplitude and if the phase of the noise goes through opposition with the phase of the signal. In the vicinity of threshold, the noise amplitude associated with a typical click event is likely to be just slightly stronger than the signal, so the click trajectory is likely to pass very close to the origin; that is, the amplitude of the resultant often can be expected to be small in the middle of a click passage.

Under conditions of near cancellation of signal by the noise, a small change of noise phase can cause a large change of resultant phase. Therefore, a click trajectory can sweep around the origin very rapidly, much more quickly than might be suggested by the restricted bandwidth of the input filter. These features of amplitude and phase have considerable bearing on the response of a PLD.

Clicks may also be examined by means of phase and frequency waveforms, as in Fig. 16.7. In the absence of noise, the fixed-signal phaser of Fig. 16.6 produces a constant resultant phase of 0°. Small noise causes small phase fluctuations about zero, while a click causes the resultant phase waveform to have a 2π step (Fig. 16.7*a*). Frequency is the time derivative of phase—the rate of rotation of the resultant phaser—and is sketched in Fig. 16.7*b*. Small phase noise produces small frequency noise, while a phase step produces a large frequency spike; this is the spike or click that is heard in an audio message or observed in the laboratory.

Click waveforms vary greatly [16.13]; the only property they have in common is that each has an area of 2π or an integer multiple thereof. Polarity of a click

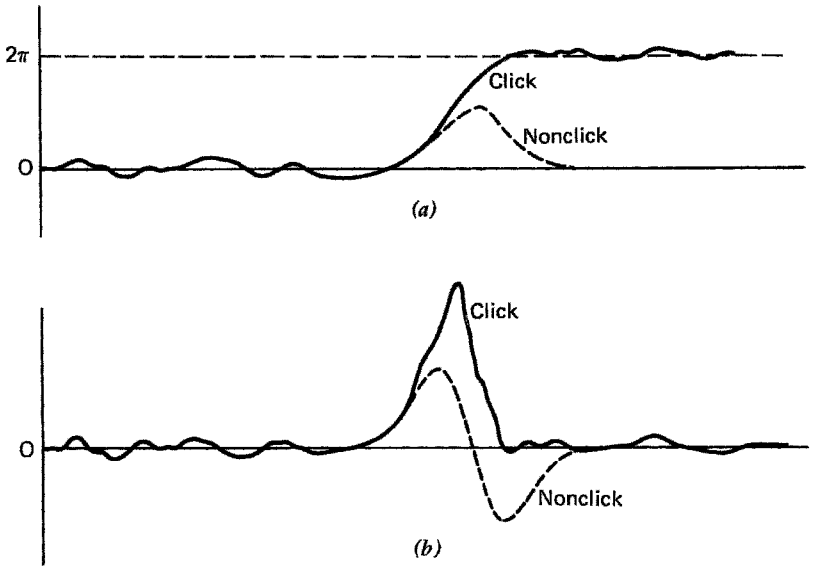


Figure 16.7 Illustrative click waveforms: (*a*) phase; (*b*) frequency.

pulse depends on whether a cycle is lost or gained. An individual click pulse is essentially unipolar. The duration of a click event is usually short compared to the reciprocal of baseband signal bandwidth. To calculate the influence of clicks on output SNR_o, it is useful to approximate the waveform as an impulse with area 2π. An impulse has a flat spectrum extending down to DC and with substantial energy in the baseband.

Figure 16.7 also shows a large phase disturbance that does not cause a click: a nonclick. The peak of the frequency pulse of the nonclick is much smaller than that of a completed click. More significantly, the frequency pulse associated with a nonclick is a doublet, which has its energy concentrated at high frequencies and falls off to zero at DC. A doublet causes much less disturbance to a lowpass system than does a unipolar pulse.

If the average click rate is known, the contribution to output noise can be calculated [16.9, 16.14]. The CNR at which clicks increase total noise 1 dB above that calculated by (16.20) alone is the customary formal definition of threshold. There is substantial energy in a click, so threshold occurs at a surprisingly small click rate. Rice [16.9] has determined the click rate for an ideal discriminator. It is a function of CNR, input passband shape, and modulation parameters. Use of his formulas provides a good prediction of threshold of an ideal discriminator.

16.3.3 Clicks in PLD

Unfortunately, no one has yet been able to analyze the output click rate of a phaselock demodulator. The click concept gives physical insight into the operation of a PLD, but a quantitative theory has been unattainable. This section summarizes the author's qualitative understanding of the problem, based on largely unpublished experimental work. (Smith [16.15, 16.16] has pursued a PLD approach similar to that presented here, but neglecting the input filter.)

Filter and Bandwidth Considerations First, examine a complete block diagram of phaselocked FM demodulator, as in Fig. 16.8. It is composed of an input bandpass filter, a phase detector, a loop filter, a VCO, and a lowpass post filter. All five elements are essential to proper operation of the PLD even though many earlier publications have ignored the input and output filters completely. The post filter contains the deemphasis networks, correction for linear-filter distortion

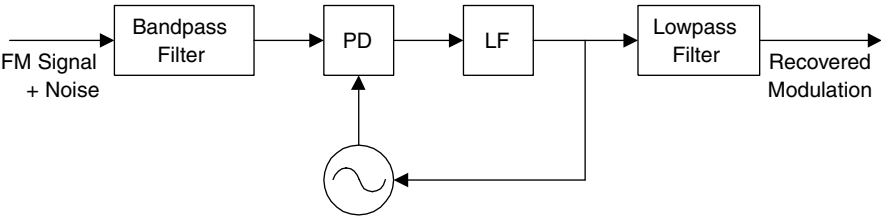


Figure 16.8 PLL FM demodulator (PLD).

by the PLL transfer function, and the main baseband filtering of the recovered message. This filtering is needed to achieve the FM improvement of (16.22). However, the post filter processes the signal only after it is recovered from the PLL; it clearly can have no influence on tracking performance and therefore does not affect threshold. Neglect of the post filter is justified in a study of threshold phenomena.

On the other hand, neglect of the input filter is completely wrong. It is commonplace to think of a PLL as a narrowband device that combats noise by means of its narrow bandwidth. This is a misconception when applied to a PLD, where the loop bandwidth must be rather large. In fact, the loop bandwidth is likely to be considerably larger than the RF signal bandwidth.

The input filter must be wide enough to avoid excessive distortion of the message; this is a very complex subject in itself and is not treated here. A lower bound on input bandwidth is established by Carson's rule [16.12]:

$$B_i > 2(\Delta f + f_m) \tag{16.23}$$

which is appropriate for sinusoidal modulation, or a modified version,

$$B_i > 2(B_m + \gamma\sigma_f) \tag{16.24}$$

which is appropriate for Gaussian modulation with lowpass bandwidth B_m, rms frequency deviation σ_f, and "crest factor" γ (see Section 5.2.4).

Experiments show that the best choice of loop bandwidth (discussed below) is substantially larger than the Carson's rule bandwidth. Loop bandwidth will so exceed the input-filter bandwidth in a well-designed PLD that the loop will provide no appreciable linear filtering of the RF noise. The only significant noise reduction is provided by the input filter. For this reason, the input filter should be as narrow as possible, consistent with signal-distortion specifications. A wider filter admits extra noise, thereby degrading performance. This statement has been tested by experiment. The results show conclusively that excessive bandwidth of the input filter increases the threshold of the PLD. The amount of degradation depends on the degree of bandwidth excess; some measured results are shown in Fig. 16.9 for sinusoidal modulation and a first-order loop.

Frequency deviation out to the edges of the input filter passband raises concerns with the response in the edges. Most filters roll off gradually at frequencies away from center; frequency deviation of the signal imposes a corresponding amplitude variation on the filter output. This incidental AM is quite noticeable as a scalloping of the signal envelope applied to the PLL.

To a first approximation, the PLL output is insensitive to amplitude effects. However, examined more closely, it becomes evident that the loop gain is proportional to signal amplitude (in a multiplier PD without any limiter), so the scalloping produces an instantaneous reduction in gain. In consequence, the tracking ability of the PLL is impaired and the loop is less capable of tracking the modulation excursions. Therefore, the scalloping worsens the threshold,

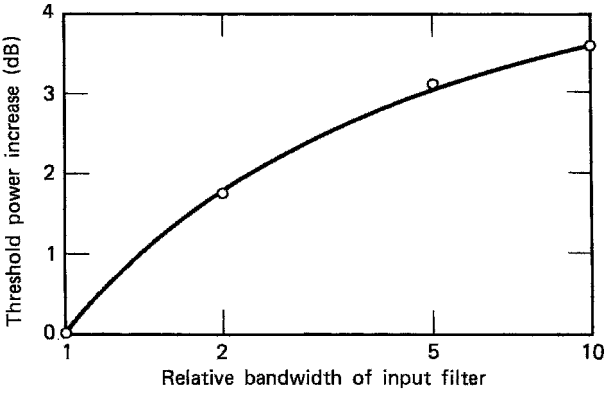


Figure 16.9 Measured threshold increase due to the excess bandwidth of an input filter. Unity relative bandwidth corresponds to Carson's rule [eq. (16.23)]. The ordinate shows the additional signal power needed at the threshold because of a widened input bandwidth.

particularly in a first-order loop where maximum loop stress occurs at maximum frequency deviation (Section 5.2.4). A desirable bandpass filter would have a flat response over the entire frequency-deviation range.

PLD Response to Input Clicks If an input filter is good for use with a PLD it is presumably also good with a conventional discriminator. Using identical input filters, a PLD can exhibit a lower threshold than a conventional discriminator. How does the improvement come about? Signal plus noise at the output of the bandpass filter will have click events as described above. [**Comment**: A click is a property of the input signal and noise, entirely independent of any demodulator.] The average rate of clicks is given by Rice's analysis. An ideal discriminator, by definition, demodulates every one of the clicks in the signal applied to it. A PLL is unable to follow some of the clicks, so its output remains closer to the original message than the output of an ideal discriminator. *It is this inability to follow some of the input clicks that accounts for the improved threshold of a PLD* [16.17].

Why does the PLL fail to follow some clicks? One reason is that the PLL is a limited-bandwidth element and a typical click is quite fast; the loop often cannot move quickly enough to follow the click around the circle. The sluggishness is emphasized when the noise nearly cancels the signal so that amplitude of the resultant is very small as the trajectory whips around the origin. In the vicinity of threshold CNR, most click events are associated with small amplitudes of the resultant. A small amplitude means reduced gain of the PLL and therefore reduced ability to follow the resultant phasor. Reduction of gain is a nonlinear effect that, in this instance, apparently improves the threshold behavior.

Now it is possible to see why a limiter would worsen threshold performance. An ideal hard limiter holds the amplitude of the resultant phasor constant,

irrespective of any possible cancellation of signal by noise. If amplitude is large as the trajectory goes around the circle, the PLL is better able to follow and more input clicks are demodulated than when a limiter is omitted. Hess [16.17] has shown the deleterious effects of a limiter by experiment and by approximate analysis. Minimizing threshold demands that a limiter be omitted. Therefore, any PD that implicitly uses a limiter should be avoided: a PD whose s-curve has a sawtooth, rectangular, or triangular shape, or any sequential PD (see Chapter 10).

Sometimes one encounters the notion that PLD threshold is somehow "caused" by the nonlinearity of the phase detector and that the threshold could be avoided, or at least reduced, if only a linear PD were possible. That is not correct. Instead, the argument here is that the reduced threshold of a PLD is at least partly due to the nonlinearity of the PD and that a linear PD would yield the same threshold as an ideal discriminator. Real-world PDs have a periodic s-curve, whereas a *wide-linear* PD would include a means of counting cycles and therefore have a straight-line s-curve extending to infinity in both directions. In equilibrium, both types of PD cause the loop to track close to the PD null.

What happens when a fast input click appears? Assume that the loop is too slow to follow the click immediately and just consider behavior after the input click has ended. The periodic PD ignores the cyclic increment and the loop continues to track as if the click had never occurred. However, the wide-linear PD would recognize that an extra cycle has been accumulated, so it would produce an output corresponding to a phase error of 2π. The loop would servo out that error by adjusting the VCO phase by 2π to return to the PD null. In other words, the PLD with a wide-linear PD would be unable to ignore input clicks and eventually would track them all, even if quite slowly. A wide-linear PD would be just as bad as an ideal discriminator. The periodic nonlinearity of a real PD contributes, in part, to the improved threshold performance of a PLD.

Clearly, to ignore as many input clicks as possible, the loop bandwidth should be as narrow as possible. If the bandwidth were very large, the loop would follow all input clicks and performance would be the same as that of an ideal discriminator. On the other hand, bandwidth must not be too narrow or else modulation will cause cycle slipping even in the absence of noise (Chapter 5). A loop that is overstressed by excessive modulation is very sensitive to slips induced by small noise disturbances. A cycle slip is indistinguishable from a demodulated input click at the demodulator output; for convenience they are all called output clicks. It seems reasonable to suppose that a compromise bandwidth will minimize the output click rate.

Measurements of PLD Clicks Figures 16.10 to 16.12 show representative click-rate data that were measured on a second-order PLL. Table 16.1 gives the experimental parameters. The click rate was measured using the instrumentation described in [16.18]. Each data point represents 100 sec of accumulating output clicks. Figure 16.10 plainly bears out the prediction of an optimum loop bandwidth. Substantial improvement in click rate can be obtained by choosing the proper loop bandwidth. Or, stated negatively, a large penalty is incurred if

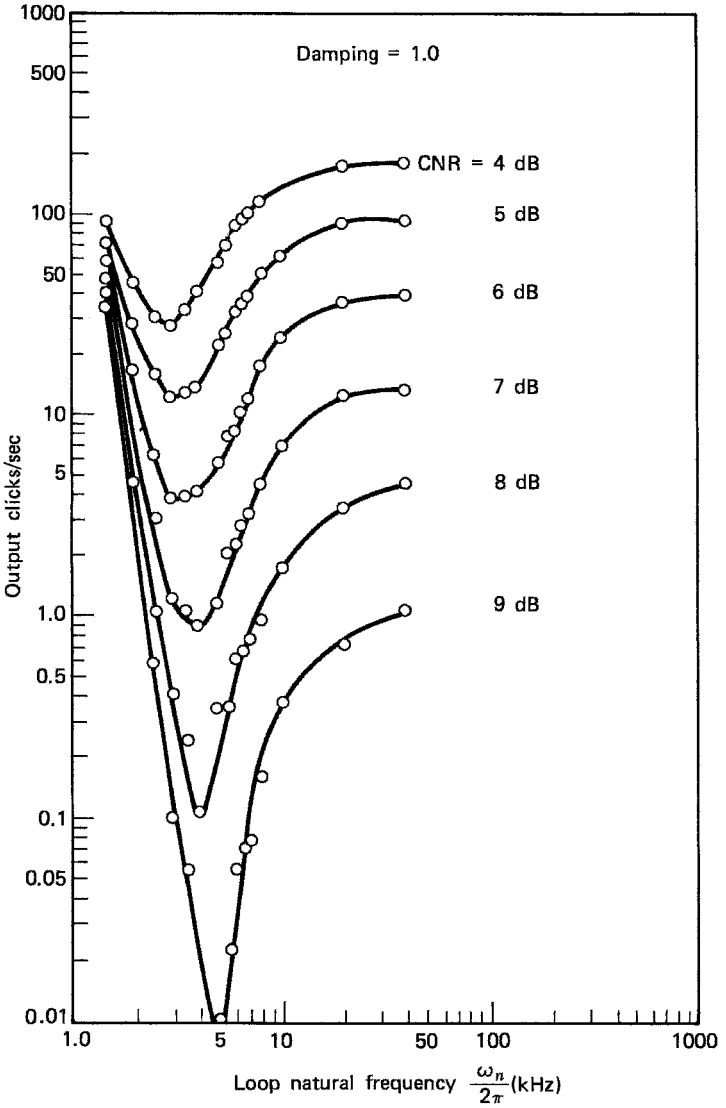


Figure 16.10 Measured PLL click rates under the conditions of Table 16.1. Damping $\zeta = 1$.

the wrong bandwidth is used. Figure 16.11 shows a broad click-rate minimum for damping in the vicinity of 1 to 2. Other data suggest that a choice of 1 to 1.5 ought to be good.

The click-rate curves of Fig. 16.12 are plotted vs. CNR, with natural frequency ω_n as a parameter. The solid black curve labeled n_r is a plot of Rice's prediction [16.9] of the click rate of an ideal discriminator with the same modulation parameters and the same input filter. Data points for a wideband PLL

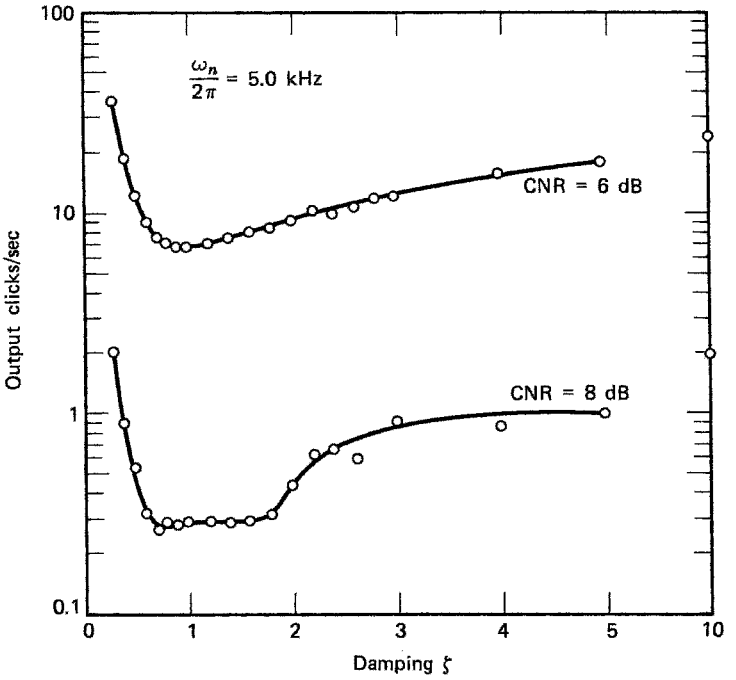


Figure 16.11 Measured PLL click rates under the conditions of Table 16.1. Natural frequency $\omega_n/2\pi = 5$ kHz.

($\omega_n/2\pi = 40$ kHz) agree very closely with the Rice prediction, supporting the statement that a wideband PLL has the same threshold as that of an ideal discriminator.

The solid straight line labeled $\overline{n}$ shows the click rate that will increase output noise of (16.20) by 1 dB. Threshold is formally defined by the intersection of $\overline{n}$ and an actual click-rate curve. Intersection for the optimum-bandwidth PLL is some 2.5 dB lower than that for the ideal discriminator. Choice of optimum bandwidth is strongly dependent on modulation parameters; a priori knowledge of message statistics is essential for designing a threshold extension PLD. Design of a conventional discriminator virtually ignores message statistics, thereby incurring a threshold penalty. On the other hand, a conventional discriminator is relatively insensitive to changes of message statistics, whereas the PLD is affected adversely and perhaps severely.

A second-order type 2 loop is not necessarily the configuration that provides a minimum threshold. If the modulation index is small (as in the examples shown), a first-order loop has nearly the same performance (as long as any steady frequency offset is small compared to the loop bandwidth K). If the modulation index is large, a PLL of higher type tracks with less phase error (Fig. 5.10). Experiments have shown that type 2 loops outperform first-order loops for large modulation index. The experimental data reported here were

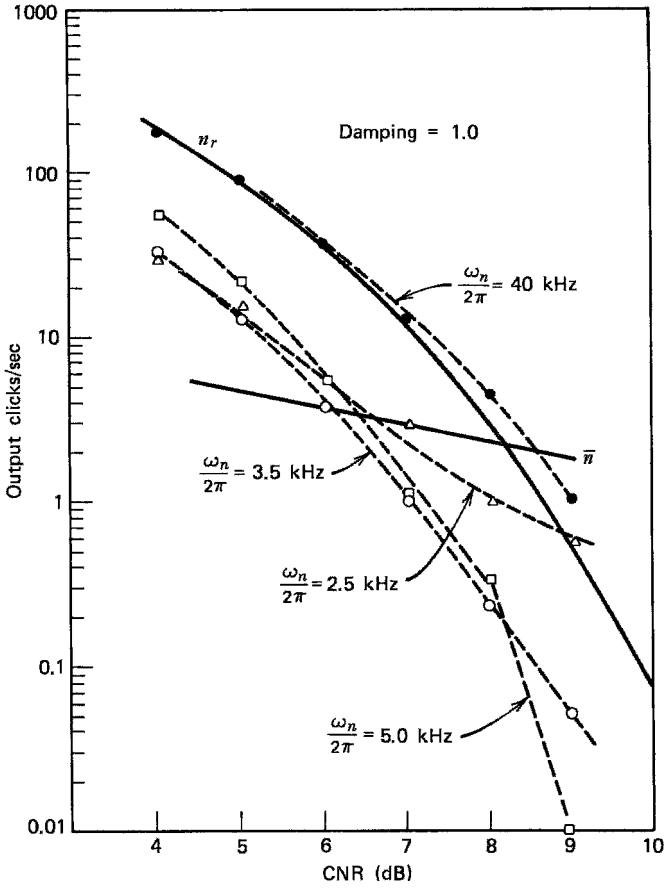


Figure 16.12 Measured PLL click rates under the conditions of Table 16.1. Damping $\zeta = 1$.

TABLE 16.1 Click-Rate Experimental Parameters

Carrier frequency: 25 MHz
Modulation type: Gaussian-noise FM
Modulation baseband spectrum: essentially flat from DC to 2.4 kHz ($B_m = 2.4$ kHz)
Deviation: $\sigma_f = 1485$ Hz, rms
RF spectral occupancy: $2(B_m + \gamma\sigma_f) = 15.2$ kHz for $\gamma = 3.5$
Passband of input filter:
  15.2 kHz at -1 dB
  18.3 kHz at -3 dB
  24 kHz at -30 dB
  Noise bandwidth = 16.4 kHz
PLL configuration:
  Second-order type 2
  Analog circuits

taken for a moderately small modulation index. Threshold extension of a PLD improves, relative to an ideal discriminator, as the modulation index increases.

16.3.4 Formal Optimization

A PLD is not the optimum FM demodulator but only an approximation thereto. The optimum demodulator would examine the entire message, even if it were of infinite duration, before producing the *maximum a posteriori* (MAP) estimate of the message. Viterbi [16.19] and Van Trees [16.20] give excellent discussions of MAP estimation applied to FM demodulation.

The integral equation of the MAP estimator is nearly identical to that of a phaselock loop. Because of the close resemblance between integral equations, many investigators have hoped that the PLL would be a good approximation to the MAP demodulator. The only difference between equations is in the limits of integration. On paper the difference seems trivial, but the PLL must track in real time—work with zero lag—and it cannot wait for the end of message before starting to process the signal.

Although barred from achieving the ultimate MAP performance, one can still ask: What is the optimum zero-lag stable PLD? To permit mathematical tractability, it is common to assume linear operation of the PLL and to determine the optimum realizable Wiener filter. Viterbi warns, emphatically, that this procedure does not lead to the MAP performance, nor does it lead necessarily to the lowest attainable threshold. Nonlinear behavior cannot be inferred from linearized analysis.

The linear approximation to variance of VCO phase caused by untracked angle modulation and additive noise is

$$\sigma_o^2 = \int_0^\infty \left[|1 - H(f)|^2 \frac{W_m(f)}{(2\pi f)^2} + W_{n'}(f)|H(f)|^2 \right] df \tag{16.25}$$

where $W_m(f)$ is the spectral density of the frequency modulation on the signal, in $(\text{rad/sec})^2/\text{Hz}$, and $W_{n'}(f)$ is the spectral density of additive noise as defined in (6.7) and (6.16). The optimum Wiener filter is a particular choice of $H(f)$ that minimizes the phase-error variance.

However, a designer is concerned with minimizing threshold: Does the Wiener filter design accomplish that also? The Wiener procedure assumes linear operation, but threshold is a strongly nonlinear phenomenon. No linear analysis can be trusted without verification. Few reports of experimental testing of Wiener-optimized loops can be found in the literature. By gleaning among several obscure sources, it was discovered that the linearly derived Wiener filter in fact does not minimize the threshold. At best, it is a starting point for an empirical search for the minimum-threshold PLD.

The potential complexity of a Wiener filter is often unacceptable. A simpler approach is to use a loop of ordinary form (e.g., a standard second-order loop) and then minimize the threshold by adjusting the loop parameters (e.g., damping and natural frequency in the second-order loop). An analytic approach could be

attempted by explicitly writing $H(f)$ in terms of its parameters in (16.25) and then minimizing by proper choice of parameters. If the spectra of modulation and noise are at all complicated, the minimization must be accomplished by computer search. Such a search was tried with a second-order loop and spectra as described in Table 16.1. The following results were obtained:

- Calculated values of phase-error variance were an oscillatory function of natural frequency, suggesting that any automatic computation procedures might get into difficulty.
- The minima were extremely shallow. This comes about because of the narrowband filter preceding the PLL; widening the loop bandwidth beyond that of the filter bandwidth adds very little additional noise.
- Experiments did not agree at all well with calculations. The natural frequency producing a minimum click rate in the laboratory did not correspond with the natural frequency that yielded a minimum calculated variance.

16.3.5 Modified PLD

Consider one other method before concluding that linearized phase variance is a poor way to approach PLD threshold optimization. Represent the loop transfer function in polar form $H(f) = |H| \exp(j\psi)$, where the frequency dependence of the polar components has been suppressed for notational convenience. Substituting this form of $H(f)$ into (16.25) and performing some algebra yields the variance

$$\sigma_o^2 = \int_0^\infty \left[(1 - 2|H| \cos\ \psi + |H|^2) \frac{W_m(f)}{(2\pi f)^2} + W_{n'}(f)|H|^2 \right] df \qquad (16.26)$$

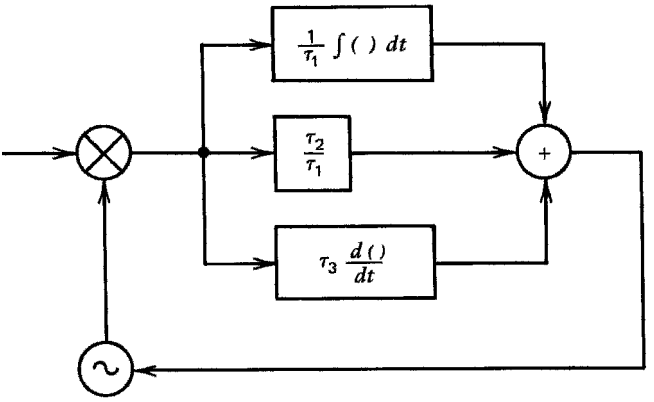
Irrespective of $|H|$, the variance will be minimized if $\cos\ \psi = +1$. (The same condition arises in the derivation of the optimum Wiener unrealizable infinite-lag filter [16.21].) That is, $\psi = 2\pi k$, where $k =$ integer.

The amplitude and phase response of a network are closely related; they cannot be specified separately. Assuming a minimum-phase network and oversimplifying somewhat, the phase condition implies that the amplitude response has a slope of $24k$ dB/octave. From (16.16), a flat response of $|H(f)|$ is needed in the frequency range of the modulation spectrum, so a choice of $k = 0$ deserves attention. In conventional second-order PLLs that have equal numbers of poles and zeros in the loop filter, the amplitude response $|H(f)|$ is flat for low frequencies and falls off at -6 dB/octave at high frequencies. The phase is zero at DC and approaches -90° asymptotically for high frequency.

A realizable network with zero phase at all frequencies must also have constant amplitude response at all frequencies; no finite-bandwidth PLL could have such a response. Neglecting any parasitic elements, the rolloff of a conventional loop is due to the integrator action of the VCO; if that could be overcome, a phase near 0° could be achieved over a much larger frequency range than is accomplished in

an ordinary loop. Building a loop with proportional-plus-integral-plus-*derivative* (PID) control achieves asymptotically zero phase at high, as well as low, frequencies. Such a PLL is shown in Fig. 16.13 and its Bode plots are shown in Figure 16.14.

Novick and Klapper [16.22, 16.23] arrived at essentially the same configuration, starting from an entirely different point of departure. They devised a variance-minimization algorithm and found that minimum variance occurs if $K\tau_1\tau_3/\tau_2 = 1$. (The notation is shown in Fig. 16.13.) Therefore, at high frequencies, $|G|$ approaches 1 and $|H|$ approaches 0.5. The phase of the closed-loop approaches zero at high and low frequencies but must exhibit a lag in the vicinity of the break in open-loop amplitude slope. Since response is flat at high frequencies, the noise bandwidth of this circuit would be infinite. Yet the authors report a significant lowering of threshold from that of a conventional PLL. If the technique proves to be generally applicable, it could be an important advance in the PLD art.



Open–loop gain:

$$G(s) = \frac{K_o K_d}{s}\left(\frac{\tau_2}{\tau_1} + \frac{1}{\tau_1 s} + \tau_3 s\right)$$

$$= \frac{K}{s^2\tau_2}\left(s^2\tau_1\tau_3 + s\tau_2 + 1\right)$$

$$\left(K = \frac{K_o K_d \tau_2}{\tau_1}\right)$$

Closed – loop gain:

$$H(s) = \frac{G(s)}{1 + G(s)} = \frac{s^2\tau_1\tau_3 + s\tau_2 + 1}{s^2\left(\frac{\tau_2}{K} + \tau_1\tau_3\right) + s\tau_2 + 1}$$

Figure 16.13 PLL with a proportional-plus-integral-plus-derivative (PID) controller.

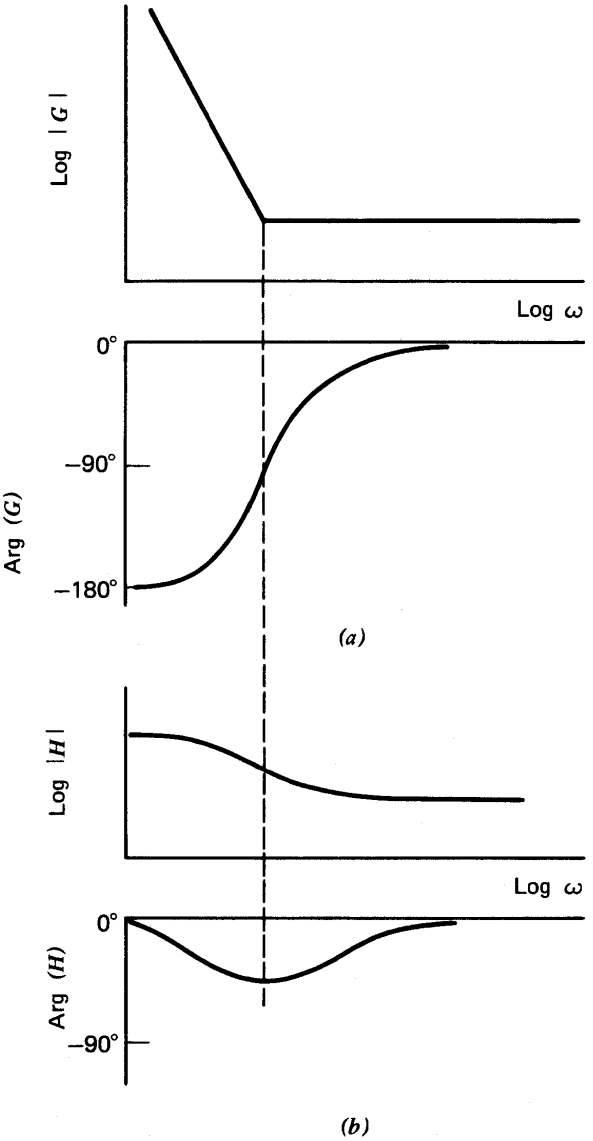


Figure 16.14 Bode plot of a PLL with a PID controller: (*a*) open loop (compare to Fig. 3.13); (*b*) closed loop.

16.3.6 FM PLD Threshold: Summary

- An input filter is an integral part of a PLD and should not be neglected.
- The bandwidth of the input filter should be the minimum consistent with acceptable message distortion. A larger bandwidth entails a threshold penalty.

- Amplitude response of the input filter should be substantially flat over the full range of frequency deviation, to avoid interaction with PLL gain.
- A limiter is not needed in a PLD, and its inclusion raises the threshold level.
- The bandwidth of the PLL must be substantially larger than that of the message and probably larger than that of the input filter.
- For any loop configuration, there exists an optimum loop bandwidth that yields a minimum output click rate. The optimum is a weak function of the input CNR and a strong function of modulation conditions.
- Present methods are inadequate to determine the optimum analytically; experiments with the actual signal and hardware to be used are perhaps the best approach open to design engineers.
- If a second-order loop is employed, experiments suggest that damping of about 1 to 1.5 is optimum.
- A standard second-order PLL is not likely to be the optimum configuration.
- Wiener optimization of a linearized loop appears not to offer guidance to the practical minimum-threshold PLD.
- Adding derivative control to a loop filter seems to be helpful.

REFERENCES

16.1 M. H. Perrott, T. L. Tewksbury III, and C. G. Sodini, "A 27-mW CMOS Fractional-*N* Synthesizer Using Digital Compensation for 2.5 Mb/s GFSK Modulation," *IEEE J. Solid-State Circuits* ***32***, 2048–2059, Dec. 1997. Reprinted in [16.4].

16.2 D. R. McMahill and C. G. Sodini, "Automatic Calibration of Modulated Frequency Synthesizers," *IEEE Trans. Circuits Syst. II* ***49***, 301–311, May 2002.

16.3 E. Hegazi and A. A. Abidi, "A 17-mW Transmitter and Frequency Synthesizer for 900-MHz GSM Fully Integrated in 0.35-mm CMOS," *IEEE J. Solid-State Circuits* ***38***, 782–792, May 2003.

16.4 B. Razavi, ed., *Phase-Locking in High-Performance Systems*, Reprint Volume, IEEE Press, New York, and Wiley, Hoboken NJ, 2003.

16.5 S. Butman and V. Timor, "Interplex: An Efficient Multichannel PSK/PM Telemetry System," *IEEE Trans. Commun.* ***COM-20***, 415–419, June 1972.

16.6 W. C. Lindsey, "Design of Block-Coded Communication Systems," *IEEE Trans. Commun.* ***COM-15***, 525–534, Aug. 1967.

16.7 I. Galton, "Analog-Input Digital Phase-Locked Loops for Precise Frequency and Phase Demodulation," *IEEE Trans. Circuits Syst. II* ***42***, 621–630, Oct. 1995.

16.8 M. G. Crosby, "Frequency Modulation Noise Characteristics," *Proc. IRE* ***25***, 472–514, April 1937.

16.9 S. O. Rice, "Noise in FM Receivers," in M. Rosenblatt, ed., *Time Series Analysis*, Wiley, New York, 1963, Chap. 25.

16.10 O. Shimbo, "Threshold Characteristics of FM Signals Demodulated by an FM Discriminator," *IEEE Trans. Inf. Theory* IT-15, 540–549, Sept. 1969. Corrections: *Trans. Inf. Theory* IT-16, 769, Nov. 1970.

16.11 O. Shimbo, "Threshold Noise Analysis of FM Signals for a General Baseband Signal Modulation and Its Application to the Case of Sinusoidal Modulation," *IEEE Trans. Inf. Theory* ***IT-16***, 778–781, Nov. 1970.

16.12 J. R. Carson, "Notes on the Theory of Modulation," *Proc. IRE* ***10***, 57–64, Feb. 1922.

16.13 D. Yavuz, "FM Click Shapes," *IEEE Trans. Commun.* ***COM-19***, 1271–1273, Dec. 1971.

16.14 J. Klapper and J. T. Frankle, *Phase-Locked and Frequency-Feedback Systems*, Academic Press, New York, 1972, Chap. 6.

16.15 B. M. Smith, "Phase-Locked Loop Threshold," *Proc. IEEE* ***54***, 810–811 May 1966.

16.16 B. M. Smith, "A Semi-empirical Approach to the PLL Threshold," *IEEE Trans. Aerosp. Electron. Syst.* ***AES-2***, 463–468, July 1966.

16.17 D. T. Hess, "Cycle-Slipping in a First-Order Phase-Locked Loop," *IEEE Trans. Commun.* ***COM-16***, 255–260, Apr. 1968.

16.18 F. M. Gardner, "A Cycle-Slip Detector for Phase-Locked Demodulators," *IEEE Trans Instrum. Meas.* ***IM-16***, 251–254, Sept. 1977.

16.19 A. J. Viterbi, *Principles of Coherent Communications*, McGraw-Hill, New York, 1966, Chaps. 5 and 6.

16.20 H. L. Van Trees, *Detection, Estimation and Modulation Theory: Part II*, Wiley, New York, 1971, Chaps. 2–4.

16.21 H. W. Bode and C. E. Shannon, "A Simplified Derivation of Linear Least Square Smoothing and Prediction Theory," *Proc. IRE* ***38***, 417–425, Apr. 1950.

16.22 W. A. Novick and J. Klapper, "Optimum Design of the Extended-Range Phase-Locked Loop," Paper 32D, *Conf. Rec. Natl. Telecommun. Conf.*, 1972.

16.23 W. A. Novick, "Investigation and Optimum Design of the Generalized Second-Order Phase-Locked Loop," *Ph.D. dissertation*, New Jersey Institute of Technology, Newark, NJ, 1976.

CHAPTER 17

MISCELLANEOUS APPLICATIONS OF PHASELOCK LOOPS

Several different applications of phaselock loops are mentioned briefly in this chapter to show how PLLs have been employed.

17.1 SYNCHRONIZATION OF DATA SIGNALS

Most data signals are generated as a stream of uniformly spaced symbols that are transmitted either directly as a baseband signal or modulated onto a carrier for transmission as a passband signal. A receiver of data signals must synchronize to the received symbols (symbol timing recovery) and, if reception is to be coherent, synchronize to the received carrier.

The subject of synchronization is much too broad to cover adequately in this book. Moreover, although PLLs are employed extensively in synchronizers, PLL issues are minor compared to the main issues of synchronization. The PLL principles enumerated in earlier chapters are readily applied to phaselock synchronizers as well. Therefore, this section gives only a cursory overview of synchronization.

In the past, almost all synchronizers were implemented in analog circuits. References [17.1–17.4] and references cited therein provide examples of the era. More recently there has been a division into two separate approaches. In one line of work, the signal formats are complicated (and growing even more complicated as the communications profession becomes ever more sophisticated); excellent, near-perfect performance is essential. The data receivers for this approach are implemented digitally to the greatest extent permitted by current technology.

Phaselock Techniques, Third Edition, by Floyd M. Gardner

Synchronizers, as essential parts of the receivers, are also implemented digitally, replacing analog implementation almost entirely. A large and intricate body of knowledge of digital synchronizers has grown up rapidly, as exemplified in [17.5–17.7].

The other approach is directed primarily toward high-speed, baseband data signals of relatively simple format (e.g., binary symbols). Signal environments are otherwise relatively benign, at least in comparison to the complicated signal formats, noise, interference, multipath, transmission dispersion, and other problems that beset the first class of signals. Excellent performance in noise often can be sacrificed for speed and simplicity. High speeds for the signals of the second class preclude digital implementation; analog circuits are needed. Numerous circuits for 10-Gbps service had been described, and the frontier was at 40 Gbps at the time this passage was written. Circuit examples have been collected in [17.8] and [17.9]. Problems and methods for high-speed PLLs are addressed in [17.10–17.12].

17.2 NETWORK CLOCKS

Digital telecommunications networks include large numbers of highly accurate clocks synchronized to a common source [17.13]. Network synchronization is arranged in a hierarchy, with timing propagated in a tree network from masters to slaves. A slave clock typically might contain a precision oscillator as part of a PLL that locks the oscillator to the timing propagated from a higher source.

Because of the high quality of a good slave clock and the need to reduce timing fluctuations, the bandwidth of the slave PLLs might be very small—an example bandwidth of one cycle per day (11.6 μHz) has been shown in the literature [17.14]. No practical analog PLL can operate with such small bandwidths. Instead, digital loop filtering is used, with its nonvolatile integration and frequency memory. The output of the loop filter is applied to a digital-to-analog converter (DAC) that provides the control voltage to adjust the frequency of the oscillator. A tuning range of 1 ppm is appropriate for highly precise oscillators. Because of the small bandwidth, ordinary noise in the oscillator, DAC, and loop filter has to be kept small since it is not tracked out by feedback. Refer to [17.14] for further details of a particular design.

17.3 VARIOUS LOCKED OSCILLATORS

A phaselock loop includes a locked oscillator, by definition. It may seem redundant to devote a separate section to locked oscillators in a book on PLLs. Nonetheless, there are applications in which the primary objective of the PLL is to lock an oscillator, usually to improve its stability or accuracy of its operating frequency or to generate a new frequency. Some of those applications are presented here.

17.3.1 Oscillator Stabilization

There are two diametrically opposed varieties of oscillator stabilization: narrowband and wideband. In the first, a narrowband PLL is employed as a filter to clean up another oscillator or other signal that is accompanied by noise. In the second, a noisy drifting oscillator is phaselocked to a clean reference to stabilize the locked oscillator.

Narrowband Stabilization Crystal oscillators used as frequency standards have their best long-term frequency stability if they are operated at extremely low RF power levels (crystal aging is slower at the low levels). However, as is noted in Section 9.3.1, the best short-term phase stability is obtained at an intermediate power level, with amplitude of the RF signal much larger than circuit noise. The best results are obtained if two separate oscillators are used: a very low-level one for good long-term stability and a second oscillator, phase-locked to the first, operated at a higher power level for good short-term stability. Bandwidth of the loop should be as narrow as possible consistent with maintaining reliable lock. Output is taken from the locked oscillator. Using the PLL is equivalent to passing the phase noise of the first oscillator through an extremely narrow filter to reduce the noise substantially. The same technique is useful for cleaning up the output of frequency synthesizers, in which harmonics and multiplier products are often present.

Wideband Stabilization Microwave oscillators with usefully large output power can be built with transistors, klystrons, backward-wave tubes, IMPATT or TRAPATT diodes, or Gunn diodes. Electronic tuning is accomplished by changing operating biases on the active device, by using a varactor diode, or by using a magnetically variable YIG resonator. These diverse oscillators share a common trait of poor phase stability. Without additional stabilization, they are unusable in narrowband applications. An effective method of stabilization is to lock the microwave oscillator to a harmonic of a stable, low-frequency source, such as a crystal oscillator. The PLL tracks out the phase fluctuations of the locked oscillator, so the output has the stability of the frequency-multiplied reference source.

One configuration of oscillator stabilization is shown in Fig. 17.1. The only novel element is the frequency multiplier needed to obtain the proper harmonic of the stable low-frequency source. Often, the multiplier is incorporated into the phase detector itself using a harmonic-lock PD as described in Section 14.2. Large

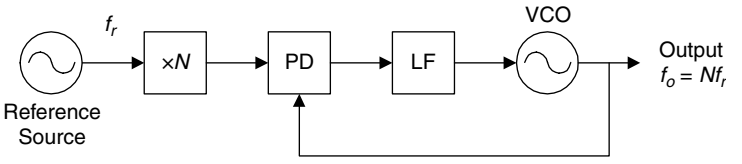


Figure 17.1 Harmonic-locked oscillator.

numbers of such phaselock oscillators are sold as complete packages with acquisition circuitry included. The packages are widely used for transmitters and for receiver local oscillators in fixed-frequency service. A phaselock source is usually more economical of power than is a multiplier string of equal output power.

17.3.2 Frequency-Multiplier PLLs

An oscillator often must be locked to a harmonic of an input reference frequency. The preceding section illustrated methods employing frequency multipliers or harmonic phase detectors to generate the harmonic in a straightforward manner. Locking to high harmonics can be troublesome, partly because of difficulties of generating high harmonics at a large-enough amplitude and partly because of the possibility of locking to the wrong harmonic when they are closely spaced compared to the tuning range of the VCO.

Another popular technique utilizes frequency dividers in the feedback path to reduce the oscillator frequency to that of the reference. The technique is especially attractive at frequencies low enough to permit the use of digital counters as the dividers. A digital counter has a unique division ratio that does not permit the loop to lock to a wrong, closeby harmonic. A frequency-multiplier PLL may be regarded as a fixed-frequency synthesizer, as treated in Chapter 15. Reference clocks are commonly distributed between modules in a computer at a comparatively low frequency and then boosted to a higher clock frequency within individual modules by means of an on-chip frequency-multiplier PLL.

17.3.3 Frequency-Translation PLLs

A frequency translator shifts an input frequency f_i by an amount f_b to an output frequency $f_i \pm f_b$. The benefits of translating by means of a PLL may be seen from an example. Suppose that a 30-MHz signal had to be offset by 1 kHz. One way to accomplish this would be by means of conventional single-sideband techniques, but good suppression of carrier and rejected sideband would depend on critical circuit adjustments. A phaselock offset can be less critical and perform better.

In the example offset PLL shown in Fig. 17.2, a VCO normally running at a frequency close to the desired output frequency is heterodyned with the incoming

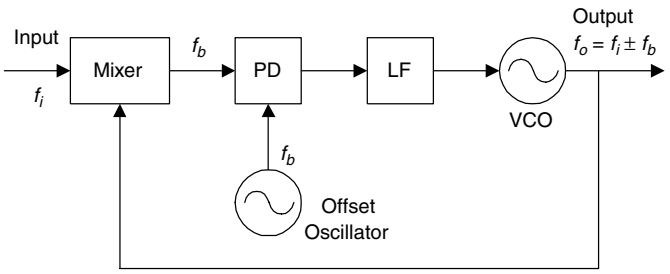


Figure 17.2 Frequency-translation PLL.

frequency f_i so that the mixer output frequency is close to the desired offset frequency f_b. Then mixer output is compared in a PD to an oscillator whose frequency is exactly f_b, and the loop is closed back to the VCO, forcing the mixer output into lock with the frequency-offset oscillator. At first appearance, it would seem that phaselock has completely eliminated the residual carrier and unwanted sidebands that remain in conventional SSB techniques. Such perfection is not really obtainable; any phase-detector ripple will modulate the VCO and produce unwanted sidebands in the output. Suppose that the PD is a perfect multiplier and that both inputs to it are pure sinusoids. The ripple output of the PD will be a sinusoid at $2f_b$ and ripple modulation of the VCO will generate a pair of sidebands at $f_o \pm 2f_b$. If the desired output frequency is $f_o = f_i + f_b$, the sidebands will be at $f_i - f_b$ and $f_i + 3f_b$. Other phase detectors will have additional ripple components in their outputs and so will cause additional unwanted sidebands to appear in the VCO output.

Ripple may be reduced to any desired extent by means of brute-force noncritical lowpass filtering in the loop (see Appendix 10A). Such filtering usually requires a narrowing of loop bandwidth. A phase detector with low inherent ripple (e.g., a PFD or a sample and hold PD; see Chapter 10) may be more effective than any practical filtering in suppressing ripple.

The PLL of Fig. 17.2 is capable of locking its VCO to either of the two frequencies $f_i + f_b$ or $f_i - f_b$. In most instances just one frequency is wanted and the other is an undesired image. Provision must be made to avoid locking to the image. No problem exists if the VCO is unable to tune to the image. More generally, if the VCO can tune both to the desired signal and to the image, other measures must be taken to avoid image lock.

One approach is to use an image-rejecting phase detector, as shown in Fig. 17.3. This circuit is a variation on conventional single-sideband mixers and is applicable in any long loop where image rejection is needed. It is a variation on the complex phase detector of Fig. 10.16. The input signal at frequency f_i is heterodyned to the offset (IF) frequency in a pair of identical mixers. Filters select the desired difference-frequency product and reject the unwanted the sum-frequency products

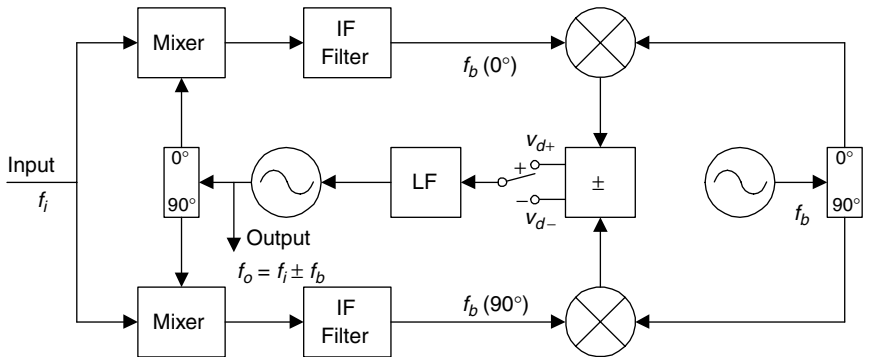


Figure 17.3 Complex-signal frequency translation.

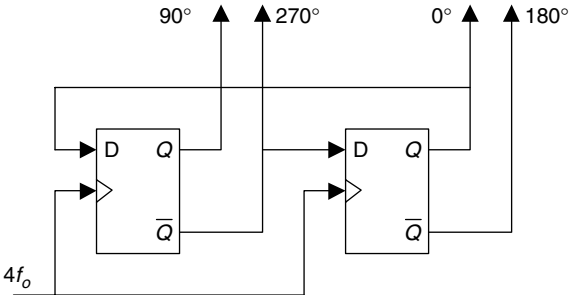


Figure 17.4 Generation of quadrature signals by digital counters.

from the mixers. Because the two mixers have quadrature local drive, the two IF outputs are 90° apart in phase. (The same 90° relation could be achieved with a single mixer and a 90° phase-splitting network at the intermediate frequency. Alternatively, it is often convenient to produce quadrature drives using binary counters as in Fig. 17.4.)

Each IF signal drives a phase detector whose other input comes from the offset oscillator at frequency f_b. A 90° phasing is imposed between the f_b drive voltages to the two phase detectors. Each phase detector must be a true multiplier, and waveforms on both input ports must be sinusoidal so that the only ripple component is a sinusoid at frequency $2f_b$. When the PLL is locked, the ripple components of the PD outputs cancel each other while the zero-frequency components add together. This feature may permit loop bandwidth to exceed the offset frequency f_b—an impossible achievement in a conventional PLL.

An ordinary PLL could lock the VCO equally well to either the desired frequency or its image, depending on which was encountered first during acquisition. An image-rejecting PLL ideally can lock to only one of the two frequencies, determined by whether the PD outputs are added or subtracted. If cancellation is perfect, the image is rejected completely. Perfect cancellation is impossible to achieve, so a weak lock may still be possible at the image frequency. Acquisition aiding, such as rapid sweep, is needed to override locking to the weak image.

Perfect cancellation also suppresses PD ripple completely, thereby suppressing the spurious image sidebands located $2f_b$ away from the desired VCO output signal. Imperfect cancellation permits residual ripple to propagate through to the VCO and generate the spurious sidebands, but with amplitude reduced by the cancellation factor. Attenuation of the image sidebands then is provided by cancellation as well as filtering in the PLL.

A complete image-rejecting PLL is often overly elaborate for merely avoiding lock to the unwanted sideband. A simple quadrature phase detector (Fig. 8.14) has one polarity of output for lock on the upper sideband and the opposite polarity on the lower sideband. This polarity information can be furnished to acquisition-aiding circuits to prevent the loop from locking up at the wrong sideband.

17.4 PLLs IN TELEVISION RECEIVERS

A television receiver might use a number of phaselock loops, for the following purposes:

- Synchronization of horizontal and vertical scans [17.15]
- Establishing a phase reference for demodulation of the color subcarrier [17.16]
- Demodulation of FM sound (Section 16.2.3)
- Coherent demodulation of a vestigial-sideband video signal (Section 16.2.1)
- Frequency synthesizer for the RF tuner (Chapter 15)

17.5 PLLs IN DIGITAL SYSTEMS

Computers and other digital systems make use of PLLs in various ways. This section describes two examples.

17.5.1 Compensation of Timing Skew

High-speed computers and other digital systems incur delays in their clock-distribution paths, delays causing timing offsets—*skew*—between data and clock. An example is shown in Fig. 17.5*a*. The reference clock is applied to chip 1 through a clock tree that has internal delay. Data are written into the output latch at instants determined by the delayed clock. Data from chip 1 are transferred to a latch in chip 2. Transfer is timed by the reference clock, which is not properly aligned with the instants of state change of the chip 1 output latch because of the clock delay internal to chip 1. Timing skew exists between the output data and the reference clock.

Figure 17.5*b* shows how a PLL can be used to compensate for the delay. The delayed clock applied to the chip 1 output latch is compared against the reference clock in a phase detector. Output from the PD goes through a loop filter and then a VCO that drives the clock tree at the frequency of the reference clock but at an adjusted phase that causes the delayed output of the clock tree to align with the reference clock. Now the output latch on chip 1 is clocked at the same instants as the input latch on chip 2; the PLL has removed the skew. Accuracy of compensation depends on the accuracy of the phase detector and the delay equality of the data and clock paths between the two chips.

17.5.2 Jitter Attenuators

Timing jitter tends to build up in digital-communications chains (see Section 14.5); or, the original digital data may be irregularly timed in some systems. Telecommunications links require stringent limits on jitter for satisfactory

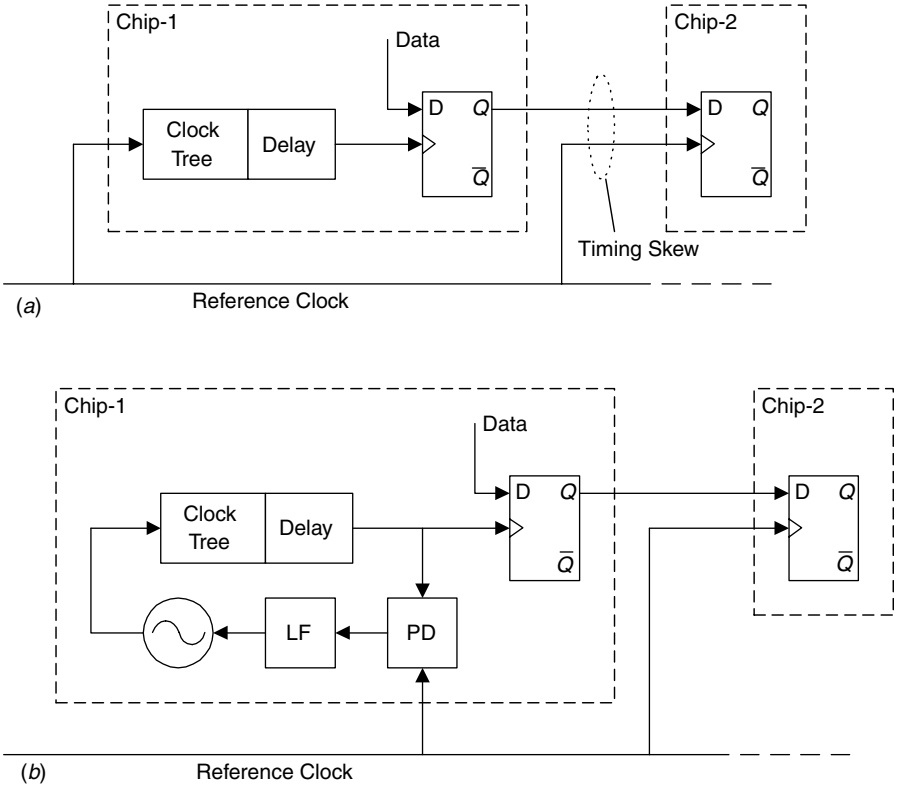


Figure 17.5 Skew correction by a PLL: (*a*) origin of skew; (*b*) PLL correction of skew.

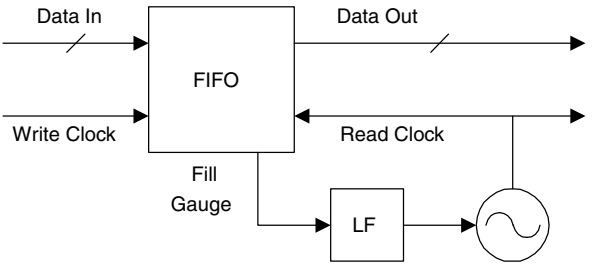


Figure 17.6 PLL jitter attenuator.

operation. Phaselock jitter attenuators, widespread throughout the telecommunications networks of the world, are employed to suppress jitter before it can build up to harmful levels. A basic jitter attenuator is shown in Fig. 17.6. Irregularly timed input data are written into the FIFO (first-in, first-out buffer memory) with the aid of the accompanying Write clock, which has the same timing irregularities as the input data. A smoothed Read clock is produced with the aid of a PLL.

The FIFO has provision for monitoring its number of occupied cells and delivering that information externally—the digital information labeled "fill gauge" in Fig. 17.6. Fill information is compared to a desired fill condition (typically half-full) to generate a fill error that serves as a phase-error measurement. (The comparison is omitted from the figure; it is a simple numerical operation.) Phase error is applied to a digital loop filter whose output controls the frequency of a DCO. A digital-to-analog converter, a filter, and a slicer produce a rectangular wave for the Read clock.

When the PLL is locked, the average frequency of the Read clock is the same as the average frequency of the Write clock. Rigorous long-term equality of the average clock frequencies is essential to avoid underflow or overflow of the FIFO, with consequent loss of data. Phaselocking assures that equality of frequencies. The size of the FIFO is chosen so that the FIFO can absorb the worst-case jitter without saturation. A FIFO tolerates far greater timing excursions than does any ordinary phase detector. The bandwidth of the PLL is chosen to be small enough to attenuate specified input-jitter conditions to acceptable output-jitter conditions. Digital implementation of the loop filter is appropriate if the bandwidth has to be small; very narrow bandwidths require inconveniently large component values for analog loop filters. The fixed oscillator (not shown in the figure) that clocks the DCO has to have good jitter properties since its jitter appears directly in the Write clock.

17.6 PLLs FOR MOTOR SPEED CONTROL

Phaselocked motor controls have been used for extremely accurate control of average speed of motors. Early publications on the subject include [17.17] and [17.18]. Presumably, numerous later articles appear in the control system literature.

17.6.1 Basic Operation

Figure 17.7 shows the essential elements of a motor-speed PLL. A reference oscillator produces a reference frequency that is an integer multiple of the desired rotation speed of the motor; high accuracy in the reference frequency is readily provided. The motor drives a *tone wheel*, an electro- or optomechanical device that produces a tone or a series of pulses whose frequency depends on the motor speed. The tone wheel might be a cog with a magnetic sensor, or it might be an optical disk with alternating transparent and opaque markings around the disk. A photodetector would be the sensor for an optical disk. The reference signal and electric output of the sensor are applied to the two inputs of the phase detector, whose output is a phase-error indication to the loop filter. A power amplifier (PA) then drives the motor. The tone frequency from the tone wheel is forced to be equal to the reference frequency when the PLL is locked, whereupon the average speed of the motor has been set to the accuracy of the reference.

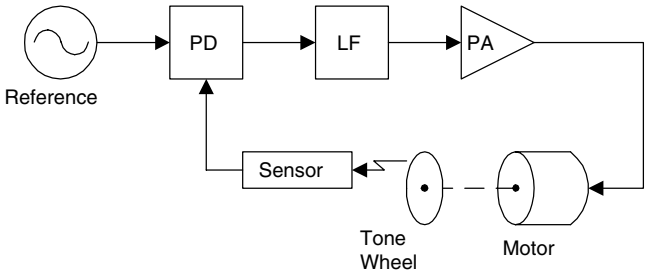


Figure 17.7 Phaselock motor-speed control.

17.6.2 Electromechanical Considerations

Superficially, the speed-control loop of Fig. 17.7 appears to be similar to an ordinary all-electronic PLL whose VCO has been replaced by a motor and tone wheel. The similarity is instructive, but important differences exist. One difference arises in the mechanical nature of the motor, in that it has an inertial load and a frictional load, both of which might change from time to time. The inertia and friction combine to insert a low-frequency lowpass pole into the loop, a pole that is usually absent or can be neglected in a VCO. Mechanical loading is a well-known issue in control system engineering, but its electronic equivalent usually is neglected, often justifiably, in PLL engineering.

Another difference is in the substantial power required to drive a motor as opposed to trivial power involved in control of the frequency of a VCO. The PA might apply a voltage drive to the motor, in which case the equilibrium speed is proportional to the applied voltage, or it might be preferable to apply a current drive, which determines the motor's torque. Either way, the PA is expected to deliver all of the power needed to run the motor—a substantial amount of power in a big motor.

In addition to the usual stability issues of a servomechanism, many phase detectors (such as the sequential phase/frequency detector of Section 10.3) effectively operate in a sampled manner. The loop becomes unstable if the sampling frequency is too low (see Section 12.4); correspondingly, the PLL speed-control loop goes unstable if the motor speed is too low [17.17].

Like most PLLs, a proportional-plus-integral control configuration (a type 2 PLL) is required in the loop filter to permit servo operation without requiring a static phase error to support the designated motor speed. Indeed, the required maximum speed might not be attainable with a type 1 loop of stable bandwidth; either a $P + I$ control has to be used or a bias must be applied to the PA. Mechanical servos commonly also have tachometer feedback (equivalent to proportional-plus-integral-plus derivative control) to improve the damping and the stability of the loop.

17.6.3 Alternative Configurations

Alternative arrangements of a motor-speed PLL can be visualized. The ideas presented are speculations but it would not be surprising to learn that they can

be found in existing motor controls. Rather than a tone wheel, an optical angle-encoder disk offers some advantages. Each time an encoder disk is sampled, it delivers a *b*-bit digital word indicating its angular position. The sampling rate can be fixed and independent of the motor speed or reference frequency. It can easily be made high enough compared to servo bandwidth to do away with time-discrete stability issues, yet be low enough not to overload ordinary digital processors.

A digital processor can calculate the sine and cosine of the angle captured by each sampling of the angle encoder. The reference frequency can be supplied by an NCO whose angle at each sample instant is converted to the corresponding sine and cosine. The complex pairs from reference and processed encoder samples can be compared in a complex phase detector such as that of Section 10.5. Because operations are digital, the desired balance in these operations is nearly perfect, limited only by the word lengths of the digital samples.

A complex phase detector can work down to zero frequency; indeed, it can work through zero into negative frequencies, as can an NCO with complex outputs. This scheme allows speed control in either direction of rotation. A digital loop filter is more versatile and uses much more compact components than does an analog loop filter for the small bandwidths of a mechanical servo. Instead of a linear power amplifier, a pulse-width-modulated switch would be far more efficient in driving the motor.

REFERENCES

17.1 W. C. Lindsey and M. K. Simon, *Telecommunications Systems Engineering*, Prentice Hall, Englewood Cliffs, NJ, 1973.

17.2 E. A. Lee and D. G. Messerschmitt, *Digital Communication*, Kluwer Academic, Boston, MA, 1988, Chaps. 14 and 15.

17.3 R. D. Gitlin, J. F. Hayes, and S. B. Weinstein, *Data Communication Principles*, Plenum Press, New York, 1992, Chap. 6.

17.4 Synchronization Special Issue, *IEEE Trans. Commun.* ***COM-28***, Aug. 1980.

17.5 J. W. M. Bergmans, *Digital Baseband Transmission and Recording*, Kluwer Academic, Boston, MA, 1996, Chaps. 9 and 10.

17.6 U. Mengali and A. N. D'Andrea, *Synchronization Techniques for Digital Receivers*, Plenum Press, New York, 1997.

17.7 H. Meyr, M. Moeneclaey, and S. A. Fechtel, *Digital Communication Receivers*, Wiley, New York, 1998.

17.8 B. Razavi, ed., *Monolithic Phase-Locked Loops and Clock Recovery Circuits*, Reprint Volume, IEEE Press, New York, 1996.

17.9 B. Razavi, ed., *Phase-Locking in High-Performance Systems*, Reprint Volume, IEEE Press, New York, and Wiley, New York, 2003.

17.10 B. Razavi, "A 2.5 Gb/s 15-mW Clock Recovery Circuit," *IEEE J. Solid-State Circuits* ***31***, Apr. 1996.

17.11 B. Razavi, "Challenges in the Design of High-Speed Clock and Data Recovery Circuits," *IEEE Communications Magazine* ***40***, 94–101, Aug. 2002.

17.12 J. Savoj and B. Razavi, "A 10-Gb/s CMOS Clock and Data Recovery Circuit with a Half-Rate Binary Phase/Frequency Detector," *IEEE J. Solid-State Circuits* ***38***, 13–21, Jan. 2003.

17.13 Bellcore, *Clocks for the Synchronized Network: Common Generic Criteria*, GR-1244-CORE, Bell Communications Research, June 1995.

17.14 E. A. Munter, "Synchronized Clock for DMS-100 Family," *IEEE Trans. Commun.* ***COM-28***, 1276–1284, Aug. 1980.

17.15 K. R. Wendt and G. L. Fredendall, "Automatic Frequency and Phase Control of Synchronization in Television Receivers," *Proc. IRE* ***31***, 7–15, Jan. 1943.

17.16 D. Richman, "Color-Carrier Reference Phase Synchronization in NTSC Color Television," *Proc. IRE* ***42***, 106–133, Jan. 1954.

17.17 J. Tal, "Speed Control by Phase-Locked Servo Systems: New Possibilities and Limitations," *IEEE Trans. Ind. Electron. Control Instrum.* ***IECI-24***, 118–125, Feb. 1977.

17.18 D. F. Geiger, *Phaselock Loops for DC Motor Speed Control*, Wiley, New York, 1981.

INDEX

Phaselock Techniques, Third Edition, by Floyd M. Gardner