

The background of the cover features a faint, light-colored circuit diagram with various components like resistors, capacitors, and transistors, along with arrows indicating signal flow. A prominent blue waveform, consisting of a regular sine wave that transitions into a single cycle of a higher-frequency wave, is overlaid on the circuit.

Phase-Lock Basics

Second Edition

William F. Egan



PHASE-LOCK BASICS



THE WILEY BICENTENNIAL—KNOWLEDGE FOR GENERATIONS

Each generation has its unique needs and aspirations. When Charles Wiley first opened his small printing shop in lower Manhattan in 1807, it was a generation of boundless potential searching for an identity. And we were there, helping to define a new American literary tradition. Over half a century later, in the midst of the Second Industrial Revolution, it was a generation focused on building the future. Once again, we were there, supplying the critical scientific, technical, and engineering knowledge that helped frame the world. Throughout the 20th Century, and into the new millennium, nations began to reach out beyond their own borders and a new international community was born. Wiley was there, expanding its operations around the world to enable a global exchange of ideas, opinions, and know-how.

For 200 years, Wiley has been an integral part of each generation's journey, enabling the flow of information and understanding necessary to meet their needs and fulfill their aspirations. Today, bold new technologies are changing the way we live and learn. Wiley will be there, providing you the must-have knowledge you need to imagine new worlds, new possibilities, and new opportunities.

Generations come and go, but you can always count on Wiley to provide you the knowledge you need, when and where you need it!

A handwritten signature in black ink that reads "William J. Pesce".

WILLIAM J. PESCE
PRESIDENT AND CHIEF EXECUTIVE OFFICER

A handwritten signature in black ink that reads "Peter Booth Wiley".

PETER BOOTH WILEY
CHAIRMAN OF THE BOARD

PHASE-LOCK BASICS

Second Edition

WILLIAM F. EGAN, Ph.D.

Lecturer in Electrical Engineering
Santa Clara University



The Institute of Electrical and Electronics Engineers, Inc., New York



WILEY-INTERSCIENCE

A JOHN WILEY & SONS, INC., PUBLICATION

MATLAB[®] is a trademark of The MathWorks, Inc. and is used with permission. The MathWorks does not warrant the accuracy of the text or exercises in this book. The book's discussion of MATLAB[®] software or related products does not constitute endorsement or sponsorship by The MathWorks, Inc. of a particular pedagogical approach or particular use of the MATLAB[®] software.

Copyright © 2008 by John Wiley & Sons, Inc. All rights reserved

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

Published simultaneously in Canada

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley Bicentennial Logo: Richard J. Pacifico

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

Egan, William F.

Phase-lock basics / by William F. Egan. – 2nd ed.

p. cm.

Includes index.

ISBN 978-0-470-11800-9 (cloth)

1. Phase-locked loops. I. Title.

TK872.P38E33 2007

621.3815'364 – dc22

2007013688

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To Henry P. Nettesheim
Emeritus Professor of Electrical Engineering
Santa Clara University

CONTENTS

PREFACE TO THE SECOND EDITION	xix
PREFACE TO THE FIRST EDITION	xxi
SYMBOLS LIST AND GLOSSARY	xxv

PART 1 PHASE LOCK WITHOUT NOISE

1 INTRODUCTION	3
1.1 What is a Phase-Locked Loop (PLL)? / 3	
1.2 Why Use a Phase-Locked Loop? / 3	
1.3 Scope of this Book / 4	
1.4 Basic Loop / 5	
1.5 Phase Definitions / 6	
1.6 Phase Detector / 8	
1.7 Combined Gain / 10	
1.8 Operating Range / 11	
1.9 Units and the Laplace Variable s / 13	

2	THE BASIC LOOP	15
2.1	Steady-State Conditions / 15	
2.2	Classical Analysis / 16	
2.2.1	Transient Response / 17	
2.2.2	Modulation Response / 19	
2.3	Mathematical Block Diagram / 21	
2.4	Bode Plot / 24	
2.5	Note on Phase Reversals / 26	
2.6	Summary of Transient Responses of the First-Order Loop / 26	
3	LOOP COMPONENTS	29
3.1	Phase Detector / 29	
3.1.1	Flip-Flop Phase Detector / 29	
3.1.2	Exclusive-OR Gate Phase Detector / 30	
3.1.3	Charge-Pump Phase Detector / 31	
3.1.4	Sinusoidal Phase Detector / 33	
3.1.4.1	<i>Phase Detection in a Simple Mixer / 34</i>	
3.1.4.2	<i>Balanced Mixers / 35</i>	
3.1.4.3	<i>Analog Multipliers / 39</i>	
3.1.4.4	<i>Integrated Circuit Doubly Balanced Mixer / 40</i>	
3.2	Voltage-Controlled Oscillator (VCO) / 41	
3.3	Loop Filter / 44	
3.3.1	Passive Loop Filter / 45	
3.3.2	Filters Driven by Current Sources / 45	
3.3.2.1	<i>Integrator / 47</i>	
3.3.2.2	<i>Integrator-and-Lead Filter / 48</i>	
3.3.2.3	<i>Lag (Low-Pass) Filter / 48</i>	
3.3.2.4	<i>Lag-Lead Filter / 49</i>	
3.3.2.5	<i>DC Transimpedance / 49</i>	
3.3.3	Active Filters / 51	
3.3.3.1	<i>Op-Amp Considerations / 52</i>	
3.3.3.2	<i>Unintended Poles / 53</i>	
3.3.3.3	<i>Reducing the Size of C / 54</i>	
3.3.3.4	<i>Impractical Loop Filters / 55</i>	
3.4	Filter Reference Voltage / 55	
3.5	Note on the Form of the Filter Equation / 56	
3.6	Capacitors in Loop Filters / 56	
3.7	Higher-Order Filters / 57	

- i3.A Appendix: Integrated Circuit Doubly Balanced Mixer—Details / 57
- i3.B Appendix: Op-Amps in Loop Filters / 57

4 LOOP RESPONSE 59

- 4.1 Loop Order and Type / 59
- 4.2 Closed-Loop Equations / 60
- 4.3 Open-Loop Equations—Lag-Lead Filter / 62
- 4.4 Loop with a Lag Filter / 65
- 4.5 Loop with an Integrator-and-Lead Filter / 66
- 4.6 Summary of Equations / 67
- 4.7 Unity-Gain Bandwidth in Highly Damped Loops / 67

5 LOOP STABILITY 71

- 5.1 Observing the Open-Loop Response / 71
- 5.2 Methods of Stability Analysis and Measures of Stability / 72
 - 5.2.1 Bode Plot / 73
 - 5.2.2 Nyquist Plot / 74
 - 5.2.3 Evans Plot (Root Locus) / 76
- 5.3 Stability of Various PLL Configurations / 76
 - 5.3.1 First-Order Loop / 77
 - 5.3.2 Second-Order Loop / 77
 - 5.3.3 Third-Order Loop / 77
 - 5.3.4 Controlling Stability over Wide Gain Ranges / 79
 - 5.3.5 Representing Delay / 80
- 5.4 Computing Open-Loop Gain and Phase / 81
- 5.5 Phase Margin Versus Damping Factor / 86
- i5.M Appendix: Stability Plots Using MATLAB[®] / 86

6 TRANSIENT RESPONSE 87

- 6.1 Step Response / 87
 - 6.1.1 Form of the Equations / 87
 - 6.1.2 Step Response Equations / 89
- 6.2 Envelope of the Long-Term Step Response / 92
- 6.3 Response to Ramp Input / 96
- 6.4 Response to Parabolic Input / 98
- 6.5 Other Responses / 100
- 6.6 Note on Units for Graphs / 101
- 6.7 Equivalent Circuit / 102

- 6.8 General Long-Term (Steady-State) Response Characteristics / 102
- 6.9 Open-Loop Equations in Terms of Closed-Loop Parameters / 103
- 6.10 More Complex Loops and State-Space Analysis / 103
 - 6.10.1 Basic Equations / 104
 - 6.10.2 Initial Conditions / 107
- 6.11 An Approximate Solution Using State-Space Variables / 108
- 6.12 Effect of an Added Pole / 108
- i6.M Appendix: Transient Responses Using MATLAB / 109

7 MODULATION RESPONSE

111

- 7.1 Phase and Frequency Modulation / 111
- 7.2 Modulation Responses / 113
- 7.3 Responses in a First-Order Loop / 113
- 7.4 Transfer Functions in a Second-Order Loop / 116
 - 7.4.1 Output Response / 117
 - 7.4.2 Error Response / 118
 - 7.4.3 Responses Near ω_L / 118
 - 7.4.4 Phase or Frequency at Inputs and Outputs / 119
- 7.5 Transient Responses Between Various Points / 120
- 7.6 Magnitude and Phase of the Transfer Functions / 120
 - 7.6.1 Output Responses / 121
 - 7.6.2 Error Responses / 121
 - 7.6.3 Effect of α / 122
 - 7.6.4 Responses for $\zeta \leq 1$ / 123
 - 7.6.5 Responses for $\zeta \geq 1$ / 125
- 7.7 Related Responses / 127
- 7.8 Modulation and Demodulation in the Second-Order Loop / 128
 - 7.8.1 Frequency Demodulation / 128
 - 7.8.2 Phase Demodulation / 130
 - 7.8.3 Frequency Modulation / 130
 - 7.8.4 Phase Modulation / 131
 - 7.8.5 Extending the Modulation Frequency Range / 131
 - 7.8.5.1 *Frequency Modulation* / 131
 - 7.8.5.2 *Phase Modulation* / 133
- 7.9 Measurement of Loop Parameters for $\alpha = 0$ or 1 from Modulation Responses / 134

- 7.10 Effect of an Added Pole / 134
- 7.M Appendix: Modulation Response Using MATLAB[®] / 135

8 ACQUISITION

137

- 8.1 Overview / 137
- 8.2 Acquisition and Lock in a First-Order Loop / 141
 - 8.2.1 Transient Time / 143
 - 8.2.2 Acquisition / 144
- 8.3 Acquisition Formulas for Second-Order Loops with Sine Phase Detectors / 146
- 8.4 Approximate Pull-In Analysis / 149
 - 8.4.1 Basic Equations / 150
 - 8.4.2 General Analysis / 153
 - 8.4.3 Pull-In Range / 157
 - 8.4.4 Approximate Pull-In Time / 158
- 8.5 Phase-Plane Analysis / 161
- 8.6 Pull-Out / 164
- 8.7 Effect of Offsets / 164
- 8.8 Effect of Component Saturation / 165
- 8.9 Hangup / 166
- 8.10 Simulation of the Nonlinear Loop / 166
- 8.A Appendix: Summary of Acquisition Formulas for Second-Order Loops / 167
- i8.M Appendix: Nonlinear Simulation / 167
- 8.S Appendix: Acquisition Spreadsheet / 167
- 8.V Appendix: Some Values in Terms of α , ζ , and ω_n / 167

9 ACQUISITION AIDS

171

- 9.1 Coherent Detection—Lock Indicator / 171
 - 9.1.1 During Acquisition / 172
 - 9.1.2 During Sweep, Locked Loop / 173
- 9.2 Changing Loop Parameters Temporarily / 174
 - 9.2.1 Coherent Automatic Gain Control / 174
 - 9.2.2 Filter Modification / 175
 - 9.2.3 Comparison of Two Types of Parameter Modifications / 176
- 9.3 Automatic Tuning of ω_c , Frequency Discriminator / 177
- 9.4 Acquisition Aiding Logic / 179

- 9.5 Sweeping ω_c , Type 2 Loop / 180
 - 9.5.1 Maximum Sweep Speed, Closed-Loop Sweeping / 180
 - 9.5.2 Open-Loop Sweeping / 181
 - 9.5.3 Combined Techniques / 185
- 9.6 Sweep Circuits / 185
 - 9.6.1 Switched Current Source / 185
 - 9.6.2 Automatic Sweep Circuit—Sinusoidal / 186
 - 9.6.3 Automatic Sweep Circuit—Nonsinusoidal / 186
- 9.A Appendix: Maximum Sweep Rate, Open-Loop vs. Closed-Loop / 187

10 APPLICATIONS AND EXTENSIONS

189

- 10.1 Higher-Order Loops / 189
- 10.2 Generalized Voltage-Controlled Oscillator / 189
 - 10.2.1 Frequency Synthesis, Frequency Division / 190
 - 10.2.1.1 *Stability* / 191
 - 10.2.1.2 *Transient Response* / 191
 - 10.2.1.3 *Response to Noise* / 192
 - 10.2.2 Heterodyning (Frequency Mixing) / 192
- 10.3 Long Loop / 193
- 10.4 Carrier Recovery / 195
 - 10.4.1 Biphase Costas Loop / 195
 - 10.4.2 *N*-Phase Costas Loop / 196
 - 10.4.3 Multiply and Divide / 196
- 10.5 Data Synchronization / 197
 - 10.5.1 Early–Late Gate Bit Synchronizer / 197
 - 10.5.2 Synchronizing to a Pseudorandom Bit Sequence / 198
 - 10.5.3 Delay-and-Multiply Synchronizer / 200
- 10.6 Clock and Data Timing Control / 200
 - 10.6.1 Phase-Locked Loops / 200
 - 10.6.2 Delay-Locked Loops / 202
 - 10.6.2.1 *For Clock Synchronization* / 202
 - 10.6.2.2 *For Data Synchronization* / 203
 - 10.6.3 Combined Loops / 204
- 10.7 All-Digital Phase-Locked Loop (ADPLL) / 204
 - 10.7.1 Basic Digital Implementation / 205
 - 10.7.1.1 *The Loop* / 205
 - 10.7.1.2 *Sampling and Stability* / 206

- 10.7.1.3 *Choice of Values* / 209
 - 10.7.1.4 *Higher-Order Loops* / 210
 - 10.7.2 OA, NCO, DDS / 211
 - 10.7.3 Implementing an ADPLL by Pulse Addition and Removal / 212
 - 10.7.3.1 *Transfer Function* / 213
 - 10.7.3.2 *Tuning Range* / 214
- 10.8 Summary / 214
- 10.A Appendix: Exact Analysis of a Special-Case Third-Order Loop / 215
 - 10.A.1 Loop Response / 217
 - 10.A.2 Final Values / 219
 - 10.A.3 Triple Roots / 220
 - 10.A.4 Step Response / 220
 - 10.A.5 Modulation Response / 224
- i10.B Appendix: Costas Loop for N Phases / 227
- i10.C Appendix: Symbol Clock Recovery / 227
- i10.D Appendix: ADPLL by Pulse Addition and Removal, Additional Material / 228
- 10.F Appendix: Fractional- N and Sigma-Delta / 228
- i10.M Appendix: MATLAB[®] Simulations / 228
- i10.T Appendix: Combined PLL and DLL / 229

PART 2 PHASE LOCK IN NOISE

- 11 PHASE MODULATION BY NOISE 233**
 - 11.1 Representation of Noise Modulation / 233
 - 11.2 Processing of Noise Modulation by the Phase-Locked Loop / 236
 - 11.3 Phase and Frequency Variance / 237
 - 11.4 Typical Oscillator Spectrums / 238
 - 11.5 Limits on the Noise Spectrum—Infinite Variances / 240
 - 11.6 Power Spectrum / 242
 - 11.6.1 Spectrum for Small m / 242
 - 11.6.2 Single-Sideband Density / 243
 - 11.6.3 When is the Modulation Small? / 245
 - 11.6.4 Modification of Spectral Shape at Higher Modulation Index / 246
 - 11.6.5 Script \mathcal{L} / 247

11.7	Frequency Multiplication and Division / 247	
11.8	Other Representations / 248	
i11.H	Shape of Output Spectrum / 249	
11.S	Appendix: Spreadsheets for Integrating Densities / 249	
12	RESPONSE TO PHASE NOISE	251
12.1	Processing of Reference Phase Noise / 251	
12.2	Processing of VCO Phase Noise / 254	
12.3	Harmful Effect of Phase Noise in Radio Receivers / 255	
12.4	Superposition / 256	
12.5	Optimum Loop With Both Input and VCO Noise / 257	
12.6	Multiple Loops / 260	
12.7	Effects of Noise Injected Elsewhere / 260	
12.8	Measuring Phase Noise / 261	
12.8.1	Using a Phase Detector / 263	
	12.8.1.1 Calibration / 263	
	12.8.1.2 Obtaining a Measurement Reference / 265	
12.8.2	Using a Frequency Discriminator / 269	
12.8.3	Using a Spectrum Analyzer or Receiver / 270	
13	REPRESENTATION OF ADDITIVE NOISE	271
13.1	General / 271	
13.2	Phase Modulation on the Signal / 273	
13.3	Multiplicative Modulation on Quadrature Carriers / 275	
13.4	Noise at the Phase Detector Output / 276	
13.5	Restrictions on the Noise Models / 277	
13.6	Does the Loop Lock to the Additive Noise? / 280	
13.7	Other Types of Phase Detectors in the Presence of Noise / 281	
	13.7.1 Triangular Characteristic / 282	
	13.7.2 Sawtooth Characteristic / 283	
13.8	Modified Phase Detector Characteristic with Phase Noise / 283	
i13.A	Appendix: Decomposition of a Single Sideband / 286	
14	LOOP RESPONSE TO ADDITIVE NOISE	287
14.1	Noise Bandwidth / 287	
14.2	Signal-to-Noise Ratio in the Loop Bandwidth / 290	
14.3	Loop Optimization in the Presence of Noise / 291	
	14.3.1 The Problem / 292	

14.3.2	Measures to be Used / 292	
14.3.3	Optimum Loop for a Phase Step Input / 293	
14.3.4	Optimum Loop for a Frequency Step Input / 294	
14.3.5	Optimum Loop for a Frequency Ramp Input / 294	
i14.A	Appendix: Integration of Eq. (6.4a) / 296	
i14.B	Appendix: Loop Optimization in the Presence of Noise / 296	
15	PHASE-LOCKED LOOP AS A DEMODULATOR	297
15.1	Phase Demodulation / 297	
15.1.1	Noise / 297	
15.1.2	Distortion of the Demodulated Signal / 299	
15.1.3	Demodulation with a Linear Phase Detector Characteristic / 300	
15.2	Frequency Demodulation, Bandwidth Set by a Filter / 301	
15.3	Frequency Discriminator, First-Order Loop / 305	
15.4	Frequency Discriminator, Second-Order Loop / 306	
15.5	Expected Phase Error / 307	
15.6	Summary of Frequency Discriminator S/N / 308	
15.7	Standard Discriminator and Click Noise / 310	
15.8	Clicks with a PLL / 313	
15.9	Noise in a Carrier Recovery Loop / 314	
15.C	Appendix: Spectrum of Clicks / 316	
15.E	Appendix: erfc / 317	
16	PARAMETER VARIATION DUE TO NOISE	319
16.1	Preview / 319	
16.1.1	Automatic Gain Control (AGC) / 319	
16.1.2	Limiter / 320	
16.1.3	Driving the Phase Detector Hard from the Signal / 321	
16.1.4	Effects of Variations / 322	
16.2	AGC / 322	
16.2.1	Types of Detectors / 322	
16.2.2	Square-Law Detection / 323	
16.3	Limiter / 325	
16.3.1	Limiting in the Presence of Small Noise / 325	
16.3.2	Limiting in the Presence of Large Noise / 326	
16.4	Effects of Gain Variation on Loop Parameters / 329	

- 16.5 Effect of AGC or Limiter on an Optimized Loop / 331
- 16.A Appendix: Modified Bessel Functions / 333

17 CYCLE SKIPPING DUE TO NOISE 335

- 17.1 Phase / 336
 - 17.1.1 Fokker–Plank Method / 336
 - 17.1.1.1 Assumption Regarding the Nature of the Noise / 337*
 - 17.1.1.2 First-Order Loop / 337*
 - 17.1.1.3 Second-Order Loop / 339*
 - 17.1.2 Experimental Results / 341
 - 17.1.3 Simulation / 342
 - 17.1.3.1 First-Order and $\alpha = 0$ / 342*
 - 17.1.3.2 Second-Order, High-Gain / 343*
- 17.2 Cycle Skipping, Mean Time / 347
 - 17.2.1 First-Order Loop / 347
 - 17.2.2 Second-Order Loop, $\alpha = 0$ / 348
 - 17.2.3 Second-Order Loop, $\alpha = 1$ / 348
 - 17.2.4 Second-Order Loop, $0 < \alpha < 1$ / 350
 - 17.2.5 Probability of Cycle Skipping in a Given Time / 350
- 17.3 Cycle Skipping, Mean Frequency / 350
- 17.4 Cycle Skipping with Mistuning / 353
 - 17.4.1 Effect of α / 353
 - 17.4.2 Comparison to Other Results / 356
- 17.5 Summary / 356
 - 17.5.1 Phase Variance / 356
 - 17.5.2 Cycle Skipping / 357
 - 17.5.3 Skip Frequency / 357
 - 17.5.4 Mistuning in High-Gain Loops ($\zeta = 0.7$) / 357
- i17.D Appendix: Additional Data / 357
- i17.M Appendix: MATLAB[®] Scripts / 357

18 NONLINEAR OPERATION IN A LOCKED LOOP 359

- 18.1 Notation / 359
- 18.2 Phase-Detector Output u_1 / 360
- 18.3 Changes in the Output Spectrum / 361
- 18.4 Gain Suppression, Quasi-Linear Approximation / 361
 - 18.4.1 Basics / 363

- 18.4.2 Phase Variance with Additive Noise / 365
- 18.4.3 Tracking the Carrier / 368
 - 18.4.3.1 *With Phase Modulation* / 368
 - 18.4.3.2 *With Phase Modulation and VCO Noise* / 369
- 18.4.4 Effect on Phase Offset / 372
 - 18.4.4.1 *With High-Frequency Phase Modulation* / 372
 - 18.4.4.2 *With Additive Noise* / 373
 - 18.4.4.3 *With Limiting* / 375
- 18.4.5 Summary of Suppression / 375
- i18.S Appendix: Additional Offset Simulation Data / 376

19 ACQUISITION AIDS IN THE PRESENCE OF NOISE 377

- 19.1 Sweeping with Plain Closed-Loop / 377
 - 19.1.1 Maximum Sweep Rate in Noise / 377
 - 19.1.2 Effect of Initial Mistuning / 380
 - 19.1.3 Determining Successful Acquisition / 382
 - 19.1.4 Optimum Sweep Parameters / 383
- 19.2 Reduction of Coherent Detector Output (Closed-Loop Sweeping) / 385
- 19.3 Closed-Loop Sweeping in Noise with Coherent Detector / 386
 - 19.3.1 Successful Acquisition / 386
 - 19.3.2 False Stops / 388
 - 19.3.3 False Restart / 389
 - 19.3.4 False Stop versus False Restart / 390
- i19.M Appendix: MATLAB[®] Script, *swpi* / 393
- i19.S Appendix: Optimum Sweep for a Fixed Noise Density / 393

20 BANDLIMITED NOISE 395

- 20.1 Signals Centered in a Noise Band / 395
 - 20.1.1 Clicks with a First-Order PLL / 395
 - 20.1.2 Simulation Results Compared to Measured Data / 396
 - 20.1.3 Skip Rate / 398
 - 20.1.4 Output Phase Variance / 403
 - 20.1.5 Effect of a Limiter / 404
 - 20.1.6 Comparison to AGC / 406
 - 20.1.7 Higher-Order Loops / 407
 - 20.1.8 Summary, Symmetrical Narrowband Noise / 407
- 20.2 Eccentric Signals / 407
 - 20.2.1 Expected Performance / 408

20.2.2	Simulating Eccentric Noise /	409
20.2.3	Some Simulation Results /	409
20.2.3.1	<i>Cycle Skipping</i> /	410
20.2.3.2	<i>Phase Offset</i> /	411
20.2.3.3	<i>Output Variance</i> /	411
20.3	Extension to Other Types of Interference /	411
i20.M	Appendix: MATLAB [®] Scripts /	413
i20.O	Appendix: Offset Interference Data Correlation /	413
i20.S	Appendix: Band Limited Simulation Data /	413

21 FURTHER INFORMATION **415**

21.1	Sources for Additional Studies in Phase Lock /	415
21.2	Sources Covering Phase-Locked Frequency Synthesis /	416
21.A	Appendix: Modulations and Spectrums /	416
21.B	Appendix: Getting Files from the Wiley Internet Site /	421

REFERENCES **423**

INDEX **429**

PREFACE TO THE SECOND EDITION

This is the second edition of a book about the basics of phase-locked loops (PLLs). Basics tend to be unchanging, but we can learn more about them and their manner of presentation can be improved and tuned to current practice.

Loop filters driven by current sources are now so common that I have given them a more prominent place, allowing them to lead into the study of filters that use op-amps, rather than the reverse. The discussion of various applications, at the end of Part 1, which is designed to show how the basics can be applied to a wide variety of circuits, as well as to broaden the overview, has been expanded to include the now common use of PLLs for timing adjustment in ICs. This discussion includes the delay-locked loop, which is not a PLL but whose common forms can still be analyzed with the basic techniques that are used for PLLs. Frequency synthesis is another important application to which the basics can be applied, and I have broadened that description by adding material on modern fractional- N and sigma-delta techniques.

The most significant change, however, is in the discussion of the operation of PLLs in the presence of large amounts of noise (including the discernment of what amounts are large). The breakthrough that makes the improvements possible is the wide use of simulations with MATLAB[®], simulations that the reader can repeat and extend.

The first edition progressed into ever-increasing noise levels by moving from linear formulas, through approximations, to a collection of experimental or simulated results from the literature. It was difficult to know when the approximations had been taken beyond their limits of applicability, and only a limited amount of data is available from the literature. The second edition uses these available results to verify the simulations but is then not limited to the conditions for which results were available. This changes the process of studying nonlinear performance to such a degree that we now begin with the simulations, learning how the loop operates under whatever conditions are

of interest, and then progress to useful approximations whose limits are now clear. The approximations are still useful—a formula has advantages over a collection of data—but they are more useful when we know their limits of applicability.

Simulations have also permitted further investigation and a clearer understanding of the limits of sweep speed in noise.

The new process has led also to a clearer understanding of the theory, including a better understanding of what defines broadband noise and how bandwidth affects the results. Simulations also permit the study of narrowband noise, which is the more fundamental process, of which broadband noise is only an approximation in the limit. A chapter has been added on narrowband noise.

In order to keep the book to a desirable size in spite of the additional information, some material, primarily appendices, has been put online at ftp://ftp.wiley.com/public/sci_tech_med/phase_lock. This sometimes provides a benefit of multicolor graphs, which can be easier to read. The numbers of a section that is online may be preceded by the letter “i” to make that clear. The problems and problem answers have also been placed on the Wiley ftp site.

WILLIAM F. EGAN

Cupertino, California
May 2007

PREFACE TO THE FIRST EDITION

The phase-locked loop (PLL) has become so important in electronics over the past three decades that a working knowledge of its principles is an essential asset for many electrical engineers. This book is here to help you gain that knowledge. It concentrates on achieving an understanding of basic principles that are applicable to a wide range of PLL circuits—old, new, and yet-to-be-invented.

In 1977 I was given the opportunity to teach a graduate course, in the Santa Clara University School of Engineering, on the design of phase-locked frequency synthesizers, an area I had been working in for about fourteen years. The text for that course was my notes, which I later expanded into *Frequency Synthesis by Phase Lock*, published by John Wiley and Sons, Inc. in 1981.

In 1985, Professor Timothy Healy asked if I would teach a two-quarter PLL course, one that he had developed and taught for several years. I have done so since then. I used Blanchard's *Phase-Locked Loops* for a text, but I had to develop problems as well as course notes and I engaged in an ongoing study of original references. By 1991 I had introduced enough new and modified material that I decided to supplant Blanchard with my own text, but I thought I would begin gradually by providing the text for the first quarter only. However, before the end of the quarter, I was surprised (shocked?) to find that Wiley had stopped publishing Blanchard and it would not be available. This resulted in a rather intensive effort that produced the rest of my text in time for the second quarter. I have been using my own text since that time, modifying it each time (every other year) based on student comments and my own re-readings and to introduce new material.

In 1995 I submitted the latest version to Wiley as part of a proposal to publish. As was true with my previous book, the editor had the manuscript reviewed by a knowledgeable professional. In this instance, the reviewer was Professor William Tranter,

currently of Virginia Tech, who suggested that I introduce a software component to the book, especially since my biography showed experience in developing computer programs related to PLLs. He recommended MATLAB[®] for consideration. Since I was then prepared to immediately publish a text that had been carefully refined, I was reluctant to modify it. Nevertheless, I began to experiment with MATLAB and, as I did my enthusiasm grew. Here was an opportunity, using the student edition of MATLAB* on a personal computer, to significantly enhance the learning of PLL fundamentals. It was almost as though the reader would now have an opportunity for hands-on experimentation, being able to modify parameters and observe the effect on loop performance. In addition, MATLAB could provide useful design aids and even open some important areas where analysis is very difficult without a computer.

As a result, I have included MATLAB programs (scripts) and other software in this book in a way that does not detract from the original text, placing them in appendices following the chapters where the corresponding material is discussed. Many of the programs are written explicitly in the book, and they, and others, are available for downloading from a Wiley Internet site. That site also provides other aids in the form of Excel spreadsheets. The MATLAB and Excel files can be read by Macintosh[®] and IBM[®]-compatible computers, at least. One other very important benefit of the Internet site: the inevitable errata can be provided and updated.

The software is relatively easy to understand and customize to better suit an engineer's particular analysis or design problem. It adds an additional dimension of instruction by illustrating the use and design of computer-aided engineering (CAE) for PLLs; the simulation programs should be especially instructive in this regard. I have found that new ways to use the software to improve my understanding of the design process continue to occur to me. One must limit the volume of such material in a book, but, when one becomes familiar with the methods by which the software can be used, one is able to apply it as the situation demands.

Phase Lock Basics can be used in many ways by both the professor and the self-directed student. One can concentrate on the development of the theory or on the examples or work the problems, all of which have answers in the back, or can experiment with the MATLAB programs. Of course, a combination will give the most complete learning experience. That experience will also depend on the reader's background. For the graduate engineer, this material not only adds important specific knowledge but it can serve to integrate and solidify previously learned fundamentals, showing how the theory is used in the design of a particular kind of circuit, the PLL. For those lacking some of the desirable background, results, in the forms of equations, graphs, and descriptions of techniques, are still there. And, for you fortunate designers who already possess an understanding of PLLs, the many graphs and the software can help make your work more productive and, I hope, enjoyable.

*The Professional Version of MATLAB, with appropriate toolboxes, may also be used. Those who cannot gain access to either, but can use a program with similar capabilities, may be guided by the text form of the scripts. However, it will be considerably easier to use some version of MATLAB directly.

The text concentrates on the second-order loop. Loops of any greater complexity have too many parameters to permit general discussion. The standard parameters for second-order systems are natural frequency ω_n and damping factor ζ . Commonly, graphs are provided for various kinds of loop responses, in which the x axis is normalized to ω_n and multiple curves are given for various common values of ζ , enabling the reader to denormalize and interpolate to find results appropriate to a particular problem. There is one other necessary parameter, however, which I call α . For practical reasons (too many parameters), when plots of loop responses are given, they must be restricted to a few values of α . Other texts have generally given response curves for type 2 loops, those employing an integrator, and sometimes for loops with low-pass filters. These loops correspond to $\alpha = 1$ and $\alpha = 0$, respectively, but a continuum of useful loop filters exists between those extremes, and I show how transient response curves representing the extreme values of α can be combined to give the response for any value of α . MATLAB can generate responses, in both the time and frequency domain, for any value of α .

Many practical loops are second-order but some are more complex. Often the more complex loops can be approximated as second order so the second-order curves still give valuable information. I show how some of these curves are affected by the introduction of an additional pole, as is common, for example, in frequency synthesis. To illustrate how all the curves are affected by all the possible additional pole frequencies would, again, require too many dimensions to be practical. However, there is a common third-order configuration that can be analyzed with relative ease, and I have provided this analysis and corresponding curves in an appendix to Chapter 10. (I have placed this in an appendix to not detract from the concentration on loops of second order, on which the main development of the theory is based.) One of the great values of the state-space method, and the MATLAB programs that are based on it, is that it can handle higher order loops so, while one learns the general characteristics of PLLs using the more easily handled second-order theory, one can also expand to more complex problems with the help of MATLAB.

Chapter 10 introduces various applications of PLLs and shows how the theory developed in the previous chapters can be applied to them. This is not a substitute for detailed study in particular areas—there are many other things required for a good knowledge of synthesizer design, for example—but does show how the fundamentals apply. Part 2 of the text introduces the effects of noise, often large amounts of noise, on the performance discussed in Part 1. I have found that Part 1 is a good prerequisite for my synthesizer course, reducing the amount of material that has to be absorbed there. The whole book is an even better preparation, since noise concepts are also very important in synthesizer design. However, the noise levels are relatively low in synthesizer design, its elimination being an important goal, whereas communications systems may be designed to operate in the presence of large amounts of noise, such as are treated in Part 2. Thus this book would also be good preparation for a more advanced study of communication systems that depend on PLLs (coherent communications) and must operate in the presence of noise.

It is not uncommon for students, and designers presumably, to make 2π errors in computing loop bandwidth and other parameters, due to the prevalence of both

radians and cycles as units of phase (and in units of frequency). By handling units in a certain way, we are able to employ whichever is more convenient and to freely mix them without succumbing to this error. Proper handling of units is also a necessity in development of the equations. This easy movement between units decreases the importance of differentiating between f and ω symbols (although we will attempt to use the more appropriate symbol), especially since we can equate a frequency in one unit to the equivalent in the other.

While the answers provided for all the problems are of obvious help to the reader studying on his or her own, instructors may prefer problems that do not have answers. I suggest that these can be developed by modification of parameters given in the problems. The devoted but uncertain student can then verify his or her problem solving using the parameters in the book before changing to the parameters specified for the homework. The answers in the back of the book should be highly accurate since most have been proved by use. I have taught electrical engineering graduate students using nine two-hour classes plus two periods of exams for each half of the text. However, the software has not been employed in these courses nor has some other new material.

I would like to thank Professor Healy and Santa Clara for the opportunity to teach the Phase-Locked Loops course over the years and the students who have taken it for the contribution they have made, by interaction in the classroom and through their homework as well as by their suggestions and comments, in the development of the material in this book. I hope it will help you to better understand and appreciate the vast and fascinating world of phase-locked loops.

WILLIAM F. EGAN

Cupertino, California
June 1998

SYMBOLS LIST AND GLOSSARY

\Rightarrow	imply, become, go to
\approx	approximately equal to
\equiv	identically equal to, rather than being equal only under some particular condition
\triangleq	is defined as
AGC	automatic gain control
AM	amplitude modulation
BM	balanced mixer
B_n	noise bandwidth
C	controlled signal in standard control system terminology; fed-back input to the summing junction
c	cycle
closed-loop transfer function	C/R with the loop operating
damping factor	ζ , a standard parameter of the second-order system, Section 4.2
dB	decibel, $10 \log_{10}(\text{power ratio})$
dBc	decibels relative to carrier power, i.e., $10 \log_{10}(\text{power}/\text{carrier power})$
dBc/Hz	ratio of sideband power density, per hertz bandwidth, to total signal power; decibels relative to “carrier” per hertz noise bandwidth
dBm	decibels relative to 1 mW
DBM	doubly balanced mixer
dB r	decibels relative to 1 rad ²

dBr/Hz	phase noise power spectral density in decibels relative to 1 rad ² per hertz bandwidth
dBV	decibels relative to 1 volt
DDS	direct digital synthesizer
DLL	delay-locked loop
DPLL	digital phase-locked loop
E	error signal in standard control system terminology; output from the summer
error	output from summer in standard control system terminology
ExOR	exclusive-OR function, equal to one if and only if the two inputs differ
$E[y]$	expected value of y
$E[y^2(\cdot)]$	expected value of y^2
\bar{F}	mean frequency or skip rate (in Hz)
frequency power spectral density	$S_f(\omega_m)$, mean square frequency σ_f^2 in a narrow frequency band divided by the bandwidth. The band is narrow enough that the value of $S_f(\omega_m)$ does not change appreciably when the bandwidth decreases
$F(s)$	frequency sensitive part of filter transfer function, which is $K_{LF}F(s)$
f_x	see ω_x for any x
F_x	see Ω_x for any x
G	transfer function
$ G $	gain (absolute value of G)
gain margin	reciprocal of open-loop gain at frequency at which open-loop phase has 180° excess phase. The gain increase necessary for oscillation of the loop
G_F	forward transfer function of the loop
G_R	reverse transfer function of the loop (1 except in Chapter 10)
H	closed-loop transfer function, C/R in Fig. 2.5
hold-in	range $\pm\Omega_H$ of mistuning Ω over which lock can be maintained as the loop is tuned (Fig. 8.2)
IC	integrated circuit
ICO	current-controlled oscillator
\Im	imaginary part of
integrator-and-lead	filter of form $k(\omega_c + s)/s$
I_v	modified Bessel function (first kind) of order v
J_v	Bessel function (first kind) of order v
$K = K_p K_{LF} K_v$	$s \times$ [open-loop gain at DC] (∞ for a type 2 loop)
$K' = K'_p K_{LF} K_v$	maximum value of K (as the phase changes)
K_{LF}	DC gain of the loop filter (∞ in a type 2 loop)
K_p	phase-detector gain, change in voltage divided by change in phase
K'_p	maximum K_p (maximum as phase changes)
K_{pd}	K_p for coherent detector

K_v	VCO gain, change in VCO frequency divided by change in tuning voltage
L	single-sideband power spectral density (noise) relative to carrier
\mathcal{L}	single-sideband power spectral density under small modulation index. When due to phase modulation, always equals $S_\varphi/2$
$\lim_{a \rightarrow b}$	limit as a approaches b
low-pass filter	filter with transfer function of the form $K_{LF}/(1 + s/\omega_p)$
L_φ	single-sideband power spectral density due to phase noise
m	modulation index, peak phase deviation
mistuning	$\pm\Omega$: difference between the input (reference) frequency and the VCO center frequency (Fig. 8.2)
n'	phase equivalent to voltage at PD output due to additive noise
natural frequency	ω_n : a standard parameter of the second-order system
NCO	numerically controlled oscillator
N_i	input noise power to a limiter
N_0	one-sided noise power spectral density
OA	output accumulator
one-sided spectrum	representation of power spectral density in which all power is shown at positive frequencies
op-amp	operational amplifier
open-loop transfer function	transfer function around the loop as if it were broken to allow the response of the opened loop to be measured
P	signal power
P_c	carrier power; main signal power
PD	phase detector
phase margin	additional phase lag to cause the loop to oscillate, to have -360° at the frequency at which open-loop gain is unity
phase power spectral density	$S_\varphi(\omega_m)$, mean square phase σ_φ^2 in a narrow frequency band divided by the bandwidth. The band is narrow enough that the value of $S_\varphi(\omega_m)$ does not change when the bandwidth decreases
PLL	phase-locked loop
PM	phase modulation
P_o	total output power from a limiter
power spectral density	power per unit frequency, the limit (at a given frequency) of the ratio of power to bandwidth as bandwidth approaches zero
PPSD	phase power spectral density (which see)
PSD	power spectral density (which see)
pull-in	process of acquiring lock. Range $\pm\Omega_{PI}$ of mistunings Ω over which lock can be acquired (Fig. 8.2)

$P_y(f')$	power spectral density of the parameter y at the frequency f'
Q	quality factor, (energy stored)/(energy dissipated in a cycle), a measure of the sharpness of a resonance curve
r	ratio of the video-equivalent noise bandwidth of the RF filter (half of the actual noise bandwidth) to the noise bandwidth of the loop. Also Eqs. (8.61), (10.A.6)
R	normalized sweep rate, $ d\Omega/dt \text{rad}/\omega_n^2$ (see Table 9.1)
R	reference signal in standard control system terminology; independent input to the summer
rad	radian
rad^2/Hz	a measure of PPSD, the density that produces a phase variance of $\delta \text{ rad}^2$ in a narrow bandwidth $\delta \text{ Hz}$
\Re	real part of
reference	input to standard control system
r_T	relative threshold, ratio of threshold voltage to maximum signal available from a detector (Section 9.5.2)
s	the Laplace variable, $\sigma + j\omega$
S	demodulated signal power
SBM	singly balanced mixer
sec	second
seize	range $\pm\Omega_s$ of mistuning Ω over which the loop will acquire lock without skipping a cycle (Fig. 8.2)
$S_f(\omega_m)$	frequency power spectral density (which see) at frequency ω_m
S_i	input signal power to a limiter
S/N	signal-to-noise (ratio), sometimes specifically at the output rather than input, also called ρ
spectrum analyzer	an instrument that displays power vs. frequency by measuring power passing through a filter whose center frequency is swept in synchronism with the abscissa of the display
$S_y(f')$	power spectral density of y at frequency f'
S_ϕ	phase power spectral density {which see}
$S_{\phi n}(\omega_m)$	phase power spectral density of noise at frequency ω_m
$S_{\phi s}(\omega_m)$	phase power spectral density of signal at frequency ω_m
S_ω	frequency power spectral density (which see)
T_m	mean time to first cycle skip
T_{PI}	pull-in time
two-sided spectrum	representation of power spectral density in which half of the power is shown at positive frequencies and half is shown at negative frequencies
u_1	phase-detector output voltage or current
u_2	VCO tuning voltage or current

v	subscripts indicate responses to noise at VCO input
V_T	threshold voltage
V_v	VCO signal amplitude
W	RF bandwidth (one-sided)
α	parameter in the numerator of the second-order response, Eq. (6.4). 0 for lag filter; 1 for integrator and lead, otherwise
	$\alpha \equiv 1 - \frac{\omega_n}{2\zeta K} \equiv \frac{1}{1 + \omega_z/K} \equiv \frac{1}{1 + \omega_z\omega_p/\omega_n^2} \quad (6.5)$
γ	radius of gyration of a noise band, Eq. (15.41)
γ_d	factor in the S/N of frequency discriminators that differs for the various kinds (Section 15.6)
$\Delta\varphi$	$\varphi_{in} - \varphi_{out}$
$\Delta\omega$	$\omega_{out} - \omega_{in} = -\omega_e$
ζ	damping factor, a standard parameter of the second-order system
η	suppression factor or efficiency, multiplies K_p to produce effective PD gain. Subscripts: A AGC, Eq. (16.5) e phase error, Eq. (18.20) L limiter, Eq. (16.14) m modulation, Eq. (18.16) n noise, Eq. (18.12) p PD efficiency as a multiplier, Eq. (13.12)
ρ	signal-to-noise ratio (S/N), sometimes specifically at the output rather than input
ρ_{L0}	signal-to-noise ratio in a bandwidth equal to the loop noise bandwidth under linear conditions, $1/\sigma_{\varphi 0}^2$
σ_{φ}^2	mean square phase deviation, variance of phase
σ_{ω}^2	mean square frequency deviation, variance of frequency
$\sigma_{\omega s}^2$	variance of the signal frequency
τ_1	$(1/\omega_p)$ time constant of the loop pole
τ_2	$(1/\omega_z)$ time constant of the loop zero
φ_0	output phase response to additive input noise, linear response assumed (loop parameters are not changed by the noise)
φ_e	phase error, change in $\Delta\varphi$ (often from the center of the PD range)
$\phi_i(t)$	modulation-related input phase
φ_{in}	loop input phase (at phase detector)
φ_n	output phase response to additive input noise, nonlinear response assumed (loop parameters are changed by the noise)
$\phi_o(t)$	modulation-related output phase

φ_{out}	loop output phase (at VCO output)
Ω	normalized modulation frequency, ω_m/ω_n (Section 7.6)
Ω	mistuning: difference between the input (reference) frequency and the VCO center frequency (Fig. 8.2)
ω_c	VCO center frequency; spectral center or carrier frequency
ω_e	frequency error, frequency at output of loop summing junction ($\omega_{in} - \omega_{out}$)
Ω_H	\pm hold-in frequency range: maximum \pm change in mistuning Ω that still permits lock to be maintained
ω_{in}	loop input frequency (at phase detector)
ω_L	unity-gain bandwidth of the loop; frequency where open-loop gain is one
ω_m	modulation frequency
ω_n	natural frequency: a standard parameter of the second-order system
ω_{out}	loop output frequency (at VCO output)
ω_p	loop-filter pole frequency
Ω_{PI}	\pm pull-in frequency range: range of mistuning Ω over which eventual lock is certain
ω_{po}	pull-out frequency, the largest frequency step that can be tolerated without cycle skipping
ω_s	sampling frequency
Ω_S	\pm seize frequency range: range of mistuning Ω over which loop locks without skipping a cycle
ω_z	loop-filter zero frequency

Subscripts (shown with x):

x_0	x under linear conditions
x_n	effective value of x , changed from x_0 due to the effects of noise
x_v	responses to noise at VCO input

PART 1

PHASE LOCK WITHOUT NOISE

CHAPTER 1

INTRODUCTION

1.1 WHAT IS A PHASE-LOCKED LOOP (PLL)?

A phase-locked loop is a circuit that synchronizes the signal from an oscillator with a second input signal, called the reference, so that they operate at the same frequency. The synchronized oscillator is commonly a voltage-controlled oscillator (VCO), so we will usually use these terms interchangeably. The loop synchronizes the VCO to the reference by comparing their phases and controlling the VCO in a manner that tends to maintain a constant phase relationship between the two. In some types of phase-locked loops (PLLs) this phase relationship is held constant. In other types it is allowed to vary somewhat. But the frequency is always synchronized—otherwise the loop is said to be “out of lock.”

1.2 WHY USE A PHASE-LOCKED LOOP?

Phase-locked loops are often used because they provide filtering to the phase or frequency of a signal that is similar to what is provided to voltage or current waveforms by ordinary electronic filters. The designer has some control over the manner in which the phase (or frequency) of the VCO follows a changing reference phase (or a changing reference frequency—one cannot occur without the other). The loop can be made to follow quickly or to follow sluggishly. This capability is particularly valuable in removing the effects of noise on the reference or on the synchronized oscillator or, with astute design, on both.

If the reference signal is supposed to be a constant-amplitude phase-modulated signal [e.g., a frequency-modulated (FM) radio signal], the spectrum of the VCO will be a cleaned-up version of the reference spectrum. The PLL, while reproducing the reference signal, rejects all amplitude modulation noise and all other noise that is separated sufficiently in frequency from the signal. It acts like a filter that tracks the signal frequency. In fact, it can provide filtering that ordinary filters cannot because it can follow a signal whose frequency varies slowly by an amount that is greater than the filter bandwidth. For example, a PLL could be designed to filter out noise that is more than 1 kHz from a signal as the signal drifts by say 10 kHz.

Another use of the PLL is in phase or frequency modulating and demodulating signals. This is possible because there exist, within the PLL, a voltage that is proportional to the frequency of the reference and another that follows its phase. Not only can these be extracted for demodulation, but the loop can be forced to produce phase or frequency changes that are proportional to voltages that are injected into the loop, and thus to provide modulation.

Phase-locked loops are also used in frequency synthesis where the VCO oscillates at a selectable multiple of the reference. In this case the oscillator is synchronized with the reference without having the same frequency. However, the combination of the controlled oscillator and the subsequent frequency divider, which outputs the reference frequency, can be looked upon as a synchronized oscillator whose frequency is equal to the reference frequency.

Other PLLs, which are seemingly very different, involve things as diverse as rotating machinery and computer programs.

1.3 SCOPE OF THIS BOOK

This text describes the fundamentals of PLLs, principles that apply to all types of PLLs. It emphasizes the details that apply to the processing of analog electronic signals, as discussed in the first three paragraphs of the previous section, leaving details that are more peculiar to frequency synthesis to other works (see Section 21.2). However, at the end of Part 1, we will illustrate how the theory developed here can be used to analyze many applications, including synthesizers and other circuits usable in wireless and telecommunications.

The text concentrates on second-order loops because higher-order loops have too many parameters to make general discussions practical and because higher-order loops can often be approximated by second-order loops. The effect of such approximations will be discussed, too. A special class of third-order loops is analyzed and described in Appendix 10.A.

While the main body of text stands by itself, additional material is available from appendices that can be downloaded from the Wiley Internet site, as explained in Appendix 21.B. This material includes additional text, problems for most chapters and the answers to those problems, MATLAB[®] programs (scripts), and Microsoft[®] Excel spreadsheets. Appendices show how to use MATLAB or Excel as aids in understanding and designing PLLs.

The MATLAB scripts can run under either the Student or Professional editions of MATLAB. (Some scripts require the Control or Signal Processing Toolbox, which will be identified near the beginning of the script.) Refer to notes in the scripts and to the instructional and help material that comes with MATLAB or is available online to learn how to use the scripts and how to enhance and expand them. Scripts tend to increase in complexity as we advance through the text. They are intended as learning and research aids. While enhancing the understanding of the text, they also depend on an understanding of that material. They are not foolproof. Some of the scripts contain user-interface features that can be incorporated into others; there is a tradeoff between user-friendliness and simplicity.

Excel is a very adaptable program that can be of great help to the engineer. The graphs of time-domain and frequency-domain loop responses in this book were generated using Excel.

Appendices that are available online are briefly described in the text. Their section numbers may be preceded by the letter “i” to quickly indicate that they are online.

1.4 BASIC LOOP

A block diagram of the basic PLL is shown in Fig. 1.1. While the synchronized oscillator is usually a VCO, one whose frequency changes in response to a control voltage, u_2 , it is also possible that u_2 could be a current and the synchronized oscillator would thus be a current-controlled oscillator (ICO). However, we will consider it a VCO—the transition between the two is easily understood.

The phase detector compares the reference to the VCO’s output and produces a signal u_1 , which changes in proportion to the difference in their phases. This is processed by the loop filter to provide the oscillator control signal u_2 . The loop filter can be as simple as a conductor ($u_2 = u_1$) or a flat amplifier ($u_2 = K_{LF} u_1$), but it is usually designed to provide some advantageous response characteristic.

If the output frequency, $\omega_{out} = d\varphi_{out}/dt$, should be greater than the reference frequency, $\omega_{in} = d\varphi_{in}/dt$, then $u_1 \sim (\varphi_{in} - \varphi_{out} - \Theta)$ would necessarily decrease with time, causing u_2 to decrease, which, in turn, would cause ω_{out} to decrease, bringing ω_{out} down toward ω_{in} . Thus the PLL provides negative feedback to keep the output

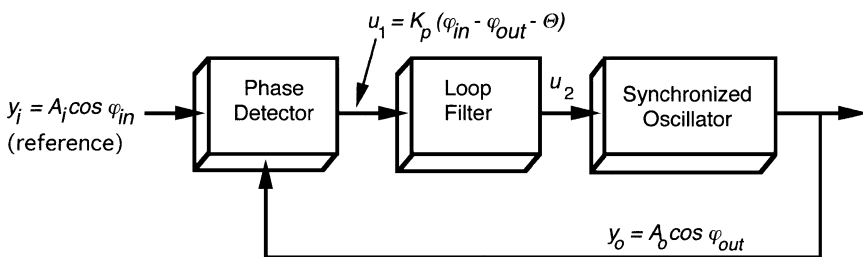


Fig. 1.1 Basic phase-locked loop; Θ is an offset that depends on the type of phase detector.

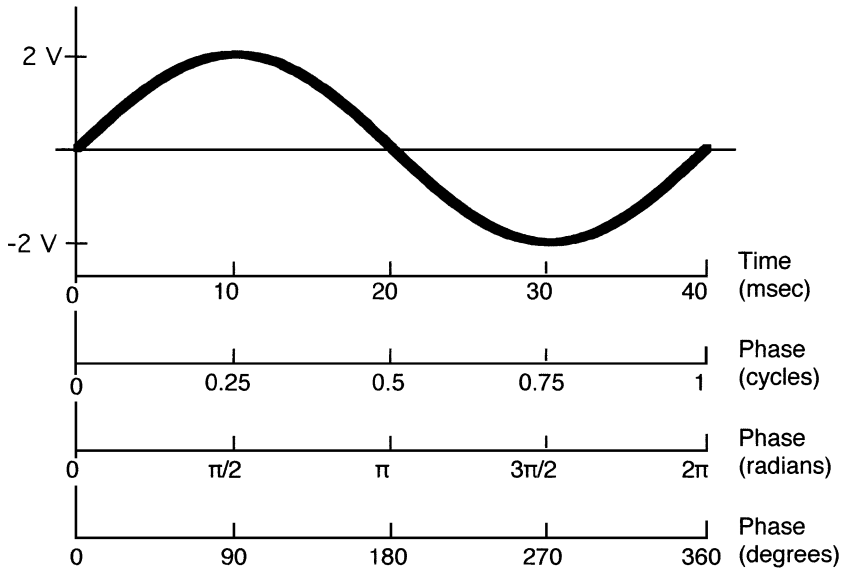


Fig. 1.2 A 2-V 25-Hz sine wave showing various measures of the abscissa.

frequency ω_{out} equal to the reference frequency ω_{in} . The output amplitude A_o is constant and independent of the input amplitude A_i .

1.5 PHASE DEFINITIONS

Figure 1.2 shows one cycle of a sinusoid and various ways of measuring the abscissa value. The first measure is time. The other measures are of phase, which indicates the time relative to the waveform's period. The figure illustrates how phase can be measured in any of three common units, cycles (c), radians (rad), or degrees (deg). Because of this we should be careful to carry the units in our calculations. Otherwise errors, typically of 2π , begin to appear in important computed parameters. For example, if φ equals $10t^2 \text{ rad/sec}^2$ (a linear frequency sweep), then the frequency is

$$\omega = \frac{d\varphi}{dt} = 20t \text{ rad/sec}^2. \quad (1.1)$$

The frequency can also be expressed as¹

$$f = (20t \text{ rad/sec}^2)(1\text{c}/2\pi \text{ rad}) = 3.18t \text{ c/sec}^2 = 3.18t \text{ Hz/sec}. \quad (1.2)$$

¹ Note that these expressions produce correct units for phase and frequency when time units are included with t . For example, at $t = 3 \text{ sec}$, $\varphi = 10(3 \text{ sec})^2 \text{ rad/sec}^2 = 90 \text{ rad}$, $\omega = 20(3 \text{ sec})\text{rad/sec}^2 = 60 \text{ rad/sec}$, and $f = 3.18(3 \text{ sec})\text{Hz/sec} = 9.54 \text{ Hz}$.

In this example the change of symbol between ω and f is a clue to the units, but it is more efficient to carry units than to memorize a different set of formulas for each set of units. The task of maintaining units is made more difficult, however, because of the widespread practice of not considering the radian a unit (i.e., considering angles measured in radians to be without units) and the common practice of not incorporating units in equations involving Laplace transforms (see Section 1.9) or derivatives and integrals of some functions. An alternative to carrying units might be to employ always a set of formulas that use one type of measure, say radians, and each time to convert to those units. This can be awkward in a field that deals so much with angle units and in which various angle units are more natural to use in various situations. In this text the usual procedure will be to carry units, with one exception that we will discuss at the end of this chapter.

Example 1.1 Preservation of Units The time derivative of $1 \text{ V} \sin kt$ is commonly said to be

$$\frac{d}{dt} 1 \text{ V} \sin kt = k \text{ V} \cos kt \quad (1.3a)$$

but the units are wrong. For example, if $k = 5 \text{ Hz}$, Eq. (1.3a) gives a time derivative at $t = 0.01 \text{ sec}$ of $[5 \text{ c-V/sec} \cos 0.05 \text{ c}]$. The cosine could conceivably be obtained from a table written in cycle units or any table could be used with conversion of 0.05 c to the correct units (e.g., 18°), but cycle-volts/second (c-V/sec) are not proper units for the slope of a voltage waveform. We are able to use Eq. (1.3a) correctly only because we know that it requires that k be in radians/second and because we know that radian units may not be carried. We would therefore begin by converting 5 Hz to $10\pi \text{ rad/sec}$. Then we would drop the radian unit to obtain the correct answer without explicitly converting units.

The formula that correctly carries units gives the derivative of $\sin kt$ as $(k/\text{radian}) \cos kt$ so that Eq. (1.3a) becomes

$$\frac{d}{dt} 1 \text{ V} \sin kt = \frac{k}{\text{rad}} \text{ V} \cos kt \Rightarrow \frac{5 \text{ c-V}}{\text{rad-sec}} \cos 0.05 \text{ c} = 10\pi \text{ V/sec} \cos 0.05 \text{ c}. \quad (1.3b)$$

Equation (1.3b) permits the use of any units.

Figure 1.3 shows how the phase difference between two sine waves is defined. At time t_x the phase difference is

$$\Delta\varphi(t_x) = \varphi_2(t_x) - \varphi_1(t_x) = \sin^{-1} v_2/A_2 - \sin^{-1} v_1/A_1. \quad (1.4)$$

This is true even if the frequencies of the two sinusoids are not the same. However, some phase detectors are sensitive to the zero crossings of the waveforms. For these the quantity measured is really

$$\Delta\varphi = (\Delta t/T) \text{ cycles}. \quad (1.5)$$

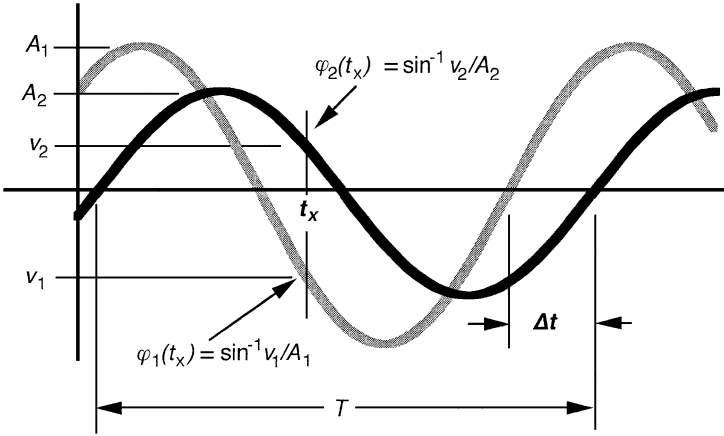


Fig. 1.3 Defining relative phase.

If the frequencies differ, the phase of one of the waveforms is used as a reference and T is taken from that waveform. In that case what we are actually measuring is the phase of the reference waveform at the zero crossing of the second waveform.

Equation (1.5) not only illustrates the carrying of units but also the efficiency of using whatever units are most convenient.

1.6 PHASE DETECTOR

Figure 1.4 shows the type of response that we would like to get from a phase detector (PD). It produces a voltage proportional to the difference in phases of the reference φ_{in} and the VCO output, which is also the loop output, φ_{out} . The constant of proportionality, K_p , is the gain of the phase detector, relating its output voltage change to its input phase change. For this case

$$K_p = A / (2\pi \text{ rad}) = A / (1 \text{ c}). \tag{1.6}$$

For example, if $A = 3 \text{ V}$, then $K_p = 3 \text{ V/c} = 0.48 \text{ V/rad}$.

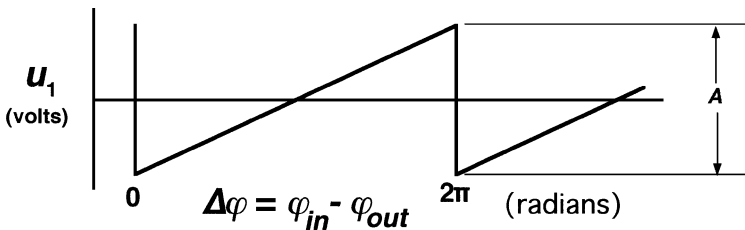


Fig. 1.4 A “Linear” phase detector characteristic. The characteristic is linear but only over one cycle of phase range.

The PD response will always involve a nonlinearity. Of course, all representations of the responses of physical things involve a nonlinearity for large enough values, but, in the case of the PD, this nonlinearity can easily occur in a region of operation, or at least of consideration. The PD in Fig. 1.4 has a linear range of 1 cycle. Often the linear range is smaller than a cycle, but some types have a linear range of 2 cycles.

A common type of PD, the balanced mixer, may produce a sinusoidal response like that shown in Fig. 1.5. The gain is the slope of the characteristic,

$$u_1 = B \cos \Delta\varphi, \tag{1.7}$$

$$K_p = \frac{du_1}{d\Delta\varphi} = \frac{d(B \cos \Delta\varphi)}{b \rightarrow B d\Delta\varphi} = \frac{-B \sin \Delta\varphi}{\text{rad}}, \tag{1.8}$$

where

$$\Delta\varphi = \varphi_{\text{in}} - \varphi_{\text{out}}. \tag{1.9}$$

The most desirable operating point is at $\Delta\varphi = -\pi/2$, where the slope is maximum and equals

$$K'_p \equiv K_p(\Delta\varphi = -\pi/2) = B/\text{rad} \tag{1.10}$$

so Eq. (1.7) can be rewritten

$$u_1 = K'_p \text{ rad} \cos \Delta\varphi. \tag{1.11}$$

Thus K'_p , the maximum value of K_p , for a sinusoid equals its amplitude per radian. Often K_p is used to mean this maximum value. Obviously, this characteristic has no truly linear range, but $K'_p \geq K_p \geq 0.7 K'_p$ for a range of $\pm 45^\circ$ about $\Delta\varphi = -\pi/2 = -90^\circ$.

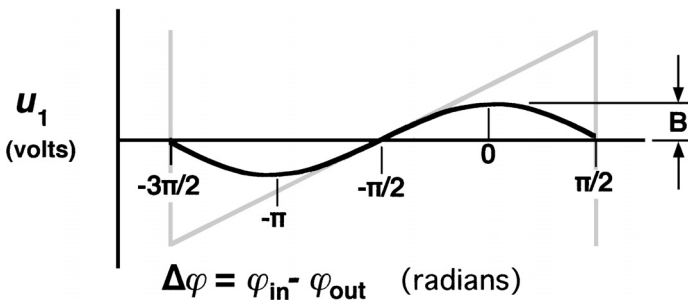


Fig. 1.5 Sinusoidal phase-detector characteristic, superimposed on the linear response. Both of these have the same gain K'_p at $\pi/2$ radians.

1.7 COMBINED GAIN

In its simplest form, the loop filter in Fig. 1.1 consists merely of an amplifier, in which case

$$\frac{du_2}{du_1} \equiv K_{\text{LF}}. \quad (1.12)$$

Just as the PD can have a nonlinear response characteristic, so can the VCO. But it too can be characterized at some operating point by the gain there:

$$\frac{d\omega}{du_2} \equiv K_v, \quad (1.13)$$

where K_v has units of radians per second per volt, or

$$\frac{df}{du_2} \equiv K_v, \quad (1.14)$$

where the units are hertz per volt.

If we break the loop's feedback path in order to determine the gain around the loop (the "open-loop gain"), we see that

$$\frac{d\omega_{\text{out}}}{d\Delta\varphi} = K_p K_{\text{LF}} K_v \triangleq K. \quad (1.15)$$

Checking the units of K_p , K_{LF} , and K_v , we can see that the units of K should be reciprocal seconds, sec^{-1} . This is not quite the entire open-loop transfer function, which is the ratio of $d\omega_{\text{out}}$ at the output of the opened loop to $d\omega_{\text{out}}$ at its input, and which has no units.

Example 1.2 Combined Gain A phase detector, in the region of operation, has a characteristic with a slope of 0.1 V per radian so

$$K_p = 0.1 \text{ V/rad.}$$

The loop filter is an amplifier with a voltage gain of 5 so

$$K_{\text{LF}} = 5.$$

The slope of the VCO tuning curve at the operating frequency is 2 MHz/V so

$$K_v = 2 \times 10^6 \text{ Hz/V.}$$

The combined gain is, from (1.15),

$$K = 10^6 \frac{\text{Hz}}{\text{rad}} = 10^6 \frac{\text{c}}{\text{sec-rad}} = 10^6 \frac{\text{c}}{\text{sec-rad}} \frac{2\pi \text{ rad}}{\text{c}} = 6.28 \times 10^6 \text{ sec}^{-1}.$$

If the input phase φ_{in} suddenly changes by 0.01 radian, the immediate result will be a change of voltage at the phase detector output of

$$0.01 \text{ rad}(K_p) = 10^{-3} \text{ V}.$$

This will cause a change K_{LF} larger at the loop filter output, that is 5×10^{-3} V. This will produce a frequency change at the VCO output of

$$5 \times 10^{-3} \text{ V}(K_v) = 10^4 \text{ Hz}.$$

Or we can also simply multiply the phase change by K to obtain the frequency change

$$0.01 \text{ rad}(6.28 \times 10^6 \text{ sec}^{-1}) = 6.28 \times 10^4 \text{ rad/sec} = 1 \times 10^4 \text{ Hz}.$$

In a locked loop, this frequency change causes the phase at the VCO output to begin to advance more rapidly than the input phase and thus to begin the process of restoring the phase difference that existed before the step in input phase.

1.8 OPERATING RANGE

Assuming a PLL that is initially locked, what limitation is there on the range of ω_{in} over which lock can be maintained? As the input frequency ω_{in} is slowly lowered, the operating point will move down the ramp in Fig. 1.4 to lower the VCO's frequency ω_{out} , and keep it equal to ω_{in} . At some point ω_{in} may become so low that lock will be lost because it is not within the capability of the PD to generate a voltage that is low enough to cause ω_{out} to equal ω_{in} . Likewise there is a limit on how high ω_{in} can go before the operating point runs off the upper extremity of the PD characteristic. The difference between these two extremes is called the "hold-in" or "synchronization" range.

From Fig. 1.4 we can see that the output of a PD with the sawtooth characteristic can vary over a total range of A . By Eq. (1.6), this equals $2\pi K_p$ rad. We can also write this as $\pm \pi K_p$ rad about the midpoint. We obtain the corresponding change in ω_{out} by multiplying the change in PD output by $K_{\text{LF}}K_v$, giving a total hold-in or synchronization range of

$$\pm \Omega_{H, \text{sawtooth}} = \pm \pi K_p K_{\text{LF}} K_v \text{ rad} = \pm \pi K \text{ rad}. \quad (1.16)$$

Note that, since the units of K are reciprocal seconds, the units of Ω_H are radians/second, as they should be. If there is significant curvature in the VCO tuning

characteristic over $\pm\Omega_H$, an average value of K_v should be used. In practice, however, the actual values of u_2 that correspond to the extremes of the PD range must be found before the region of operation of the VCO can be determined so it is as easy to find Ω_H from the tuning curve as it is to determine the average value of K_v . This is illustrated in Fig. 1.6.

From Fig. 1.5 we can see that the output of a PD with a sinusoidal characteristic can vary over a total range of $2B$. By Eq. (1.11), this equals $2K'_p$ rad. The hold-in range for this PD is therefore

$$\pm\Omega_{H, \text{ sinusoidal}} = \pm K' \text{ rad}, \tag{1.17}$$

where $K' = K'_p K_{LF} K_v$, the maximum value of K . Thus, for the same maximum gain, the hold-in range of the sawtooth PD is π larger than that of the sinusoidal PD. Assuming other components are linear, we can write the forward transfer function for a loop with this phase detector as

$$\Delta f = K' \cos \Delta\varphi. \tag{1.18}$$

Here the Δf and $\Delta\varphi$ are deviations from midrange. In practice, the range of other components, such as the VCO or an operational amplifier, can be more limiting than that of the phase detector. However, these limitations are not fundamental—they can normally be increased, whereas the range of any phase detector has an inherent limit.

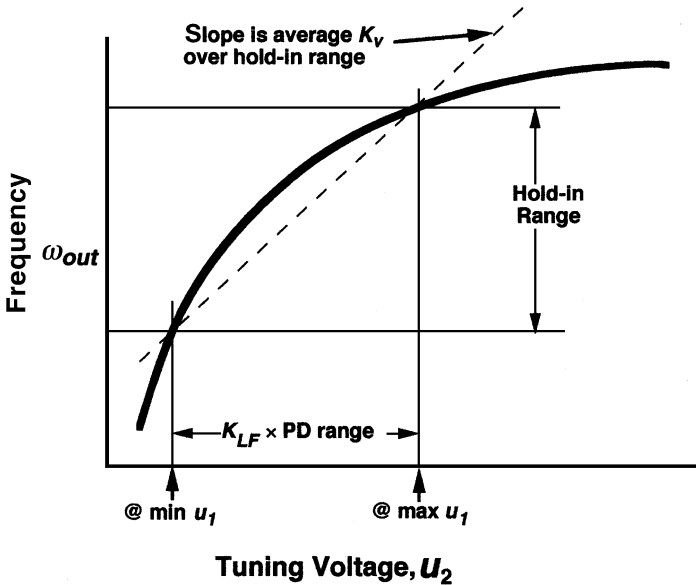


Fig. 1.6 Typical VCO tuning curve with the average slope over the hold-in range shown.

For this reason, the phase detector is assumed to be the range-limiting element in most PLL analyses.

While we have assumed a simple loop filter, Eqs. (1.16)–(1.18) apply also for more complex loops whose low-frequency filter gains are K_{LF} .

Example 1.3 Hold-in Range The parameters in Example 1.2 describe a loop when the phase is in the center of the phase detector's range (output midway between maximum and minimum). Assume the other components have linear responses.

If the phase detector is sinusoidal, (1.17) indicates a hold-in range of

$$\pm K' \text{ rad} = \pm(6.28 \times 10^6 \text{ sec}^{-1}) \text{ rad} = \pm 6.28 \times 10^6 \text{ rad/sec} = \pm 1 \text{ MHz.}$$

Thus the VCO can be tuned ± 1 MHz without loss of lock.

If the phase detector has a sawtooth characteristic, (1.16) gives the hold-in range as

$$\begin{aligned} \pm \pi K \text{ rad} &= \pm 3.14(6.28 \times 10^6 \text{ sec}^{-1}) \text{ rad} \\ &= \pm 1.97 \times 10^7 \text{ rad/sec} = \pm 3.14 \text{ MHz.} \end{aligned}$$

1.9 UNITS AND THE LAPLACE VARIABLE s

The Laplace variable s represents d/dt , so it should have units of reciprocal seconds. When $1/s$ represents $\int dt$, again, s should have units of reciprocal seconds. But then, for consistency of units, $s = j\omega + \sigma$ should be written $s = j\omega \text{ rad} + \sigma/\text{neper}$. The open-loop transfer function should be written $-G = -K/(j\omega/\text{rad})$. How can we prevent the confusion and wrong answers that can be caused by not carrying units and yet not be overburdened by nontraditional rules involving radian units? Perhaps the best solution is to *drop or add radians, and only radians, as required*. For example, if I want the hold-in range in hertz and I use Eq. (1.16) but drop the “rad,” the hold-in range is $2\pi K$. The unit of K is reciprocal seconds, but this is not a proper unit for frequency. To obtain proper units, we add “rad,” obtaining $2\pi K \text{ rad}$ with units of radians/second for hold-in range. We cannot mistake this for the hold-in range in hertz because we do not treat cycles as arbitrarily as we do radians. To obtain hold-in range in hertz, we must convert Eq. (1.16) as follows:

$$2\Omega_{H, \text{ sawtooth}} = 2\pi K \text{ rad} \frac{c}{2\pi \text{ rad}} = Kc. \quad (1.19)$$

This has units of cycles/second = hertz ($c/\text{sec} = \text{Hz}$), as desired. In fact, even though the equalities above are valid, Ω , which we might assume from the symbol to be in radians/second, has here been equated to hold-in range in hertz. This gives a hint of the danger in relying on the symbol to differentiate between radians/second and hertz.

CHAPTER 2

THE BASIC LOOP

We begin by studying the basic phase-locked loop (PLL), whose loop filter is represented by a frequency-independent gain. We will look at its transient response and its frequency response in the same way that we will later look at those for the more complicated cases, but it is important that we first gain a clear understanding by studying this simpler case.

2.1 STEADY-STATE CONDITIONS

The center frequency of the voltage-controlled oscillator (VCO) f_c is the frequency that occurs when the phase detector is in the center of its output range or the frequency midway between the frequencies that occur at the edges of the phase detector range. If the transfer function from the phase detector to the VCO output frequency is linear, these are the same thing. The output from the phase detector is not necessarily zero at that point; it depends on the particular phase detector realization. Even if it is zero, a voltage shift may be implemented between the phase detector and the VCO. In fact, most VCOs are operated with a tuning voltage that is always positive or always negative. While these details are of considerable practical concern, they are easily understood, and our studies will concentrate on changes from some initial or average condition.

Figure 2.1 is similar to Fig. 1.6 except that the independent variable has been changed from the tuning voltage u_2 to the phase difference $\Delta\varphi$. It illustrates the

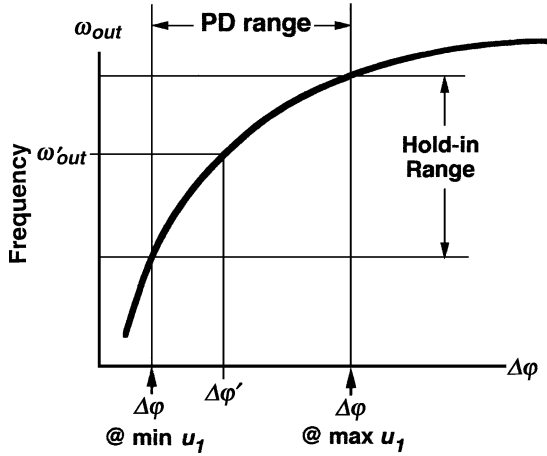


Fig. 2.1 Loop output frequency vs. phase difference at the phase detector with a typical VCO. The hold-in range, over which the frequency will vary as the phase detector covers its output range, is indicated.

one-to-one correspondence between the VCO’s frequency and its phase (relative to the reference phase) in steady state. In this simple first-order, loop, the one-to-one relationship holds not only in steady state but always. If we know one variable, we know the other because there is no storage, no memory, in the filter. The loops that we will study later will be different in this regard; the filter will permit the relationship between the VCO’s frequency and phase to change with time. That will make those loops more versatile, and more challenging.

2.2 CLASSICAL ANALYSIS

Before using Laplace transforms to analyze the loops’ dynamics, let us first determine the response by more directly solving Eq. (1.15). We will find that the PLL responds to changes in phase or frequency in a manner similar to that in which other simple circuits respond to changes in voltage or current. Thus we find that, whereas, in other circuits voltage and current are the variables that describe the state of the circuit and phase and frequency are parameters of those variables, in the PLL frequency and phase are state variables. They too, nevertheless, have frequencies and phases, which are their modulation parameters. For example, Eq. (2.1),

$$\omega_{out} = \omega_c + A \sin(\omega_m t + \theta_0), \tag{2.1}$$

describes the state variable ω_{out} , which represents the output frequency of the loop, but ω_{out} can be seen to possess a frequency (its modulation frequency) ω_m and a

modulation phase, θ_0 . These parameters must also be parameters of u_1 and u_2 , other state variables in the loop, since u_1 , and u_2 are proportional to ω_{out} .

2.2.1 Transient Response

Equation (1.15) can be rewritten as

$$d\omega_{\text{out}} = K d\Delta\varphi. \quad (2.2)$$

The time derivative of this equation is

$$\frac{d\omega_{\text{out}}}{dt} = K \frac{d\Delta\varphi}{dt} = K(\omega_{\text{in}} - \omega_{\text{out}}), \quad (2.3)$$

where Eq. (1.9) was used. We rewrite (2.3) as

$$d\omega_{\text{out}} = K(\omega_{\text{in}} - \omega_{\text{out}}) dt. \quad (2.4)$$

If ω_{in} is constant, it may be subtracted from ω_{out} on the left side, since the differential will not be affected by inclusion of a constant. This gives

$$d(\omega_{\text{out}} - \omega_{\text{in}}) = -K(\omega_{\text{out}} - \omega_{\text{in}}) dt \quad (2.5)$$

or

$$d\Delta\omega = -K\Delta\omega dt, \quad (2.6)$$

where

$$\Delta\omega \equiv \omega_{\text{out}} - \omega_{\text{in}}. \quad (2.7)$$

We now write Eq. (2.6) as

$$\frac{d\Delta\omega}{\Delta\omega} = -K dt \quad (2.8)$$

and integrate both sides to obtain

$$\int \frac{1}{\Delta\omega} d\Delta\omega = - \int K dt \quad (2.9)$$

with the solution

$$\ln \Delta\omega = -Kt + C. \quad (2.10)$$

Taking the exponential of both sides, we obtain

$$\Delta\omega = e^{-Kt+C} = e^C e^{-Kt}. \quad (2.11)$$

We can determine the value of e^C by evaluating this expression at $t = 0$. When we do so and substitute the results in Eq. (2.11), we obtain

$$\Delta\omega(t) = \Delta\omega(0)e^{-Kt}. \quad (2.12a)$$

The output frequency is the input frequency plus $\Delta\omega$,

$$\omega_{\text{out}} = \omega_{\text{in}} + \Delta\omega(t) = \omega_{\text{in}} + \Delta\omega(0)e^{-Kt}. \quad (2.12b)$$

Thus we see that this simple loop responds to an initial frequency error (difference between the output and reference frequencies) $\Delta\omega(0)$ by exponentially decreasing the error with a time constant equal to $1/K$. As a result, the step response shown in Fig. 2.2a is obtained. This is the same response to a frequency step that a single-pole low-pass circuit has to a voltage step (Fig. 2.2b).

We might intuit this exponential response through the following considerations. After the input frequency steps, there is a frequency error $\Delta\omega$ that causes $\Delta\phi$ to change. This change in $\Delta\phi$ (multiplied by K) causes the VCO frequency to change in such a direction as to decrease the frequency error (Fig. 2.3). This, in turn, decreases the rate of phase change and thus the rate at which the VCO frequency changes. As always, when the rate of change of a variable is proportional to the value of the variable, it changes exponentially. Of course, the exponential shape will be perturbed to the degree that there is curvature in the gain characteristics over the operating region.

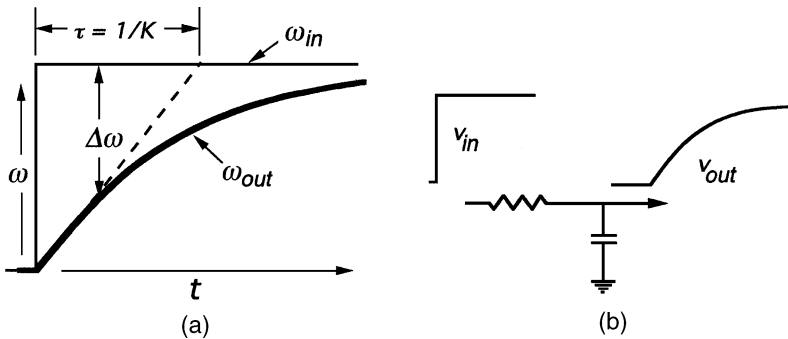


Fig. 2.2 Step response of (a) a simple loop compared to the step response of (b) a low-pass filter. The filter in this loop is a simple amplifier. The frequency of the loop responds like the voltage of the RC circuit.

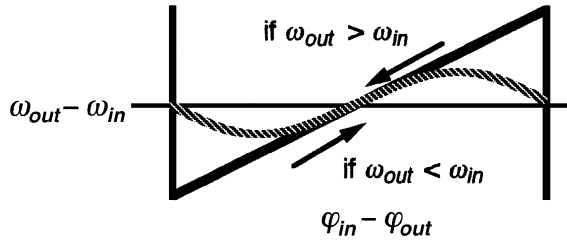


Fig. 2.3 Output frequency vs. phase for two phase detector characteristics. Since frequency is the derivative of phase, the direction of change is determined by the sign of the frequency difference.

Example 2.1 Transient Response When the input frequency to a particular loop steps to a new frequency, we require that the output achieve that frequency, to within an accuracy of 1%, in 1 msec. What are the range of K values that allow a simple loop to comply with this requirement?

Restating the problem, we require

$$\Delta\omega(1 \text{ msec})/\Delta\omega(0) \leq 0.01.$$

From (2.12a) this requires

$$(e^{-K \cdot 10^{-3} \text{ sec}}) \leq 0.01 \Rightarrow K \geq -\ln(0.01)/10^{-3} \text{ sec} = 4.6 \times 10^3 \text{ sec}^{-1}.$$

2.2.2 Modulation Response

Rearranging Eq. (2.3) we can write

$$\frac{1}{K} \frac{d\omega_{out}}{dt} + \omega_{out} = \omega_{in}. \tag{2.13}$$

If the input is frequency modulated, then we expect the response to be frequency modulated also, so we assume a response of the form

$$\omega_{out}(t) = \omega_c + A \sin(\omega_m t), \tag{2.14}$$

where ω_c is the center, or carrier, frequency, A is the peak frequency deviation, and ω_m is the modulation frequency. Putting Eq. (2.14) into (2.13), we obtain

$$A \frac{\omega_m}{K} \cos(\omega_m t) + \omega_c + A \sin(\omega_m t) = \omega_{in}. \tag{2.15}$$

We rewrite (2.14) and (2.15) in terms of deviations from the mean:

$$\Delta\omega_{out} \equiv \omega_{out}(t) - \omega_c = A \sin(\omega_m t), \tag{2.16}$$

$$\Delta\omega_{in} \equiv \omega_{in} - \omega_c = A \frac{\omega_m}{K} \cos(\omega_m t) + A \sin(\omega_m t). \tag{2.17}$$

The amplitude of $\Delta\omega_{\text{out}}$ is A and that of $\Delta\omega_{\text{in}}$ is

$$|\Delta\omega_{\text{in}}| = A\sqrt{\left(\frac{\omega_m}{K}\right)^2 + 1}, \quad (2.18)$$

so the magnitude of the frequency response is

$$|H(\omega_m)| = 1/\sqrt{\left(\frac{\omega_m}{K}\right)^2 + 1}. \quad (2.19)$$

Since $\Delta\omega_{\text{in}}$ contains a component equal to $\Delta\omega_{\text{out}}$ plus one that leads it by 90° and is larger by ω_m/K , the phase angle of $\Delta\omega_{\text{in}}$ relative to $\Delta\omega_{\text{out}}$ is $\tan^{-1}[\omega_m/K]$ and the transfer phase is

$$\angle H(\omega_m) \equiv \varphi_{\text{out}}(\omega_m) - \varphi_{\text{in}}(\omega_m) = -\tan^{-1}[\omega_m/K]. \quad (2.20)$$

As with the time response, the response of this loop to a sinusoidally modulated frequency is the same as the response of a low-pass filter to a sinusoidal voltage. In each case the corner (-3 dB) frequency is $\omega = K$.

Example 2.2 Modulation Response A simple loop has a sinusoidal phase detector characteristic (Fig. 1.5) and $K = 1000 \text{ sec}^{-1}$. The input (reference) frequency is frequency modulated at an unknown rate with a peak deviation of 50 Hz. What is the phase shift between the VCO modulation and the input modulation when the deviation at the output is 10 Hz?

We will use the gain to determine the modulation frequency and then use the modulation frequency to determine the phase shift. The closed-loop gain is

$$|H(\omega_m)| = 10 \text{ Hz}/50 \text{ Hz} = 0.2.$$

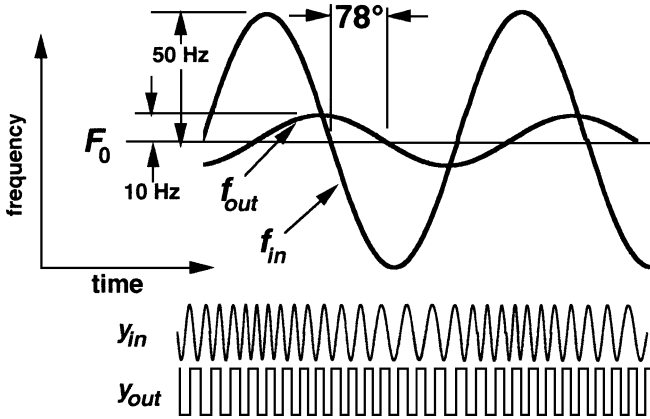
From (2.19), this occurs at

$$\sqrt{\left(\frac{\omega_m}{K}\right)^2 + 1} = 5,$$

from which we obtain

$$\frac{\omega_m}{K} = \sqrt{24},$$

giving the modulation frequency as $\omega_m = 4899 \text{ rad/sec}$. Inserting ω_m/K into (2.20), we obtain the phase shift as $\angle H(\omega_m) = -\tan^{-1}(4.9) = -78.5^\circ$. Thus the output modulation lags the input modulation by 78.5° .



The usual input and output state variables (y for voltage or current) are shown below the plot of instantaneous frequency. The input happens to be a sinusoid while the output is a square wave. These are intended to illustrate the meaning of the frequency deviations plotted above. Note that the magnitude of the output frequency deviation is smaller than that of the input deviation and how the times of occurrence of maximum frequency (minimum period) correspond to the peaks in the frequency plots.

2.3 MATHEMATICAL BLOCK DIAGRAM

The mathematical relationships in the loop can be shown by means of a block diagram such as Fig. 2.4. Here the input variable is the reference, or input, instantaneous phase φ_{in} , which is, in general, a function of time, perhaps a sinusoidal function such as

$$\varphi_{in}(t) = \varphi_i + A_i \sin(\omega_m t + \theta_i) \tag{2.21}$$

with average value φ_i , peak deviation A_i , modulation frequency ω_m , and phase θ_i . Generally, as in any control system, responses are given in terms of the deviation from

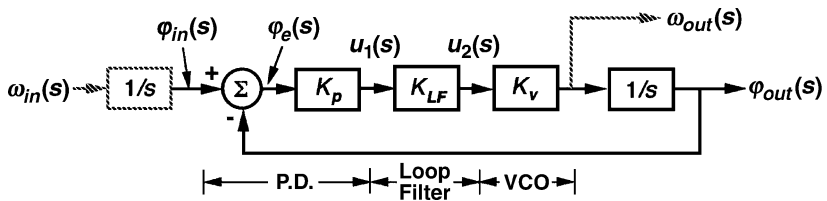


Fig. 2.4 Mathematical block diagram of the simple loop with phase variables as input and output. The frequency state variables are shown to indicate how they relate to the phase variables.

the average or initial or steady-state value, and φ_i is only of interest in establishing the steady-state operating point. A block is shown preceding φ_{in} in order to establish the integral relationship ($1/s$) between $\omega_{in}(s)$ and $\varphi_{in}(s)$. The output from the summer (a subtracter in this case) is the phase error $\varphi_e(s)$,¹ the difference between the input phase and the VCO phase, φ_{out} . The phase error is converted to voltage in the phase detector, which is represented by the gain $K_p = u_1(s)/\varphi_e(s)$. For this simple case the loop filter is merely an amplifier with gain K_{LF} ; K_v represents the tuning sensitivity $\omega_{out}(s)/u_2(s)$ of the VCO. The output from this block is the output frequency ω_{out} . However, φ_{out} is needed to complete the loop so $\omega_{out}(s)$ is integrated (multiplied by $1/s$) to produce it. Then the loop is completed by subtracting $\varphi_{out}(s)$ from $\varphi_{in}(s)$ in the summer. Note that the minus sign at the summer represents -180° of phase shift around the loop that does not appear in the transfer functions of the individual blocks.

The generic control system block diagram is shown in Fig. 2.5. The well-known equations describing its transfer function are the response of the controlled variable C to the reference R ,

$$\frac{C}{R} = \frac{G_F}{1 + G_F G_R}, \quad (2.22)$$

where G_F and G_R are forward and reverse transfer functions, respectively, and the response of the error E to the reference²

$$\frac{E}{R} = \frac{1}{1 + G_F G_R}. \quad (2.23)$$

The correspondence between the usual designations for loop signals, C , R , and E , and state variables of the phase-locked loop depends on the input and output of interest. Most often, in this text, we will be interested in the configuration shown in Fig. 2.4 so we identify C as $\varphi_{out}(s)$ and R as $\varphi_{in}(s)$. Because the feedback path has unity gain, we have also $G_R = 1$ and $-G_F$ is the entire open-loop transfer function $-G(s)$,

$$G(s) = K_p K_{LF} K_v / s \equiv K / s. \quad (2.24)$$

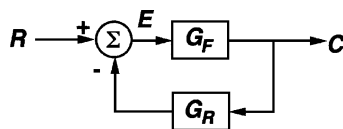


Fig. 2.5 Generic control system diagram.

¹ With reference to Eq. (1.9), $\varphi_e(s) = \Delta\varphi(s)$ except that $\varphi_e(s)$ is used to refer to a change without reference to the steady-state phase difference.

² Note: $-G_F G_R$ is the open-loop gain and C/R and E/R are closed-loop gains.

Thus from Eq. (2.22) we obtain

$$\frac{\varphi_{out}(s)}{\varphi_{in}(s)} = \frac{G(s)}{1 + G(s)} \quad (2.25a)$$

$$= \frac{K/s}{1 + K/s} = \frac{1}{1 + s/K}, \quad (2.25b)$$

while from Eq. (2.23) we obtain

$$\frac{\varphi_e(s)}{\varphi_{in}(s)} = \frac{1}{1 + G(s)} \quad (2.26a)$$

$$= \frac{1}{1 + K/s} = \frac{s}{s + K}. \quad (2.26b)$$

When we are interested in inputs or outputs at points other than those used above, we can rewrite the equations or, what is often simpler, relate the desired input or output to that shown in Fig. 2.4 by the transfer function of the segment that connects them.

Equation (2.25) represents a low-pass characteristic with a cutoff (-3 dB) frequency of $\omega = K$. This is analogous to the low-pass filter of Fig. 2.6a in which the voltages have been given names corresponding to the phases. Equation (2.26) says that the error phase has a high-pass characteristic, analogous to Fig. 2.6b. It is generally true, even in more complex loops, that the output has a low-pass relationship to the input while the error has a high-pass relationship. This reflects the fact that the error responds immediately to a change in the input while, and because, the output response is delayed.

Figure 2.7 is essentially the same as Fig. 2.4 except that the two $1/s$ blocks have been moved forward through the summer; since both integration and summing are linear operations, either can be done first. The input and output are now frequencies rather than phases, illustrating that the response of output frequency to input frequency is the same as the response of output phase to input phase. We can show the same

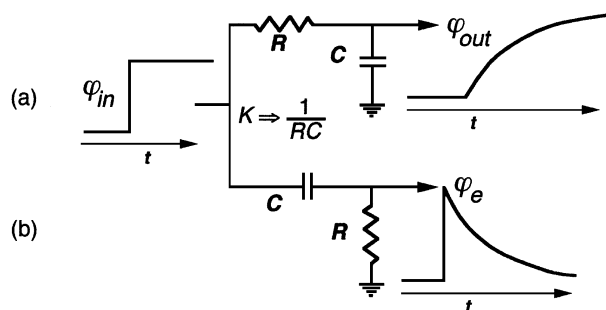


Fig. 2.6 Electrical analogs to the PLL. Voltages have names of analogous loop variables.

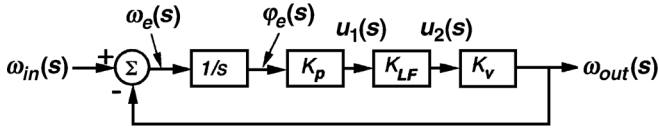


Fig. 2.7 Mathematical block diagram of the simple loop with frequency variables as input and output.

results as follows:

$$\frac{\varphi_{out}(s)}{\varphi_{in}(s)} = \frac{s\varphi_{out}(s)}{s\varphi_{in}(s)} = \frac{\omega_{out}(s)}{\omega_{in}(s)}. \tag{2.27}$$

A similar relationship holds for error responses.

It is apparent from Figs. 2.4 and 2.7 that

$$\varphi_{in} = \varphi_{out} + \varphi_e \tag{2.28a}$$

and

$$\omega_{in} = \omega_{out} + \omega_e \tag{2.28b}$$

and thus

$$1 = \frac{\omega_{out}}{\omega_{in}} + \frac{\omega_e}{\omega_{in}} = \frac{\varphi_{out}}{\varphi_{in}} + \frac{\varphi_e}{\varphi_{in}} \tag{2.28c}$$

so that the output can easily be obtained from error and vice versa in most cases.

Example 2.3 Modulation Response Use (2.25b) to obtain the results of Example 2.2. At $s = j4899/\text{sec}$, we have

$$\frac{\varphi_{out}}{\varphi_{in}} = \frac{1}{1 + j4899/1000} = 0.2 \angle -78.5^\circ$$

as before.

2.4 BODE PLOT

One of the most important tools with which we will work is a logarithmic plot of open-loop gain $|G_F G_R|$ and phase shift $\angle G_F G_R$ versus modulation frequency, $s \rightarrow \omega_m$. This plot, called the Bode plot, is shown for this simple loop in Fig. 2.8. Sometimes we will forgo the phase plot since the phase shift is implied by the slope of the gain.

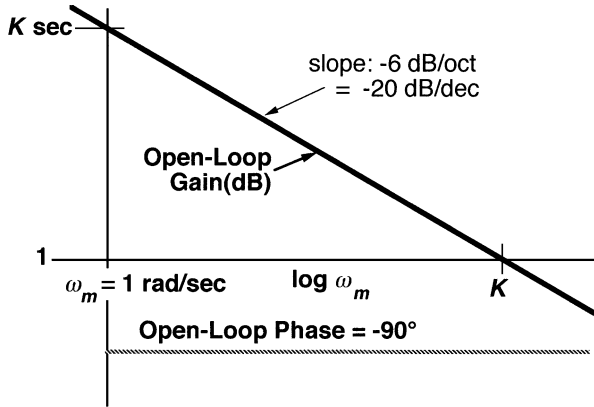


Fig. 2.8 Bode plot for the simple loop. Note that the abscissa axis is marked with the values of ω_m (e.g., K), rather than the actual distance, $\log(\omega_m \text{ sec/rad})$. Similarly, the ordinate axis is marked with gain rather than its logarithm or value in dB.

A linear increase in gain with ω_m originates in a term such as $s = j\omega$ and therefore implies both a gain that is increasing proportionally with frequency and a 90° phase shift. In general the segments of the Bode plot correspond to approximations of the form

$$F(\omega) = C(j\omega)^n = C\omega^n \angle n(90^\circ). \tag{2.29}$$

Thus, for example, when $n = -2$, the gain drops as the square of the frequency and the phase is -180° . The slope is often expressed logarithmically. The magnitude is described by

$$|F(\omega)| \Rightarrow 20 \text{ dB} \log_{10}(C\omega^n) = 20n \text{ dB} \log_{10}(\omega) + C'. \tag{2.30}$$

Note that we precede \log_{10} by 20 dB rather than 10 dB because $G_F G_R$ is a voltage gain. We can break the loop at a point where the variable is voltage and $G_F G_R$ will describe the open-loop gain there.

In each decade of frequency change the function changes by

$$|F(10\omega)/F(\omega)| \Rightarrow 20n \text{ dB} \log_{10}(10) = 20n \text{ dB} \tag{2.31}$$

so the slope is $20n$ dB per decade. Similarly, a two-to-one change in frequency corresponds to

$$|F(2\omega)/F(\omega)| \Rightarrow 20n \text{ dB} \log_{10}(2) = 6n \text{ dB} \tag{2.32}$$

or $6n$ dB per octave (an octave is a two-to-one change).

Figure 2.8 also illustrates that K is both the gain at $\omega = 1$ and the radian frequency at which the open-loop gain is unity. As has been shown, this is the frequency at which the closed-loop gain is -3 dB. At frequencies well below $\omega = K$, Eq. (2.25) can be approximated as

$$\frac{\varphi_{\text{out}}(s)}{\varphi_{\text{in}}(s)} = 1, \quad (2.33)$$

whereas at frequencies well above $\omega = K$ the equation is approximately

$$\frac{\varphi_{\text{out}}(s)}{\varphi_{\text{in}}(s)} = \frac{K}{s} = -j \frac{K}{\omega}. \quad (2.34)$$

This equals the open-loop transfer function, Eq. (2.24). Thus, at low frequencies, where the gain is high, the output follows the input faithfully, whereas, at high frequencies, where the gain becomes low, the loop is essentially open and the response is as if there were no loop (except that the low-frequency gain keeps it locked—otherwise none of these equations would be valid). At the loop corner frequency, $\omega = K$, it is easy to show from Eq. (2.25b) that the closed-loop gain is $1/\sqrt{2}$, or -3 dB, and the phase shift is -45° , just as in the case of the low-pass filter.

Similarly, the high-pass characteristic of Eq. (2.26) approaches unity at high frequencies and zero at low frequencies and, at $\omega = K$, also has a gain of -3 dB, but the phase shift is $+45^\circ$.

The more complex loops that we will study also have these general tendencies, but we will be able to shape their characteristics more exactly to our needs because of the increased number of parameters at our disposal.

2.5 NOTE ON PHASE REVERSALS

It is not uncommon that the transfer functions of the various blocks in Fig. 2.4 have negative signs. It is only necessary that the total phase shift around the loop not be altered. The negation of an even number of transfer functions has no important effect. In fact, when a balanced mixer is used, it will automatically operate on the proper slope to produce a correct number of phase reversals.

2.6 SUMMARY OF TRANSIENT RESPONSES OF THE FIRST-ORDER LOOP

Here we will summarize the transient responses of the first-order loop based on previous material and simple extensions of that material.

The unit step response is given in Eq. (2.12b) and shown in Fig. 2.2. Since a unit ramp t is the time integral of a unit step, we can obtain the time response to

TABLE 2.1 Summary of First-Order Loop Transient Responses

Input			Response ("sec" not shown)	
	$f(t)$	$F(s)$	Error	Output
Step	1	$\frac{1}{s}$	e^{-Kt}	$1 - e^{-Kt}$
Ramp	$\frac{t}{\text{sec}}$	$\frac{1}{s^2}$	$\frac{1}{K}(1 - e^{-Kt})$	$t - \frac{1}{K}(1 - e^{-Kt})$
Parabola	$\frac{t^2}{2 \text{ sec}^2}$	$\frac{1}{s^3}$	$\frac{1}{K}[t + \frac{1}{K}(e^{-Kt} - 1)]$	$\frac{t^2}{2} - \frac{1}{K}[t + \frac{1}{K}(e^{-Kt} - 1)]$

t by integrating the time response to the unit step. Similarly, the integral of t is a parabola $t^2/2$, and we can obtain the response to this by integrating the ramp response. Alternately, we can take the inverse Laplace transform of the transform obtained by multiplying the transfer function of Eq. (2.25b) by $1/s$, $1/s^2$, and $1/s^3$ to obtain step, ramp, and parabola responses. (The parabolic input is important because it represents the phase corresponding to a frequency ramp.)

The error response can be obtained by following the above process but starting with equations for the error response, Eq. (2.12a) or the Laplace transform of the error response, Eq. (2.26b).

Either the output response or the error response can be obtained from the other by subtraction according to Eq. (2.28c).

Table 2.1 summarizes these responses. If an input is a 5-Hz step, since that is 5 Hz times the given unit input, the output or error will be 5 Hz times the given output or error. If the input is a 2-rad or 2° step, then the output will be 2 rad or 2° times the given response. A 10-Hz step is also a 10-cycle/sec ramp (10 cycles \times t/sec), 10 cycles times the given ramp input, so the output phase will be 10 cycles times the given ramp response. If the input is a frequency ramp of 100 rad/sec/sec [(100 rad/sec) \times t/sec], the output frequency will be obtained by multiplying the ramp output response by 100 rad/sec. However, we can also obtain the output phase by considering the input as a parabolic phase [100 rad \times $t^2/(2 \text{ sec}^2)$] and obtaining the phase of the response by multiplying the given response by 100 radians.

CHAPTER 3

LOOP COMPONENTS

In this chapter we consider the characteristics of the major components of a phase-locked loop (PLL).

3.1 PHASE DETECTOR

We will consider several of the many types of phase detectors (PDs) that are in common use [Egan, 1981 and 2000, Chapter 5]. We begin with three types that employ logic circuits because they are both useful and relatively easy to understand. The most important type for the processing of analog signals, however, is the balanced mixer, so we will then study its operation in some detail.

3.1.1 Flip-Flop Phase Detector

Figure 3.1 illustrates the operation of the flip-flop PD. The symbol for an RS flip flop is shown in Fig. 3.1a. This type changes state in response to state change at either input, going to the $Q = 1$ state with a 1 input at A and to the $Q = 0$ state with a 1 input at B . When both inputs are 1, the state is undefined,¹ so the input waveforms are shown as very narrow pulses to avoid overlap of 1 states at the two inputs. The duration of the $Q = 1$ state depends on the time from the A input to the B input, as can be seen in

¹ The state is undefined in general but can be discerned for a particular realization of the circuit.

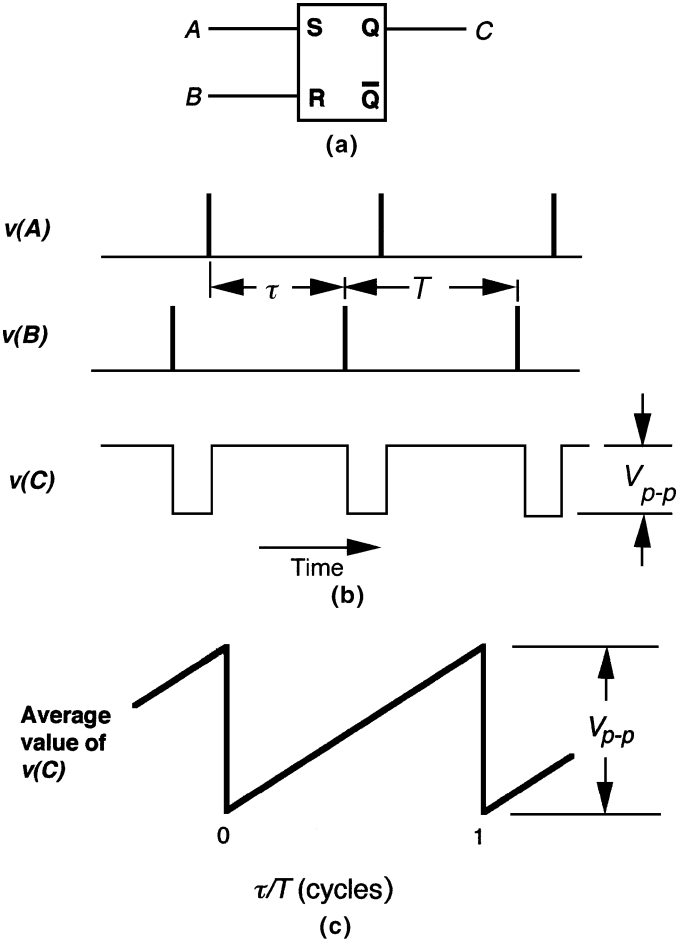


Fig. 3.1 Flip-flop phase detector: (a) the symbol for an RS flip-flop, (b) waveforms, and (c) phase detector characteristic.

Fig. 3.1*b*. Therefore, the duration of the 1 state, and thus the average output voltage, is proportional to the time difference between inputs and thus to their phase difference. The average value is the useful phase detector output. Higher-frequency components must be filtered out. The average voltage has a sawtooth-shaped characteristic versus phase, as shown in Fig. 3.1*c*. The linear phase range is 2π radians.

One useful realization employs a flip-flop with an edge triggered clock for one input. That input need not be narrow because the flip-flop changes states on one of the transitions of the input, but the other input must still be kept narrow.

3.1.2 Exclusive-OR Gate Phase Detector

Figure 3.2 illustrates the operation of the Exclusive-OR (ExOR) phase detector. The ExOR gate outputs 1 when the two input states differ and outputs 0 when they are

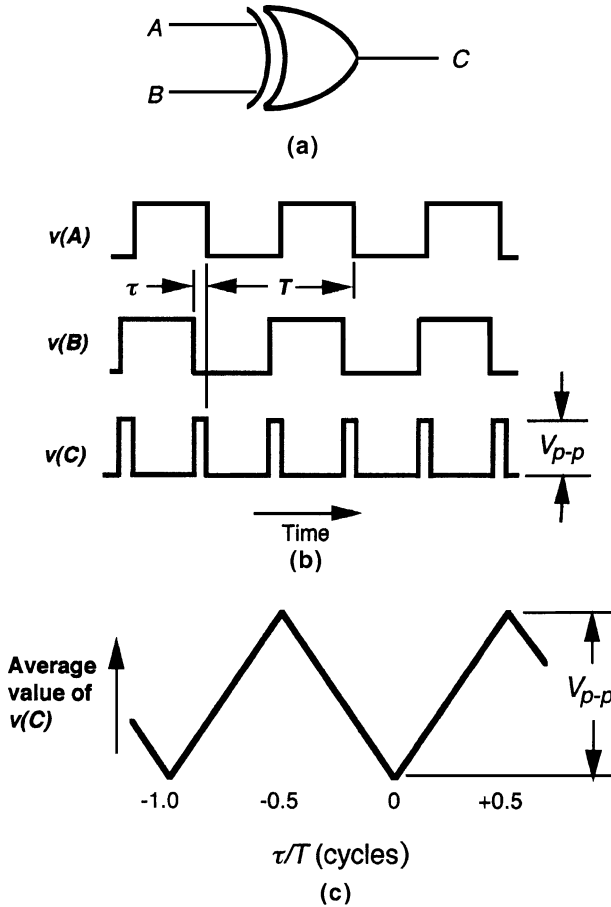


Fig. 3.2 Exclusive-OR phase detector: (a) the symbol for an Exclusive-OR gate, (b) waveforms, and (c) the phase detector characteristic.

the same. Since the part of the time when they are the same (or different) indicates their relative phase, the ExOR is useful as a phase detector. However, deviation from a perfect in-phase condition (or a perfect out-of-phase condition) in either an increasing or a decreasing direction of phase change, produces the same change in duty factor, and thus in average voltage. Therefore, the characteristic is triangular and each linear slope extends for only half a cycle, so the linear phase range is π radians.

3.1.3 Charge-Pump Phase Detector

One widely used type of PD has a characteristic like the flip-flop PD, as shown in Fig. 3.1c, except for one very important addition [Gardner, 1980; Egan, 1981, pp. 115–123; Egan, 2000, pp. 197–211]. It generates two different pulse outputs, one when the transition of $v(B)$ lags that of $v(A)$, as shown in Fig. 3.1b, and another when it

leads. (The initial definition of whether a particular transition is leading or lagging is arbitrary, but the loop will eventually establish a phase relationship such that one or the other of these outputs is being produced during lock.) The normal operating point is near-zero phase difference ($\tau = 0$). As τ increases, the characteristics represented by Figs. 3.1*b* and *c* occur. But, if τ decreases from 0, $v(C)$ stays low, and another output generates pulses that extend from the $v(B)$ transition to the $v(A)$ transition, thus beginning very narrow for small $-\tau$ and widening as $v(B)$ occurs earlier and earlier. Effectively we see the Q output of the flip-flop when $v(B)$ lags and the \overline{Q} output when it leads. If we just added these two outputs, we would get something like the characteristic of the Exclusive-OR PD in the vicinity of zero phase difference—not a useful operating point. But, by converting one of the outputs to a positive analog signal and the other to a negative signal and adding them, we obtain a characteristic that is linear, being positive or negative depending on the sign of the phase and passing through zero at zero phase difference.

The characteristic is further illustrated by Fig. 3.3. Here B initially leads A but has a lower frequency so eventually achieves a lagging position, as in Fig. 3.1. Here we see that $v(U)$ is similar to Fig. 3.1 when A lags, but, when A leads, its pulses disappear and are replaced by pulses at $v(D)$. Voltage v_p is produced by combining U and D using the appropriate sign [$V(U) - V(D)$]. Note how the average value of this waveform produces a ramp as the phase difference increases, illustrating the linear relationship between average voltage or current and phase. Moreover, this linear relationship has a range of $\pm 2\pi$ (Fig. 3.4).

These pulses are used in various ways to “pump charge” into the loop filter. A phase detector that provides this kind of output is often called a charge-pump phase detector and is commonly part of a phase-frequency detector (see Section 9.4). Sometimes an actual switched current source is used, as illustrated in Fig. 3.5*a*. Sometimes voltages are switched and converted to current by a resistor, as illustrated in Fig. 3.5*b*. In the first case the PD gain is $K_p = I/\text{cycle}$, since the average current would be $\pm I$ if the phase error were ± 1 cycle and would be 0 for 0 phase error, and the characteristic is linear between these points. Similarly, in the second case, if $V_x = 0$, the PD gain

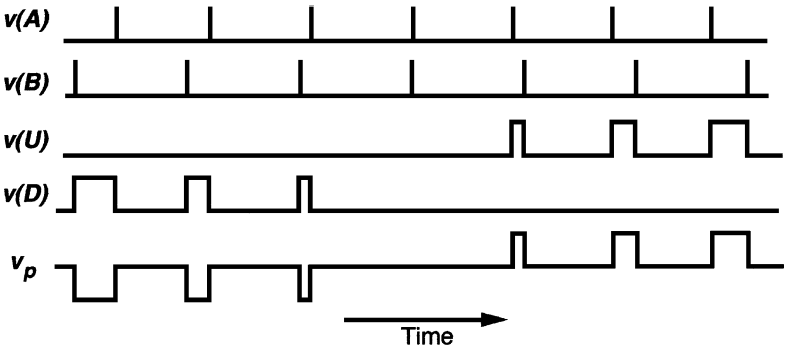


Fig. 3.3 Charge-pump waveforms.

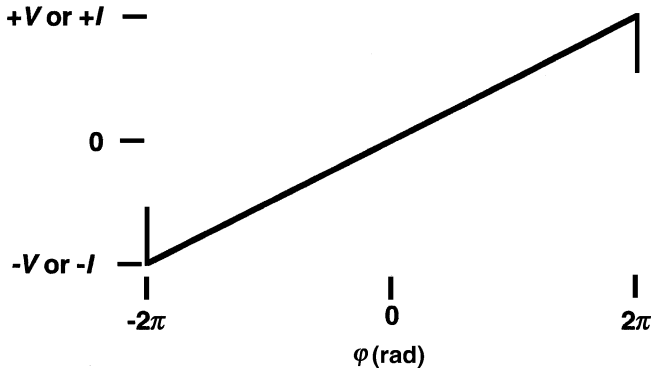


Fig. 3.4 Charge-pump phase detector characteristic.

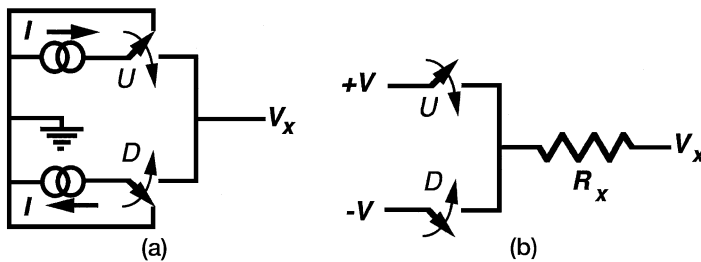


Fig. 3.5 Charge pumps: (a) switched current sources and (b) switched voltage sources.

is $K_p = V/\text{cycle}$. This circuit should not ordinarily be used where V_x does not stay midway between the two voltage sources (near ground for $\pm V$ sources as shown); otherwise the gain will differ for positive and negative phase errors and will vary with V_x . Gain K_p could also be written in current units as $(V/R_x)/\text{cycle}$. Resistor R_x can be considered part of the PD or part of the loop filter (but not both). Gains K_p and K_{LF} will both change if R_x is moved conceptually from one block to the other.

3.1.4 Sinusoidal Phase Detector

The PD with a sinusoidal response of voltage to phase was introduced in Chapter 1. This response can be obtained from any mixer.² The object of a mixer is to provide signals at the sum or difference of the two frequencies that are injected into it,³ both the sum and difference being provided simultaneously. In the case of a phase detector, for which the two input frequencies are the same, the difference frequency

² A mixer is a nonlinear device used for frequency conversion in radios, for example.

³ Occasionally, the desired output involves a harmonic of one of the input signals, but the results, for our present purposes, are fundamentally the same as would be obtained by multiplying the frequency of that signal and then mixing to obtain the sum or difference. The PLL theory is the same.

goes to zero and the output at this “low” frequency becomes the desired voltage that is proportional to phase. Note, in Fig. 1.5, that if $\varphi_{\text{in}} = \omega_{\text{in}}t$ and $\varphi_{\text{out}} = \varphi_{\text{out}}t$, then $\Delta\varphi = -\Delta\omega t$, and the output voltage will be a sinusoid at frequency $\Delta\omega$. Thus the same characteristic that describes the PD output also describes the output at the difference frequency.

Other signals will accompany the desired low- or zero-frequency output and must be suppressed by a low-pass filter. For a balanced mixer (BM) the strongest of these will normally be at the sum frequency. Of next greatest concern will be leakage at the input frequency. There will also be various harmonics of the input signals. In many applications the harmonics can be easily filtered. When undesired components are a problem, which is often the case in synthesizer applications, other types of phase detectors that have better rejection of these components are used.

The undesired output of most concern in a PD is that which is due to rectification of the input signals because it is a direct current (DC) signal and cannot be rejected by a filter without the filter also rejecting the desired signal. Of course, compensation is possible—the rectification term can be canceled—but variations with temperature and with input amplitude sometimes make this difficult. Balanced mixers greatly reduce this rectification component as well as some of the other undesired signals and therefore are generally used for PDs. However, simpler mixers also have phase detection capability, and they form the components of which the BMs are composed, so we begin by discussing the simpler types and work our way to the BM.

3.1.4.1 Phase Detection in a Simple Mixer. Figure 3.6 shows a simple mixer composed of a resistor and a diode. The nonlinearity of the diode produces the desired difference-frequency signal when two sinusoids are injected through the resistor. We will assume that the driving voltage v_i consists of two sinusoids, $A \cos \varphi_1(t)$ and $B \cos \varphi_2(t)$, where, in each case, $\varphi_i(t) = \omega_i(t) + \theta_i$. It is also likely that there would be a significant load, used to extract the desired energy from the diode, but the combination of the diode and a shunt load resistor can still be represented as a nonlinearity.

The important feature of the diode is that it is nonlinear. We use the general MacLaurin expansion to represent the nonlinearity of the circuit:

$$v_o = a + bv_i + cv_i^2 + dv_i^3 + \dots \quad (3.1)$$

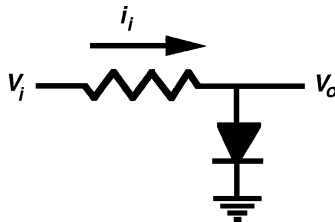


Fig. 3.6 Simple mixer.

with

$$v_i = A \cos \varphi_1(t) + B \cos \varphi_2(t). \tag{3.2}$$

The output component of greatest interest to us is that generated by the square-law term:

$$v_{oc} = c[A \cos \varphi_1 + B \cos \varphi_2]^2 \tag{3.3}$$

$$= c \left[\frac{A^2}{2}(1 + \cos 2\varphi_1(t)) + 2AB \cos \varphi_1(t) \cos \varphi_2(t) + \frac{B^2}{2}(1 + \cos 2\varphi_2(t)) \right] \tag{3.4}$$

$$= c \left\{ \frac{A^2 + B^2}{2} + AB[\cos(\varphi_1(t) + \varphi_2(t)) + \cos(\varphi_1(t) - \varphi_2(t))] + \dots \right\} \tag{3.5}$$

The second-harmonic terms in Eq. (3.4) are to be filtered and have not been shown in Eq. (3.5). The first term in Eq. (3.5) is the rectification term, the highly undesirable DC component that can be confused with the desired output. The next term comes from the middle term in Eq. (3.4) by a trigonometric identity. It contains the sum-frequency term, which, at phase lock, is at the same frequency as the second-harmonic terms and will be filtered with them,⁴ and the desired difference-frequency term that contains the phase information.⁵

3.1.4.2 Balanced Mixers. Since the rectification term cannot be filtered, it must be reduced by balancing. The process is illustrated in Fig. 3.7, which shows a single-balanced mixer, (SBM). The difference is taken between the outputs of two identical single-diode mixers. Each has the same two input signals except that the phase of one of them has been inverted at one diode relative to the other. Since the rectification

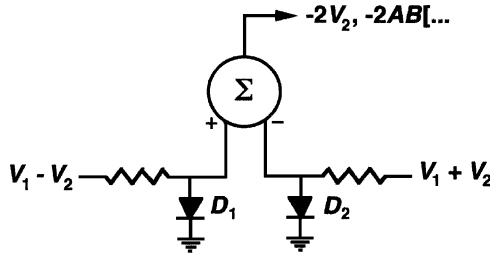


Fig. 3.7 Balanced mixer concept.

⁴ Before filtering, the sum frequency term is often the largest of the undesired terms, having an amplitude equal to the desired difference frequency term.

⁵ Direct current and sum and difference frequency outputs are also available from higher even-ordered terms in Eq. (3.1). Their strengths, relative to those produced by the second-order nonlinearity, are dependent on the details of the nonlinearity and the strengths of the signals, but their effects are essentially the same as described here.

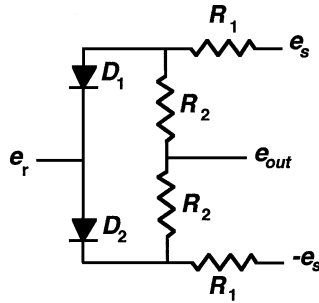


Fig. 3.8 Possible balanced mixer realization.

term in Eq. (3.5) is not affected by the signs of A or B , it will be the same for both diodes and thus will cancel in the subtraction. The desired term is affected, however, being proportional to the product of A and B and therefore changing signs with either of them. Thus, the desired difference-frequency term from each diode will have an opposite sign, and subtraction of the two outputs will cause the desired terms to add.

A more practical realization of the SBM is shown in Fig. 3.8. The results are similar, with e_s in Fig. 3.8 injected into the two diodes in phase (positive polarity to the anodes), acting as v_1 in Fig. 3.7, and e_r being injected out of phase (to the anode of one and the cathode of the other), acting as v_2 . Resistor R_2 takes the difference of the diode voltages (averaging the anode voltage of one with the cathode voltage of the other). A still more practical realization is shown in Fig. 3.9 where radio frequency (RF) transformers are use to accomplish the same summations.

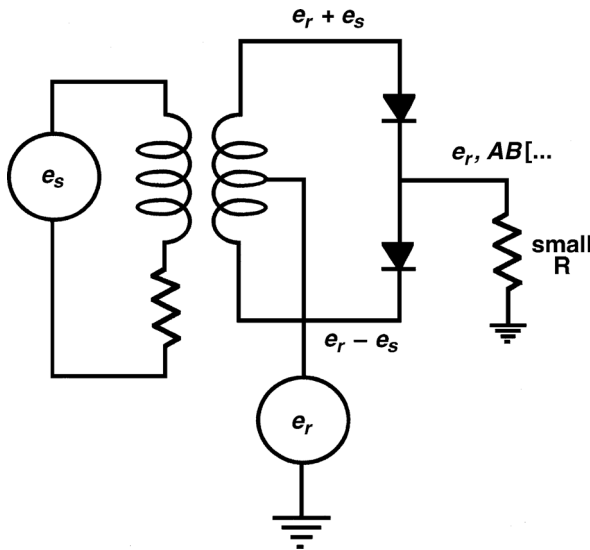


Fig. 3.9 Practical singly balanced mixer.

It is common, in frequency mixing applications, to operate with a large, local oscillator (LO) input and a weak “signal” input. In that case the LO is often sufficiently strong that the diodes may be considered to be switches that change between conduction and nonconduction as the LO signal reverses polarity. Thus, in the circuit of Fig. 3.9, if e_s were the strong input, the diodes could be considered to connect and disconnect the output load to e_r under control of e_s . With e_s driving current through the diodes, assuming a well-balanced circuit, the center tap of the transformer would be at the same voltage as the center point between diodes, thus passing e_r to the output. When the polarity of e_s would reverse, the back-biased diodes would disconnect the load. Thus e_r would be passed to the output only half of the time, with the exact part of its waveform that is transmitted depending on the phase relationship between e_r and e_s . This is illustrated in Fig. 3.10, where e_r is multiplied by 1, starting at a phase Δ and ending one-half cycle later; Δ would here be the phase difference between the two signals. After low passing to eliminate the high-frequency outputs from the PD, the DC or low-frequency component would remain. This would be proportional to the time average over a cycle:

$$V_{\text{avg.}} = \frac{1}{2\pi} \int_{\Delta}^{\Delta+\pi} \sin \theta d\theta = \frac{1}{2\pi} [\cos \Delta - \cos(\Delta + \pi)] = \frac{1}{\pi} \cos \Delta. \quad (3.6)$$

Again we see the cosinusoidal output as a function of the phase difference between the two signals. Similar arguments could be made if e_r were stronger.

While a large difference between signal strengths is common for frequency mixers, this is not the best operating condition for PDs. This is because the desired output is proportional to AB while the highly undesired rectification component is proportional to $(A^2 + B^2)$. As one input drops much below the other, the desired output will fall proportionally to this dropping input, but the undesired component will become approximately proportional to the square of the unchanging stronger input and

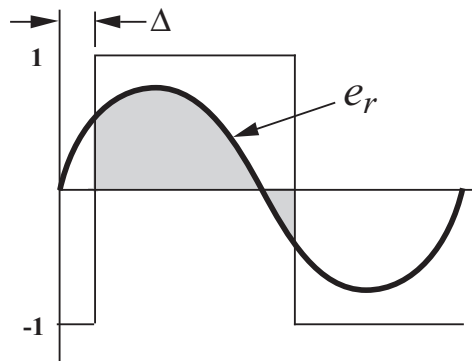


Fig. 3.10 Averaging of mixer output. The shaded area is used for a singly balanced mixer. In a doubly balanced mixer, the area under the rest of the sinusoid is subtracted from this.

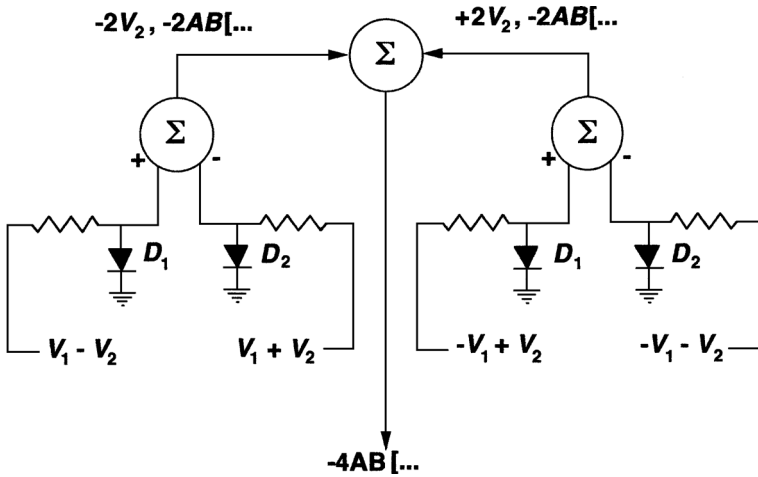


Fig. 3.11 Doubly balanced mixer.

will not drop appreciably. Therefore, it is optimum for a PD to have the two inputs approximately of equal strength ($A \approx B$).

While the SBM balances out the e_s term, the e_r term is not balanced out and appears quite strongly in the output. To balance it also, doubly balanced mixers (DBMs) are common. A DBM is shown conceptually in Fig. 3.11 and schematically in Fig. 3.12. It is necessary that the BM selected for a PD application have a DC output, like that shown in Fig. 3.12. Usually one port will be DC connected and, while interchanging ports may be acceptable in some frequency mixing applications, only the DC-connected port can be chosen as the output port for a PD.

If one signal is much stronger than the other, it can be thought of as switching the polarity of the weaker signal as it is seen at the output. This is apparent from

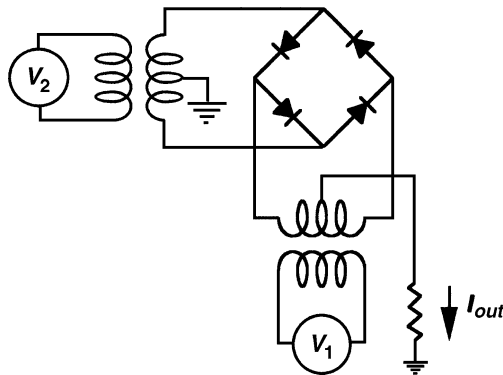


Fig. 3.12 Practical doubly balanced mixer.

Fig. 3.12; the strong signal will cause current to flow through one half of the diode bridge at a time, depending on its polarity. Since each half of the bridge is connected to a different polarity of the weaker signal, the voltage delivered at the output will be proportional to the weaker signal multiplied by the polarity of the stronger. We could easily repeat Eq. (3.6) for this case and come up with a similar result. If the weaker signal is a square wave, rather than a sinusoid as assumed until now, the DBM can be seen to perform an ExOR function; when the two signals have the same polarity (state), one output occurs, and when they have different polarities, the other occurs. When the two input signals are sinusoidal but have similar amplitudes, the ExOR's triangular characteristic is also approached, [Blanchard, 1976, p. 14], as illustrated in Fig. 3.13.

The sinusoidal shape is applicable under the usual frequency mixing conditions where one signal (the LO) is at some relatively high level appropriate to the particular mixer design and the other signal is relatively weak. Under such conditions K_p is relatively independent of the strength of the stronger signal but is directly proportional to the amplitude of the weaker signal. The relative independence of K_p from the stronger signal is related to the higher-order terms in Eq. (3.1).

3.1.4.3 Analog Multipliers. The desired term in Eq. (3.5) is the result of multiplication of the two input sinusoids. An analog multiplier can perform this function without producing the additional undesired components that are created when the even-order terms in the expansion of the diode characteristic are used. Such multipliers are available in integrated circuit (IC) form [Gray and Meyer, 1977, pp. 561–575]. Their maximum operating frequency is, however, limited, whereas that of the diode mixer is practically unlimited.

Analog multipliers generally make use of the diode current–voltage characteristic. The diode current can be expressed as

$$I = I_s(e^{v/V_T} - 1). \quad (3.7)$$

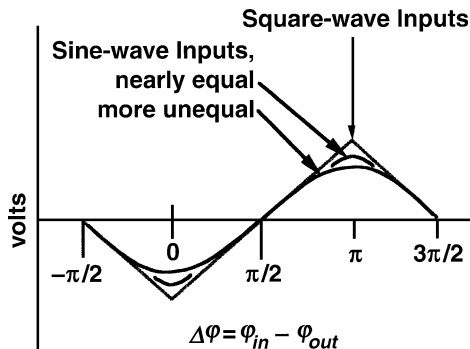


Fig. 3.13 PD responses for various drive levels, DB mixer.

For $I \gg I_s$ this implies

$$\ln \frac{I}{I_s} \approx \frac{v}{V_T}. \tag{3.8}$$

As an example of how this equation might be used for multiplication, consider a circuit that sums the voltage across two diodes, each with a different current. Then the total voltage can be written

$$v = v_1 + v_2 = V_T \left[\ln \frac{I_1}{I_{s1}} + \ln \frac{I_2}{I_{s2}} \right] = V_T \left[\ln \frac{I_1 I_2}{I_{s1} I_{s2}} \right]. \tag{3.9}$$

Note that an expression has been obtained that can contain the product of two signals represented by the currents I_1 and I_2 , and v can then be linearized by various means. For example, a current might be passed throughout a third diode and controlled in such a manner that the diode voltage would equal v . Then the current could be written, from Eqs. (3.7) and (3.9), as

$$\frac{I}{I_s} = \left(\frac{I_1}{I_{s1}} \frac{I_2}{I_{s2}} - 1 \right), \tag{3.10}$$

thus providing a signal, in the form of the current I , that contains the product of two other signals, I_1 and I_2 .

3.1.4.4 Integrated Circuit Doubly Balanced Mixer. To understand the IC BM we begin with the bipolar transistor differential pair in Fig. 3.14. For small

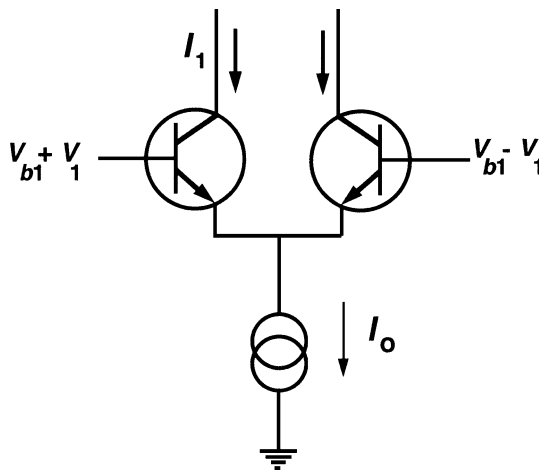


Fig. 3.14 Basic circuit in the IC doubly balanced mixer.

excursions of the base voltage, the collector current I_1 can be expressed as

$$I_1 = \frac{I_0}{2} + \frac{v_1}{r_e} = \frac{I_0}{2} \left(1 + \frac{v_1}{V_T} \right), \quad (3.11)$$

where r_e is the differential emitter resistance, obtained by differentiating I with respect to v in Eq. (3.7), $r_e = V_T/I$. The objectives will be to make one signal proportional to v_1 , which is easily done, and the other proportional to I_0 , a little more complicated, and to eliminate all but the product term. The details of how this is accomplished are given in Appendix 13.A. Here we will note that a circuit such as shown in Fig. 3.14 can be used to generate the current I_0 for a second such circuit.

3.2 VOLTAGE-CONTROLLED OSCILLATOR (VCO)

The voltage-controlled oscillator (VCO) is an essential part of every PLL. While we will not study oscillator design [Manassewitsch, 1987 pp. 379–407; Rhode, 1997, pp. 197–204] in this text, we will here consider the various types of oscillators that are useful in PLLs and discuss their properties.

The VCO may, in fact, be a current-controlled oscillator, an ICO. There is no difference to the extent that the response of the loop is concerned. At low frequencies a current-controlled astable multivibrator may be used to produce square waves at a frequency that is proportional to the control current over a wide range. Such a circuit is illustrated in Fig. 3.15.

When a negative transition occurs at the collector of Q_2 , it is coupled to the base of Q_1 which it turns off. The cross-coupling capacitor at the base of Q_1 then recharges through the controlled current source Q_3 . When it recharges to the point where the base of Q_1 becomes forward biased, the collector of Q_1 moves negative, causing the base of Q_2 to go negative and its collector to go positive. This reinforces the positive-going waveform at the base of Q_1 . Thus a regenerative state change occurs. The time between such transitions is the time required for the base voltage to rise to a diode drop above ground. The rate of change is proportional to the charging current, $I_c = I_3 = I_4$, and the time required for a transition is therefore $T = C \Delta V / I_c$. Thus the frequency, $f = 2/T$, is directly proportional to I_c , which can be controlled either by a voltage or a current at the tuning control input, but it will be somewhat more linear relative to a tuning current, I_T . Since both I_T and I_c pass through a resistor and a diode with a common voltage across them, they tend to be proportional, but the tuning voltage is less so because of the base-to-emitter voltage of Q_3 and Q_4 .

While the circuit is conceptually simple and good linearity of frequency versus control current can be obtained, it has relatively poor spectral purity. That is, its phase (and therefore its frequency) is modulated by noise, producing a broader spectral line. In addition, this simple version has a potential startup problem—if both of the cross-coupled transistors should be saturated at the same time, they will remain so—which may require additional circuitry to solve.

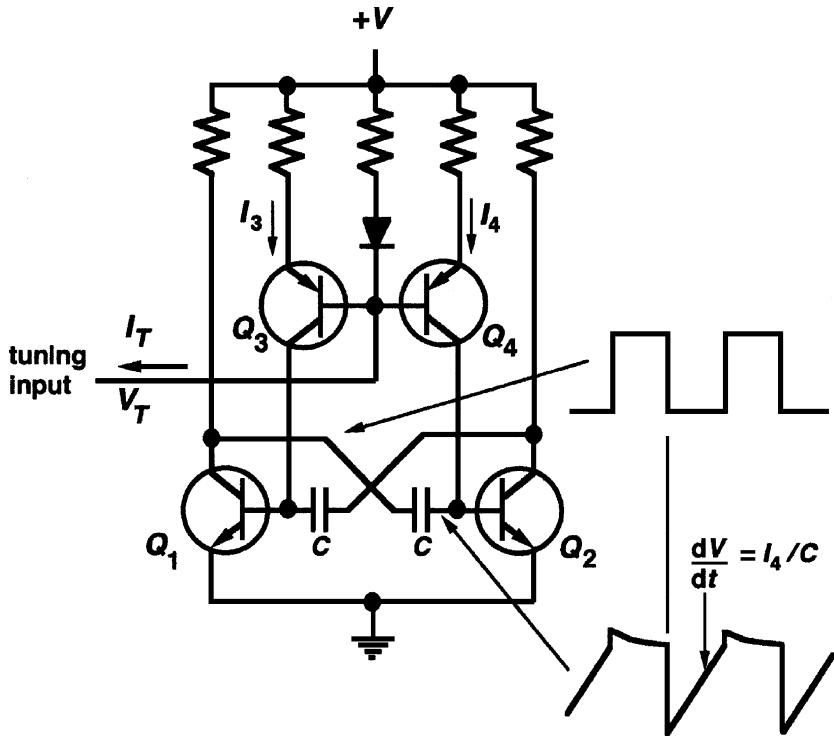


Fig. 3.15 Astable multivibrator. Auxiliary circuitry may be needed to ensure that both transistors do not saturate and cause the oscillator to lock up.

A second form of ICO, used at microwave frequencies, is the YIG-tuned oscillator. Like the astable, this has an inherently linear tuning characteristic and can easily cover more than an octave of frequency range. In addition, the tunable resonator, a YIG (yttrium iron garnet) sphere, has a very high Q . High oscillator Q is necessary for good spectral purity, and this requires high component Q . The resonant frequency of the sphere is proportional to the strength of the magnetic field in which it is immersed, and that is in turn proportional to the control current. Disadvantages are that the magnetic circuitry is relatively bulky and susceptible to stray magnetic fields. Isolation in the form of bulky shielding and physical separation from potential interfering fields may be necessary to prevent frequency modulation of the oscillator.

Solid-state oscillators can be tuned by control of bias (supply) voltages that alter active-device capacitances or, more commonly, by changing bias on a varactor, a diode designed for use of its voltage-variable junction capacitance, incorporated into the tuning circuit. The varactor-tuned oscillator sometimes has a quite nonlinear tuning characteristic, especially when tuned over a wide frequency range. A linearizer circuit can be used in series with the VCO if linearity is important. Hyperabrupt junction varactors can give a much more linear tuning curve than ordinary varactors, but they tend to have lower Q .

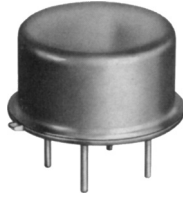


Fig. 3.16 VCOs may be integrated or separately packaged. This TO-8 can is 0.27 in. high by 0.45-in. diameter. (Courtesy of VARI-L Company, Inc.)

A representative VCO package and tuning curve are shown in Figs. 3.16 and 3.17. Figure 3.18 shows a representation of a VCO circuit. The output is fed back through R_{FB} to the shunt resonator “tank” circuit consisting of L_0 , C_0 , and C_{TUNE} (assuming C_{BLOCK} is large enough to be ignored as a frequency-determining element). A tap on the inductor L_0 (or sometimes a capacitive divider) sends a portion of the resonator voltage to the active element G by which it is amplified to provide the output, thus completing the loop. Oscillation occurs at the resonant frequency of the tank; there the voltage across the tank, and at the tap, is in phase with the output voltage. The gain G compensates for the loss in the feedback circuit and the inductor tap ratio to give unity gain around the loop with zero phase shift at the resonant frequency. Changing the tuning voltage V_{TUNE} changes the depletion capacitance of the varactor diode

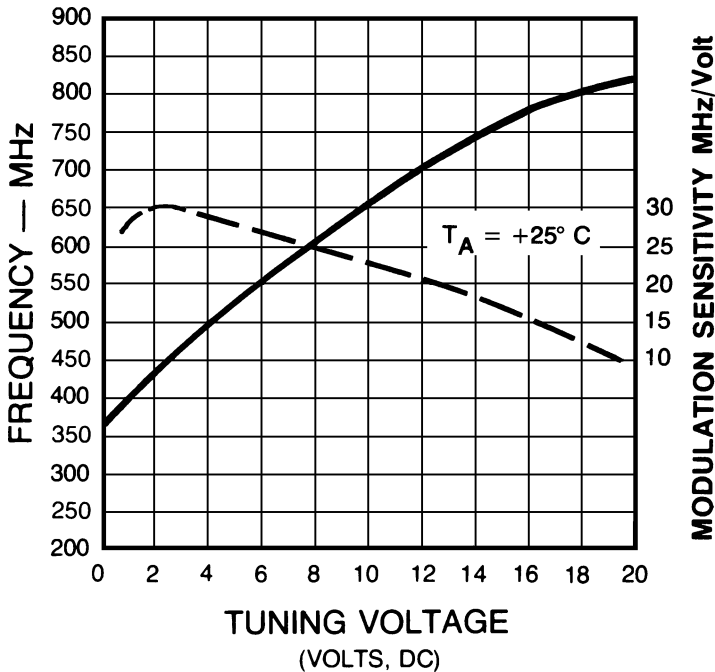


Fig. 3.17 VCO tuning curve (solid line). (Courtesy of VARI-L Company, Inc.)

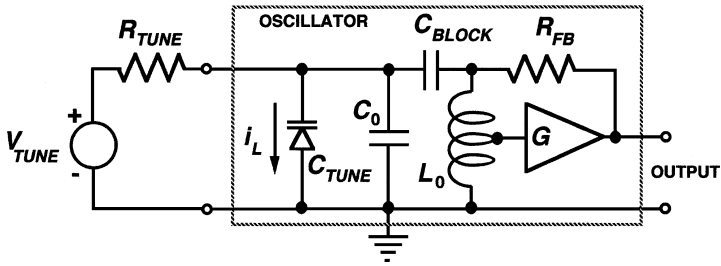


Fig. 3.18 VCO circuit representation.

C_{TUNE} , thus modifying the resonant frequency of the tank circuit and the frequency of oscillation.

Some impedance is needed at R_{TUNE} to prevent the tank circuit from being shorted, but, if it is too large (e.g., $\geq 1 \text{ k}\Omega$; experiment or consult the manufacturer), the noisy leakage current i_L from the varactor diode junction will produce a voltage across R_{TUNE} that will modulate the oscillator frequency and can thereby produce significant additional phase noise.

Other methods of oscillator tuning also exist, particularly in microwave tubes, klystrons, magnetrons, and backward wave oscillators (BWOs). Where the required range is only a few hundred parts per million or less, a voltage-controlled crystal oscillator (VCXO) can be used. This provides many of the advantages of the crystal oscillator, high Q and long-term frequency stability, although to a degree that is lessened by the voltage control. Voltage control is achieved by modifying the resonant frequency, which is primarily determined by the crystal, by varying the capacitance of a varactor that is also part of the resonant circuit.

Other high- Q circuits also exist, dielectric resonator and cavity oscillators at microwave frequencies and surface-acoustic-wave (SAW) resonator oscillators in the region between these and the highest crystal oscillator frequencies ($\approx 100 \text{ MHz}$). These oscillators, like the VCXO, are usually tuned by perturbing a high- Q resonator with a varactor, and, like the VCXO, their tuning range is therefore limited; however, not, on a relative basis, to the extent of the VCXO.

In ICs, the ring oscillator is a very common type of VCO. It consists of several amplifiers connected in a series loop, or ring, with a net phase inversion around the loop. The speed at which the (approximately square) waveform travels around the ring is proportional to the delays in the amplifiers; these delays, and thus the frequency of the oscillator, can be controlled by circuit bias voltage. The ring oscillator is very useful because of its ease of integration but it tends to be noisy because, much like the multivibrator, it has no high- Q resonator.

3.3 LOOP FILTER

We have considered the operation of the loop with a simple amplifier as the loop filter. Now we will consider the filters that give us more flexibility in controlling the characteristics of the PLL. These will be of two types, passive and active.

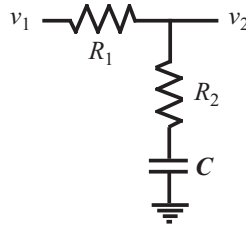


Fig. 3.19 Passive lag-lead loop filter.

3.3.1 Passive Loop Filter

A general form of the passive filter that we will discuss is shown in Fig. 3.19. Here u_1 and u_2 are voltages and the transfer function is

$$F(s) = \frac{v_2(s)}{v_1(s)} = \frac{R_2 + 1/(Cs)}{R_1 + R_2 + 1/(Cs)} = \frac{1 + R_2Cs}{1 + (R_1 + R_2)Cs} \quad (3.12)$$

$$= \frac{1 + \tau_2 s}{1 + \tau_1 s} = \frac{1 + s/\omega_z}{1 + s/\omega_p}, \quad (3.13)$$

where

$$\tau_1 = 1/\omega_p = (R_1 + R_2)C \quad (3.14)$$

and

$$\tau_2 = 1/\omega_z = R_2C. \quad (3.15)$$

The gain and phase of $F(\omega)$ are shown in Fig. 3.20.

If $R_2 \Rightarrow 0$, this becomes a low-pass filter (Fig. 3.21) with transfer function

$$F(s) = \frac{1}{1 + R_1Cs} = \frac{1}{1 + s/\omega_p}, \quad (3.16)$$

and response as shown in Fig. 3.22.

3.3.2 Filters Driven by Current Sources

A phase detector that acts as a current source (e.g., Fig. 3.5a) drives a filter whose transfer function is the ratio of its output voltage to its input current, a transimpedance. This transimpedance is just the impedance of a network connected to ground, Z_F in Fig. 3.23a. Figures 3.23b–g are various implementations of Z_F . While Z_F is passive,

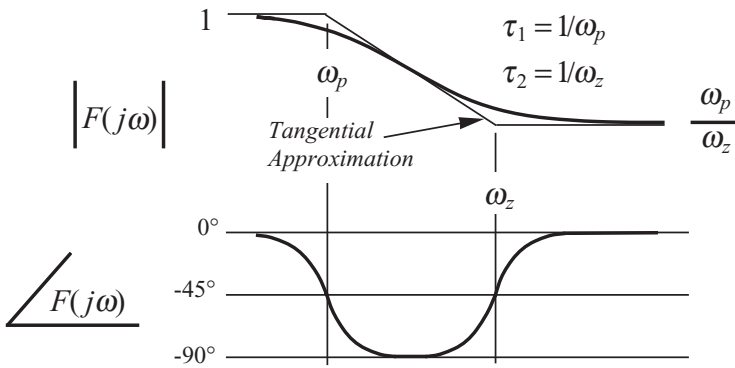


Fig. 3.20 Gain and phase of the passive loop filter vs. frequency, with tangential gain approximation shown.

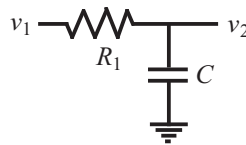


Fig. 3.21 Low-pass loop filter.

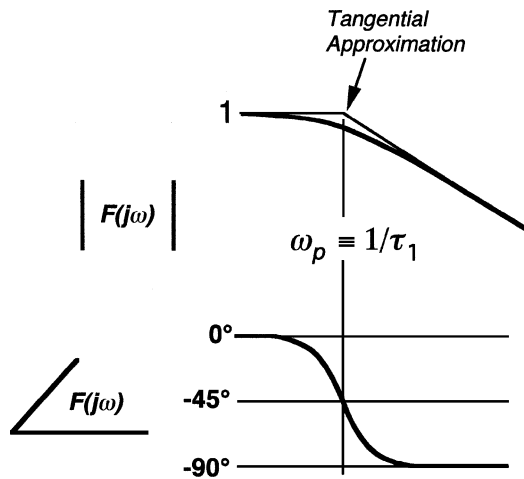


Fig. 3.22 Gain and phase of the low-pass loop filter vs. frequency, with tangential gain approximation shown.

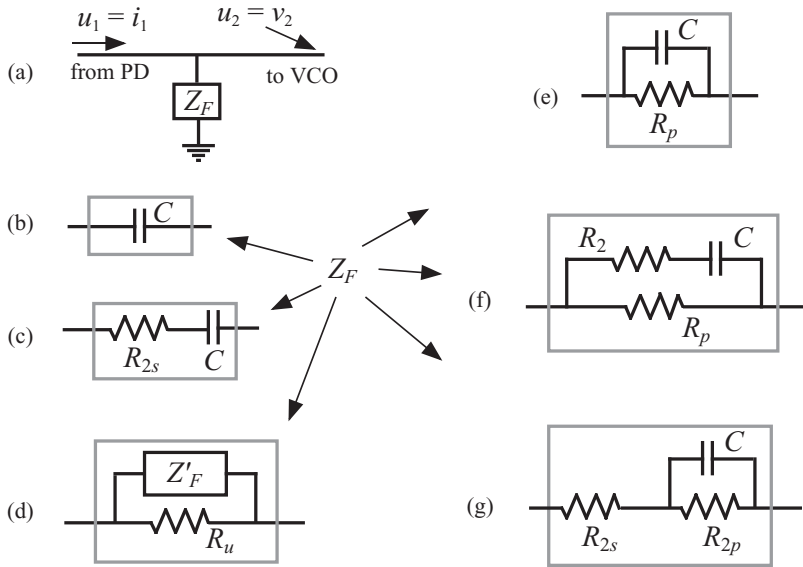


Fig. 3.23 Transimpedance loop filter. The generic filter is shown in part *a*, while particular realizations of Z_F are shown in parts *b–g*.

the transimpedance filter cannot be realized without the active circuit that comprises the current source in the phase detector, so we will reserve the term “passive filter” for those previously considered. The transimpedance filter can also implement Eq. (3.13), but, in addition to possibly providing gain, a wider range of values for ω_p and ω_z can be more easily obtained.

The transfer function of the transimpedance filter is

$$K_{LF}F(s) = \frac{u_2(s)}{u_1(s)} = \frac{v_2(s)}{i_1(s)} = Z_F(s). \tag{3.17}$$

3.3.2.1 Integrator. The transfer function with the impedance in Fig. 3.23*b* is

$$K_{LF}F(s) = Z_F(s) = \frac{1}{Cs}, \tag{3.18}$$

which represent integration, so this is called an integrator. This is a function that cannot be obtained with a passive filter. The pure integrator is not a practical filter because it produces a phase shift of -90° at all frequencies and, as we will see, this would cause the loop to be unstable. Therefore, a resistor is commonly added in series with the capacitor.

3.3.2.2 Integrator and Lead Filter. The integrator will revert to an amplifier at high frequencies if a resistor R_2 is placed in series with C . This filter (Fig. 3.23c) has

$$Z_F = R_2 + \frac{1}{Cs} = \frac{1 + R_2Cs}{Cs} \quad (3.19)$$

so that Eq. (3.17) becomes

$$K_{LF}F(s) = \frac{1}{C} \frac{1 + s/\omega_z}{s}, \quad (3.20)$$

where

$$1/\omega_z \equiv \tau_2 = R_2C. \quad (3.21)$$

This is an integrator-and-lead filter, the phase lead being produced by the zero at $\omega_z = 1/\tau_2$, the same frequency that is given for the passive filter by Eq. (3.15). Its response is shown in Fig. 3.24.

3.3.2.3 Lag (Low-Pass) Filter. A resistor R_p placed across the capacitor (Fig. 3.23e) limits the gain to a constant at low frequencies and produces a low-pass filter,

$$Z_F = \frac{1}{Cs + \frac{1}{R_p}} = \frac{R_p}{1 + R_pCs}, \quad (3.22)$$

$$K_{LF}F(s) = K_{LF} \frac{1}{1 + s/\omega_p}, \quad (3.23)$$

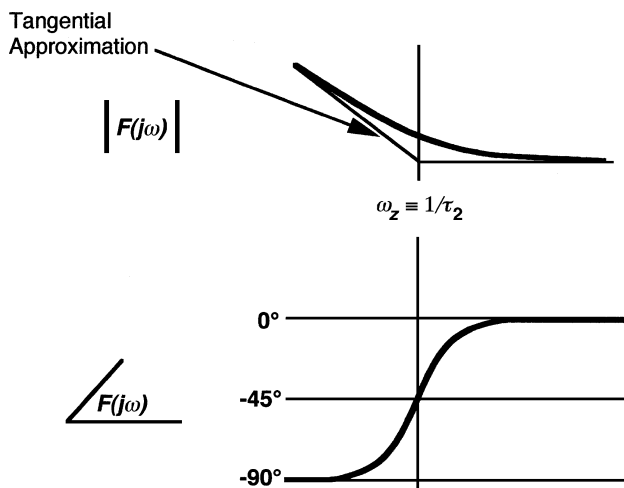


Fig. 3.24 Gain and phase of the integrator-and-lead filter vs. frequency with tangential gain approximation.

where

$$K_{LF} = R_p \quad (3.24)$$

and

$$1/\omega_p \equiv \tau_1 = R_p C. \quad (3.25)$$

The resulting low-pass response of $F(s)$ is as shown in Fig. 3.22 but is now multiplied by K_{LF} .

3.3.2.4 Lag-Lead Filter. We can obtain the lag-lead filter by inserting resistance in series with the capacitor of Fig. 3.23e, using either Fig. 3.23f or 3.23g. The result is $K_{LF}F(s)$, where K_{LF} is given by Eq. (3.24) and $F(s)$ is given by Eq. (3.13),

$$K_{LF}F(s) = K_{LF} \frac{1 + s/\omega_z}{1 + s/\omega_p}. \quad (3.26)$$

For Fig. 3.23f, the corners are at

$$1/\omega_p \equiv \tau_1 = (R_2 + R_p)C \quad (3.27)$$

and

$$1/\omega_z \equiv \tau_2 = R_2 C. \quad (3.28)$$

For Fig. 3.23g, they are at

$$1/\omega_p \equiv \tau_1 = R_{2p} C \quad (3.29)$$

and

$$1/\omega_z \equiv \tau_2 = \frac{R_{2s} R_{2p}}{R_{2s} + R_{2p}} C \quad (3.30)$$

and R_p in Eq. (3.24) is given by

$$R_p = R_{2s} + R_{2p}. \quad (3.31)$$

3.3.2.5 DC Transimpedance. There will always be some stray resistance across Z_F due to a combination of component leakage, board or substrate leakage, and finite current-source output impedance. Leakage current is likely to also have a voltage-independent component and the impedance of the current source can change dynamically, depending on its state (on or off). Nevertheless, we represent an undesired

effective resistive component by R_u in Fig. 3.23*d*, where Z'_F represents one of the other configurations of Z_F . It limits the magnitude of the transimpedance to

$$|K_{LF}F(s)| \leq R_u \quad (3.32)$$

and lowers the effective value of R_p in Figs. 3.23*e* and 3.23*f*. It turns the integrator in Fig. 3.23*b* into the low-pass of Fig. 3.23*e* and turns the integrator-and-lead of Fig. 3.23*c* into the lag-lead of Fig. 3.23*f*, moving the pole at zero frequency to

$$\omega_p = 1/(R_u C) \quad (3.33)$$

in the first case and to $1/([R_2 + R_u]C)$ in the second. Thus, the integrator and integrator-and-lead are not attainable in practice but they are, nevertheless, valuable configurations for consideration when the first pole approaches zero frequency and we are interested in performance well above its corner frequency. A value of R_u that is significantly lower than the stray value can be used to gain more control over the frequency of this first pole.

Example 3.1 Effect of Current Source Impedance

- a. A current source has an effective output resistance of $10 \text{ M}\Omega$. How large must C in Fig. 3.23*b* or 3.23*c* be so that the undesired pole created by this finite impedance is below 1 rad/sec ?

In Fig. 3.23*d*, $R_u = 10 \text{ M}\Omega$. This converts the integrator in Fig. 3.23*b* into the low-pass filter in Fig. 3.23*e* with pole given by Eq. (3.25) as

$$\omega_p = \frac{1}{10 \text{ M}\Omega \times C} < 1 \text{ rad/sec}, \quad C > 1.1 \text{ }\mu\text{F}.$$

The filter in Fig. 3.23*c* becomes the filter in Fig. 3.23*f* with pole given by Eq. (3.27) but this becomes approximately Eq. (3.25) when $R_2 \ll R_p$, giving the same answer.

This is a very large capacitor for some circuit realizations and indicates why the current-source impedance can be an important consideration.

- b. If the current source output impedance is known to be between $1 \text{ M}\Omega$ and $5 \text{ M}\Omega$, how large should be C in Fig. 3.23*e* in order that the pole frequency not vary more than $\pm 2\%$ from 1 kHz due to the uncertainty?

Now R_u in Fig. 3.23*d* is effectively in parallel with R_p in Fig. 3.23*e*. The conductance of the current source, $1/R_{pd}$, varies from 10^{-6} to $0.2 \times 10^{-6} \text{ S}$, that is $(0.6 \pm 0.4) \times 10^{-6} \text{ S}$. The total conductance,

$$\frac{1}{R} = \frac{1}{R_p} + (0.6 \pm 0.4) \times 10^{-6} \text{ S},$$

must vary less than 2% so

$$0.4 \leq 0.02(0.6 + 10^6 \Omega/R_p)$$

$$10^6 \Omega/R_p \geq 20 - 0.6 = 19.4$$

$$R_p \leq 10^6 \Omega/19.4 = 52 \text{ k}\Omega$$

$$C = \frac{1}{R_p \omega_p} \geq \frac{1}{52 \times 10^3 \Omega \times 2\pi \times 10^3 \text{ rad/sec}} = 3060 \text{ pF}$$

3.3.3 Active Filters

An active filter employs an operational amplifier (op-amp) in the configuration shown in Fig. 3.25. Besides the inverting input shown at v in Fig. 3.25, op-amps have a noninverting input. The output depends on the difference between the two inputs. For now, we assume zero at the noninverting input.

Ideally, the gain $|G_a|$ of the op-amp approach infinity. As a result of the high negative feedback, the voltage at v is held near zero and the impedance at the input port is such that all of the current flowing through R_1 also flows through Z_F . This current produces the same voltage across Z_F as in the transimpedance filter, but here the current enters at a virtual ground so the output voltage has the opposite polarity. For the sake of uniformity, we will ignore the sign reversal when writing the filter transfer function, effectively assuming an additional series inverting amplifier with unity gain. In practice, we would usually reverse the sign of K_p or K_v rather than add the inverting amplifier just to change the sign.

If the active filter is driven by a current source ($u_1 = i_1$), the transfer function is a transimpedance as described above, except for the sign. There is no need for R_1 in this case and, without it, the finite impedance of the current source is unimportant because the voltage across it is fixed.

Usually the active filter is driven by a voltage ($u_1 = v_1$). The current into the impedance is then

$$i_1 = v_1/R_1 \quad (3.34)$$

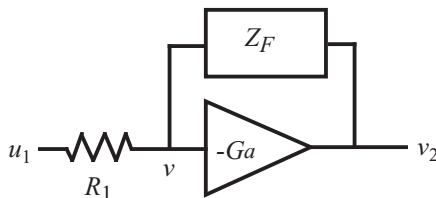


Fig. 3.25 Active loop filter. Particular realizations of Z_F are shown in Fig. 3.23b–g.

and

$$[K_{LF}]_{\text{Active}} = [K_{LF}]_{\text{Transimpedance}}/R_1. \quad (3.35)$$

That is, all the transfer functions of the previous section are just divided by R_1 . In the process, they lose all units.

3.3.3.1 Op-Amp Considerations. The approximation of infinite gain fails when the gain required from the active filter exceeds the gain available from the op-amp—that is, when $|K_{LF}F(s)|$ becomes greater than $|G_a|$. At that point, the gain is limited to the smaller of the two and $K_{LF}F(s) \Rightarrow G_a$. Equivalently, the op-amp acts like an ideal (infinite-gain) op-amp with the feedback impedance in Fig. 3.23*d* and $R_u = R_1G_a$. (Our symbols here imply G_a real, which is appropriate at low frequencies.) Without Z'_F , that has a transfer function

$$K_{LF} = R_u/R_1 = G_a \quad (3.36)$$

and thus represents the real op-amp. In other words, all of the Z_F values in Fig. 3.23 should be considered to be in parallel with $R_u = R_1G_a$. This will not affect performance as long as

$$|Z_F/R_1| \ll |G_a|. \quad (3.37)$$

Otherwise, the effects are as described in Section 3.3.2.5.

Example 3.2 Effect of Op-Amp Gain

- a. An op-amp has an input resistor $R_1 = 1 \text{ k}\Omega$ and a DC gain of 100 dB. How large must be C in Fig. 3.23*b* or 3.23*c* so the undesired pole created by this finite gain is below 1 rad/sec?

From Eq. (3.36), this is equivalent to having $R_u = 10^5 \times 1 \text{ k}\Omega = 10^8 \Omega$ in Fig. 3.23*d*. Since this is 10 times larger than the value in Example 3.1*a*, we know that C must be $0.1 \mu\text{F}/10 = 0.01 \mu\text{F}$.

- b. If the DC gain may vary from 100 to 114 dB, how large would C in Fig. 3.23*e* be in order that the pole frequency not vary more than 2% from 1 kHz due to the uncertainty?

Again, from Eq. (3.36), this is equivalent to $10^8 \Omega \leq R_u \leq 5 \times 10^8 \Omega$. These values are 100 times the values in Example 3.1*b*, resulting in R_p 100 times larger and C 100 times smaller, so $C \geq 30.6 \text{ pF}$.

A more detailed discussion of op-amp performance and stability is in Appendix i.3.B.

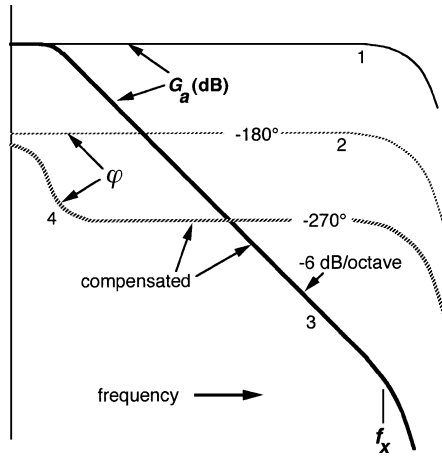


Fig. 3.26 Op-amp open-loop gain (1,3) and phase (2,4), without roll off (1,2) and with it (3,4).

3.3.3.2 Unintended Poles. Since Z_F is usually highest at low frequencies, $|K_{LF}F(s)| > |G_a|$ occurs at a low frequency, moving the zero-frequency pole of an integrator to a nonzero frequency. However, for reasons of stability in the op-amp circuit, G_a is generally rolled off with frequency (Fig. 3.26). For this reason, if $F(s)$ becomes constant at higher frequencies, as it does for all of the filters excepting the low-pass, the relationship (3.37) will again fail at a higher frequency, producing an unintended high-frequency pole where $|Z_F/R_1| = |G_a|$ (f_y in Fig. 3.27). Eventually, $|G_a|$ will fall more rapidly due to multiple unintended poles within the op-amp

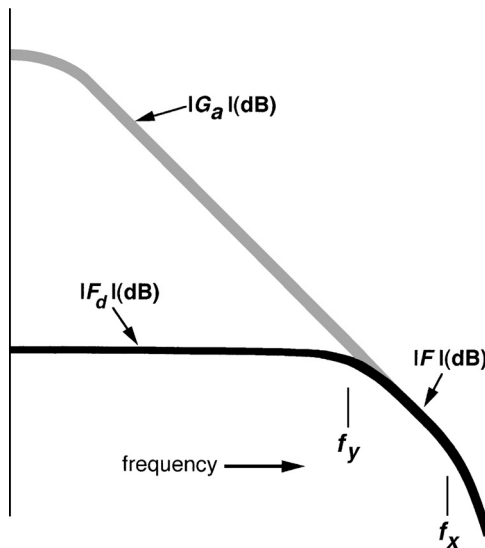


Fig. 3.27 Op-amp closed-loop gain and phase.

(f_x in Figs. 3.26 and 3.27) and this will be accompanied by severe phase lag. Generally, these unintended poles must be sufficiently high in frequency so they do not impact the PLL design.

3.3.3.3 Reducing the Size of C. Suppose the largest resistor available under some condition or with some technology is R_{\max} . The smallest capacitor that we can use to attain a pole frequency ω_p with the circuit of Fig. 3.21 or 3.23e is then $C = 1/(\omega_p R_{\max})$. But, if we make $R_1 = R_{\max}$ in Fig. 3.25 and use just a capacitor as feedback Z_F , the effective resistance in parallel with the capacitor will be $R_u = G_a R_{\max}$ and we can use $C = 1/(\omega_p R_{\max} G_a)$, a huge reduction in capacitor size by a factor G_a . [Looking at it another way, the gain, $1/(R_1 \omega C)$, increases with lowering ω but stops when it reaches G_a , creating the low-frequency corner there at ω_p .] Of course, K_{LF} would also increase by G_a , so gain must be reduced elsewhere in the circuit to obtain the same results as with the passive circuit. Unfortunately, variations in G_a now produce proportional variations in ω_p .

Sometimes the required value of a capacitor is excessive because a low-frequency zero is required and the maximum value of the series resistor (R_2 in Fig. 3.23c) is restricted (e.g., due to technology or noise considerations). In that case, we can create the zero (Fig. 3.24) by adding the response of an integrator circuit (Fig. 3.23b) to a linear response, as shown in Fig. 3.28. Here the total sum current is

$$i_{\text{sum}}(s) = \varphi_{\text{in}}(s) \left[\frac{K_{p1}}{R_1 C s} + \frac{K_{p2} R_2}{R_3} \right] = \varphi_{\text{in}}(s) K_{LF} \frac{1 + s/\omega_z}{s}, \tag{3.38}$$

where

$$K_{LF} = \frac{K_{p1}}{R_1 C} \tag{3.39}$$

and

$$\omega_z = \frac{K_{p1} R_3}{K_{p2} R_2} \frac{1}{R_1 C}. \tag{3.40}$$

Then the zero frequency can be lowered by increasing the ratio of K_{p2} to K_{p1} or of R_2 to R_3 , rather than by increasing the $R_1 C$ product [De Muer and Steyaert, 2003,

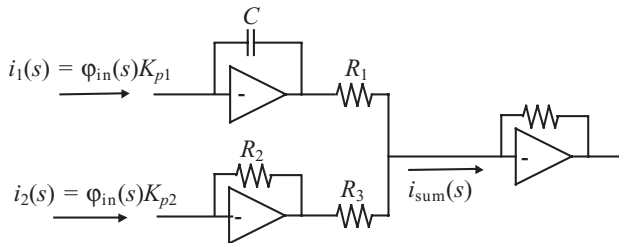


Fig. 3.28 Two currents, each proportional to phase, are summed.

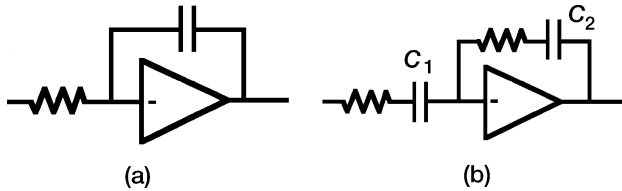


Fig. 3.29 Loop filters to be avoided: (a) pure integrator and (b) filter with capacitor in the input circuit.

pp. 143, 144]. Note that the input currents could be created by driving input resistors from a voltage source.

3.3.3.4 Impractical Loop Filters. The circuits shown in Fig. 3.29 have serious problems for use as loop filters. While consideration of the pure integrator (Fig. 3.29a) was of academic value, the integrator is not a practical loop filter because there is no frequency at which it does not have a transfer phase of -90° . As we will see in the following chapters, this would inevitably put the loop on the verge of instability.

A transfer function with $F(s)$ from Eq. (3.13) could be implemented by the configuration shown in Fig. 3.29b if the op amp could be prevented from “hitting the rails” (i.e., for the output to reach an extreme value set by the power supply voltages). However, it should come as no surprise that a DC connection is necessary to maintain steady state conditions. Assuming that phase lock is somehow established, the op amp’s input bias current or board leakage I_b will charge C_1 while the loop causes the circuit input voltage (PD output) to change in order to maintain zero input voltage at the op amp. Eventually the circuit input voltage will run out of range and will no longer be able to compensate. At this point the loop will be out of lock. The leakage current will then charge C_2 until the op amp output hits a rail. The slope of the voltage ramp at the filter input, I_b/C_1 , can vary widely, depending on components employed, but it would be an unusual application that would cause us to be satisfied with a loop with such a built-in failure mechanism, especially where the alternative discussed in Section 3.3.2.4 is available.

3.4 FILTER REFERENCE VOLTAGE

In the passive filters of Figs. 3.19 and 3.21, there is no DC connection to the ground reference, so its value makes little difference. In active filters, it can be important. Since both inputs to an op-amp are maintained at the same voltage, the reference on the $+$ input (v_+ in Fig. 3.30) should be at the voltage where we wish to establish steady-state output of the phase detector, usually at mid-range. For some phase detectors, this may be other than ground potential.

If a charge pump drives one of the filters of Fig. 3.23e, 3.23f, or 3.23g, it must provide, in steady state, average current that is proportional to the difference between

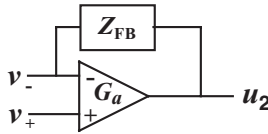


Fig. 3.30 Op-amp showing both inputs.

the filter output voltage and the reference at the other end of the filter. In general, this results in repeated pulses at the reference frequency, which can produce FM sidebands on the VCO. For this reason, charge pumps usually drive the filter in Fig. 3.23c, creating a loop that has zero phase error; once the capacitor has attained its steady-state value, no further current input is required or allowed. Unfortunately, charge-pump phase detectors have been known to be nonlinear in a small region near zero phase [Egan, 2000, pp. 203–213]. Techniques used to overcome this problem usually cause one or both pumps to be on for a short time, reducing the effective output resistance.

The charge pump is connected to the filter output voltage and, due to finite impedances of the current sources, K_p varies somewhat with changes in that voltage. Moreover, the gains for the up and down pumps are affected differently, producing nonlinearity in the phase detector characteristic, which can be problematic in applications where the phase varies dynamically (e.g., due to phase modulation). For these reasons, it is often desirable for the charge pump to drive an op-amp, which presents it with a fixed voltage, enabling the up and down pumps to be adjusted for equal currents at that voltage.

3.5 NOTE ON THE FORM OF THE FILTER EQUATION

Note that Eq. (3.26) can be written in another form:

$$K_{LF}F(s) = K_{LF} \frac{1 + s/\omega_z}{1 + s/\omega_p} = \left(K_{LF} \frac{\omega_p}{\omega_z} \right) \frac{s + \omega_z}{s + \omega_p}. \quad (3.41)$$

It is important to recognize that the constant term on the right-hand side of Eq. (3.41) is not K_{LF} . If $F(s)$ is written in the form on the right, the constant there is likely to be mistaken for K_{LF} . Then equations (which we will derive) that contain K_{LF} , or K , will yield incorrect answers. The appropriate form of $F(s)$, except in the limiting cases of an integrator with or without lead, gives $F(0) = 1$.

3.6 CAPACITORS IN LOOP FILTERS

There are two deviations from the ideal that we should be aware of in choosing capacitors for use in loop filters. The first is leakage current, which can cause a resistor

to effectively appear in parallel with the capacitor, especially where electrolytic capacitors (e.g., tantalum) are used in order to obtain large values in a small size and especially at higher temperatures. This could effectively turn the integrator-and-lead filter employing the feedback circuit of Fig. 3.23*c* into a lag-lead filter with the feedback circuit shown in Fig. 3.23*g*.

The second is dielectric absorption [Pease, 1982; Buchanan, 1983], which can cause the voltage across a capacitor to continue to change after current has stopped flowing through it. High-dielectric-constant ceramic capacitors (e.g., CK05, CK06), in particular, exhibit this phenomenon. It can be a problem if the PLL is required to settle rapidly.

The need for large capacitance values in a small package tends to aggravate both problems. Plastic capacitors should be considered as alternatives where necessary.

3.7 HIGHER-ORDER FILTERS

Many practical filters have additional poles at higher frequencies. For example, another high-frequency pole can be created by a placing a small capacitor across the filter in Fig. 3.23*c* or the op-amp output might drive a high-frequency low-pass (Fig. 3.21). Sometimes a high-frequency low-pass filter precedes the op-amp to keep it from being driven nonlinear by the rapid rise and fall of pulses from a phase detector. For now, however, we will concentrate on filters with only one pole.

i3.A APPENDIX: INTEGRATED CIRCUIT DOUBLY BALANCED MIXER—DETAILS

This appendix is available from the Wiley Internet site.

i3.B APPENDIX: OP-AMPS IN LOOP FILTERS

This appendix provides details about the use of op-amps in loop filters, including equations describing the operation of voltage- and current-feedback op-amps, unintended poles, stability of the filter, and response to the noninverting input. It is available from the Wiley Internet site.

CHAPTER 4

LOOP RESPONSE

In this chapter we will develop the equations that describe both the frequency and transient responses for loops that use the filters studied in Chapter 3 (Fig. 4.1).

4.1 LOOP ORDER AND TYPE

The order of the loop equals the number of poles in the open-loop transfer function $G(s)$ [Klapper and Frankle, 1972, pp. 85–86]. This is also the highest power of s in its denominator. Its type is the number of such poles that are at the origin ($s = 0$). Thus the simple loop described in Chapter 2 by (2.24) is a first-order, type 1, loop. Because of this pole at the origin and an additional pole contained in the loop filter, the loops that we will study here, and throughout the remainder of the book, are second order. This restriction will permit us to make use of the long and well-developed theory of second-order systems.

The pole due to conversion from frequency to phase is truly the only pole at the origin so the phase-locked loop (PLL) is strictly always a type 1 loop. However, the active filter can have a pole at a very low frequency, which is often approximated as being at the origin, giving the loop then two such poles and thereby making it a type 2 loop. A primary difference between the two types lies in the nature of the phase error. In response to a frequency step, the steady-state phase error of a type 1 loop will change but that of a type 2 loop will not—after the transient settles, the phase error will have returned to zero. (If such an error could exist in a type 2 loop, the integrator

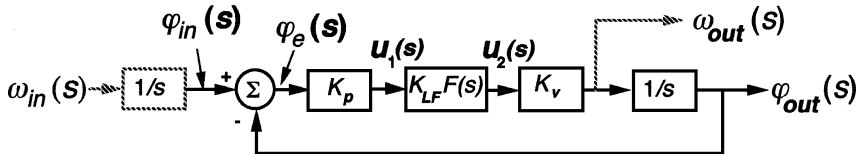


Fig. 4.1 Mathematical block diagram of a loop with a filter. This is the same as Fig. 2.4 except for the filter transfer function $F(s)$.

in the loop filter would continue to change its output and steady state would therefore not exist.) When following a frequency ramp, a type 2 loop has a steady-phase error whereas a type 1 loop has an ever-increasing phase error.

While PLLs may actually be of an order higher than second, performance estimates can often be made based on similar second-order loops. (See examples in Sections 6.12 and 7.10.) In practical circuits, multiple poles exist at high frequencies where the gains of the various elements begin to fall. Fortunately, these may be of little importance in determining performance in the range of interest. The phase transfer function has been given by

$$\frac{\varphi_{out}(s)}{\varphi_{in}(s)} = \frac{G(s)}{1 + G(s)}. \quad (2.25a)$$

At low frequencies, where $G(\omega)$ is very large, this function is approximately equal to one, regardless of the details of $G(\omega)$. At high frequencies, where $G(\omega)$ is small, the function is approximately equal to $G(\omega)$, as if there were no feedback. Thus, frequency response can be easily predicted there, and transient response, which is affected by the various frequencies of which the transient wave is composed, tends to be affected relatively little because of the low amplitude of $G(\omega)$. Only in the transition region, near the frequency where $G(\omega) \approx 1$, is the gross performance in question. Therefore, if $G(\omega)$ is approximately the same in that region as for some second-order loop, we expect the gross performance of the two loops to be the same. The two will differ in some fine details, for example, how long is required to settle to a very small fraction of a step change or the value of a small error in following some low-frequency signal that is being followed very closely, but gross performance will be similar.

Similar statements can be made concerning the error response, Eq. (2.26a), except it is approximately one at high frequencies and approximately equal to $1/G(\omega)$ at low frequencies.

4.2 CLOSED-LOOP EQUATIONS

The equations describing the loop will be the same as Eqs. (2.25) and (2.26) except that the gain K is replaced by $KF(s)$ to account for the filter. Therefore the open-loop

transfer function is

$$G(s) = \frac{KF(s)}{s} \quad (4.1)$$

$$= \frac{K}{s} \left(\frac{1 + s/\omega_z}{1 + s/\omega_p} \right), \quad (4.2)$$

and the closed-loop transfer function is

$$H(s) \triangleq \frac{\varphi_{\text{out}}(s)}{\varphi_{\text{in}}(s)} = \frac{G(s)}{1 + G(s)} = \frac{1}{1 + s/[KF(s)]} \quad (4.3)$$

or

$$1 - H(s) = \frac{\varphi_e(s)}{\varphi_{\text{in}}(s)} = \frac{1}{1 + G(s)} = \frac{s}{s + KF(s)}. \quad (4.4)$$

Substituting Eq. (4.2) into (4.3), we obtain

$$H(s) = \left[1 + \frac{s}{K} \left(\frac{1 + s/\omega_p}{1 + s/\omega_z} \right) \right]^{-1} = K\omega_p \frac{1 + s/\omega_z}{s^2 + s\omega_p(1 + K/\omega_z) + K\omega_p} \quad (4.5)$$

The standard second-order equation form for the denominator is

$$s^2 + 2\zeta\omega_n s + \omega_n^2 = s^2 + \omega_p(1 + K/\omega_z)s + K\omega_p, \quad (4.6)$$

where ζ is the damping factor and ω_n is the natural frequency.

From this it is apparent that

$$\omega_n^2 = K\omega_p \quad (4.7)$$

and

$$2\zeta\omega_n = \omega_p \left(1 + \frac{K}{\omega_z} \right) = \omega_p + \frac{\omega_n^2}{\omega_n}, \quad (4.8)$$

$$\zeta = \frac{1}{2} \left(\frac{\omega_p}{\omega_n} + \frac{\omega_n}{\omega_z} \right). \quad (4.9)$$

In terms of Eqs. (4.7) and (4.9), Eq. (4.5) can be written

$$H(s) = \omega_n^2 \frac{1 + s/\omega_z}{s^2 + 2\zeta\omega_n s + \omega_n^2}. \quad (4.10)$$

The loop error response can also be written from Eq. (4.4) as

$$1 - H(s) = \frac{s [s + \omega_n (2\zeta - \omega_n/\omega_z)]}{s^2 + 2\zeta\omega_n s + \omega_n^2} = \frac{s (s + \omega_p)}{s^2 + 2\zeta\omega_n s + \omega_n^2}, \tag{4.11}$$

where Eq. (4.9) was used in the last simplification.

4.3 OPEN-LOOP EQUATIONS—LAG-LEAD FILTER

Now that we have used the closed-loop response to establish the values of the traditional second-order parameters in terms of the parameters of our loop, we will determine how the open-loop transfer function can be interpreted in these terms. Equations (4.1) and (4.2) at $s = j\omega$ become

$$G(\omega) = \frac{K}{j\omega} F(j\omega) = \frac{K}{j\omega} \left(\frac{1 + j\omega/\omega_z}{1 + j\omega/\omega_p} \right). \tag{4.12}$$

The straight-line approximation of the magnitude of the filter response $F(\omega)$ and of the open-loop transfer function from Eq. (4.12) [$G(\omega)$] are plotted in Fig. 4.2 on logarithmic scales.¹ The open-loop gain has the same shape as the filter response except for the additional $1/\omega$ term. The filter gain is flat to $\omega = \omega_p$, at which point it begins to fall linearly with ω . After ω reaches ω_z , the two frequency-dependent terms dominate and the filter response becomes approximately ω_p/ω_z . When $\omega = 1$, the filter gain is K_{LF} and the open-loop gain is K . This assumes that ω_p and ω_z are

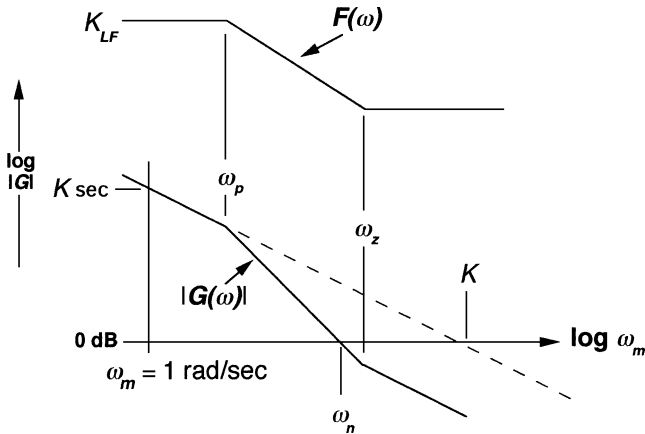


Fig. 4.2 Gain of the open-loop transfer function $G(\omega)$ and of the corresponding lag-lead filter $F(\omega)$. Tangential approximations are shown. Values of $|G|$ and ω are marked on the axes although distances are logarithms of $|G|$ and ω in rad/sec.

¹ For compactness, we will use the notation $Q(\omega)$ to represent $Q(s)$ with $s \Rightarrow j\omega$, although $Q(j\omega)$ would show the relationship more exactly.

greater than 1; otherwise K and K_{LF} are along the straight-line extension of the plots from the lowest frequencies. Since the extension of the lowest frequency portion of the open-loop gain curve falls proportionally to frequency, the fact that it intersects $\omega = 1$ rad/sec at $|G(\omega)| = K$ sec implies that it also intersects $|G(\omega)| = 1$ at $\omega = K$. This can be verified by setting the part of Eq. (4.12) that dominates at low frequencies, $|K/\omega|$, equal to unity and solving for ω . Note that the steep part of the open-loop gain curve crosses unity gain at ω_n . This is because, by Eq. (4.7), ω_n is the geometric mean of K and ω_p and therefore is midway between them on a logarithmic plot. In other words, from geometry,

$$\log(\omega_n) - \log(\omega_p) = \log(K) - \log(\omega_n), \quad (4.13)$$

$$2 \log(\omega_n) = \log(K) + \log(\omega_p), \quad (4.14)$$

$$\omega_n^2 = K \omega_p. \quad (4.7)$$

If this -12 dB/octave portion of the open-loop gain curve does not cross the unity gain level, the same geometric arguments show that ω_n occurs at the intersection of the extension of this region, as illustrated in Fig. 4.3.

We can also relate the damping factor ζ to the open-loop gain curve. The first term in (4.9) dominates the second if

$$\frac{\omega_p}{\omega_n} \gg \frac{\omega_n}{\omega_z} \Rightarrow \omega_n \ll \sqrt{\omega_p \omega_z} \equiv \omega_{\text{mid}}. \quad (4.15a)$$

In other words, the natural frequency is significantly to the left of the midpoint ω_{mid} between ω_p and ω_z in the log plot of Fig. 4.3, as in curve c . Conversely, if the natural

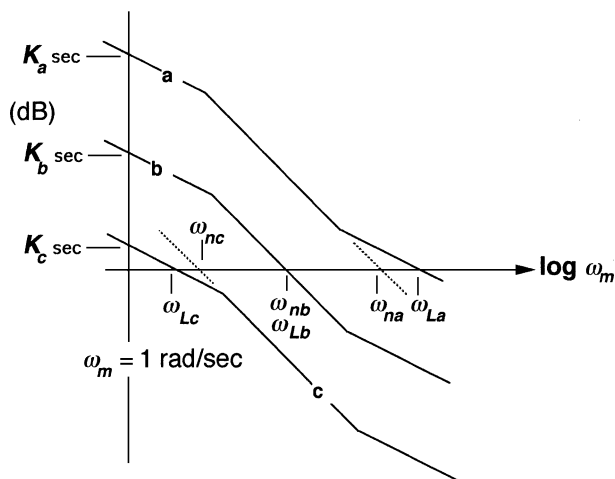


Fig. 4.3 Gain vs. modulation frequency ω_m for several values of K ; ω_L is the frequency at which $|G(\omega)|$ is one.

frequency is significantly to the right of the midpoint, as in curve *a*, the second term dominates (4.9).

For the former case (curve *c*) we have

$$\zeta |_{\omega_n \ll \omega_{mid}} \approx \frac{1}{2} \left(\frac{\omega_p}{\omega_n} \right), \tag{4.15b}$$

whereas in the latter case (curve *a*) we have

$$\zeta |_{\omega_n \gg \omega_{mid}} \approx \frac{1}{2} \left(\frac{\omega_n}{\omega_z} \right). \tag{4.16}$$

Writing these last two equations in logarithmic form, we obtain

$$\log(2\zeta |_{\omega_n \ll \omega_{mid}}) \approx \log(\omega_p) - \log(\omega_n) \tag{4.17}$$

and

$$\log(2\zeta |_{\omega_n \gg \omega_{mid}}) \approx \log(\omega_n) - \log(\omega_z). \tag{4.18}$$

This is illustrated in Fig. 4.4. From the graph can be seen that, when a critical frequency, either a pole or a zero, is at unity gain, the damping factor is 0.5 (the logarithm

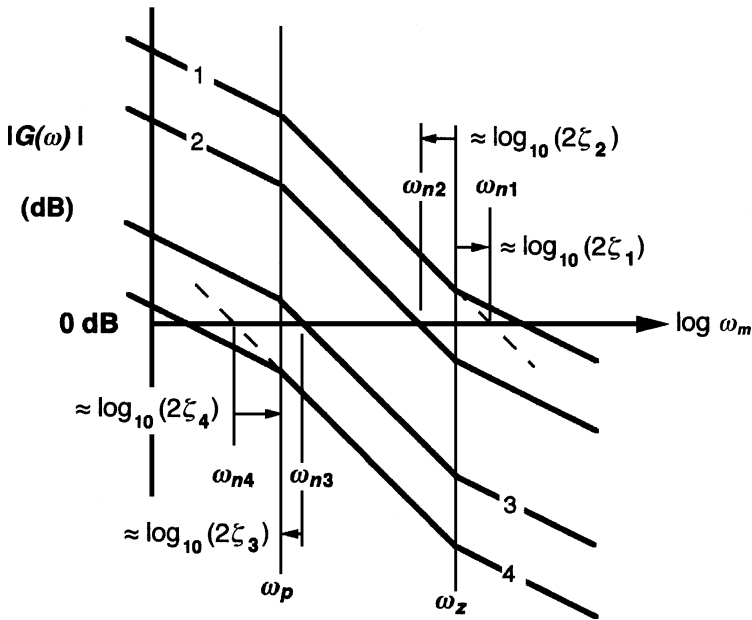


Fig. 4.4 Natural frequency ω_n and damping factor ζ shown graphically for several values of gain factor K . The graphical representation of ζ is accurate when $\omega_z \gg \omega_p$ and unity gain occurs far from ω_{mid} .

of 2ζ is zero) approximately. The accuracy of the approximation requires significant separation between ω_p and ω_z . Otherwise the inequality in (4.17) or (4.18) will not hold with the critical frequency at unity gain and ζ will be larger than 0.5 there. Assuming these inequalities, as the unity-gain frequency ω_L moves into the steep region between ω_p and ω_z , the logarithm becomes negative (arrow pointing toward the left), corresponding to a fractional value of 2ζ , that is, $\zeta < 0.5$. Along this steep slope, once the frequency is far removed from both ω_p and ω_z , the total phase shift in the open-loop transfer function approaches -180° , the value for instability. In this region the loop is underdamped. As the unity-gain frequency ω_L increases beyond ω_z , or decreases below ω_p , ζ increases until it equals one. At $\zeta = 1$, the loop is said to be critically damped. At higher or lower values, respectively, of ω_L it is overdamped.

4.4 LOOP WITH A LAG FILTER

If the zero frequency goes to infinity, then the open-loop transfer function, Eq. (4.2), becomes

$$G(s) = \frac{KF(s)}{s} = \frac{K}{s} \left(\frac{1}{1 + s/\omega_p} \right), \tag{4.19}$$

the Bode plot is as shown in Fig. 4.5, the filter is represented by Eq. (3.16), and the

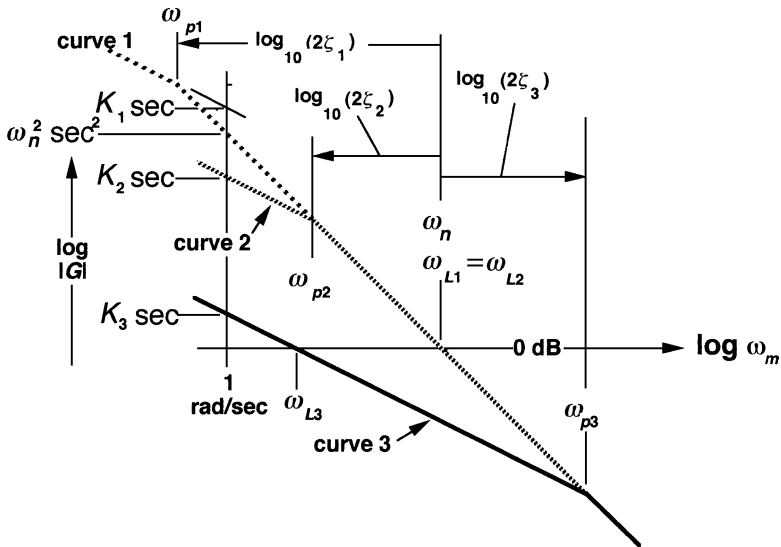


Fig. 4.5 Gain of the open-loop transfer function $G(\omega)$, loop with a lag filter, at three values of K .

closed-loop transfer function, Eq. (4.10), becomes

$$H(s) = \omega_n^2 \frac{1}{s^2 + 2\zeta\omega_n s + \omega_n^2}, \quad (4.20)$$

where ω_n is still given by Eq. (4.7) and Eq. (4.9) becomes Eq. (4.15b) exactly. The loop error is obtained from Eq. (4.11) as

$$1 - H(s) = \frac{s(s + 2\zeta\omega_n)}{s^2 + 2\zeta\omega_n s + \omega_n^2} = \frac{s(s + \omega_p)}{s^2 + 2\zeta\omega_n s + \omega_n^2}. \quad (4.21)$$

4.5 LOOP WITH AN INTEGRATOR-AND-LEAD FILTER

If the pole moves toward zero frequency, Eq. (4.2) becomes, for frequencies well beyond ω_p ,

$$G(s) \approx \frac{K\omega_p}{s^2} \left(1 + \frac{s}{\omega_z}\right). \quad (4.22)$$

This form illustrates that an integrator can only be an approximation—there really is a pole back there somewhere.² However, once we have lost track of, or interest in, the position of that pole, we should be able to express $K\omega_p$ as a single constant. Therefore, we use Eq. (4.7) to write G in a form that does not explicitly involve ω_p ,

$$G(s) \approx \frac{\omega_n^2}{s^2} \left(1 + \frac{s}{\omega_z}\right). \quad (4.23)$$

This is represented by Fig. 4.6. By substituting $K_p K_v K_{LF}$ for K in Eq. (4.1) and then substituting $(1 + s/\omega_z)$ divided by Cs or $R_1 Cs$ for $K_{LF} F(s)$ in that result [according to Eqs. (3.20) and (3.35)] and comparing the result to Eq. (4.23), we see that

$$\omega_n^2 = \frac{K_p K_v}{R_1 C}, \quad (4.24)$$

where $R_1 \Rightarrow 1$ for a transimpedance filter. Equation (4.9) becomes Eq. (4.16) exactly. Equation (4.10) remains a valid expression for $H(s)$ and Eq. (4.11) becomes

$$1 - H(s) = \frac{s^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}. \quad (4.25)$$

² Blanchard (1976) uses this form. It has the advantage that the expression for ω_n remains the same for the various filter configurations.

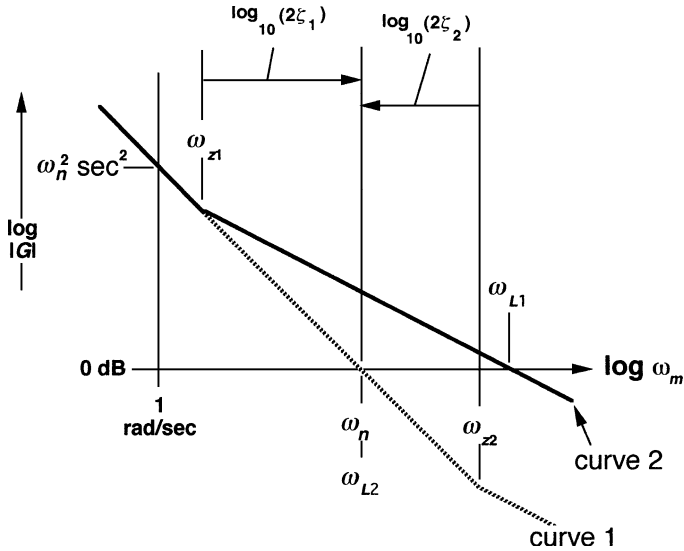


Fig. 4.6 Gain of the open-loop transfer function $G(\omega)$, loop with an integrator-and-lead filter, at two values of ω_z .

4.6 SUMMARY OF EQUATIONS

The equations for the second-order loop parameters and closed-loop responses that were developed above are summarized in Table 4.1. Figure 4.7 shows the corresponding Bode gain plots.

4.7 UNITY-GAIN BANDWIDTH IN HIGHLY DAMPED LOOPS

The approximate value of the frequency ω_L where the open-loop gain is one can be easily obtained for highly damped loops ($\zeta \gg 0.5$). In such loops, ω_L occurs well into a region where $|G| \sim 1/\omega$, so we can depend on the tangential approximation of the open-loop gain to give relatively accurate results. This might be the condition for the lower curve of Fig. 4.5 (depending on the scale) in which case

$$\omega_L = K \tag{4.26}$$

for a loop with a low-pass loop filter (just as for the first-order loop in Fig. 2.8).

From the upper curve in Fig. 4.6,

$$\log(\omega_L) - \log(\omega_n) = \log(\omega_n) - \log(\omega_z), \tag{4.27}$$

$$\log(\omega_L) = 2 \log(\omega_n) - \log(\omega_z) = \log(\omega_n^2/\omega_z), \tag{4.28}$$

$$\omega_L = \omega_n^2/\omega_z, \tag{4.29}$$

for loop with an integrator-and-lead filter.

TABLE 4.1. Summary of Second-Order Loop Equations

Filter Type	(a) Lag	(b) Lag-Lead	(c) Integrator and Lead
${}^a\alpha =$	0	$0 < \alpha < 1$	1
$F(s) =$	$\frac{1}{1 + s/\omega_p}$	$\frac{1 + s/\omega_z}{1 + s/\omega_p}$	$\frac{1 + s/\omega_z}{R_1 C s}$
$\zeta =$	$\frac{1}{2} \left(\frac{\omega_p}{\omega_n} \right)$ Eq. (4.15b)	$\frac{1}{2} \left(\frac{\omega_p}{\omega_n} + \frac{\omega_n}{\omega_z} \right)$ Eq. (4.9)	$\frac{1}{2} \left(\frac{\omega_n}{\omega_z} \right)$ Eq. (4.16)
$\omega_n^2 =$	$K \omega_p$ Eq. (4.7)		$\frac{K_p K_v}{R_1 C}$ ^b Eq. (4.24)
$H(s) =$	$\frac{1}{\omega_n^2 s^2 + 2\zeta\omega_n s + \omega_n^2}$ Eq. (4.20)	$\frac{1 + s/\omega_z}{\omega_n^2 s^2 + 2\zeta\omega_n s + \omega_n^2}$ Eq. (4.10)	
$1 - H(s) =$	$\frac{s(s + \omega_p)}{s^2 + 2\zeta\omega_n s + \omega_n^2}$ Eq. (4.11)		$\frac{s^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$ Eq. (4.25)

^a α will be defined by Eq. (6.5)

^b $R_1 \Rightarrow 1$ for a transimpedance filter

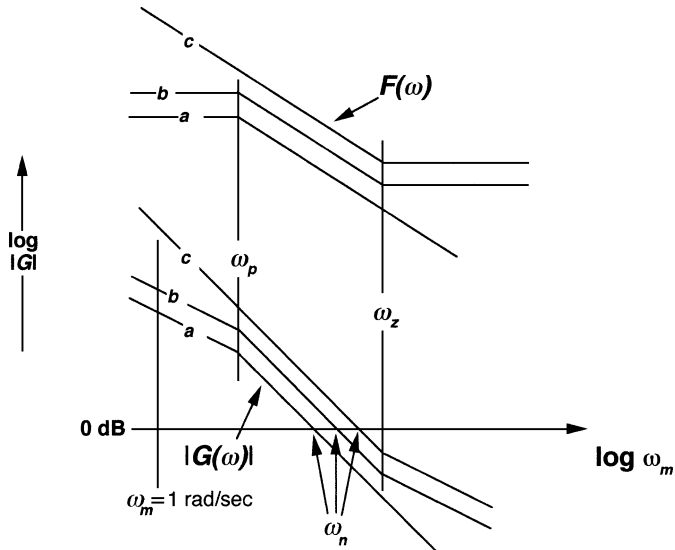


Fig. 4.7 Gain of the open-loop transfer function $G(\omega)$ and of the corresponding loop filter $F(\omega)$ for: (a) lag, (b) lag-lead, and (c) integrator-and-lead filter. Tangential approximations are shown.

Equation (4.26) also applies to the *low-gain loop with a lag-lead filter*, the lower curve in Fig. 4.3. For the upper curve, the gain is K reduced by the ratio of ω_z to ω_p so

$$\omega_L = K \omega_p / \omega_z \quad (4.30)$$

for a *high-gain loop with a lag-lead filter*. [Note that this approaches the equation for a first-order loop (Fig. 2.8) as $\omega_z \Rightarrow \omega_p$.]

CHAPTER 5

LOOP STABILITY

The shapes of the closed-loop frequency and transient responses indicate the stability of the loop, but during design, stability is assessed from the open-loop characteristics.

5.1 OBSERVING THE OPEN-LOOP RESPONSE

In general, open-loop response is observed by injecting a signal at some point in a loop and observing the response around the loop to the same point. At least conceptually, the loop is broken so that the observed response can be differentiated from the injected excitation. This presents a problem for loops of type 1 and above because they can have no steady-state “position” error. Any deviation of the integrator input from its steady-state value will cause it to integrate. For example, in the simplest phase-locked loop (PLL) (Fig. 2.7), which is type 1, if the loop is broken at the voltage-controlled oscillator (VCO) input, there will always be some error in the injected direct current (DC) voltage relative to the steady-state value that would be there with the loop closed. As a result, the VCO frequency will be different from the reference frequency by some amount ($\omega_e \neq 0$), causing the phase difference (φ_e) to change with time. This will eventually drive some component out of its linear range, usually so quickly that testing is not practical.

Practically, the open-loop response must be ascertained through closed-loop testing. Figure 5.1 illustrates this. Here a signal is injected at the VCO input, and the

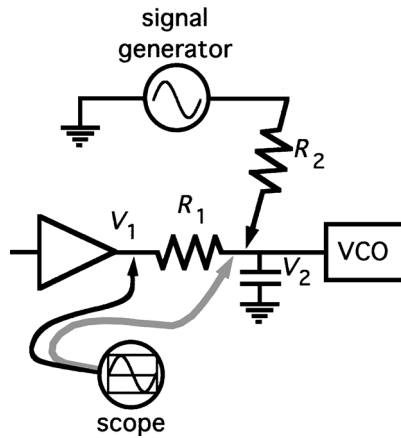


Fig. 5.1 Measuring open-loop response with the loop closed.

response is observed at the output of a low-impedance amplifier, possibly the op-amp used in the loop filter. If R_2/R_1 is large enough, R_2 will not appreciably affect the loop response, and if R_1 is large enough compared to the amplifier output impedance, the injected signal will not appreciably affect v_1 . The ratio v_1/v_2 is not quite the open-loop response, however. The transfer function of the low-pass must be accounted for. This can be done by multiplying v_1/v_2 by the theoretical response of the low-pass, but often its cutoff frequency will be high enough to have little effect.

Type 2 loops are even more difficult to test. Not even the loop filter alone can be tested open loop because any error at that point will cause the output of the integrator eventually to saturate.

Responses involving the integrating filter can be measured by reducing the op-amp's gain, by placing a resistor in the feedback path (R_u in Fig. 3.23*d*), to the point where the DC level at its input u_1 can be controlled well enough to keep it from saturating. Of course, the proper DC level must be applied along with the test signal to do this.

5.2 METHODS OF STABILITY ANALYSIS AND MEASURES OF STABILITY

We will consider three commonly used methods—Bode plot, Nyquist diagram, and Evans (root locus) plot—as they apply to the PLL. Throughout the text we will concentrate on the Bode plot, which seems to be the most useful for PLL design, but at this point we will compare it to the other two. This should be helpful to those who are familiar with either of the other methods. See the References for more complete treatments.

The most common measures of stability are phase margin and gain margin. Both unity gain and 180° excess phase shift (i.e., more negative than the -180° that is built in with negative feedback) are necessary for oscillation so the margins relative to this pair of conditions become a measure of stability. Phase margin is the additional open-loop phase shift necessary to give 180° excess phase when the open-loop gain is unity. All PLLs have -90° phase shift due to the $1/s$ term, so excess phase is 90° even before any effect from the loop filter is considered. Gain margin is the additional gain necessary to give unity open-loop gain when the open-loop excess phase is 180° . While a loop will not achieve steady-state oscillation with a gain that exceeds unity at 180° , it may enter a condition wherein oscillations grow until average gain is reduced to unity due to saturation of components or the loop may lose lock due to excessive excursions of the phase error. Whether a loop will oscillate when its gain exceeds unity at 180° is more easily seen from the Nyquist plot or from the root locus plot than from the Bode plot. However, in most cases it will not operate properly under these conditions, and the Bode plot is adequate.

5.2.1 Bode Plot

Figure 5.2 illustrates a Bode plot. Open-loop gain and phase are plotted against modulation frequency, and the gain and phase margins can be seen in the figure. Tangential approximations for gain and phase are often used; we will use the tangential approximation for gain only, computing phase as accurately as necessary. When more

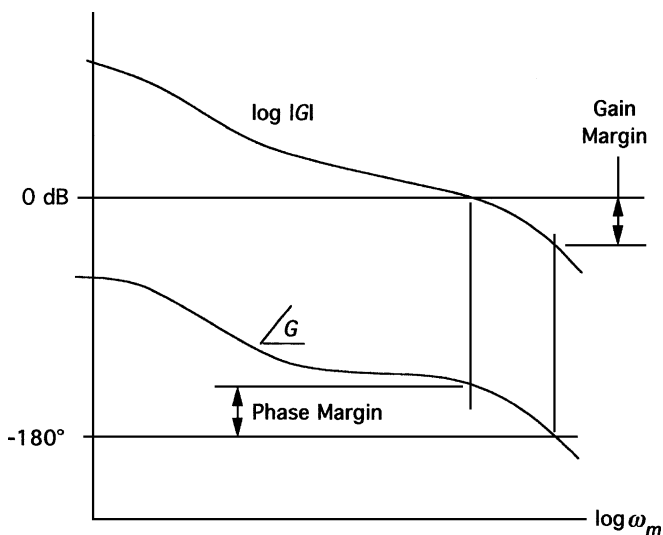


Fig. 5.2 Bode plot, open-loop gain on a log scale and open-loop phase plotted against modulation frequency on a log scale.

accuracy is required exact calculations can be made, but much can be learned by the tangential approximations.

5.2.2 Nyquist Plot

Figure 5.3 illustrates a Nyquist plot [Kuo, 1987, pp. 556–590]. Normally, the shaded lines are not used, the information relative to the whole plot being easily discerned from the half shown in solid lines. However, we will consider the whole plot at this point.

To make the plot, the value of s is varied along the path shown at Fig. 5.3a. The object is to determine if the closed-loop transfer function $H(s)$ in Eq. (4.3) has any poles with positive real parts, corresponding to rising exponentials and instability. These would occur when the denominator $G(s) + 1$ would have one or more zeros with positive real parts, corresponding to a time response factor

$$e^{st} = e^{(j\omega + \sigma)t} = e^{j\omega t} e^{\sigma t}, \quad \sigma > 0. \quad (5.1)$$

Assume that the open-loop function $G(s)$ is itself everywhere finite (stable) so that $G(s)$ [and so $(1 + G(s))$] has no poles in the right-half plane. The locus in the s plane is drawn in Fig. 5.3a in such a manner as to surround the nonallowed area, the right-half plane, which contains values of $s = \sigma + j\omega$ that have positive values of σ . As s traverses this locus, the corresponding value of $G(s)$ is plotted in Fig. 5.3b. If the $G(s)$ curve surrounds the point -1 , where the value of $H(s)$ becomes infinite,

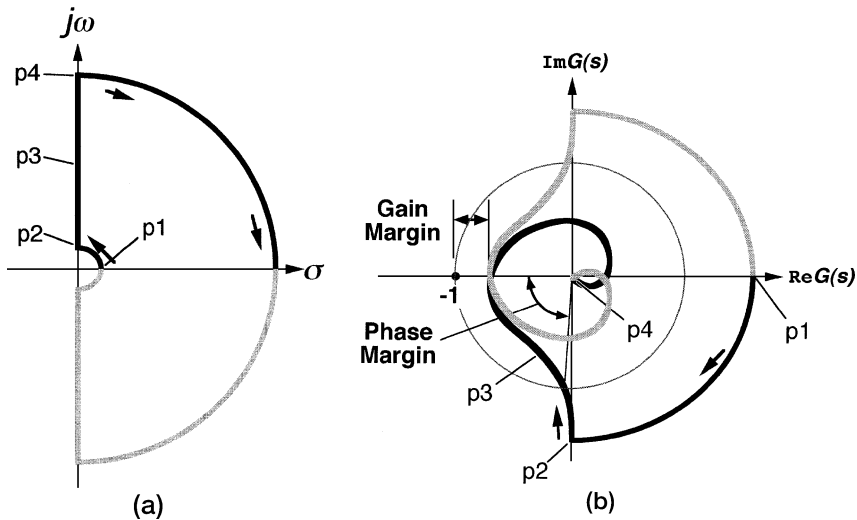


Fig. 5.3 Nyquist plot, a plot of the complex open-loop transfer function as the locus in the s plane encloses the right half of the plane. The shaded portion is a mirror image of the other half and is often not shown explicitly: (a) s plane and (b) $G(s)$ plane.

n times, then the s -plane locus surrounds n zeros. But these are then in the right-half plane, so the pole(s) of $H(s)$ [at $G(s) = -1$] would occur where σ , the real part of s , is positive, corresponding to a rising exponential and instability.

Let us study the illustration to see why it represents a PLL and why this diagram indicates stability. To avoid the allowable pole at zero, the locus in the s plane begins at an infinitesimal radius ($p1$). It moves from $\omega = 0$ to positive values of ω ($p2$) and then up the $j\omega$ axis. If any poles are encountered on the $j\omega$ axis, the locus passes these allowed singularities on the right. At the beginning of this locus, when s is very small, $|G(s)|$ is very large since, at low frequencies, $G(s) \Rightarrow K/s$, so it approaches infinity as the radius of the beginning circle in the s plane approaches zero. Thus the locus in the $G(s)$ plane begins at infinite radius. Also, since $G(s) \Rightarrow K/s$, $G(j\omega) \Rightarrow K/(j\omega) = -jK/\omega$ and the $G(s)$ locus reaches the negative imaginary axis when s reaches the positive imaginary axis ($p2$). As s increases along this axis, $|G(s)|$ decreases, as in the Bode plot, and the phase of $G(s)$ becomes more negative than -90° ($p3$). For the function $G(s)$ represented here, the phase goes beyond -360° as $|G(s)|$ continues to shrink ($p4$). With $|s|$ very large, the locus in the s plane is brought clockwise to the negative $j\omega$ axis, leaving $G(s)$ near zero. The s -plane locus is then completed and the second half of the $G(s)$ locus is symmetrical to the first because the $\Re G(s)$ is an odd function of ω and $\Im G(s)$ is an even function. The important fact is that the locus does not surround the point $-1 + j0$ so poles of $H(s)$ are in the left half of the s plane and that corresponds to the stable region. The gain and phase margins can also be seen from the Nyquist diagram and are shown in Fig. 5.3b.

One of the benefits of the Nyquist diagram relative to the Bode plot is the ease of determining if a system where $|G(s)|$ exceeds unity at 180° is stable. Figure 5.4 represents a conditionally stable system. The locus of $G(s)$ does not surround -1 so it is stable. It is “conditionally stable” because a reduction of gain, such as might occur due to saturation or at power on, could cause it to enter the region of instability. Once there, it is likely to stay in a state where the waveforms are clipped due to limiting

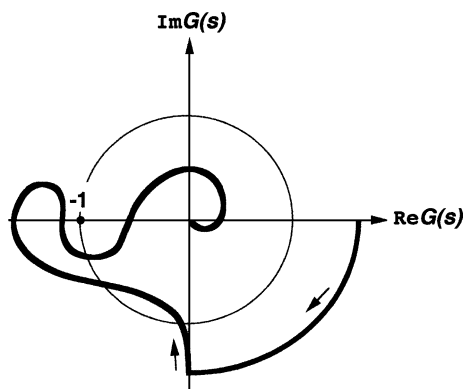


Fig. 5.4 Nyquist plot of a conditionally stable system. A reduction in the gain factor can cause the point $-1, j0$ to be enclosed, an indication of instability.

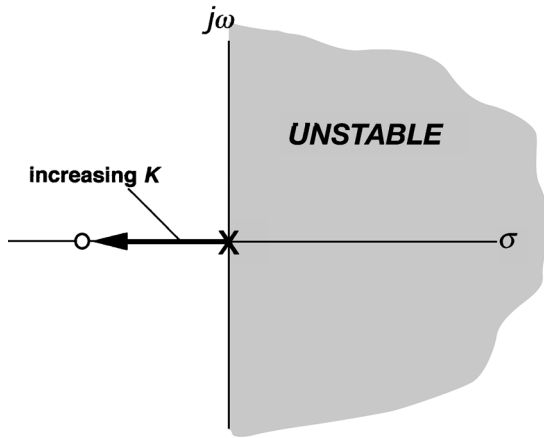


Fig. 5.5 Evans plot for a very simple system. The closed-loop poles are plotted in the s plane as the gain factor K increases.

and the average gain is unity. In a phase-locked loop it may exceed the linear range of the phase detector and may lose lock. With complicated loci it may be beneficial to determine whether a line with a fulcrum at $-1 + j0$ rotates a net nonzero number of times (clockwise under the assumptions above) when its other end follows the locus around its path. If it does, the locus encloses -1 .

One minor problem with the Nyquist plot is that it can be hard to draw in a manner that shows its features close to the critical -1 point as well as for large values. The representations that follow are only approximate. In many cases the radial distance has something like a logarithmic scale, giving an expanded presentation near the origin compared to further out. In fact we often use the greatest radius to represent infinity.

5.2.3 Evans Plot (Root Locus)

The third common method of stability analysis is to plot the poles of $H(s)$, which are the zeros of $[1 + KF(s)/s]$, as K increases to determine at what value of K they enter the right-half plane, taking on positive real parts σ [Kuo, 1987, pp. 384–456]. This is done according to a set of rules that permits the locations of the closed-loop poles to be determined from those of the open-loop poles and zeros. Figure 5.5 shows an Evans plot for a simple stable system (not for a practical PLL).

5.3 STABILITY OF VARIOUS PLL CONFIGURATIONS

In this section we will consider the stability of PLLs, having various types of loop filters, using all three methods of analysis.

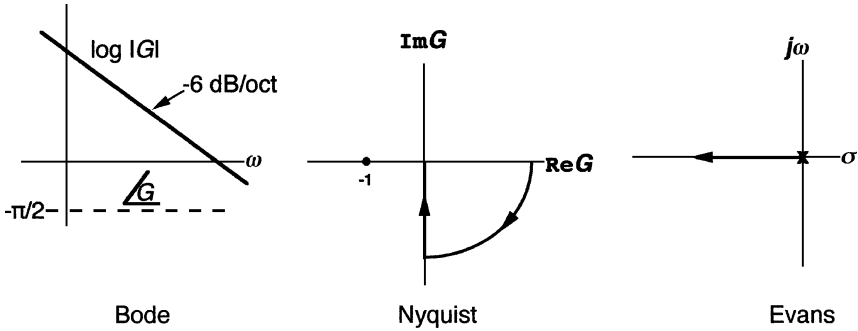


Fig. 5.6 Stability plots for a first-order loop.

5.3.1 First-Order Loop

The simple loop described in Chapter 2 is represented in Fig. 5.6. In the Bode plot the gain falls in proportion to modulation frequency while the phase shift is constant at -90° , offering a fixed phase margin of 90° . The gain margin is not meaningful since the phase never reaches -180° . In any practical circuit, however, phase shift will continue to become more negative at sufficiently high frequencies due to parasitic poles and delays.

The Nyquist plot is also very simple.

The root locus begins at the only open-loop pole and moves along the real axis to infinity, moving further from the region of instability at increasing values of K .

5.3.2 Second-Order Loop

We will find that, like the first-order loop, all second-order loops are inherently stable because the phase shift only *approaches* -180° . Again, gain margin is therefore meaningless (infinite), but phase margin is important because it indicates the sluggishness or “ringiness” of the responses and because parasitic effects can turn a theoretically stable loop unstable.

The loop with a low-pass filter is represented in Fig. 5.7, that with an integrator and lead is in Fig. 5.8, and Fig. 5.9 represents the loop with a lag-lead filter. Phase margin ϕ_m is shown in each case.

5.3.3 Third-Order Loop

We tend to stay away from higher-than-second-order loops in analysis because the number of variables is too large for compilation of standard response curves. Often we approximate higher-order loops as second order by ignoring a pole that is further removed from the critical unity-gain region than the other poles. However, higher order loops do occur in practice, and the stability analyses, according to the methods here, are easily applied. Figure 5.10 illustrates the plots for a third-order loop, showing

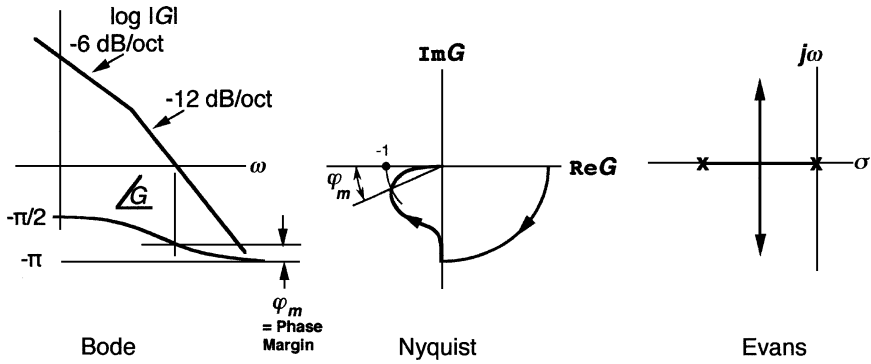


Fig. 5.7 Stability plots for a second-order type 1 loop with a low-pass filter.

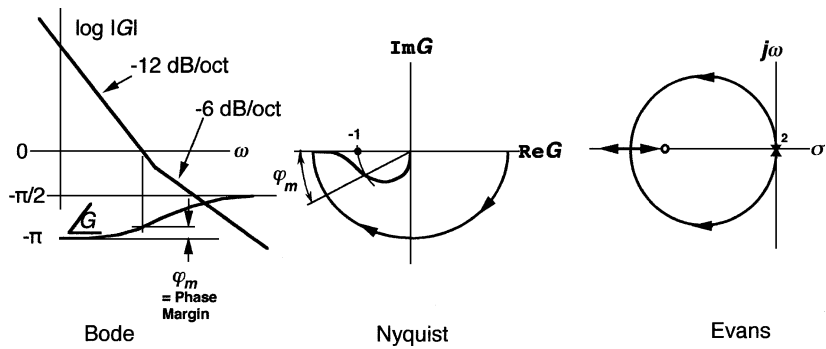


Fig. 5.8 Stability plots for a second-order type 2 loop.

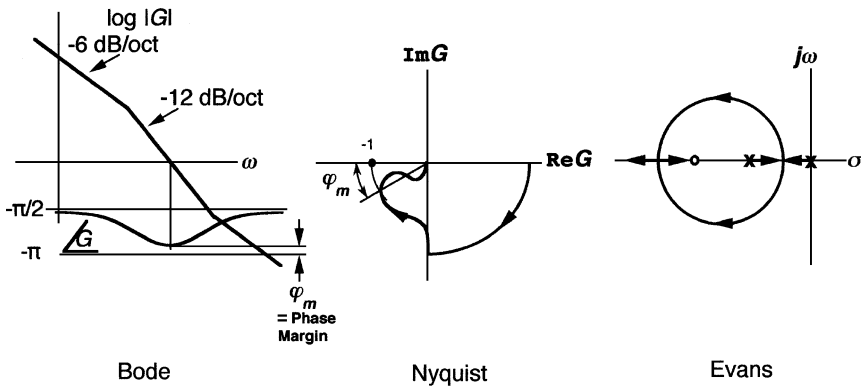


Fig. 5.9 Stability plots for a second-order type 1 loop with a lag-lead filter.

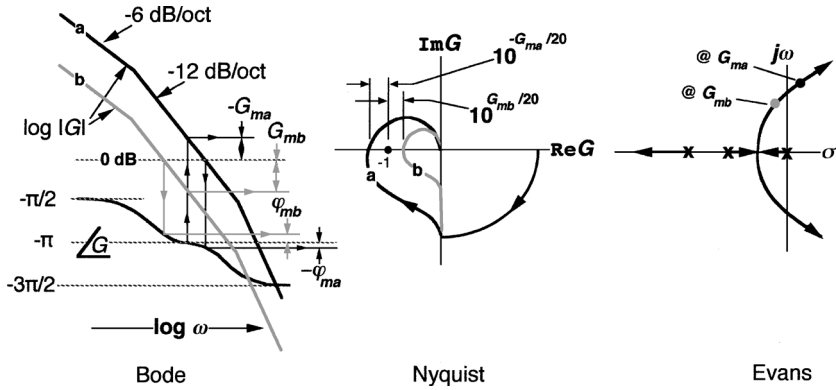


Fig. 5.10 Stability plots for a third-order type 1 loop at two values of gain factor K . This could represent a low-pass filter with a parasitic pole at a higher frequency.

gain and phase margins for both a stable value of gain (curve b) and an unstable value (curve a).

Figure 5.11 illustrates a conditionally stable third-order loop. Here a gain factor between K_a and K_b results in instability.

5.3.4 Controlling Stability over Wide Gain Ranges

Sometimes the gain of a loop varies over wide ranges. If the gain is not “rolled off,” it may be too high at modulation frequencies where parasitic phase shifts begin to appear. However, if it is rolled off too fast, there will be excessive phase shift, also leading to instability or poor performance. A solution is illustrated in Fig. 5.12. Here

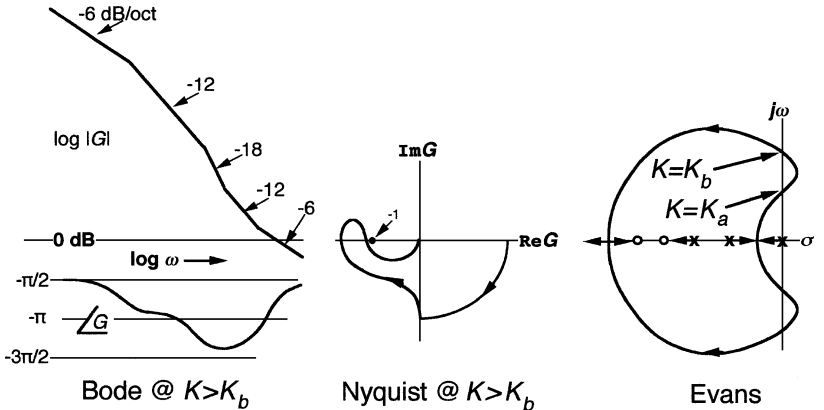


Fig. 5.11 Stability plots for a conditionally stable third-order loop. The loop is unstable for values of gain factor K between K_a and K_b .

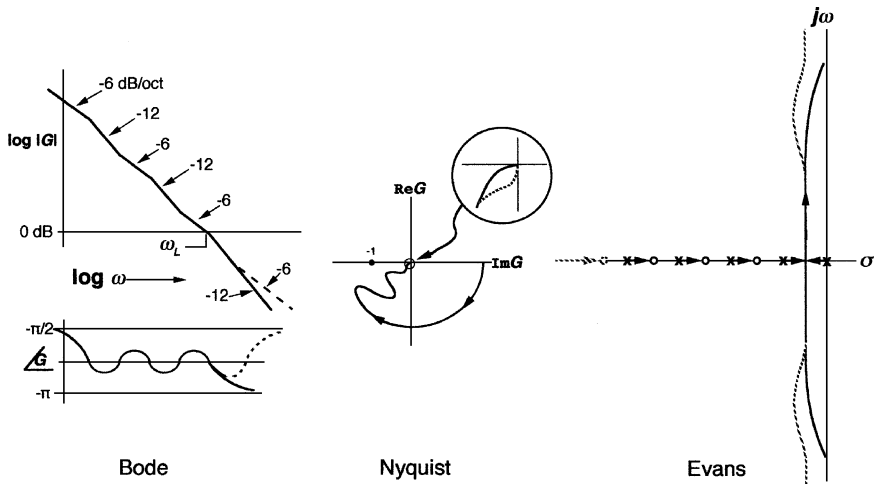


Fig. 5.12 Stability plots for a loop that is unconditionally stable over a wide range of frequencies.

a number of poles and zeros are alternated in the loop filter to give a roll-off rate faster than would be obtained with a first-order loop while maintaining the phase margin. The phase margin is never allowed to become too small, even though the gain can vary, moving the curve vertically and moving ω_L along the frequency axis. Loci are shown for the case where the highest frequency singularity is a pole and for the case where it is a zero.

5.3.5 Representing Delay

Usually any delay (transportation lag) in a loop will be much less than the reciprocal of the loop bandwidth and will therefore have little influence on stability or response.

A delay T_d is represented in Laplace transforms by

$$\exp(-sT_d) \Rightarrow \exp(-j\omega T_d), \tag{5.2}$$

a pure phase shift φ_T proportional to frequency. Phase, frequency, and voltage modulations, which apply to the loop's state variables, will be delayed by T_d and therefore

$$\varphi_T \Rightarrow -\omega_m T_d \tag{5.3}$$

should be placed in the loop where appropriate. Note that, while Eq. (5.2) may give a very large phase shift at high carrier (signal) frequencies, that phase shift does not enter into the loop model. The modulation frequency is the parameter used in the

loop [as in (5.3)], so it is wide loop bandwidth, rather than high carrier frequency, that aggravates the problem.

A phase shift proportional to ω_m can easily be introduced into the Bode phase plot. At the unity-gain frequency ω_L , the added phase shift will be $-\omega_L T_d$ and the phase margin will be correspondingly reduced. φ_T can also be added to the Nyquist diagram by shifting each point proportionally to the frequency represented. φ_T is more difficult to add to the Evans diagram; a number of poles and zeros must be used to approximate the frequency-dependent phase shift in regions of importance [Truxal, 1955, pp. 546–554].

5.4 COMPUTING OPEN-LOOP GAIN AND PHASE

We will use Figure 5.13 as an example to show how open-loop gain and phase can be computed at any frequency based on a knowledge of the critical frequencies and the known gain at some frequency [Kuo, 1987, pp. 693–706].

First note that a term that is proportional to the n th power of frequency (kf^n) is represented in decibels as

$$Q(f) = 20 \text{ dB} \log(kf^n) = 20 \text{ dB } n \log f + k'$$

Therefore, for each octave (2 : 1) change in frequency

$$\Delta Q(f) = 20 \text{ dB } n \log 2 = 6n \text{ dB}$$

and for each decade (10 : 1) change in frequency

$$\Delta Q(f) = 20 \text{ dB } n \log 10 = 20n \text{ dB}$$

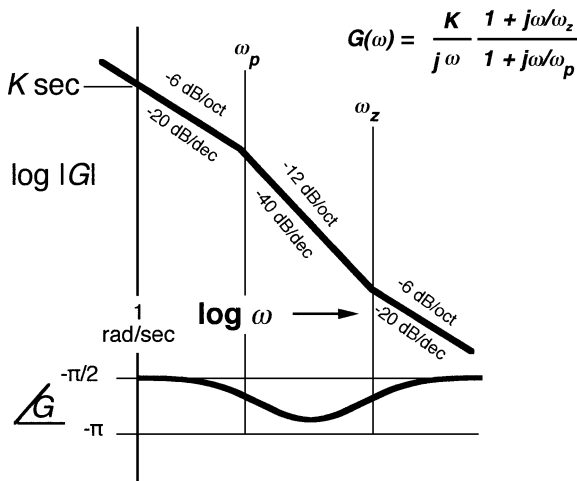


Fig. 5.13 Bode representation and corresponding equation.

We begin with the tangential approximation to open-loop gain. The value at $\omega = 1$ rad/sec is given as K sec. The slope there is -6 dB/octave. That is, gain is inversely proportional to ω . Therefore, the gain at any frequency ω along this curve is

$$|G(\omega)| = K/\omega. \quad (5.4)$$

Notice that this gives the required gain at $\omega = 1$ rad/sec. If the gain is known instead to be K_a at some frequency ω_a , the equation is

$$|G(\omega)| = K_a\omega_a/\omega, \quad (5.5)$$

giving the known value at $\omega = \omega_a$. If the value is known instead along the -12 dB/octave slope, where the gain is therefore inversely proportional to ω^2 , the gain is

$$|G(\omega)| = K_b(\omega_b/\omega)^2. \quad (5.6)$$

In each case the correct dependence on ω is provided (e.g., $1/\omega^2$), this is then multiplied by a constant (e.g., ω_b^2) to cause the resulting product to be unity at the known point, and the result is multiplied by the known gain at that point (e.g., K_b).

Suppose that the gain is desired at some frequency ω_c on the third segment of the curve. Then the gain is successively computed at each break point and finally at ω_c , basing each point upon the previous one. Gain at ω_p can be obtained from Eq. (5.4),

$$|G(\omega_p)| = K/\omega_p. \quad (5.7)$$

Then, using Eq. (5.6) with $\omega \Rightarrow \omega_z$, $\omega_b \Rightarrow \omega_p$, and $K_b \Rightarrow |G(\omega_p)|$, the value at ω_z can be found:

$$|G(\omega_z)| = |G(\omega_p)|(\omega_p/\omega_z)^2 = (K/\omega_p)(\omega_p/\omega_z)^2. \quad (5.8)$$

Finally the value at frequency ω_c is found using Eq. (5.5) with $\omega \Rightarrow \omega_c$, $\omega_a \Rightarrow \omega_z$, and $K_a \Rightarrow |G(\omega_z)|$:

$$|G(\omega_c)| = (K/\omega_p)(\omega_p/\omega_z)^2(\omega_z/\omega_c). \quad (5.9)$$

Notice also that this simplifies to

$$|G(\omega_c)| = (\omega_p/\omega_z)K/\omega_c, \quad (5.10)$$

which is the same as for a continuous -6 dB/octave ($= -20$ dB/decade) slope from ($\omega = 1$, $|G| = K$ sec) except that the gain has been reduced by the ratio of the frequencies at the ends of the -12 dB/octave ($= -40$ dB/decade) segment to take into account the additional $-20 \log_{10}(\omega_p/\omega_z)$.

The exact gain can be calculated from the formula,

$$|G(\omega)| = \left| \frac{K}{j\omega} \right| \left| \frac{1 + j\omega/\omega_z}{1 + j\omega/\omega_p} \right| = \frac{K}{\omega} \sqrt{\frac{1 + (\omega/\omega_z)^2}{1 + (\omega/\omega_p)^2}}. \quad (5.11)$$

Phase can be obtained from the formula as

$$\angle G(\omega) = \angle \frac{K}{j\omega} + \angle \left(\frac{1 + j\omega/\omega_z}{1 + j\omega/\omega_p} \right) \quad (5.12a)$$

$$= -\frac{\pi}{2} + \tan^{-1} \frac{\omega}{\omega_z} - \tan^{-1} \frac{\omega}{\omega_p}. \quad (5.12b)$$

The phase shift due to ω_z varies from $\tan^{-1}(0) = 0$ at $\omega = 0$ to $\tan^{-1}(1) = 45^\circ$ at $\omega = \omega_z$ to $\tan^{-1}(\infty) = 90^\circ$ at $\omega \Rightarrow \infty$. The phase shift is within about 18° of its final value, either 0° or 90° , when ω/ω_z equals $\frac{1}{3}$ or 3, respectively. The same is true for every zero. Poles produce similar phase shifts except for a change of sign. Generally the phase shift due to a critical frequency (pole or zero) at a frequency that is many times the frequency at which it is being computed can be ignored.

Example 5.1 Bode Plot Open-loop gain at 10 Hz is 60 dB. The loop filter is an integrator-and-lead filter with the zero at 200 Hz. There are additional poles at 3 kHz and 10 kHz. Find the phase and gain margins.

The equation for open-loop gain has the form of (4.22) except there are two more poles. Writing it explicitly for the given critical frequencies we obtain

$$G(s) = \frac{A(1 + s/\omega_z)}{s^2(1 + s/\omega_{p1})(1 + s/\omega_{p2})} \Rightarrow \frac{AF_b}{-(2\pi f_m/\text{cycle})^2 F_c F_d} \quad (5.13a)$$

where

$$F_b \equiv 1 + j \frac{f_m}{200 \text{ Hz}}, \quad F_c \equiv 1 + j \frac{f_m}{3000 \text{ Hz}}, \quad F_d \equiv 1 + j \frac{f_m}{10,000 \text{ Hz}} \quad (5.13b)$$

We begin by drawing the Bode tangential gain approximation plot as in Fig. Ex5.1. We start where the gain is given, at point a , and slope down at -40 dB/decade because of the $1/f^2$ dependency. In this frequency range, below the zero at b , F_b , F_c , and F_d are approximated as unity since the frequency-dependent term is smaller than the constant term in each factor.

As the modulation frequency increases, this approximation fails at the first zero encountered at b , where the two terms in F_b have equal magnitude. At higher frequencies the second term is larger so we approximate F_b by the magnitude of the second term, $f_m/200$ Hz. Therefore, beyond 200 Hz the net frequency dependence is f^{-1} and the slope is reduced to -20 dB/decade. Similarly, at c , the approximation for F_c

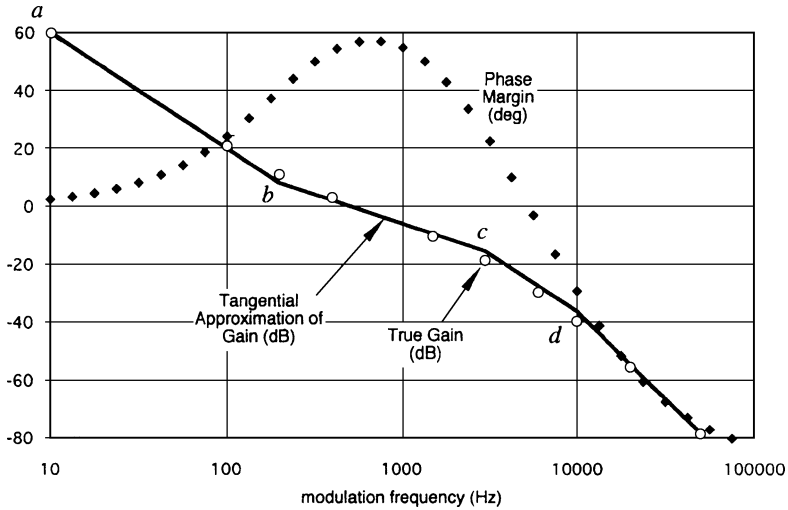


Fig. Ex5.1

transitions from unity to a term proportional to frequency, returning the net frequency dependence to f^{-2} and giving a -40 dB/decade slope. The same thing occurs at d , giving an ultimate slope of -60 dB/decade.

The tangential approximation is poorest at the corners, being off by 3 dB at simple singularities. The error is slightly greater than 3 dB at c and d since these poles are not sufficiently separated for the effect of one not to be seen at the other.

According to the tangential approximation, the gain at b is $[60 \text{ dB} - 40 \log_{10}(200/10) =] 8 \text{ dB}$. It crosses unity gain at a frequency that is higher than the zero by a factor of 8 dB or $10^{8/20} = 2.5$, thus 500 Hz. The phase shift at 500 Hz, due to the poles at 3 and 10 kHz, is

$$-\tan^{-1}(500/3000) - \tan^{-1}(500/10,000) = -9.46^\circ - 2.86^\circ = -12.32^\circ. \quad (5.14)$$

The zero at 200 Hz increases the phase by $\tan^{-1}(500/200) = 68.2^\circ$. The net from these three singularities is $+55.9^\circ$. This is also the phase margin because the two poles at the origin contribute -180° .

We can obtain the unity-gain frequency more accurately by trial and error using (5.13). A spreadsheet program can be very helpful in making such calculations. (The plots above were made from such a program.) Doing this, we find that the straight-line approximation is accurate, at 10 Hz, to one part in 1000 of $|G|$ and that the unity-gain frequency is actually 526 Hz rather than 500 Hz. At 526 Hz the phase margin is 0.3° greater than the approximate calculation showed. Thus the errors due to the tangential gain approximation are quite small.

By trial and error we find zero phase margin (180° excess phase shift) at 5235 Hz. The gain at that frequency is -27.5 dB . The gain margin is therefore $+27.5 \text{ dB}$.

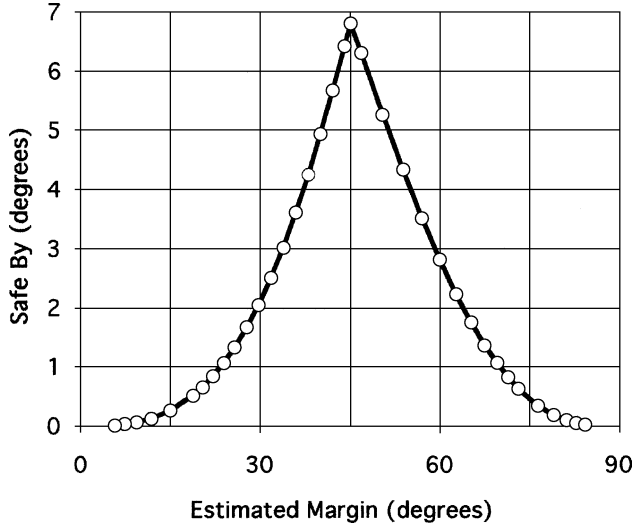


Fig. 5.14 Error in phase margin due to tangential gain approximation, single corner. This curve applies to both a lag filter (Fig. 5.7) and an integrator-and-lead filter (Fig. 5.8) in a PLL. (This applies specifically to a PLL because the additional $1/s$ term affects the result.)

When we use the Bode tangential gain approximation to compute the unity-gain frequency, the inexactness of the computed frequency will cause an inaccuracy in the computed phase margin. To gain an idea of how large this error might be, we will consider some specific cases.

Figure 5.14 shows how large that error is in the vicinity of the pole of a lag filter or the zero of an integrator and lead filter. The errors are the same in the two cases when

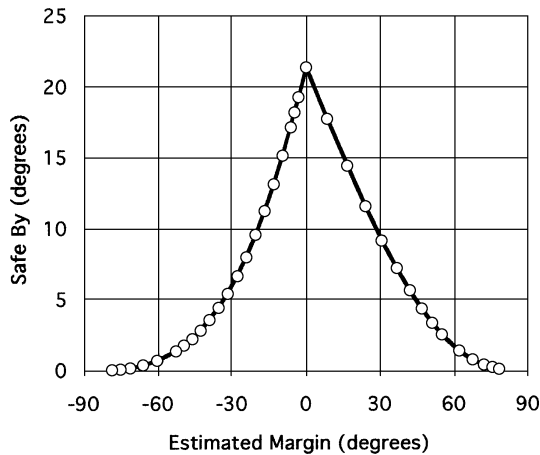


Fig. 5.15 Error in phase margin due to tangential gain approximation, double-pole loop filter. This curve applies to a PLL that has a lag filter with two superimposed poles.

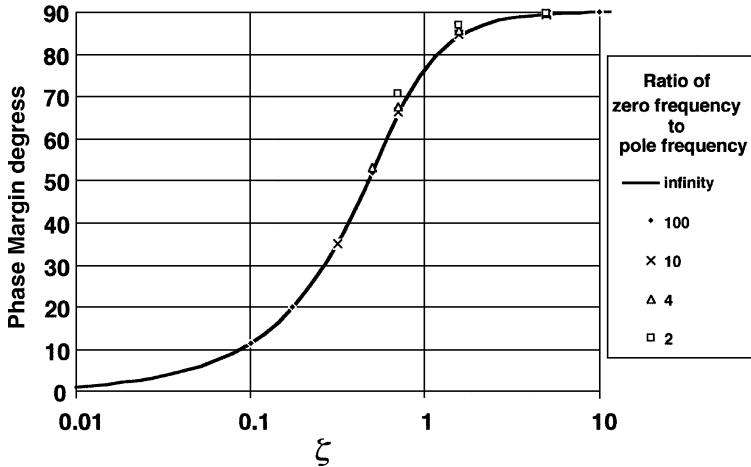


Fig. 5.16 Relationship between phase margin and damping factor, second-order loop. Solid curve applies to both integrator-and-lead and to low-pass loop filters. Data points are for lag-lead loop filters.

they are given as a function of the phase margin that is estimated from the tangential gain plot. The estimated margins are pessimistic. This plot is mainly of value for giving an indication of the magnitude of the inaccuracy because additional singularities are required for instability, and they would change the assumed conditions.

Figure 5.15 shows similar data in the vicinity of the double pole of a two-pole lag filter. Here the estimate is again pessimistic but by a larger amount. This is a severe case because the error at the corner is accentuated by the presence of two superimposed poles. However, it does represent a configuration that can be unstable without the addition of other singularities.

5.5 PHASE MARGIN VERSUS ζ DAMPING FACTOR

Figures 4.5 and 4.6 indicate that small phase margin and small damping factor tend to go together. Figure 5.16 confirms this. It shows phase margin versus damping factor. The solid curve applies to a loop with either a low-pass or an integrator-and-lead filter. Separate points are shown for lag-lead filters having various ratios of zero to pole frequencies. For each of these ratios, a point is shown at the lowest ζ that can occur with that zero-pole ratio and at some higher values of ζ . The solid curve applies well for large ratios of zero to pole frequencies.

i5.M APPENDIX: STABILITY PLOTS USING MATLAB®

All of the stability plots in this chapter can easily be generated using MATLAB, as shown in this appendix, available from the Wiley Internet site at ftp://ftp.wiley.com/public/sci_tech_med/phase_lock.

CHAPTER 6

TRANSIENT RESPONSE

One important aspect of phase-locked loop (PLL) performance is the manner in which it responds to a change in input frequency or phase, and one of the most important measures is its response to a step change of the input. Does the output reach its final value quickly or slowly? Does it overshoot and, if so, how long does it take to come finally within a specified offset from the final value? How great is the phase error during a frequency step? Does it surpass the range of the phase detector? To answer such questions we need a set of curves that show time responses under various conditions.

6.1 STEP RESPONSE

For a linear system, the step response can be obtained by use of Laplace transforms. The transform of a unit step, $1/s$, multiplied by $H(s)$ from Eq. (4.10) or (4.20), can be separated, by finding the roots and employing the Heaviside (partial fraction) expansion theorem [Thomas and Finney, 1984], into parts for which transform pairs are available. The error response, $[1 - H(s)]/s$, can be similarly obtained or can be simply taken as $[1/s - H(s)/s]$. Step responses can then be plotted for various combinations of loop parameters. Either output or error can be plotted since one is easily obtainable from the other by Eqs. (2.28).

6.1.1 Form of the Equations

The challenge is to find a reasonably small set of curves that covers a large number of cases of practical importance. If we restrict our attention to linear performance,

as we will in this chapter, the response to any size input step will be proportional to that step, so a curve of unit step response will suffice to indicate the response to all step sizes. Another device for limiting the number of curves is to restrict our study to loops of order 1 or 2. We have already shown the step response for a first-order loop (Figs. 2.2 and 2.6). There, just one curve suffices if its abscissa is normalized to the open-loop gain at $\omega = 1$ rad/sec, that is, if $x = Kt$. For second-order loops, curves can be plotted against normalized time, $\omega_n t$, for various damping factors ζ . However, Eq. (4.10) has an additional variable, ω_z , that threatens to add an additional, potentially intractable, dimension. As a first step, for integrator-and-lead filters ω_z can be eliminated from Eq. (4.10). This is done using Eq. (4.16) to give

$$H(s) = \omega_n^2 \frac{1 + s(2\zeta/\omega_n)}{s^2 + 2\zeta\omega_n s + \omega_n^2}. \quad (6.1)$$

The expression now contains a larger number of individual variables, but it is characterized entirely by two parameters, ω_n and ζ . Thus two sets of step response curves can be generated, one for a lag filter, based on Eq. (4.20), and the other for the integrator and lead, based on Eq. (6.1). Fortunately, responses for the more general lag-lead filter can be obtained from a combination of these, as we shall now see.

From Eq. (4.9) $1/\omega_z$ can be written in terms of ζ :

$$\frac{1}{\omega_z} = \frac{2\zeta\omega_n - \omega_p}{\omega_n^2}, \quad (6.2)$$

and ω_p can be eliminated by using Eq. (4.7) to give

$$\frac{1}{\omega_z} = \frac{2\zeta}{\omega_n} - \frac{1}{K}. \quad (6.3)$$

Substituting this expression into Eq. (4.10), we obtain

$$H(s) = \omega_n^2 \frac{1 + s(2\alpha\zeta/\omega_n)}{s^2 + 2\zeta\omega_n s + \omega_n^2} \quad (6.4a)$$

and

$$1 - H(s) = \frac{s^2 + 2\zeta\omega_n(1 - \alpha)s}{s^2 + 2\zeta\omega_n s + \omega_n^2}, \quad (6.4b)$$

where

$$\alpha \equiv 1 - \frac{\omega_n}{2\zeta K} \quad (6.5a)$$

$$\equiv \frac{1}{1 + \omega_z/K} \quad (6.5b)$$

$$\equiv \frac{1}{1 + \omega_z\omega_p/\omega_n^2}. \quad (6.5c)$$

Equation (6.5c) shows that $\alpha \Rightarrow 1$ as the geometric mean of the pole and zero frequencies becomes small compared to the natural frequency. Thus, in Fig. 4.3, α increases from curve *c* to curve *a*. For a lag filter, $\omega_z \Rightarrow \infty$ so $\alpha \Rightarrow 0$ and Eq. (6.4a) becomes the same as Eq. (4.20). For an integrator-and-lead filter, $\omega_p \Rightarrow 0$ (and $K \Rightarrow \infty$) so $\alpha \Rightarrow 1$.

Lag filter	$\alpha \Rightarrow 0,$
Integrator-and-lead filter	$\alpha \Rightarrow 1.$

But Eq. (6.4a) can also be written

$$H(s)|_{\text{lag-lead}} = (1 - \alpha)H(s)|_{\text{lag}} + \alpha H(s)|_{\text{integrator-and-lead}}, \quad (6.6)$$

which can be easily verified by substituting Eq. (6.4a) with $\alpha = 0$ for $H(s)|_{\text{lag}}$ and Eq. (6.4a) with $\alpha = 1$ for $H(s)|_{\text{integrator-and-lead}}$ in Eq. (6.6) to obtain Eq. (6.4a). Therefore the step response with any lag-lead filter can be obtained from curves for a loop with a lag filter and curves for a loop with an integrator-and-lead filter, each for the same ζ as the desired curve, by combining their values at each (normalized) time according to Eq. (6.6). The same relationship holds for the error response, as can be seen by subtracting Eq. (6.6) from 1:

$$\begin{aligned} 1 - H(s)|_{\text{lag-lead}} &= (1 - \alpha) + \alpha - H(s)|_{\text{lag-lead}} \\ &= (1 - \alpha)[1 - H(s)]_{\text{lag}} + \alpha[1 - H(s)]_{\text{integrator-and-lead}}. \end{aligned} \quad (6.7)$$

Equations (6.6) and (6.7) are impulse responses; they can be multiplied by $1/s$ to give step response, or by $1/s^2$ to give the response to ramp inputs, and so on. Thus a relatively small set of curves can indicate the transient response for common second-order loops regardless of filter type.

6.1.2 Step Response Equations

The step error response of the PLL in the time domain could represent the phase error $\varphi_e(t)$ in response to a unit phase step at the input or the frequency error $\omega_e(t)$ in response to a unit frequency step at the input. It is obtained from the inverse Laplace transform of $(1/s)[1 - H(s)]$, giving (for $t > 0$):

$$\begin{aligned} Y_{u, \text{error}} &= \frac{1}{2} \left[\left(\zeta \frac{2\alpha - 1}{\sqrt{\zeta^2 - 1}} + 1 \right) e^{-\omega_n t (\zeta + \sqrt{\zeta^2 - 1})} \right. \\ &\quad \left. - \left(\zeta \frac{2\alpha - 1}{\sqrt{\zeta^2 - 1}} - 1 \right) e^{-\omega_n t (\zeta - \sqrt{\zeta^2 - 1})} \right]. \end{aligned} \quad (6.8)$$

This form is most useful for $\zeta > 1$, for which an alternate form is

$$Y_{u, \text{error}}(t) = e^{-\zeta\omega_n t} \left[\cosh(\omega_n t \sqrt{\zeta^2 - 1}) + \zeta \frac{1 - 2\alpha}{\sqrt{\zeta^2 - 1}} \sinh(\omega_n t \sqrt{\zeta^2 - 1}) \right]. \quad (6.9)$$

For $\zeta = 1$ Eq. (6.8) becomes

$$Y_{u, \text{error}}(t) = e^{-\omega_n t} [1 + (1 - 2\alpha)\omega_n t], \quad (6.10)$$

while for $\zeta < 1$ it is

$$Y_{u, \text{error}}(t) = e^{-\zeta\omega_n t} \left[\cos(\omega_n t \sqrt{1 - \zeta^2}) + \zeta \frac{1 - 2\alpha}{\sqrt{1 - \zeta^2}} \sin(\omega_n t \sqrt{1 - \zeta^2}) \right]. \quad (6.11)$$

In each case the output response, that is, the phase $\varphi_{\text{out}}(t)$ at the voltage-controlled oscillator (VCO) output after a unit step in $\varphi_{\text{in}}(t)$ at the loop input or the frequency $\omega_{\text{out}}(t)$ at the VCO output after a unit frequency step in $\omega_{\text{in}}(t)$ at the loop input, is

$$Y_{u, \text{out}} = 1 - Y_{u, \text{error}}. \quad (6.12)$$

Figure 6.1 shows the output response to a step for $\zeta = 1$ and five values of α . It is apparent from this how each curve is a weighted average of the curves for $\alpha = 0$ and $\alpha = 1$, according to Eq. (6.6).

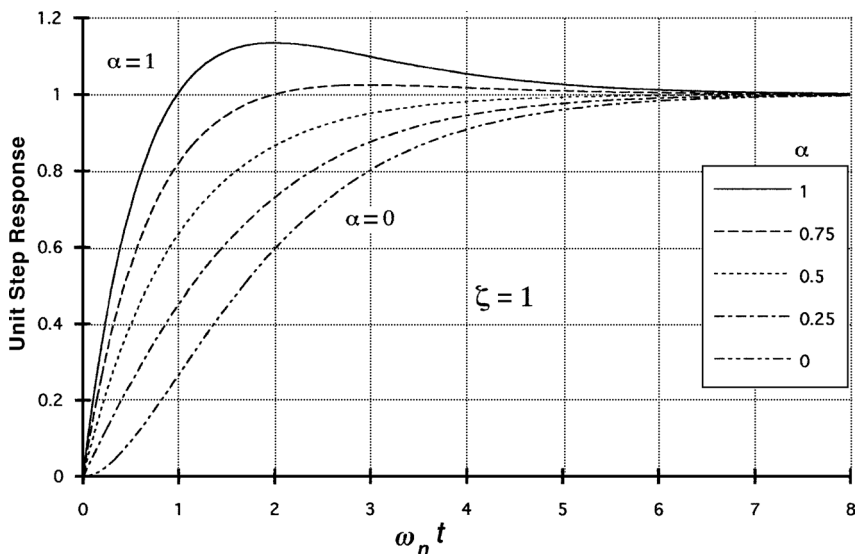


Fig. 6.1 Output response to a step input with $\zeta = 1$ for five values of α .

Example 6.1 Step Response, $0 < \alpha < 1$

- a. The frequency input to a loop steps 10 kHz. What will be the change in output frequency 5 msec after the input step if $\omega_n = 600$ rad/sec, $\zeta = 1$, and $\alpha = 0.25$, where $\omega_n t$ equals 3?

Look at Fig. 6.1 at $\omega_n t = 3$ and $\alpha = 0.25$, the second curve from the bottom. The unit step response there is 0.88 so the output equals that number times the magnitude of the input step, thus 8.8 kHz. If we use (6.10), for improved accuracy compared to reading the curves, we obtain 8.753 kHz.

- b. Repeat for $\alpha = 0.9$.

Again use Fig. 6.1. For $\alpha = 0$, the response is 0.8. For $\alpha = 1$, the response is 1.1. From (6.6), the step response for $\alpha = 0.9$ is

$$Y_{u,\text{out}} = 0.9 \times 1.1 + 0.1 \times 0.8 = 1.07.$$

Thus the output would overshoot to 10.7 kHz at the specified time. The equation gives 10.697 kHz.

- c. Repeat for $\alpha = 0.4$ and $\zeta = 0.707$.

We will first obtain the error response from Fig. 6.3, since no output response is shown for this case. The responses at $\omega_n t = 3$ for α equal to zero and one are, respectively, 0.04 and -0.16 . Multiplying the latter by 0.4 and the former by $(1 - 0.4)$ we obtain a sum of -0.040 so the error at that time is -0.040×10 kHz or -400 Hz. Since the output at zero error would be 10 kHz, the output has stepped 10.40 kHz at 5 msec. Equation (6.9) gives 10.423 kHz.

- d. Using the same α and ζ , what is the output phase change 5 msec after a 20° phase step at the input?

As before, the step output is 1.042 times the step input, that is, 20.84° .

Figure 6.2 shows the step error response for $\zeta \geq 1$ and Fig. 6.3 shows it for $\zeta < 1$. It is desirable to plot the curves against a variable that is normalized in a way that keeps the curves close together. Not only does this make more feasible the plotting of multiple curves on the same graph, but it facilitates interpolation between curves. This goal is not satisfied especially well in Fig. 6.2. For $\zeta \geq 1$, it is better to plot the curves against the unity-gain frequency ω_L , as can be seen in Fig. 6.4. (ω'_L is the unity-gain frequency of the tangential approximation of the Bode plot and is approximately equal to ω_L , especially for large ζ .) Here the curves become almost coincident for large ζ . With this normalization, a single curve is sufficient for ζ greater than about 2. In fact, that curve is given, for both $\alpha = 0$ and $\alpha = 1$, by $\exp(-\omega_L t)$, the equation

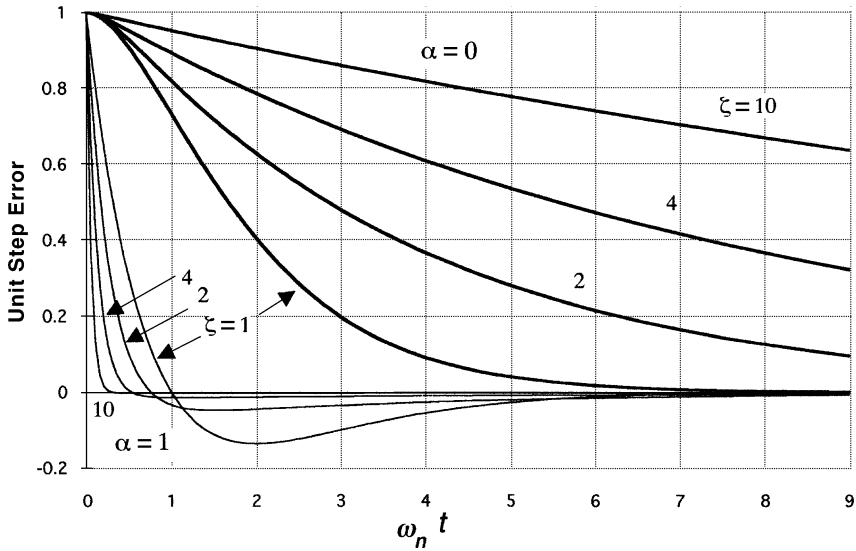


Fig. 6.2 Error response to a step input with $\zeta \geq 1$ versus $\omega_n t$.

for the step response of a first-order loop. [This is identical to Eq. (2.12a).] We might have expected this because, for $\zeta > 1$, ω'_L occurs where the slope of the Bode plot is -6 dB/octave, as for a first-order loop, and, as ζ increases, the -12 dB/octave region moves further from ω_L (curve 3 in Fig. 4.5 and 2 in 4.6). As noted in Section 4.1, when two loops have similar $G(\omega)$ in the region of ω_L , their gross performance is similar.

The approximate value for ω_L for $\zeta \geq 1$ can be obtained from the geometry illustrated in Figs. 4.5 and 4.6. From Fig. 4.5, for the low-pass filter,

$$\omega_L \approx \omega'_L = \omega_n / (2\zeta), \tag{6.13}$$

while, from Fig. 4.6, for the integrator-and-lead filter,

$$\omega_L \approx \omega'_L = 2\zeta \omega_n. \tag{6.14}$$

For $\zeta < 0.5$, $\omega_L \approx \omega_n$ so Fig. 6.3 is already plotted approximately against the unity-gain frequency.

6.2 ENVELOPE OF THE LONG-TERM STEP RESPONSE

Figures 6.5 and 6.6 are log plots of the error response that are used to show the error, as the value becomes small, after many time constants. Figure 6.5 is for $\zeta \leq 1$.

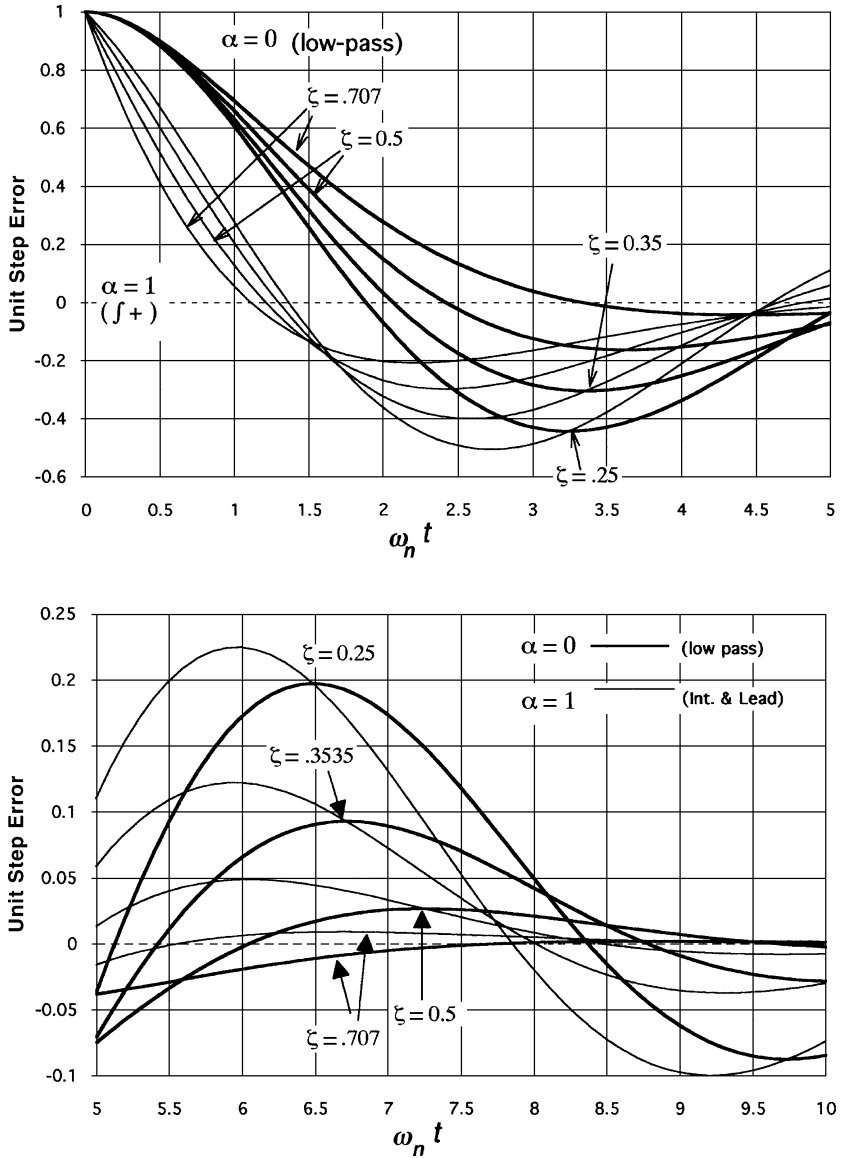


Fig. 6.3 Error response to a step input with $\zeta < 1$.

Since the response for $\zeta < 1$ is oscillatory, it is only practical to show the envelope of the response. This is obtained as the magnitude of the vector sum of two normal phasors having the amplitudes of the sine and cosine components in Eq. (6.11). For small ζ the response depends little on α , so only one curve is shown in those cases.

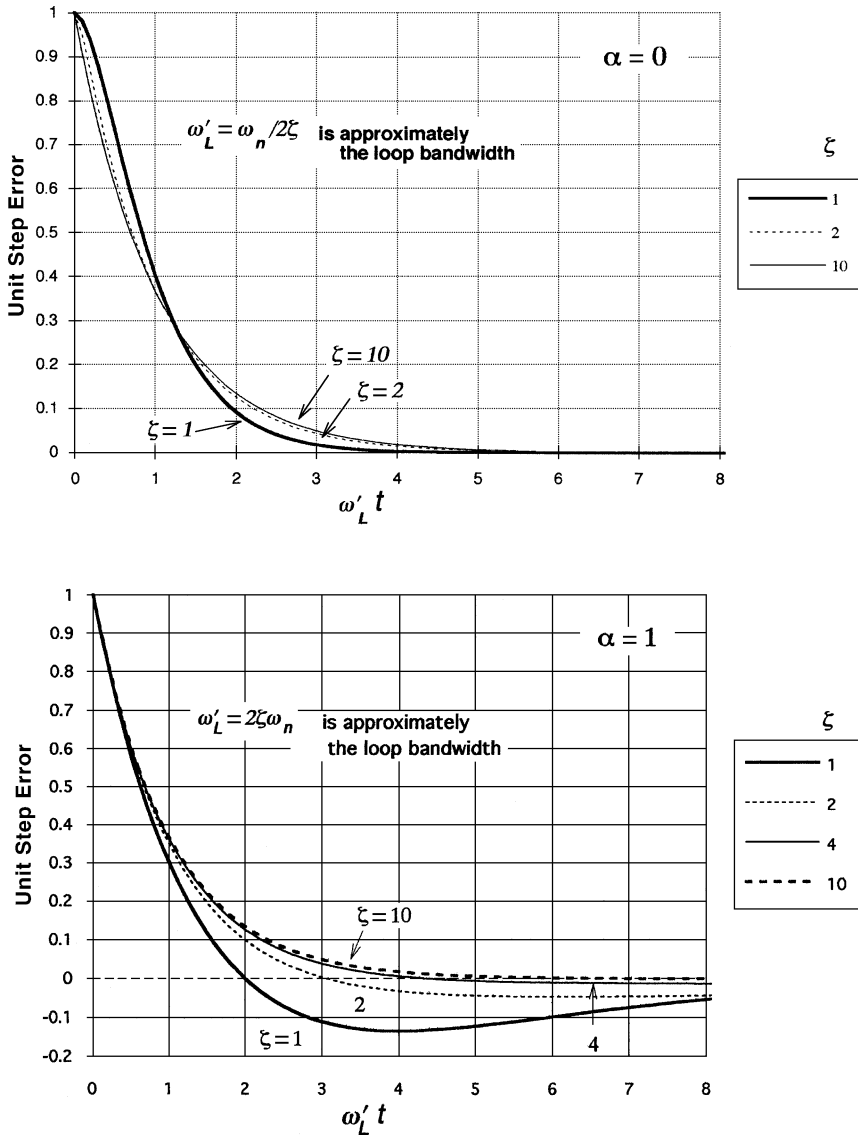


Fig. 6.4 Error response to a step input with $\zeta \geq 1$ versus $\omega'_L t$.

Example 6.2 Envelope of Step Response How long will it take for the error after a frequency step to be reduced to 10^{-6} of the step size if $\zeta = 0.5$ and $\omega_n = 10^4$ rad/sec?

Figure 6.5 shows that $\omega_n t \approx 27.7$ for $\zeta = 0.5$, regardless of the value of α , when the envelope has been reduced to 10^{-6} of the step size. This occurs at 2.77 msec.

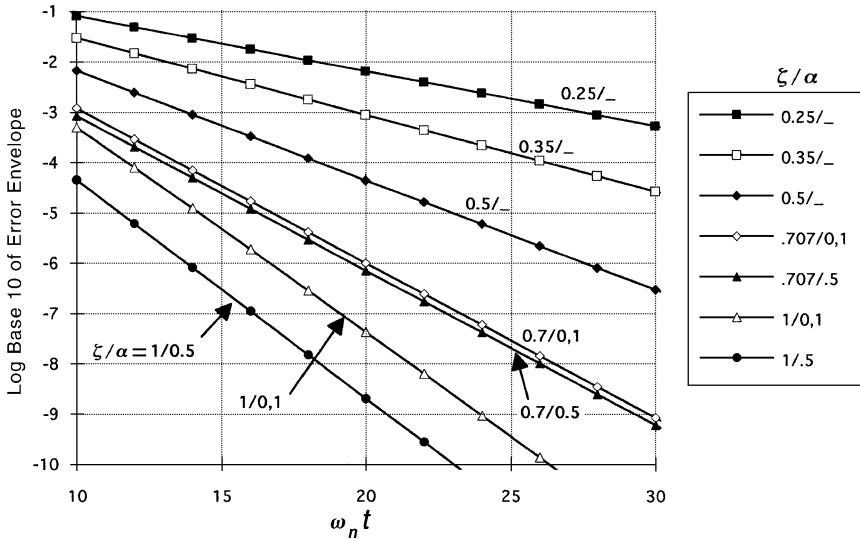


Fig. 6.5 Log envelope of the error response to a step input for large $\omega_n t$ with $\zeta \leq 1$. Curves are labeled with ζ and α , separated by a slash.

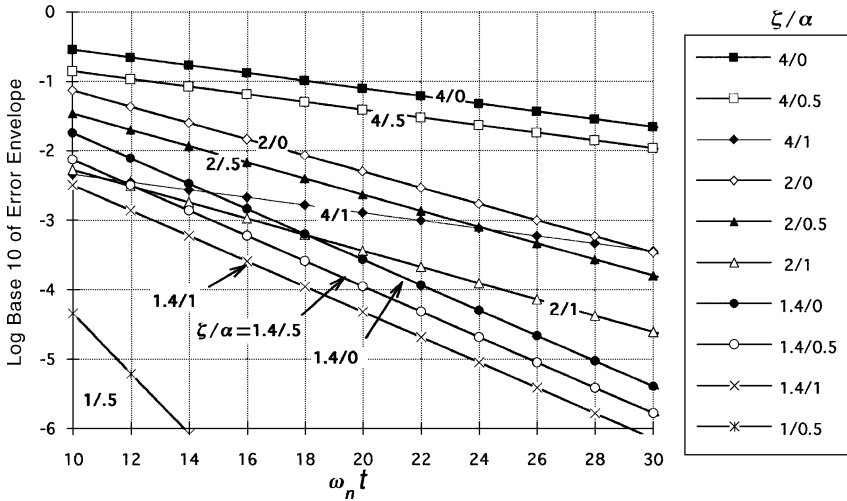


Fig. 6.6 Log envelope of the error response to a step input for large $\omega_n t$ with $\zeta > 1$. Curves are labeled with ζ and α , separated by a slash.

The error may be smaller than its envelope, depending on where the peaks of the overshoots occur, but cannot be larger.

6.3 RESPONSE TO RAMP INPUT

The integral of Eq. (6.8) is an important quantity because it gives the phase error in response to a frequency step, and this is important in determining whether operation is confined to the linear portion of the phase detector curve, as is assumed in developing the linear equations. It can be obtained by using $1/s^2$ as the forcing function—that is, taking the inverse Laplace transform of $(1/s^2)[1 - H(s)]$ —or by integration in the time domain. (Since integration is a linear function, the integral of the response to a given input is the same as the response to the integral of that input.) The same response applies to either frequency or phase when the input is a ramp of either frequency or phase, respectively. The phase response to a phase ramp is also the phase response to a frequency step.

By integrating Eq. (6.8), we obtain the response to a unit ramp (t , starting at zero). The forms for $\zeta > 1$, $\zeta = 1$, and $\zeta < 1$, respectively, are:

$$Y_{r, \text{error}}(t) = 2\zeta \frac{1 - \alpha}{\omega_n} - \frac{e^{-\zeta\omega_n t}}{\omega_n} \times \left[2\zeta(1 - \alpha)\cosh(\sqrt{\zeta^2 - 1}\omega_n t) + \frac{2\zeta^2(1 - \alpha) - 1}{\sqrt{\zeta^2 - 1}}\sinh(\sqrt{\zeta^2 - 1}\omega_n t) \right], \quad (6.15)$$

$$Y_{r, \text{error}}(t) = \frac{1}{\omega_n} \{2(1 - \alpha) - e^{-\omega_n t} [2(1 - \alpha) + (1 - 2\alpha)\omega_n t]\}, \quad (6.16)$$

$$Y_{r, \text{error}}(t) = 2\zeta \frac{1 - \alpha}{\omega_n} - \frac{e^{-\zeta\omega_n t}}{\omega_n} \times \left[2\zeta(1 - \alpha)\cos(\sqrt{1 - \zeta^2}\omega_n t) + \frac{2\zeta^2(1 - \alpha) - 1}{\sqrt{1 - \zeta^2}}\sin(\sqrt{1 - \zeta^2}\omega_n t) \right]. \quad (6.17)$$

It is convenient to normalize these equations to show the response to a ramp of slope ω_n (by multiplying the equations by ω_n). These responses are shown in Figs. 6.7 through 6.9.

Example 6.3 Ramp Response

- a. The input frequency to a PLL is changing at a uniform rate of 1 kHz/sec. What is the error at the output 20 msec after the ramp starts if $\zeta = 0.5$, $\alpha = 0$, and $\omega_n = 100$ rad/sec?

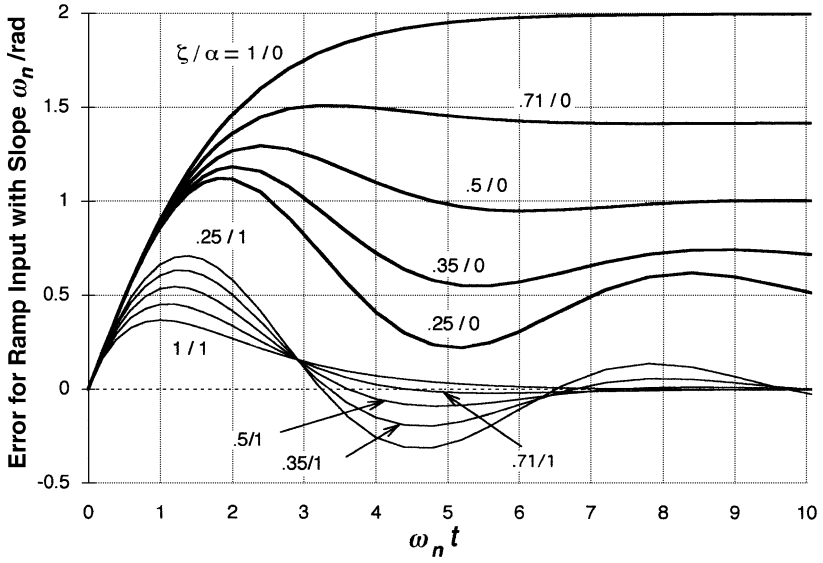


Fig. 6.7 Error response to a ramp input with $\zeta \leq 1$. Curves are labeled with ζ and α , separated by a slash.

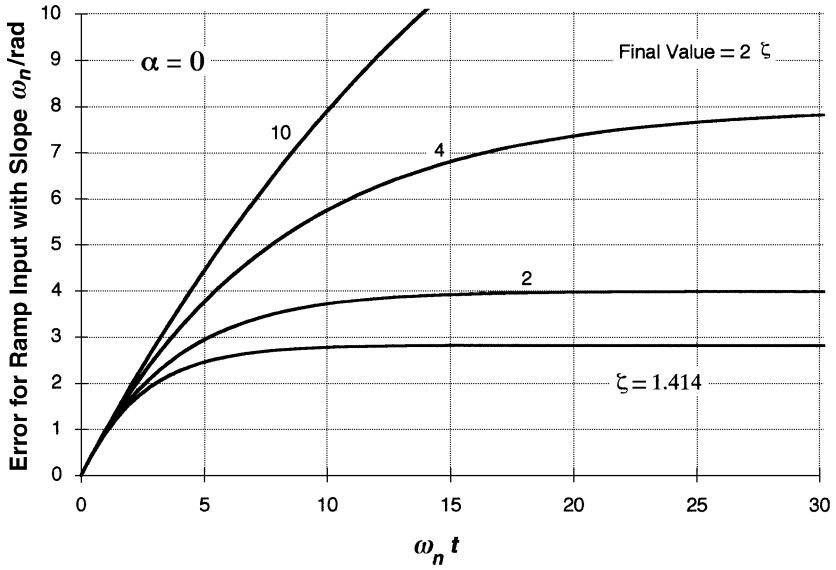


Fig. 6.8 Error response to a ramp input with $\zeta > 1$ for $\alpha = 0$.

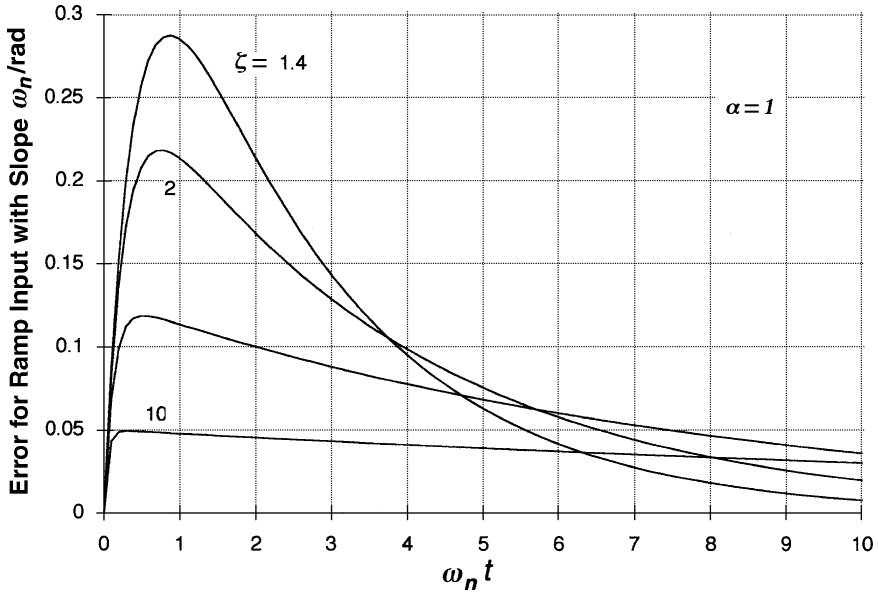


Fig. 6.9 Error response to a ramp input with $\zeta > 1$ for $\alpha = 1$.

Use Fig. 6.7 at $\zeta = 0.5$, $\omega_n t = 2$. There, according to the graph, the output error will be 1.26 if the input slope is $\omega_n = 100/\text{sec}$. We scale the input and response, multiplying them by 10 Hz, to obtain the given input ramp. The output error is then 12.6 Hz. Thus the output will lag the input ramp by 12.6 Hz at 20 msec. Equation (6.17) gives 12.687 Hz.

- b. Using the same parameters, what will be the output phase error at 20 msec if the input frequency steps 1 kHz?

A 1-kHz frequency step is a phase ramp of 1000 cycles/sec (by definition). Therefore, the problem is basically the same as the previous problem except we scale by 10 cycles, rather than 10 Hz, this time. As a result, the output phase will lag the input phase by 12.6 cycles at 20 msec.

6.4 RESPONSE TO PARABOLIC INPUT

The response to a parabolic input, $t^2/2$ beginning at zero, can be obtained by integrating the above responses to a ramp input (or by taking the inverse Laplace transform of $(1/s^3)[1 - H(s)]$). If the input frequency is represented by a unit ramp t , then the input phase is the integral of that frequency with respect to time, $t^2/2$. The phase output in response to this input is, for $\zeta > 1$, $\zeta = 1$ and $\zeta < 1$,

respectively:

$$\begin{aligned}
 Y_{p, \text{ error}}(t) = & \frac{1 - 4\zeta^2(1 - \alpha)}{\omega_n^2} + 2\zeta \frac{1 - \alpha}{\omega_n} t \\
 & + \frac{e^{-\zeta\omega_n t}}{\omega_n^2} \left[\zeta \frac{2(1 - \alpha)(2\zeta^2 - 1) - 1}{\sqrt{\zeta^2 - 1}} \sinh(\sqrt{\zeta^2 - 1}\omega_n t) \right. \\
 & \left. + [4\zeta^2(1 - \alpha) - 1] \cosh(\sqrt{\zeta^2 - 1}\omega_n t) \right], \tag{6.18}
 \end{aligned}$$

$$Y_{p, \text{ error}}(t) = \frac{t}{\omega_n} [2(1 - \alpha) - e^{-\omega_n t}(1 - 2\alpha)] + \frac{3 - 4\alpha}{\omega_n^2} (e^{-\omega_n t} - 1), \tag{6.19}$$

$$\begin{aligned}
 Y_{p, \text{ error}}(t) = & \frac{1 - 4\zeta^2(1 - \alpha)}{\omega_n^2} + 2\zeta \frac{(1 - \alpha)}{\omega_n} t \\
 & + \frac{e^{-\zeta\omega_n t}}{\omega_n^2} \left[\zeta \frac{2(1 - \alpha)(2\zeta^2 - 1) - 1}{\sqrt{1 - \zeta^2}} \sin(\sqrt{1 - \zeta^2}\omega_n t) \right. \\
 & \left. + [4\zeta^2(1 - \alpha) - 1] \cos(\sqrt{1 - \zeta^2}\omega_n t) \right]. \tag{6.20}
 \end{aligned}$$

Figures 6.10 and 6.11 show the error in response to a parabolic input $0.5(\omega_n t)^2$, which is obtained by multiplying the above equations by ω_n^2 , causing them to become functions of $\omega_n t$. Note that, unless $\alpha = 1$, the response eventually climbs.

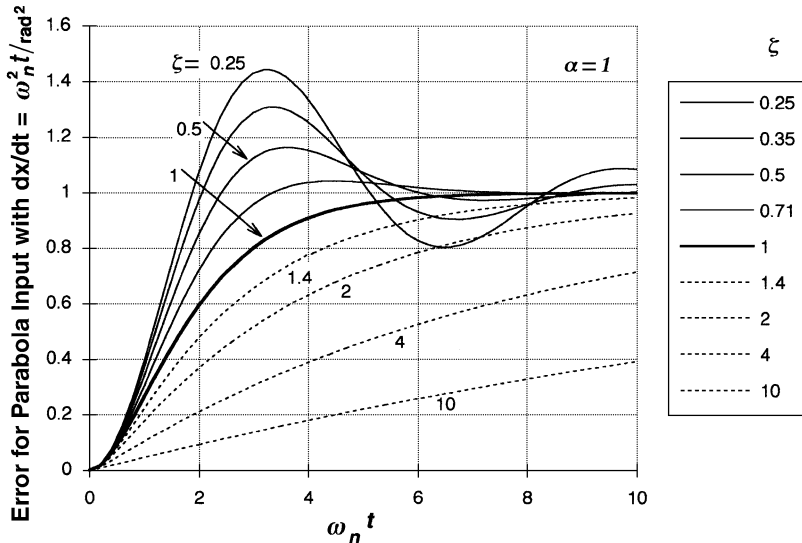


Fig. 6.10 Error response to a parabolic input with $\alpha = 1$.

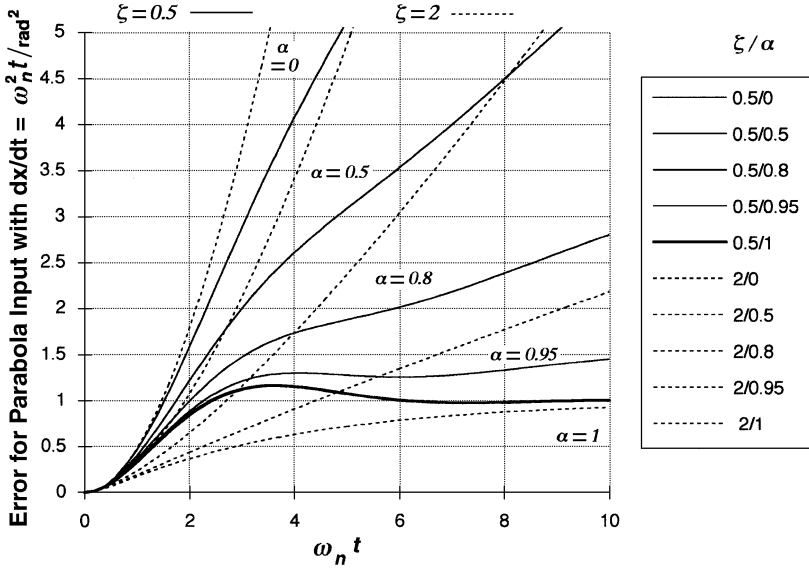


Fig. 6.11 Error response to a parabolic input with $\zeta = 0.5$ and 2.

Example 6.4 Response to a Parabolic Input The frequency into a loop is increasing at a rate of 1 kHz/sec. What is the output phase error 1 msec after the start of this frequency ramp if $\zeta = 0.5$, $\alpha = 0.5$, and $\omega_n = 7$ krad/sec?

We are seeking the phase error when the input phase has a time derivative (frequency) that is proportional to time so we use Fig. 6.11. At $\omega_n t = 7$ the error is 4. This is in response to an input with a time derivative of $\omega_n^2 t / \text{rad}^2 = 49 \times 10^6 t / \text{sec}^2$. We multiply the input and output by 10^3 cycles/ 49×10^6 to scale the graph to our problem. (The input now is a frequency ramp with a slope of t kHz/sec, which is also a phase parabola with a second derivative 10^3 cycles/sec.) The output, 1 msec after the start of the frequency ramp, is thus

$$(10^3 \text{ cycles}/49 \times 10^6)(4) = 8.2 \times 10^{-5} \text{ cycles.}$$

Equation (6.20) gives 8.1×10^{-5} cycles.

If the loop filter had been an integrator-and-lead ($\alpha = 1$, Figs. 6.10 and 6.11), the phase error would have settled to 2×10^{-5} cycles whereas it will continually increase otherwise.

6.5 OTHER RESPONSES

The closed-loop transfer function, from any point of signal injection to a point just prior to the injection, is $-H(s)$. The closed-loop transfer function, from any point of signal injection to a point just after the injection, is $1 - H(s)$. The usual point of

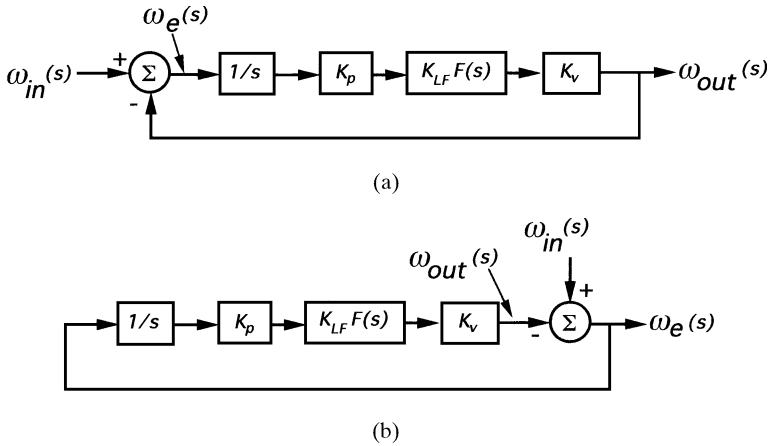


Fig. 6.12 Two different injection points, (a) Represents excitation at the reference input. (b) Represents a change in the open-loop VCO frequency. The observable frequency is shown as output in both cases.

injection is to the summer that represents part of the phase detector, as shown in Fig. 6.12a. However, another important injection point is just after the VCO, as shown in Fig. 6.12b.

This might represent, for example, a change in the open-loop frequency of the VCO (i.e., the center frequency) due to some undesired influence such a power supply voltage change. Then, with the loop closed, the VCO's output is represented by the error response, $1 - H(s)$, to the disturbance, whereas $-H(s)$ represents the ratio of the tuning voltage (in terms of equivalent frequency) to the injected disturbance. See Sections 7.4 and 7.3 for further discussion.

6.6 NOTE ON UNITS FOR GRAPHS

The graphs are normalized. Most are for a unit step so the ordinate must be multiplied by the magnitude of the step that is actually encountered. For example, if the step is 0.1 rad, the value given for the response on the curve must be multiplied by 0.1 rad. If the step is 10 Hz and the response value given by the graph at the appropriate value of $\omega_n t$ is 0.4, that actually represents $0.4 \times 10 \text{ Hz} = 4 \text{ Hz}$. In the case of the response to an input ramp (e.g., Fig. 6.9), the graph is given for an input slope of ω_n . If the input slope is 0.1 rad/sec and the graph value at the desired $\omega_n t$ is 0.6, then the actual response is $0.6 \times (0.1 \text{ rad/sec})/(\omega_n/\text{rad})$. If $\omega_n = 10 \text{ rad/sec}$ and the ramp were a frequency ramp with a slope of 10 Hz/sec, then 0.6 would represent $0.6 \times (10 \text{ Hz/sec})/(10/\text{sec}) = 0.6 \text{ Hz}$, a frequency error in response to the frequency ramp. Similarly, a parabolic input with a slope kt must be multiplied by $k \text{ rad}^2/\omega_n^2$ (see Figs. 6.10 and 6.11).

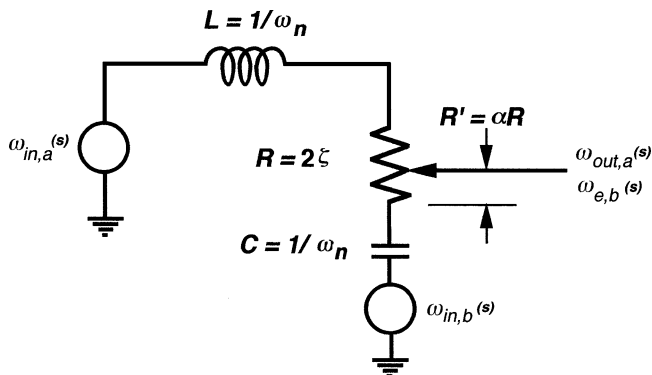


Fig. 6.13 Equivalent circuit for a second-order loop. $H(s) = \omega_{out,a}/\omega_{in,a}$; $1 - H(s) = \omega_{e,b}/\omega_{in,b}$.

6.7 EQUIVALENT CIRCUIT

As was done for the first-order loop in Chapter 2, an analogous circuit can be drawn. Figure 6.13 shows a circuit whose voltages are proportional to the frequencies or phases of the second-order PLL. In order that the same output can be used to show both $H(s)$ and $1 - H(s)$, two different injection points are used and are labeled with subscripts, a and b , corresponding to the part of Fig. 6.12 that shows the same injection point. The voltage at the bottom of the resistor represents the output for a PLL with a lag filter, that at the top of the resistor represents the output from a PLL with an integrator-and-lead filter, and the tap has a voltage representative of the output from a PLL with a lag-lead filter.

6.8 GENERAL LONG-TERM (STEADY-STATE) RESPONSE CHARACTERISTICS

These eventual, or steady-state, changes generally depend on the loop type (see Section 4.1). When the phase changes to a new value at the input to any PLL, the output eventually follows exactly (in the limit, i.e., at infinite time); see Figs. 6.1 through 6.6. Since there is no absolute phase reference, there is really no way for the loop to differentiate between different phases on a long-term basis. And all PLLs must have equal input and output frequencies in steady state.

A step of input frequency, which is equivalent to a phase ramp, causes the phase error to change in order to drive the VCO to a new frequency. As the DC zero-frequency gain approaches infinity, the long-term phase error approaches zero, so the only exception to the requirement for a change in steady-state phase error is with the integrator-and-lead filter (type 2 loop), which, theoretically, has infinite zero-frequency gain; see Figs. 6.9 and 6.7 for $\alpha = 1$. With a low-pass filter, the steady-state

phase error in following a unit frequency step is $2\zeta/\omega_n$; see Fig. 6.8. This is easily understood since, by Eqs. (4.7) and (4.15b), $2\zeta/\omega_n = 1/K \equiv d\phi_e/d\omega_{\text{out}}$.

When driven by an input frequency ramp, the phase error in a type 2 loop takes on a constant value that causes the integrator output, and thus the VCO, to ramp and maintain the constant phase offset from the input ramp; see Fig. 6.10. Since the phase error becomes constant, the frequency error must become zero. With a type 1 loop the phase error increases continuously, implying that the output takes on a constant frequency offset. See Fig. 6.11.

6.9 OPEN-LOOP EQUATIONS IN TERMS OF CLOSED-LOOP PARAMETERS

Now that we have defined closed-loop parameters ω_n , α , and ζ , it might be useful to have our open-loop equations available in those terms. Solving Eq. (4.4) for $G(s)$ in terms of $H(s)$, we obtain

$$G(s) = \frac{1}{1 - H(s)} - 1 = \frac{H(s)}{1 - H(s)}. \quad (6.21)$$

Using Eqs. (6.4a) and (6.4b), this becomes

$$G(s) = \frac{2\alpha\zeta\omega_n}{s} \frac{s + \omega_n/(2\alpha\zeta)}{s + 2(1 - \alpha)\zeta\omega_n}. \quad (6.22)$$

Appendix 8.V gives also K , ω_p , and ω_z in these terms.

6.10 MORE COMPLEX LOOPS AND STATE-SPACE ANALYSIS

In Chapter 2 we obtained transient responses for a simple loop using classical differential equations. In previous sections of this chapter, we used Laplace transform techniques to obtain responses for more complex loops. The solution of $H(s)$ in terms of its poles becomes generally quite difficult for more complex loops. Adding even one more pole can create this problem. Therefore, as the complexity increases, it becomes more desirable to use computer-aided solutions. We can use CAD to find the roots of the response equation [Kroupa, 2003, p. 103]. MATLAB[®], for example, provides the function `residue`, which, given the numerator and denominator polynomials, will provide all of the information needed for a partial-fraction expansion. The resulting sum of poles can be easily converted to a sum of time functions employing a common transform pair. This process can provide solutions for a particular set of parameter values but it is still quite difficult to solve in terms of the symbolic parameters, as we have done for the second-order loop. This is done, however, for a third-order loop with a restricted, but very useful, set of symbolic parameters in Appendix 10.A.

Another process for obtaining the response, given an arbitrary transfer function (with numerical values), which is perhaps more straightforward, employs the state-space method [Dorf, 1965; Chen, 1970]. To illustrate how this method is applied to the PLL problem, we will consider the solution of the transient response for the second-order system, which we have already solved using Laplace transforms, but the application of the method to more complex loops should be apparent. Hopefully we can get a computer to do most of the work for us, so increasing the complexity is not of such great concern.

6.10.1 Basic Equations

We write the closed-loop transfer function for a general second-order system as

$$H(s) = \frac{\varphi_{\text{out}}(s)}{\varphi_{\text{in}}(s)} = \frac{n_1 s^2 + n_2 s^1 + n_3 s^0}{s^2 + d_2 s^1 + d_3 s^0}. \quad (6.23)$$

[For simplicity we have set $d_1 = 1$. This is always possible without restricting the value of $H(s)$.] This is consistent with the output and input being written as

$$\varphi_{\text{out}}(s) = [n_1 s^2 + n_2 s + n_3] X_3(s) \quad (6.24)$$

and

$$\varphi_{\text{in}}(s) = [s^2 + d_2 s^1 + d_3] X_3(s). \quad (6.25)$$

Interpreting the Laplace operator s as time differentiation, we can write the output as a function of time as

$$\varphi_{\text{out}}(t) = n_1 \ddot{x}_3 + n_2 \dot{x}_3 + n_3 x_3, \quad (6.26)$$

where $x_i \equiv x_i(t)$ and $\dot{x}_i = dx_i/dt$.

However, since the values of each of the orders of derivative are necessary to define the state of the system at any time, we consider each separately and write

$$\varphi_{\text{out}}(t) = n_1 x_1 + n_2 x_2 + n_3 x_3 \quad (6.27)$$

with

$$x_2 \triangleq \dot{x}_3 \quad (6.28)$$

and

$$x_1 \triangleq \dot{x}_2 = \ddot{x}_3, \quad (6.29)$$

or, in general,

$$x_i \triangleq \dot{x}_{i+1}. \quad (6.30)$$

Following the same procedure we can obtain

$$\varphi_{\text{in}}(t) = x_1 + d_2 x_2 + d_3 x_3 \quad (6.31)$$

or, writing this in terms of the highest derivative,

$$x_1 = \varphi_{\text{in}}(t) - d_2 x_2 - d_3 x_3. \quad (6.32)$$

We can now write the vector

$$\mathbf{X} = \begin{bmatrix} x_2 \\ x_3 \end{bmatrix} \quad (6.33)$$

and its derivative

$$\dot{\mathbf{X}} = \begin{bmatrix} \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \varphi_{\text{in}}(t) - d_2 x_2 - d_3 x_3 \\ x_2 \end{bmatrix} \quad (6.34)$$

and incorporate these in a state-vector equation,

$$\dot{\mathbf{X}} = \mathbf{A}\mathbf{X} + \mathbf{B}u_{\text{in}}, \quad (6.35)$$

with

$$\mathbf{A} = \begin{bmatrix} -d_2 & -d_3 \\ 1 & 0 \end{bmatrix}, \quad (6.36)$$

$$\mathbf{B} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad (6.37)$$

and

$$u_{\text{in}} = \varphi_{\text{in}}(t). \quad (6.38)$$

In order to more easily see its equivalence to Eq. (6.34), let us expand (6.35) as

$$\begin{bmatrix} \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} -d_2 & -d_3 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \varphi_{\text{in}}(t). \quad (6.39)$$

This matrix equation is then solved (by a computer program, preferably) to give the state variables in \mathbf{X} as a function of time.

We still must obtain the desired output as a function of \mathbf{X} . We will do this for $n_1 = 0$, which is true for the loop of interest, and for the outputs φ_{out} and $\dot{\varphi}_{\text{out}}$. Equation (6.27) then becomes

$$\varphi_{\text{out}}(t) = n_2 x_2 + n_3 x_3, \quad (6.40)$$

and its derivative is

$$\dot{\varphi}_{\text{out}}(t) = n_2 \dot{x}_2 + n_3 \dot{x}_3 = n_2 x_1 + n_3 x_2 \quad (6.41)$$

$$= n_2[\varphi_{\text{in}}(t) - d_2 x_2 - d_3 x_3] + n_3 x_2 \quad (6.42)$$

$$= n_2 \varphi_{\text{in}} + [n_3 - n_2 d_2] x_2 - n_2 d_3 x_3. \quad (6.43)$$

Here we have used Eq. (6.32) to eliminate x_1 .

The output is now expressed in the form of a second vector equation,

$$\mathbf{Y} = \mathbf{C}\mathbf{X} + \mathbf{D}u_{\text{in}} \quad (6.44)$$

where

$$\mathbf{Y} = \begin{bmatrix} \dot{\varphi}_{\text{out}} \\ \varphi_{\text{out}} \end{bmatrix}, \quad (6.45)$$

$$\mathbf{C} = \begin{bmatrix} (n_3 - n_2 d_2) & -n_2 d_3 \\ n_2 & n_3 \end{bmatrix}, \quad (6.46)$$

and

$$\mathbf{D} = \begin{bmatrix} n_2 \\ 0 \end{bmatrix}. \quad (6.47)$$

As before, we can write this out as

$$\begin{bmatrix} \dot{\varphi}_{\text{out}}(t) \\ \varphi_{\text{out}}(t) \end{bmatrix} = \begin{bmatrix} (n_3 - n_2 d_2) & -n_2 d_3 \\ n_2 & n_3 \end{bmatrix} \begin{bmatrix} x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} n_2 \\ 0 \end{bmatrix} \varphi_{\text{in}}(t), \quad (6.48)$$

which we can compare to (6.40) and (6.43). If only one output [e.g., $\varphi_{\text{out}}(t)$] were desired, \mathbf{C} and \mathbf{D} would have only one row each.

Now we can apply these equations more specifically to $H(s)$ as given by Eq. (6.4a). Comparing Eq. (6.4a) to Eq. (6.23), the following equivalencies become apparent:

$$n_1 = 0; \quad n_2 = 2\alpha\zeta\omega_n; \quad n_3 = \omega_n^2; \quad d_1 = 1; \quad d_2 = 2\zeta\omega_n; \quad d_3 = \omega_n^2. \quad (6.49)$$

With these equations we can obtain not only $\varphi_{\text{out}}(t)$ and $\omega_{\text{out}}(t) = \dot{\varphi}_{\text{out}}(t)$ but also a phase-plane plot $\omega_{\text{out}}(\varphi_{\text{out}})$, which we will encounter later in Chapter 8.

Fortunately, there are programs, such as MATLAB[®], that make the process described above easy, not only solving the matrix equations but converting from numerator and denominator coefficients of $H(s)$ to the required matrix form.

6.10.2 Initial Conditions

The transient responses we have worked with up to now have had zero initial conditions. That is, the state variables are initially zero because we begin at steady state. Sometimes we may know the frequency and phase at some time and want to know what they are at some later time. Initial conditions can be accommodated by the state-space method, but we must know how to get the values of the initial state variables from the initial phase and frequency.

We can solve (6.40) and (6.43) to get

$$x_2 = \frac{\varphi_{\text{out}}n_2d_3 + \dot{\varphi}_{\text{out}}n_3 - \varphi_{\text{in}}n_2n_3}{\text{denom}} \quad (6.50)$$

and

$$x_3 = \frac{\varphi_{\text{out}}(n_3 - n_2d_2) - \dot{\varphi}_{\text{out}}n_2 + \varphi_{\text{in}}n_2^2}{\text{denom}}, \quad (6.51)$$

where

$$\text{denom} = n_3^2 - n_2n_3d_2 + n_2^2d_3. \quad (6.52)$$

Therefore, if we put the initial values $\omega_{\text{out}}(t = 0+)$, $\varphi_{\text{out}}(t = 0+)$, and $\varphi_{\text{in}}(t = 0+)$ in the above equations and begin with the resulting state variables, we obtain the response starting with those initial conditions.

Something interesting that we find by doing this is that the step response, for a given ω_n and ζ , starting with a given set of initial conditions, is independent of α . If a phase step is applied, α determines the initial value of frequency, but the response from then is determined by that initial frequency and the initial phase [$\varphi_{\text{out}}(t = 0+) = 0$].

We may be able to obtain the initial value of \mathbf{X} more easily if we have access to a computer program that will perform inverses. If \mathbf{Y} contains n independent variables, where n is the order of the transfer function [2 in Eq. (6.23)], then, if we multiply (6.44) by \mathbf{C}^{-1} , the inverse of \mathbf{C} (which will be square), we obtain

$$\mathbf{C}^{-1}\mathbf{Y} = \mathbf{X} + \mathbf{C}^{-1}\mathbf{D}u_{\text{in}}. \quad (6.53)$$

Rearranging and applying to the initial condition, we obtain

$$\mathbf{X}(0) = \mathbf{C}^{-1}[\mathbf{Y}(0) - \mathbf{D}u_{\text{in}}(0)] \quad (6.54)$$

or

$$\begin{bmatrix} x_2(0) \\ x_3(0) \end{bmatrix} = \mathbf{C}^{-1} \left[\begin{bmatrix} \dot{\varphi}_{\text{out}}(0) \\ \varphi_{\text{out}}(0) \end{bmatrix} - \begin{bmatrix} n_2 \\ 0 \end{bmatrix} \varphi_{\text{in}}(0) \right]. \quad (6.55)$$

The inverse for the case we are studying is

$$\mathbf{C}^{-1} = \frac{1}{\text{denom}} \begin{bmatrix} n_3 & n_2 d_3 \\ -n_2 & (n_3 - n_2 d_2) \end{bmatrix}, \quad (6.56)$$

which we can verify, or discover, by using (6.56) in (6.55) and comparing the result to Eqs. (6.50) and (6.51). We can also verify it by computing $\mathbf{C}^{-1}\mathbf{C}$ and observing that it equals the unity matrix.

6.11 AN APPROXIMATE SOLUTION USING STATE-SPACE VARIABLES

Another method [Gill, 1981], produces an approximate, but often adequate, solution that does not require complex matrix computations. New values for the state variables are computed at the end of a time interval based on the values at the beginning of that interval (end of the previous interval). Each computation interval starts with the solution of Eq. (6.32), based on the values of the state variables (e.g., x_2, x_3) and input (e.g., φ_{in}) at the end of the previous interval. Then the other state variables are found, based on Eq. (6.30), by integration:

$$x_{i+1}(t_j) = x_{i+1}(t_{j-1}) + \int_{t_{j-1}}^{t_j} x_i dx_i \approx x_{i+1}(t_{j-1}) + \frac{x_i(t_{j-1}) + x_i(t_j)}{2} (t_j - t_{j-1}). \quad (6.57)$$

This is done in sequence, starting with x_2 , until all of the state variables have been computed. Then the output is computed using (6.48). Accuracy gets better as the time interval $t_j - t_{j-1}$ gets smaller.

6.12 EFFECT OF AN ADDED POLE

Figure 6.14 illustrates the effect of a pole added to what would otherwise be a type 2, second-order, loop with $\zeta = 0.5$. We see little effect on the transient response when the frequency ω_a of the added pole is 10 times ω_n , which approximately equals ω_L when $\zeta = 0.5$ (they are equal in the tangential approximation). As ω_a comes closer to ω_n , the effect grows until an undamped oscillation occurs when they coincide. At that point the added pole cancels the zero, giving a constant 180° excess phase and producing instability.

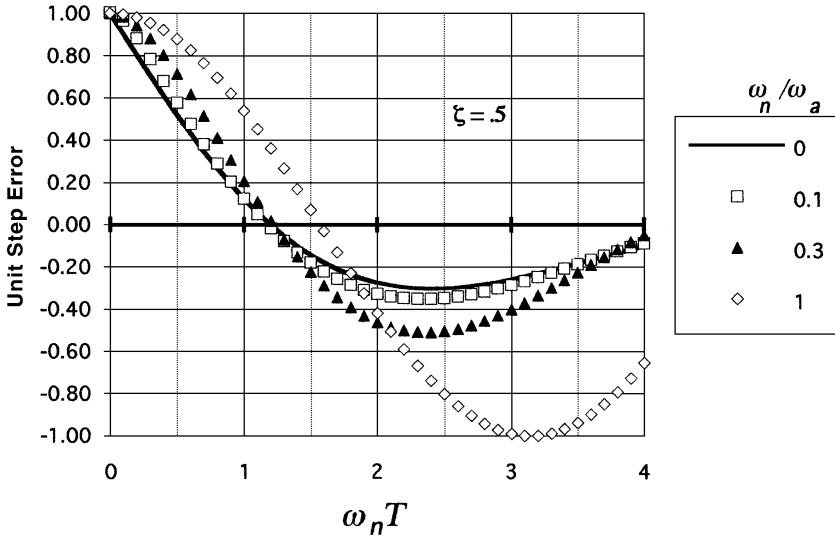


Fig. 6.14 Effect of a pole added to type 2 loop ($\zeta = 0.5$).

i6.M APPENDIX: TRANSIENT RESPONSES USING MATLAB

This appendix, which is available from the Wiley Internet site, shows various ways to simulate transient responses using MATLAB. The methods range from the simple use of a built-in function to the use of state-space equations. Scripts that can show the effect of an added pole (Fig. 6.14), display the response in the phase plane (frequency versus phase), or include initial conditions are described.

CHAPTER 7

MODULATION RESPONSE

Since phase-locked loops (PLLs) can be used as phase and frequency modulators and demodulators, the transfer functions from phase or frequency to voltage and from voltage to phase or frequency are of obvious importance. We will want to know how the loop filters the modulation. We will also be interested in how the phase error responds since we will need to know whether the loop is remaining linear (and locked). As with the transient response, the equations that we need have already been developed in Chapter 4, but we will here interpret them in terms of their implications for modulation. As was done for the transient response, the modulation response will be computed for the output and the error. And, as was done for the transient, important related responses will also be considered.

7.1 PHASE AND FREQUENCY MODULATION

A sinusoid can be written

$$v = V \sin[\psi(t)], \quad (7.1)$$

where we have chosen to use symbols appropriate to a voltage. The instantaneous phase is

$$\psi(t) = \omega_c t + \varphi(t), \quad (7.2)$$

where the phase consists of a steadily advancing component, $\omega_c t$, and a modulation component,

$$\varphi(t) = m \sin(\omega_m t + \theta_m). \quad (7.3)$$

Here m is the peak phase deviation, also called the modulation index. This implies an instantaneous frequency

$$\omega(t) \equiv \frac{d\psi(t)}{dt} = \omega_c + \frac{d\varphi(t)}{dt} \quad (7.4)$$

$$= \omega_c + m\omega_m \cos(\omega_m t + \theta_m) \quad (7.5)$$

$$= \omega_c + \Delta\omega \sin(\omega_m t + \theta_m + \pi/2), \quad (7.6)$$

where $\Delta\omega = m\omega_m$. Thus the peak phase and frequency deviations are related by

$$m = \Delta\varphi = \frac{\Delta\omega}{\omega_m} = \frac{\Delta f}{f_m} \quad (7.7)$$

while, from Eqs. (7.3) and (7.6), the frequency deviation leads the phase deviation by 90° .

Note that, if the signal in Eq. (7.1) is phase modulated, it is also frequency modulated. So, what is the difference? Why do we have two names for the same process? The difference is only significant when the modulation represents information at multiple frequencies. For example, with phase modulation, a modulating signal that has constant amplitude as its frequency changes will produce frequency-independent phase deviation, but by Eq. (7.7), the frequency deviation will be proportional to modulation frequency f_m . Conversely, with frequency modulation, the same signal would produce constant frequency deviation Δf , but the phase deviation $m = \Delta\varphi$ would decrease with modulation frequency. This is illustrated in Fig. 7.1.

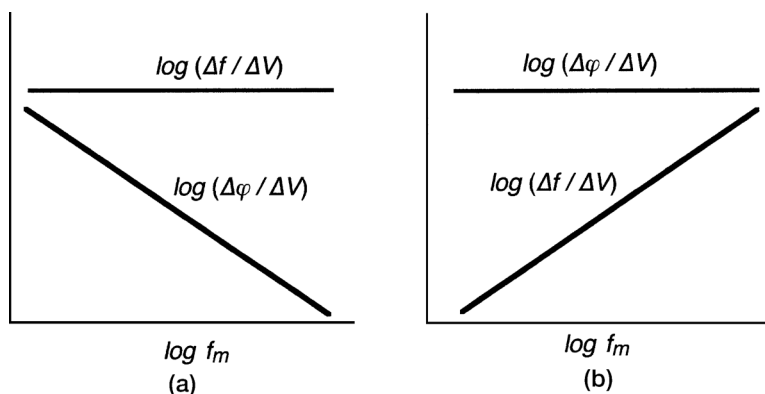


Fig. 7.1 Comparison of frequency (a) and phase (b) modulations.

7.2 MODULATION RESPONSES

Suppose Eqs. (7.1) through (7.7) represent the reference input to a PLL. A similar set of equations could then be written to describe its output. With phase modulation, the phase change at the output is related to the phase change at the input by [see Eq. (4.3)]

$$\varphi_{\text{out}}(\omega_m) = \varphi_{\text{in}}(\omega_m)H(\omega_m), \quad (7.8)$$

which also implies

$$|\varphi_{\text{out}}(\omega_m)| = |\varphi_{\text{in}}(\omega_m)||H(\omega_m)| \quad (7.9)$$

and

$$\angle\varphi_{\text{out}}(\omega_m) = \angle\varphi_{\text{in}}(\omega_m) + \angle H(\omega_m). \quad (7.10)$$

Then the equation like (7.3) that describes the phase modulation at the output would be

$$\varphi_{\text{out}}(t) = m|H(\omega_m)| \sin[\omega_m t + \theta_m + \angle H(\omega_m)]. \quad (7.11)$$

Similarly, the output frequency modulation would be described by an equation like (7.6) [see Eq. (2.27)]

$$\omega_{\text{out}}(t) = \omega_c + \Delta\omega|H(\omega_m)| \sin[\omega_m t + \theta_m + \angle H(\omega_m) + \pi/2]. \quad (7.12)$$

The error responses could be similarly written as

$$\varphi_e(t) = m|1 - H(\omega_m)| \sin\{\omega_m t + \theta_m + \angle[1 - H(\omega_m)]\} \quad (7.13)$$

and

$$\omega_e(t) = \Delta\omega|1 - H(\omega_m)| \sin\{\omega_m t + \theta_m + \angle[1 - H(\omega_m)] + \pi/2\}. \quad (7.14)$$

7.3 RESPONSES IN A FIRST-ORDER LOOP

We begin the study of modulation and demodulation with the first-order loop shown in Fig. 7.2. We have previously shown loops with phase change input (Fig. 2.4) and with frequency change input (Fig. 2.7). Figure 7.2 shows both types of inputs simultaneously. Since we are considering the loop to be linear in this chapter, superposition applies.

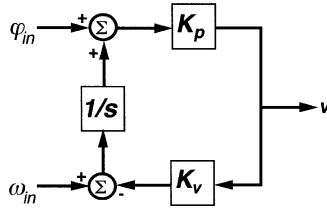


Fig. 7.2 Demodulation in a first-order loop.

In this case the only voltage variable v appears at the output of the phase detector, which is also the input to the voltage-controlled oscillator (VCO). Phase or frequency modulation, represented by φ_{in} or ω_{in} , respectively, is introduced into the phase detector. The phase appearing after the summer is the error response, so

$$v(\omega_m) = \varphi_e(\omega_m)K_p = \varphi_{in}(\omega_m)[1 - H(\omega_m)]K_p. \quad (7.15)$$

Using Eq. (2.26b) for φ_e , with $s \Rightarrow j\omega$, we obtain

$$v(\omega_m) = \varphi_{in}(\omega_m)j\omega_m K_p / (j\omega_m + K_p K_v). \quad (7.16)$$

Voltage v is also related to ω_{in} by

$$v(\omega_m) = \omega_{out}(\omega_m) / K_v = \omega_{in}(\omega_m)H(\omega_m) / K_v. \quad (7.17)$$

Using Eq. (2.25b), this is

$$v(\omega_m) = \omega_{in}(\omega_m)K_p / (j\omega_m + K_p K_v). \quad (7.18)$$

Since the two inputs are connected by $1/s$, it is not surprising that the two responses differ by a factor of $j\omega_m$. These responses are illustrated in Fig. 7.3.

We see that at low frequencies Eq. (7.18) becomes

$$v(\omega_m \ll \omega_L) = \omega_{in} / K_v \quad (7.19)$$

so that v is proportional to the input frequency and provides frequency demodulation. At low frequencies the gain is high enough to force the frequency of the VCO to equal that of the input, and thus the VCO's tuning voltage is proportional to the input frequency.

At high frequencies Eq. (7.16) becomes

$$v(\omega_m \gg \omega_L) = \varphi_{in}(\omega_m)K_p, \quad (7.20)$$

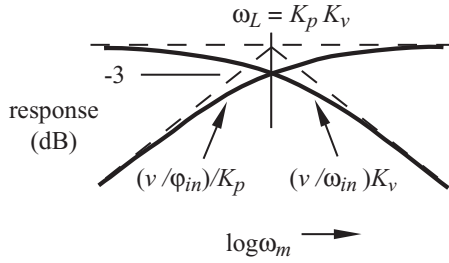


Fig. 7.3 Responses of the first-order loop.

providing phase demodulation. At high frequencies the loop is essentially open so the VCO’s phase does not vary. Thus the phase error equals the input phase, and the voltage at the output of the phase detector is therefore proportional to input phase.

Moreover, if, as in Fig. 7.4, a voltage is introduced at the phase detector output,

$$\omega_{out} = v[1 - H(\omega_m)]K_v \tag{7.21}$$

$$= vj\omega_m K_v/[j\omega_m + K_p K_v] \tag{7.22}$$

and

$$\varphi_{out}(\omega_m) = vH(\omega_m)/K_p \tag{7.23}$$

$$= vK_v/(j\omega_m + K_p K_v). \tag{7.24}$$

At low frequencies Eq. (7.24) becomes

$$\varphi_{out}(\omega_m \ll \omega_L) = v/K_p \tag{7.25}$$

so the phase is proportional to v , providing phase modulation. Because the gain is high, the phase is forced to follow v .

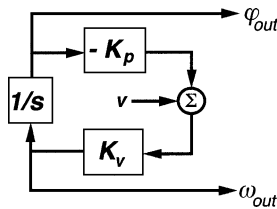


Fig. 7.4 Modulation in a first-order loop. The K_p block includes the usual summer, which contains the minus sign required to produce phase reversal in the loop. The reference input is not shown because it is constant.

At high frequencies Eq. (7.22) becomes

$$\omega_{\text{out}}(\omega_m \gg \omega_L) = \nu K_\nu, \quad (7.26)$$

enabling us to obtain frequency modulation. Here the low gain permits the VCO to be modulated as if the loop were open, but the loop keeps the average [direct current] value of the frequency equal to the reference frequency (which is not shown because it is not relevant to the discussion of the modulator).

Example 7.1 Modulation in a Simple Loop

- a. Choose an appropriate bandwidth for a first-order loop to act as a frequency modulator and for a second such loop to serve as demodulator. The tones to be transmitted have modulation frequencies between 1 and 5 kHz.

We apply the tones at ν in Fig. 7.4 and obtain the demodulated tones at ν in Fig. 7.2. We must choose a bandwidth for the loop in Fig. 7.4 that is low compared to the tone frequencies so the feedback does not attempt to counter ν and its full value will be impressed on the VCO. Conversely, we must choose a bandwidth in Fig. 7.2 that is high enough so the VCO will follow the input and ν will be accurately proportional to frequency. Thus we might select the frequency at unity loop gain f_L to be 200 Hz in Fig. 7.4 and 10 kHz in Fig. 7.2, for example.

- b. What will be the effect on the communication link if the bandwidths are interchanged between the two loops?

Then the existence of high gains at the modulation frequencies in Fig. 7.4 will cause the second input to the summer to follow ν , implying that φ_{out} is also proportional (by $1/K_p$) to ν . Thus we will produce phase modulation rather than frequency modulation. Similarly, the narrow bandwidth in Fig. 7.2 will cause ν to be proportional to φ_{in} . We will thus have constructed a *phase-modulated (PM) link rather than a frequency modulated (FM) link*.

7.4 TRANSFER FUNCTIONS IN A SECOND-ORDER LOOP

The responses in a second-order loop are generally similar to those in the first-order loop, but, because of the filter block, there are more isolated points to be considered and the response shape is more variable.

The closed-loop transfer function $H(s)$ and the error function $[1 - H(s)]$ in Table 4.1 describe the response of the loop's output phase φ_{out} or phase error φ_e , respectively, to the input phase φ_{in} . Figure 4.1 illustrates these relationships. However, the same functions describe other responses in the loop. Figure 7.5 illustrates all of these combinations. Here a disturbance $\Delta x(s)$ is introduced into the loop through a summer at any point where two of the loop's functional blocks are joined. The

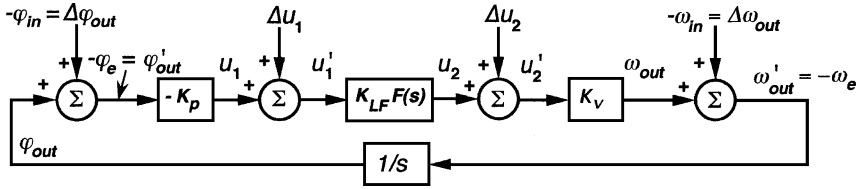


Fig. 7.5 Loop with the various injection points for disturbances shown.

disturbance could be noise or an intended signal. All the variables are functions of s , even though the dependency is not shown explicitly.

The variable x has been primed to differentiate it from the output that precedes the summer. When $\Delta x = 0$, $x = x'$. Since, in what we have studied so far, Δx has been zero for u_1 and u_2 , we have not used the primed variables. Since we have concentrated on φ and ω (see Fig. 4.1) as points of disturbance, the associated variables were given special names for the disturbance (e.g., φ_{in}) and the error (e.g., φ_e). Rather than invent new special names for all the other disturbances and errors, we will use the convention shown, and we may refer to the φ and ω disturbances and errors in the same manner to keep the presentation symmetrical and to point out the similarity in the treatment of the different variables. There is one awkward difference between the errors in Fig. 7.5 and those considered before, however. A 180° phase shift, or minus sign, must be placed somewhere in the loop and is customarily placed before the summer. We could do that for any one of the summers in Fig. 7.5 but only for one. So we place the minus sign approximately in its usual location, before K_p but not before the summer. To maintain the usual definitions for φ_{in} , ω_{in} , φ_e , and ω_e , we then place minus signs before them.

There are two ways to relate a variable to a disturbance. We can multiply Δx by $(1 - H)$ and then by the gain to the point of interest or we can multiply Δx by $-H$ and then divide by the gain from the point of interest to the point of disturbance. For example, the value of u_2 in response to a disturbance Δu_1 , can be obtained as

$$u_2 = u'_1 K_{LF} F(s) = \Delta u_1 [1 - H(s)] K_{LF} F(s) \tag{7.27}$$

or as

$$u_2 = u_1 / (-K_p K_v / s) = \Delta u_1 H(s) s / (K_p K_v). \tag{7.28}$$

7.4.1 Output Response

The output response, which precedes the summer, can be obtained by multiplication of the disturbance by $-H(s)$:

$$x(s) = -\Delta x(s) H(s). \tag{7.29}$$

Here $x(s)$ may equal u_1 , u_2 , ω_{out} , or φ_{out} .

At modulation frequencies that are low compared to ω_L , $H(\omega_m) \approx 1$, so it then may be convenient to express various responses in terms of $H(\omega_m)$. For example, a power supply transient might cause a change $\Delta\omega_{\text{out}}$ in the VCO frequency when the loop is open. The disturbance can be represented by the injection of $\Delta\omega_{\text{out}}$ as shown in Fig. 7.5, and the closed-loop response at ω_{out} can be obtained by multiplying $-\Delta\omega_{\text{out}}(\omega_m)$ by $H(\omega_m)$, which is conveniently unity at low frequencies. Then u_2 can be obtained by dividing by K_v so $u_2(\omega_m) \approx -\Delta\omega_{\text{out}}(\omega_m)/K_v$. If also $\omega_m \ll \omega_p$, then $u_1(\omega_m) \approx -\Delta\omega_{\text{out}}(\omega_m)/(K_v K_{\text{LF}})$.

At modulation frequencies that are high compared to ω_L , $H(\omega_m) \approx G(\omega_m)$. Therefore a similar procedure can be followed except that the response includes the open-loop $G(\omega_m)$ and, where it is used, $K_{\text{LE}} \Rightarrow K_{\text{LF}}F(\omega_m)$. If $\omega_m \gg \omega_z$, then $F(\omega_m) \Rightarrow \omega_p/\omega_z$. For example, we may wish to know the effect, upon the output frequency, of a high-frequency signal coupled into the circuit at the output of the phase detector. If $\omega_m \gg \omega_L$, ω_z , then $\omega_{\text{out}}(\omega_m) \approx \Delta u_1(\omega_p/\omega_z) K_{\text{LF}}K_v$.

7.4.2 Error Response

The response that follows the disturbance is the error and can be obtained by multiplying $\Delta x(s)$ by $[1 - H(s)]$.

$$x'(s) = \Delta x(s)[1 - H(s)]. \quad (7.30)$$

At modulation frequencies that are high compared to ω_L , $[1 - H(\omega_m)] \approx 1$, so it may be convenient to express responses in terms of $[1 - H(\omega_m)]$. For example, the phase error ϕ_e in response to an input ϕ_{in} can be obtained by multiplying $\phi_{\text{in}}(\omega_m)$ by $[1 - H(\omega_m)]$, which is unity at high frequencies. Then u_1 can be obtained by multiplying by K_p to give $u_1(\omega_m) \approx \phi_{\text{in}}(\omega_m)K_p$. Then u_2 can be obtained by multiplication by $K_{\text{LF}}F(\omega_m)$ and, if $\omega_m \gg \omega_z$, we have $F(\omega_m) \approx \omega_p/\omega_z$.

At modulation frequencies that are low compared to ω_L , $[1 - H(\omega_m)] \approx 1/G(\omega_m)$. Therefore a similar procedure can be followed except that the response includes $1/G(\omega_m)$ and, where it is used, $K_{\text{LF}}F(\omega_m) \Rightarrow K_{\text{LF}}$ if $\omega_m \ll \omega_p$.

7.4.3 Responses Near ω_L

Where the approximations above cannot be used because ω_m is commensurate with ω_L , it is necessary to know $H(\omega_m)$ or $[1 - H(\omega_m)]$. Table 7.1 is a compilation of transfer functions and error functions relating the variables in Fig. 7.5. The disturbances are shown at the bottom, and the transfer functions to the primed (error) and unprimed (output) variables are shown in the table on the same line as the variable. The diagonal split separates transfer functions that relate to the errors from those that relate to the outputs.

Of the two ways to link the response to the disturbance, the function that avoids $F(s)$ has been chosen for Table 7.1 because $F(s)$ is highly variable and relatively complex in form. Given responses for $H(s)$ and $[1 - H(s)]$, it is a simple matter to multiply or divide them by constants. It is somewhat more complicated to multiply

TABLE 7.1. Transfer Functions Between Loop Variables^a

	[y/x](s)				
u'_1	$-(1 - H)K_p K_v/s$	$-(1 - H)K_p/s$	$-(1 - H)K_p$	$(1 - H)$	$\Downarrow y$
$\varphi'_{out}(-\varphi_e)$	$(1 - H)K_v/s$	$(1 - H)/s$	$(1 - H)$	$-H$	u_1
ω'_{out}	$(1 - H)K_v$	$(1 - H)$	$-H$	H/K_p	φ_{out}
u'_2	$(1 - H)$	$-H$	$-Hs$	Hs/K_p	ω_{out}
$y \Uparrow$	$-H$	$-H/K_v$	$-Hs/K_v$	$Hs/(K_p K_v)$	u_2
	Δu_2	$\Delta \omega_{out}(-\omega_{in})$	$\Delta \varphi_{out}(-\varphi_{in})$	Δu_1	
		x			

^aThose above the diagonal refer to the leftmost column. Those below refer to the rightmost column.

or divide them by s , but it is still simpler than taking into account the details of the filter (which are, nevertheless, embedded in the form of H or $[1 - H]$).

Example 7.2 Use of Table 7.1 What is the transfer function between the input frequency and (a) the phase error? (b) the output frequency?

- a. Using the second column for $x = -\omega_{in}$, we go up to the second row where $y = -\varphi_e$ appears on the left. At the intersection we see the transfer function $(1 - H)/s$.
- b. Since ω_{out} appears in the fourth row on the right, the transfer function appears in the fourth row of column two, $-H$. However, this is the transfer function from $-\omega_{in}$ to $+\omega_{out}$ so the desired function is H .

7.4.4 Phase or Frequency at Inputs and Outputs

As mentioned above, $\Delta\omega_{out}$ can represent the value of a disturbance in the open-loop frequency of the VCO. However, $\Delta\omega_{out}$ might also represent an input frequency transient or modulation. In that case $\omega'_{out} = -\omega_e$ is the frequency error at the input and ω_{out} is the loop's output frequency. Similarly, while $\Delta\varphi_{out}$ may represent $-\varphi_{in}$, it may also represent an open-loop transient in the VCO. In the former case, the summer output is $-\varphi_e$ whereas, in the latter case, it is the phase of the output of the loop. The reason that these symbols can represent either input or output is that there is no physical separation between the input and output. The term $1/s$ is a mathematical rather than a physical entity. It represents an integration that converts frequency to phase, and this can be considered to occur at input or output, whichever is convenient. If we want to know the output phase, we consider the integration to occur between the VCO and the output, converting output frequency to output phase. If we want to

consider the input to be a frequency, we consider the integration to occur after the input summer, converting input frequency difference to input phase difference before entry into the phase detector.

7.5 TRANSIENT RESPONSES BETWEEN VARIOUS POINTS

While we have been discussing the modulation response between various points in the loop, Table 7.1 also applies to transient responses. We will discuss this briefly before continuing with our discussion of modulation response.

We have developed equations and graphs for error response to steps, ramps, and parabolas. These are represented in Laplace transforms by $[1 - H(s)]/s$, $[1 - H(s)]/s^2$, and $[1 - H(s)]/s^3$. From these we can also obtain the output responses, related to the driving functions by $H(s)$. Therefore we can easily find $y(t)$ in Table 7.1 when it is related to the $x(t)$ by $[1 - H(s)]$ or $H(s)$ and $x(t)$ is a step, ramp, or parabola.

However, we can also find values of $y(t)$ that are related to $x(t)$ by $(1 - H)/s$ if $x(t)$ is a step or a ramp by using the ramp or parabola response, respectively. This is because $1/s$ represents integration, and the integral of a response to some input equals the response to the integral of that input. For example, if $\Delta\omega_{\text{out}}$ is a step, from the table,

$$u'_1(s) = -\Delta\omega_{\text{out}}(s)\{[1 - H(s)] K_p/s\} = -(1/s)\{[1 - H(s)] K_p/s\} \quad (7.31)$$

$$= -\{(1/s^2)[1 - H(s)]\} K_p. \quad (7.32)$$

But the expression in $\{ \}$ in Eq. (7.32) is the error response to a unit ramp and that is shown in Figs. 6.7 through 6.9. Thus we can find the voltage at the output of the phase detector in response to a frequency step by multiplying the value given by one of those figures by $-K_p$.

Responses in Table 7.1 containing $[H(s)s]$ represent differentiation of responses given by $H(s)$. Therefore we can obtain such a response to a ramp or parabola by using the responses developed for a step or ramp, respectively. For example, to find u_2 in response to a ramp at φ_{in} , we can use the step response from Figs. 6.2 through 6.4, subtracting them from one to get $H(s)$ and dividing by K_v . Since we have not developed graphs for impulse responses, however, we must differentiate step responses to find $[H(s)s]$ in that case.

7.6 MAGNITUDE AND PHASE OF THE TRANSFER FUNCTIONS

The variables can be written in complex form as

$$x = |x| \exp(j\angle x) \quad (7.33)$$

so

$$y = Hx \Rightarrow |y| \exp(j\angle y) = \{|H| \exp(j\angle H)\} \{|x| \exp(j\angle x)\} \quad (7.34)$$

As a result, we can write separate equation for the magnitudes and the phases:

$$|y| = |H||x|, \quad (7.35)$$

$$\angle y = \angle H + \angle x. \quad (7.36)$$

Thus, the magnitude of the output modulation (i.e., its peak phase or frequency deviation) equals the magnitude of the corresponding input modulation multiplied by the magnitude of the transfer function. Also, the phase of the transfer function indicates by how much the phase of the output modulation is delayed compared to the input modulation. Similarly, for the errors we can write

$$|y'| = |(1 - H)||x|, \quad (7.37)$$

$$\angle y' = \angle(1 - H) + \angle x. \quad (7.38)$$

7.6.1 Output Responses

We write $H(\omega_m)$ from Eq. (6.4a) with $s \Rightarrow j\omega_m$, normalizing the modulation frequency to the natural frequency. [See Hgeneric Excel file.]

$$H(\Omega) = \frac{1 + j2\alpha\zeta\Omega}{(1 - \Omega^2) + j2\zeta\Omega}, \quad (7.39)$$

where

$$\Omega \triangleq \omega_m/\omega_n. \quad (7.40)$$

The magnitude and phase are then

$$|H(\Omega)| = \sqrt{\frac{1 + (2\alpha\zeta\Omega)^2}{(1 - \Omega^2)^2 + (2\zeta\Omega)^2}}, \quad (7.41)$$

$$\angle H(\Omega) = \tan^{-1}(2\alpha\zeta\Omega) - \tan^{-1}[2\zeta\Omega/(1 - \Omega^2)]. \quad (7.42)$$

7.6.2 Error Responses

The error response can be written [see 1minHgen Excel file]

$$1 - H(\Omega) = \frac{-\Omega^2 + j2(1 - \alpha)\zeta\Omega}{(1 - \Omega^2) + (j2\zeta\Omega)}. \quad (7.43)$$

$$|1 - H(\Omega)| = \sqrt{\frac{\Omega^4 + [2(1 - \alpha)\zeta\Omega]^2}{(1 - \Omega^2)^2 + (2\zeta\Omega)^2}}, \quad (7.44)$$

$$\angle[1 - H(\Omega)] = -\tan^{-1}[2(1 - \alpha)\zeta/\Omega] - \tan^{-1}[2\zeta\Omega/(1 - \Omega^2)]. \quad (7.45)$$

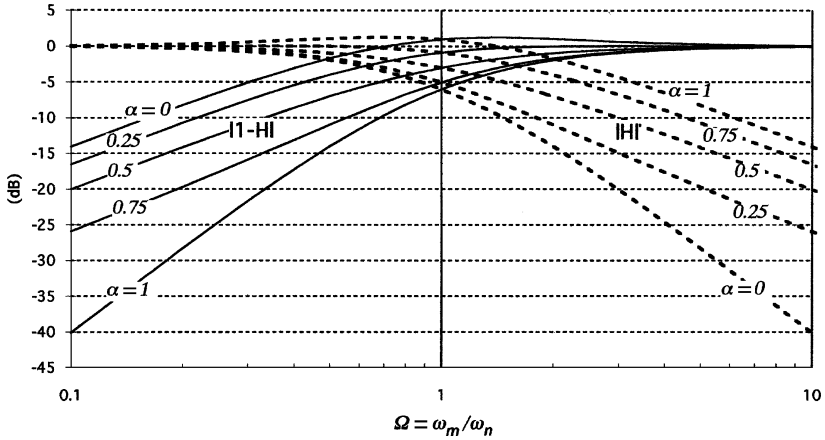


Fig. 7.6 Magnitude of $H(\Omega)$ and $[1 - H(\Omega)]$ for $\zeta = 1$.

7.6.3 Effect of α

Figure 7.6 shows both the output and error amplitude responses for $\zeta = 1$ at various values of α . Theoretically, we can add a weighted response for $\alpha = 0$ to one for $\alpha = 1$ to produce the response for any desired value of α , according to Eq. (6.6), as we did for transient responses. However, we would require that the responses be given separately for real and imaginary parts so that we could add the two real parts and the two imaginary parts separately and then combine them to get magnitude and phase. It is probably easier to compute the values from the equations above. However, knowing the way in which these responses combine can help us to predict the results in a rough way from analysis of the curves that will be given.

Figure 7.7 shows the corresponding phases. Note that the phase of the output response $H(\Omega)$ with a low-pass filter ($\alpha = 0$) is identical to that of the error response $[1 - H(\Omega)]$ with an integrator-and-lead filter ($\alpha = 1$).

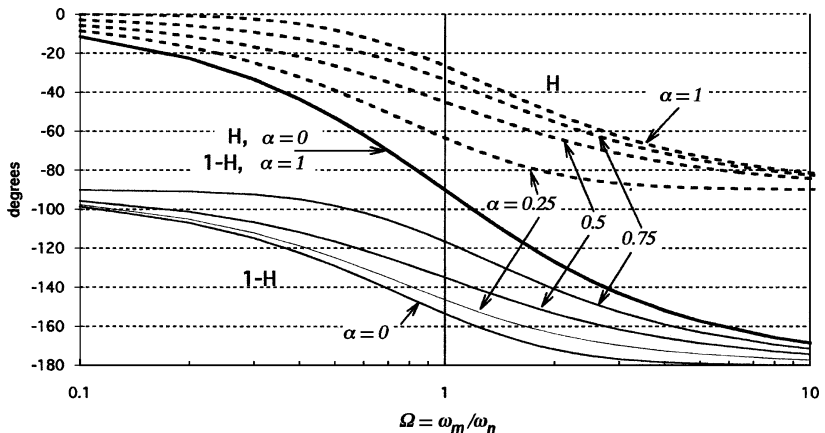


Fig. 7.7 Phase of $H(\Omega)$ and $[1 - H(\Omega)]$ for $\zeta = 1$.

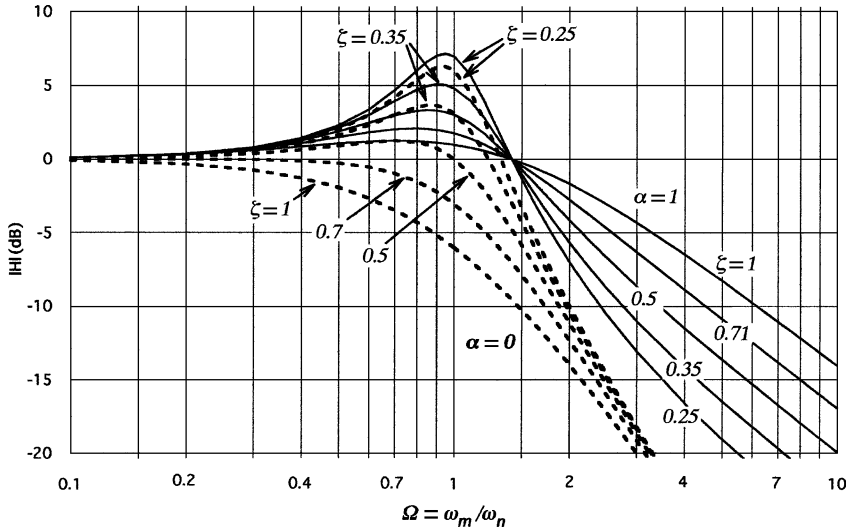


Fig. 7.8 Magnitude of $H(\Omega)$ for $\zeta \leq 1$.

7.6.4 Responses for $\zeta \leq 1$

Figures 7.8 and 7.9 show amplitude responses for output and error, respectively. Figures 7.10 and 7.11 show corresponding phases.

According to Eq. (4.20) or (6.4a), when $\zeta = 1/\sqrt{2} \approx 0.71$, the two closed-loop poles are at $(-1 \pm j)\omega_n/\sqrt{2}$ and therefore lie on lines that are 45° from the real axis. When also $\alpha = 0$, these are the only critical frequencies; there are no zeros.

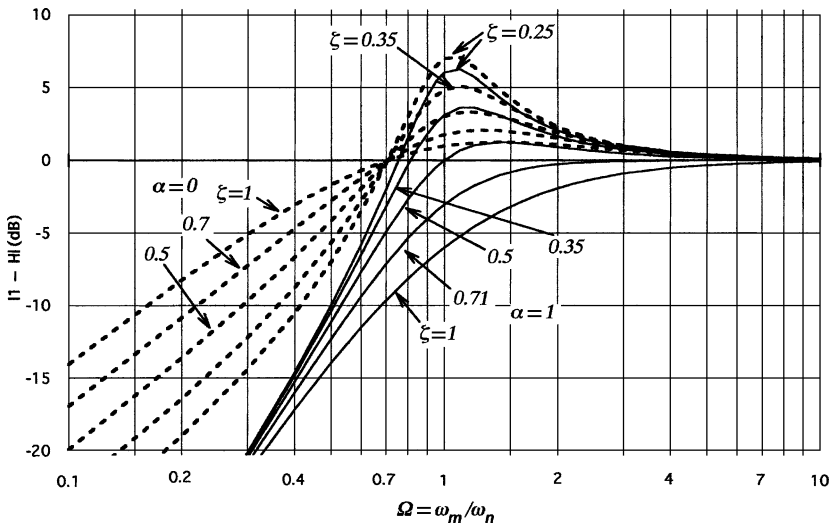


Fig. 7.9 Magnitude of $[1 - H(\Omega)]$ for $\zeta \leq 1$.

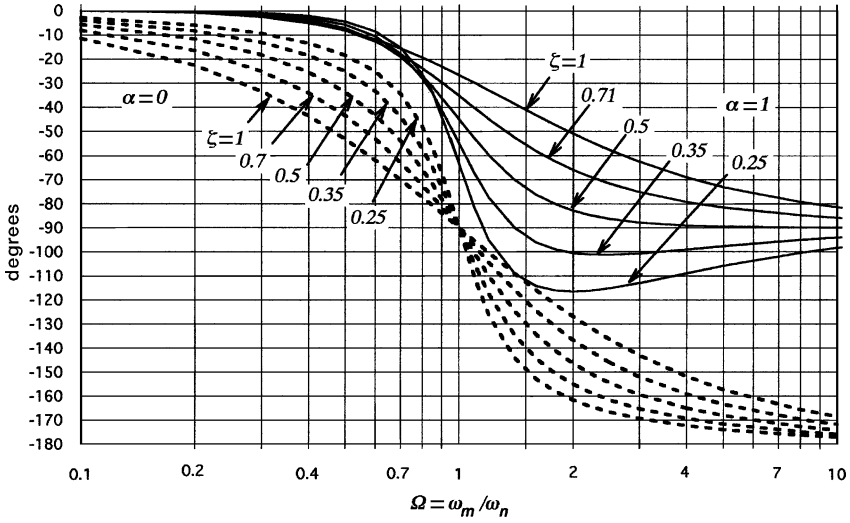


Fig. 7.10 Phase of $H(\Omega)$ for $\zeta \leq 1$.

Such a set produces a two-pole “maximally flat” (Butterworth) response, which has zero-valued first and second derivatives of magnitude relative to ω^2 at $\omega = 0$. The flatness of this response is evident in Fig. 7.8. Note also that the 3-dB bandwidth equals ω_n .

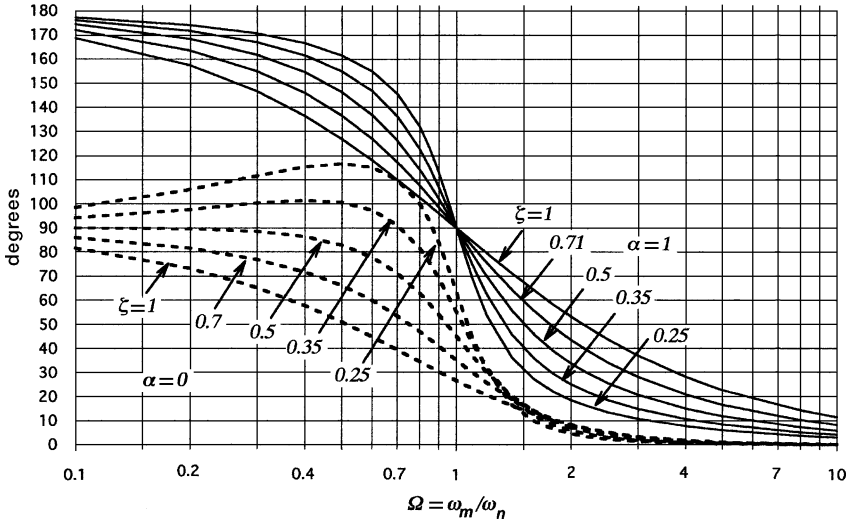


Fig. 7.11 Phase of $[1 - H(\Omega)]$ for $\zeta \leq 1$.

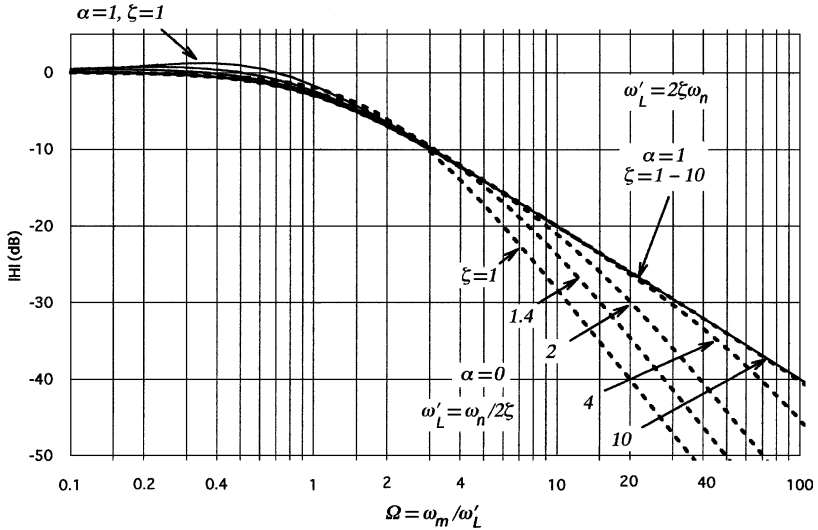


Fig. 7.12 Magnitude of $H(\Omega)$ for $\zeta \geq 1$.

7.6.5 Responses for $\zeta \geq 1$

Figures 7.12 and 7.13 show amplitude responses for output and error, respectively, and Figs. 7.14 and 7.15 show the corresponding phases. These are normalized to ω'_L [see Eqs. (6.13) and (6.14)]. As with the step response, we find that the modulation responses are more closely related to the unity open-loop gain frequency ω_L than to ω_n , and we therefore plot the responses versus ω'_L , an approximation to ω_L .

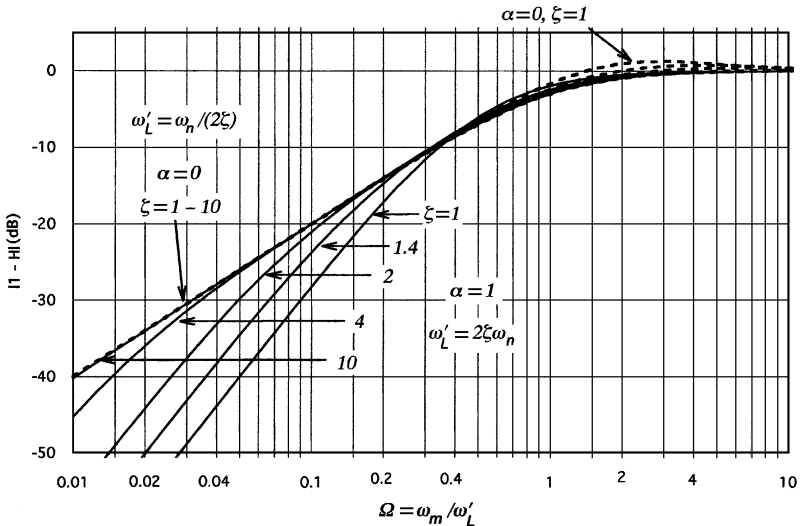


Fig. 7.13 Magnitude of $[1 - H(\Omega)]$ for $\zeta \geq 1$.

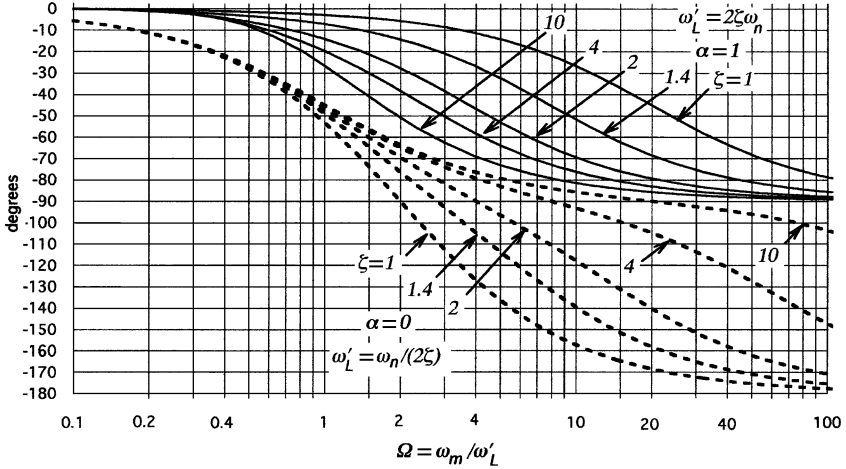


Fig. 7.14 Phase of $H(\Omega)$ for $\zeta \geq 1$.

Example 7.3 Modulation Response Curves A sinusoidal phase modulation is impressed on the reference signal to a PLL with $\alpha = 1$, $\zeta = 0.5$, and $\omega_n = 2\pi \cdot 300$ rad/sec. The input has a peak phase deviation of 40° and a modulation frequency of 600 Hz. Describe the modulation of the loop's output signal.

From Fig. 7.8, at $\Omega = 600 \text{ Hz}/300 \text{ Hz} = 2$ and $\zeta = 0.5$, $|H| = -4 \text{ dB}$. The output modulation amplitude is therefore $40^\circ \times 10^{-4/20} = 25.2^\circ$. From Fig. 7.10, this modulation lags the input modulation by 82° . Here 25.2° is the amplitude of the phase modulation and -82° is its phase. Both are parameters of the state variable, output phase.

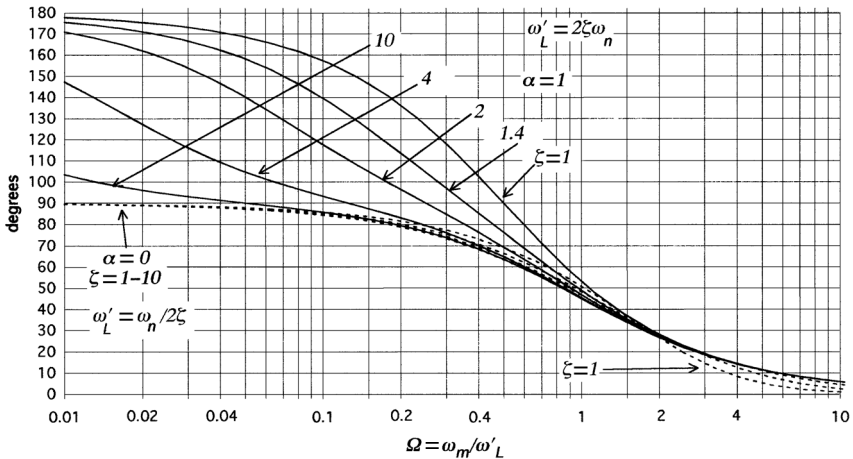


Fig. 7.15 Phase of $[1 - H(\Omega)]$ for $\zeta \geq 1$.

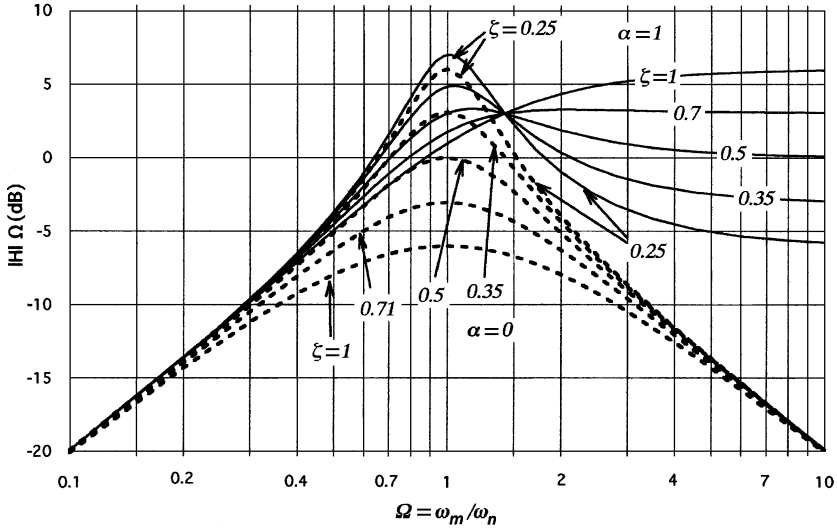


Fig. 7.16 Normalized magnitude of $\omega_m H(\omega_m)$ for $\zeta \leq 1$. Note: The peak value, for $\alpha = 0$, is $|H|_{\Omega = 1/(2\zeta)}$.

7.7 RELATED RESPONSES

The two other amplitude responses required for Table 7.1 are shown in Figs. 7.16 and 7.17. Phase is not shown since the phase of Hs is just the phase of H plus 90° and the phase of $(1 - H)/s$ is the same as that of $(1 - H)$ less 90° . Therefore the phase curves previously given can be used again.

Example 7.4 Use of Related Modulation Response Curves Describe the frequency of the loop’s output signal in Example 7.3.

Since frequency is the time derivative of phase, the frequency is $j\omega_m$ times the phase (in radians). Using Fig. 7.16, we find $|H|_{\Omega = 2} = 2$ dB at $\Omega = 2$ input amplitude, or

$$|H|\left(\frac{\omega_m}{\omega_n}\right) = 10^{2/20} = 1.26,$$

$$|H|\omega_m = 1.26\omega_n \Rightarrow 1.26(2\pi)(300 \text{ rad/sec})\left(\frac{\text{cycle}}{2\pi \text{ rad}}\right) = 378 \text{ Hz}.$$

The magnitude of the frequency modulation is 264 Hz , the product of the input, $[40^\circ(\pi \text{ rad}/180^\circ) =]0.7 \text{ rad}$, and the response. The phase angle of the frequency modulation is 90° plus the phase angle of the phase modulation (obtained in Example 7.3) or $+ 8^\circ$.

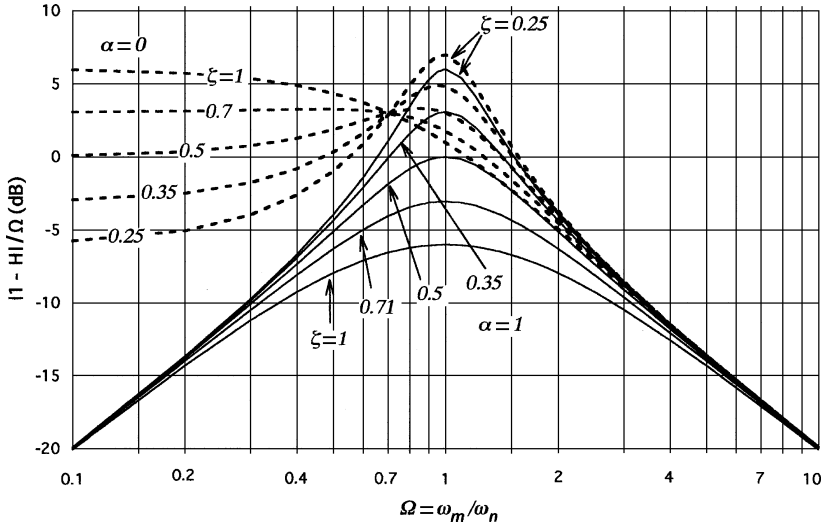


Fig. 7.17 Normalized magnitude of $|1 - H(\omega_m)|/\omega_m$ for $\zeta \leq 1$. Note: The peak value, for $\alpha = 1$, is $|1 - H|/\Omega = 1/(2\zeta)$ [Blanchard, 1976, p. 125].

7.8 MODULATION AND DEMODULATION IN THE SECOND-ORDER LOOP

Modulation and demodulation are performed in a manner similar to what was described for the first-order loop.

7.8.1 Frequency Demodulation

By Table 7.1,

$$u_2 = \omega_{in}H(\omega_m)/K_v, \tag{7.46}$$

which is like Eq. (7.17) and, as for the first-order loop, leads to Eq. (7.19). Thus the voltage at the input to the VCO equals the input frequency divided by K_v at low frequencies, and the shape of the magnitude response is given by the curves in Figs. 7.8 and 7.12 while the phase is shown in Figs. 7.10 and 7.14.

The usable range of modulation frequencies can be extended beyond the limits indicated by Eq. (7.46) by compensation of the demodulated output to counter the effects of the loop response near and beyond ω_L . Thus a circuit, external to the loop, with response $C/H(\omega_m)$, over a sufficient range of ω_m , could be driven by the tuning voltage, producing an output $\omega_{in}C/K_v$, where C is a constant. Of course, the signal level u_2 becomes increasingly weaker at higher ω_m , so it will eventually drop into thermal noise, or other circuit noise, and the compensated signal will not be useful for ω_m/ω_L sufficiently large.

Again according to Table 7.1,

$$|\varphi_e(\omega_m)| = \omega_{in}|1 - H(\omega_m)|/\omega_m \quad (7.47)$$

$$= (\omega_{in}/\omega_n)|1 - H(\omega_m)|/\Omega. \quad (7.48)$$

The magnitude of the phase error must be kept small enough to be within the range of the phase detector and, if the loop parameters that affect H are to be approximately constant, within its more or less linear portion (say $\pm\pi/4$ rad for a balanced mixer). Figure 7.17 can be used to predict this error.

Example 7.5 Frequency Demodulator Let $\alpha = 1$, $\zeta = 0.35$, $\omega_n = 2\pi(10^4)$ rad/sec, $K_v = 10$ kHz/V. At low frequencies, Eq. (7.19) gives

$$u_2 = \omega_{in}/(10^4 \text{ Hz/V}) = f_{in}10^{-4} \text{ V/Hz}.$$

Thus an FM signal with peak deviation 1 kHz will produce a waveform with a peak deviation of 0.1 V at low frequencies. At higher frequencies the response is modified according to the response curve shown in Fig. 7.8 for the given parameters and thus peaks 5 dB higher at approximately $\omega_m = 0.82\omega_n \Rightarrow f_m = 0.82 f_n = 8200$ Hz. The response is 3 dB below the low-frequency response at

$$\omega_m = 1.7\omega_n \Rightarrow f_m = 1.7f_n = 17,200 \text{ Hz}.$$

By Eq. (7.47) and Fig. 7.17, the maximum phase deviation will occur at $\omega_m = \omega_n$ and will be 3 dB above (ω_{in}/ω_n) at that frequency. Thus Eq. (7.48) leads to

$$|\varphi_e| \leq (\omega_{in}/\omega_n)\sqrt{2}.$$

To keep this error below $\pi/4$, we must have

$$(\omega_{in}/\omega_n)\sqrt{2} < \pi/4.$$

Using the given ω_n , f_{in} must therefore be kept below $[\pi \times 10^4/(4\sqrt{2})]$ Hz = 5550 Hz to keep the phase detector characteristic in its more linear range.

When the loop has a low-pass filter ($\alpha = 0$), $u_2 = K_{LF}u_1$ at low frequencies, so a frequency demodulation response can be obtained from u_1 also, as shown by Eq. (7.48) and Fig. 7.17. There are similarities between this response at u_1 when $\alpha = 0$ and the response at u_2 when $\alpha = 1$ (Fig. 7.8). Both are flat at low frequencies and fall off -6 dB/octave at high frequencies. We can obtain a stronger filtering action by passing either response through a series low-pass filter to produce a -12 dB/octave fall-off at high frequencies. This can be done quite naturally when $\alpha = 0$ by simply taking the response from u_2 . Then the responses with the steeper high-frequency skirts, shown in Fig. 7.8 for $\alpha = 0$, including the maximally flat response (see Section 7.6.4), are

obtained. It is also possible to obtain these same responses from loops where $\alpha \neq 0$ by adding the proper low-frequency filter, as we will now show.

It is apparent from Table 4.1 that $H(s)$ for $\alpha \neq 0$ (Eq. 4.10) can be changed to $H(s)$ for $\alpha = 0$ [Eq. (4.20)] through multiplication by $1/(1 + s/\omega_z)$, the response for a low-pass filter whose pole frequency is the same as the zero frequency of the loop. While this can be done by adding such a low-pass filter in series with the demodulated output at u_2 , we can obtain the desired response without adding a filter if the loop is using a passive lag-lead filter. From Fig. 3.19 and Eq. (3.15) we see that the desired filtering of u_2 is available across C in that loop filter. Thus the demodulated voltage taken there is

$$v = u_2 \frac{1}{1 + s/\omega_z} = \frac{\omega_{in}}{K_v} \left\{ \frac{H(\omega_m)_{\alpha \neq 0}}{1 + s/\omega_z} \right\} = \frac{\omega_{in}}{K_v} H(\omega_m)_{\alpha=0}. \quad (7.49)$$

7.8.2 Phase Demodulation

By Table 7.1,

$$u_1 = \varphi_{in}[1 - H(\omega)]K_p, \quad (7.50)$$

which is like Eq. (7.15) and leads to Eq. (7.20). Thus the voltage at the phase detector output equals the input phase multiplied by K_p at high frequencies, and the response shape is given by Figs. 7.9 and 7.13. As with frequency demodulation, the fall-off of the response beyond ω_L can be compensated.

By Table 7.1, the output frequency that accompanies the input phase, and which must be accommodated by the VCO, is

$$|\omega_{out}| = |\varphi_{in}H(\omega_m)\omega_m| = |\varphi_{in}|\omega_n[|H(\Omega)|\Omega] \quad (7.51)$$

and $|H(\omega_m)|\Omega$ can be obtained from Fig. 7.16.

7.8.3 Frequency Modulation

By Table 7.1,

$$\omega_{out} = \Delta u_2[1 - H(\omega_m)]K_v, \quad (7.52)$$

which is like Eq. (7.21) and leads to Eq. (7.26). Thus the output frequency equals Δu_2 , multiplied by K_v at high frequencies and the response shape is given by Figs. 7.9 and 7.13.

The accompanying phase error is given by Table 7.1 as

$$\begin{aligned} -\varphi_e &= \varphi'_{out} = -j\Delta u_2 K_v [1 - H(\omega_m)]/\omega_m \\ &= -j(\Delta u_2 K_v/\omega_n)[1 - H(\omega_m)]/\Omega, \end{aligned} \quad (7.53)$$

and $|1 - H(\omega_m)|/\Omega$ can be obtained from Fig. 7.17.

7.8.4 Phase Modulation

By Table 7.1,

$$\varphi_{\text{out}} = \Delta u_1 H(\omega_m) / K_p, \quad (7.54)$$

which is like Eq. (7.23) and leads to Eq. (7.25). Thus the output phase equals Δu_1 divided by K_p at low frequencies, and the response is given by Figs. 7.8 and 7.12.

The tuning range of the VCO would be of concern. The required range can be obtained from Table 7.1, where the output frequency is given by

$$\omega_{\text{out}} = \Delta u_1 H(\omega_m) j \omega_m / K_p = j \Delta u_1 (\omega_n / K_p) H(\omega_m) \Omega, \quad (7.55)$$

and $|H(\omega_m)|\Omega$ can be obtained from Fig. 7.16.

7.8.5 Extending the Modulation Frequency Range

By introducing properly related modulations at both Δu_1 and Δu_2 , a combined response can be obtained that has bandwidth much wider than would normally be permitted by the loop bandwidth.

7.8.5.1 Frequency Modulation. If modulation is introduced at both Δu_1 and Δu_2 , the combined response will be the sum of Eqs. (7.52) and (7.55), giving

$$\omega_{\text{out}} = \Delta u_2 [1 - H(\omega_m)] K_v + j \Delta u_1 H(\omega_m) \omega_m / K_p. \quad (7.56)$$

If Δu_1 is processed such that

$$\Delta u_1 = -j \Delta u_2 K_p K_v / \omega_m, \quad (7.57)$$

Eq. (7.57) becomes

$$\omega_{\text{out}} = \Delta u_2 K_v \quad (7.58)$$

independently of modulation frequency. The process indicated by Eq. (7.57) would involve integration of Δu_2 , as illustrated in Fig. 7.18a. Equating $\Delta u_1 / \Delta u_2$ in (7.57) to $KF(s)$ for an integrator [Eqs. (7.57) and (3.35)], we see that

$$RC = \frac{1}{K_p K_v}. \quad (7.59)$$

The integration could be obtained approximately by a low-pass filter, operating beyond

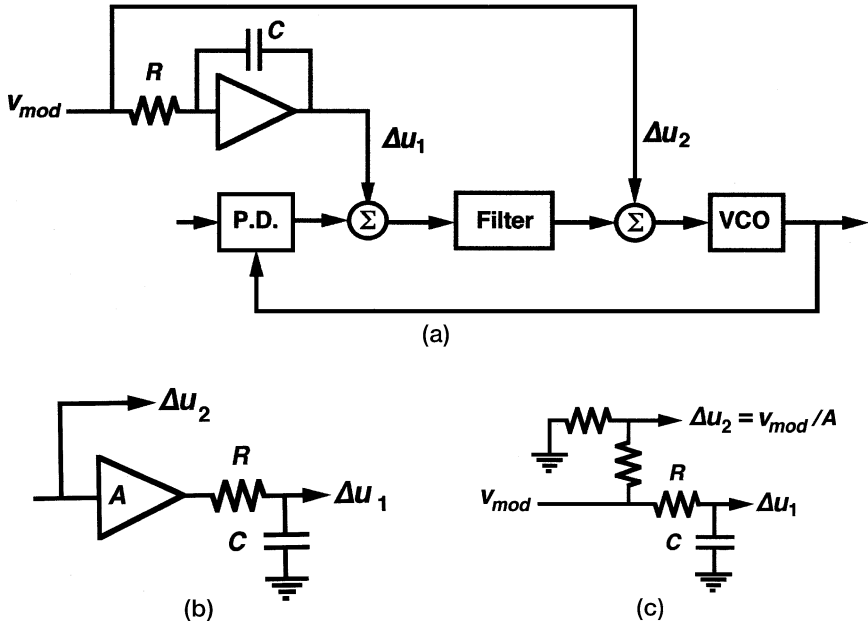


Fig. 7.18 Expanding the bandwidth of a frequency modulator.

its cutoff frequency, over a limited range of frequencies. This is illustrated in Fig. 7.18b where

$$\Delta u_1 = \Delta u_2 A / [1 + j\omega_m RC] \Rightarrow -j\Delta u_2 A / [\omega_m RC] \tag{7.60}$$

for

$$\omega_m \gg 1/(RC). \tag{7.61}$$

Equation (7.57) will be satisfied if also

$$RCK_p K_v = A. \tag{7.62}$$

An amplifier is not needed if the last two equations can be satisfied without it or if A is achieved effectively by attenuating the modulation source to obtain Δu_2 , as shown in Fig. 7.18c.

We must be aware that φ_e will continually increase as ω_m falls [Eq. (7.7)], assuming constant frequency deviation, so that we will run into linearity problems at low enough modulation frequencies.

7.8.5.2 Phase Modulation. A similar extension can be performed for phase modulation. Combining Eqs. (7.54) and (7.53) we obtain

$$\varphi_{out} = \Delta u_1 H(\omega_m)/K_p - j \Delta u_2 K_v [1 - H(\omega_m)]/\omega_m. \tag{7.63}$$

Use

$$\Delta u_2 = j \Delta u_1 \omega_m / (K_p K_v) \tag{7.64}$$

to give a composite of

$$\varphi_{out} = \Delta u_1 / K_p. \tag{7.65}$$

Equation (7.64) requires a differentiator but can be realized by a high-pass circuit with cutoff frequency sufficiently higher than the modulation frequencies of interest. This is illustrated in Fig. 7.19 where

$$\Delta u_2 = A/[1 + 1/(RCj\omega_m)]\Delta u_1 \Rightarrow j \Delta u_1 \omega_m ARC \tag{7.66}$$

for

$$\omega_m \ll 1/(RC). \tag{7.67}$$

Equation (7.65) will be satisfied also if

$$ARCK_p K_v = 1. \tag{7.68}$$

Comments in the previous section concerning the need for an amplifier apply here also.

We must also be aware that the required VCO tuning range will continually increase as ω_m increases [Eq. (7.7)], assuming constant phase deviation.

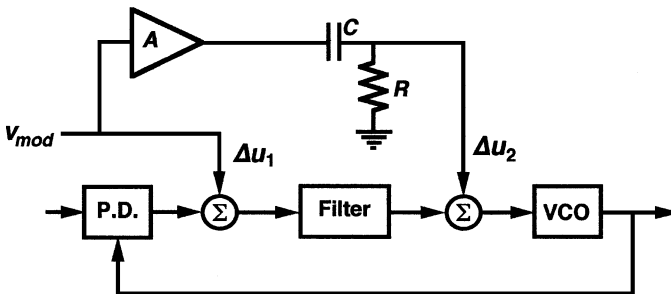


Fig. 7.19 Expanding the bandwidth of a phase modulator.

7.9 MEASUREMENT OF LOOP PARAMETERS FOR $\alpha = 0$ OR 1 FROM MODULATION RESPONSES

One can verify the design loop parameters by comparing measured responses to the expected responses using the curves in this chapter. A more direct method is available for determining ω_n and ζ for a loop when α is 0 or 1. With an integrator-and-lead filter, we can frequency modulate the input, sweep the modulation frequency (constant frequency deviation), and observe the phase detector output. By Table 7.1 the response shape will be like the curves shown in Fig. 7.17 for $\alpha = 1$. It can be shown, by algebraic manipulation of Eq. (4.23) that [Blanchard, 1976, pp. 124–127], if ω_- and ω_+ are the two values of ω_m where the response is 3 dB below the peak response, then

$$\omega_n^2 = \omega_+ \omega_- \quad (7.69)$$

and

$$\omega_+ - \omega_- = 2\zeta\omega_n. \quad (7.70)$$

Comparing Eqs. (4.25) and (4.20) (see Table 4.1), it is apparent that $|[1 - H(\omega_m)]/\omega_m|$ for $\alpha = 1$ has the same shape as does $|H(\omega_m)/\omega_m|$ for $\alpha = 0$. This can also be seen by comparing Figs. 7.16 for $\alpha = 0$ and 7.17 for $\alpha = 1$. Therefore, by Table 7.1, Eqs. (7.69) and (7.70) also apply to a loop with a low-pass filter if we phase modulate the input, sweep the modulation frequency (constant phase deviation), and determine the modulation frequencies, ω_+ and ω_- , at which the deviation of the tuning voltage is 3 dB below its peak value.

7.10 EFFECT OF AN ADDED POLE

Figure 7.20 illustrates the effect of a pole added to what would otherwise be a type 2, second-order, loop with $\zeta = 0.5$. We see little effect on the modulation response when the frequency ω_a of the added pole is 10 times ω_n , which approximately equals ω_L when $\zeta = 0.5$ (they are equal in the tangential approximation). As ω_a comes closer to ω_n , the effect grows until the response becomes infinite at the natural frequency when $\omega_a = \omega_n$. At that point the added pole cancels the zero, giving a constant 180° excess phase and producing instability at the frequency where open-loop gain is one. Compare the effects of the added pole on the modulation response to its effect on the transient response (see Section 6.12).

7.M APPENDIX: MODULATION RESPONSE USING MATLAB[®]

`FreqRsp` draws the magnitude and phase of the output and error responses versus log of frequency, normalized (to ω_n) in most cases. It can draw curves like those in Figs.

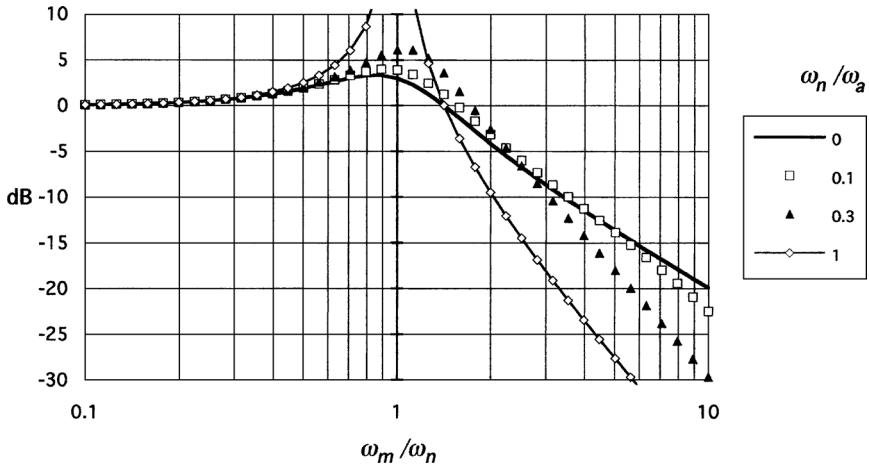


Fig. 7.20 Effect of a pole added to type 2 loop ($\zeta = 0.5$).

7.6 through 7.15 or like Fig. 7.20 as well as first-order loop response. We can select the kind of loop and its parameters dynamically. While the multiple curves presented in the text have some advantages, the advantage of the program is that it gives us more freedom to choose parameters, especially α . We have shown how transient responses for any α can be obtained by combining response for $\alpha = 0$ and $\alpha = 1$ in the correct proportion, but this is not easily done with the phase or magnitude of the responses.

FreqRsp

```
%FREQUENCY RESPONSE
% Choose loop and parameters dynamically.
%   Compare first-order to Fig. 2.6,
%       second-order to Fig. 7.6–7.15
%       third-order to Fig. 7.20
r2deg = 180/pi; % for use in changing radians to degrees
w = logspace (-1,1,51); % Frequency range
Wn = 1 % Natural frequency normalized

go = 2;
while (go ~= 0)
    order = input ('Enter order (1-3), 0 to quit: ','s');
    if strcmp(order, '0'),
        go = 0;
    elseif strcmp(order, '1'),
        disp('First Order');
        K = input('First Order. Input K:');
        out = [0 K]; % from Eq. (2.25b)
        er = [1 0]; % from Eq. (2.26b)
        den = [1 K];
    elseif strcmp(order, '3'),
```

```

rn = input('Third Order, Type 2, zeta = 0.5. Enter
           Wn/W+: ');
out = [0 0 1 1]; % as in SS Added Pole, paragraph 6.M.5
er = [rn 1 0 0]; % from Eq. (6.M.4)
den = [rn 1 1 1];
elseif (go~=0),
    alpha = input('Second Order. Input alpha: ');
    zeta = input('Input zeta: ');
    out = [2*alpha*zeta 1]; % from Eq. (6.4a)
    er = [1 2*zeta*(1-alpha) 0]; % from Eq. (6.4b)
    den = [1 2*zeta 1];
end
hout = freqs(out, den, w);
magout = 20*log10(abs(hout));
phout = r2deg*angle(hout);
her = freqs(er, den, w);
mager = 20*log10(abs(her));
pher = r2deg*angle(her);

clf
subplot(211), semilogx(w, magout, 'r', w, mager, ':r')
grid
title('Frequency Response Magnitude. solid = output
      dotted = error')
xlabel('Omega / Omega-n')
ylabel('dB')
subplot(212), semilogx(w, phout, 'g', w, pher, ':g')
grid
title('Frequency Response Phase. solid = output dotted
      = error')
xlabel('Omega / Omega-n')
ylabel('degrees')
end % while go

```

In FRspDorI, the second-order response is modified to produce the derivatives and integrals, as shown in Figs. 7.16 and 7.17, by shifting the numerator vector elements (see Section i6.M.1).

CHAPTER 8

ACQUISITION

So far we have considered linear performance almost exclusively. Of course, we normally seek to ensure that the loop will operate in this fashion. However, it is essential to know under what conditions the loop will not lock, and it is valuable to understand what is involved in the acquisition of phase lock [Blanchard, 1976, Chapter 10; Gardner, 1979, Chapter 5; Gardner, 2005, Chapter 8].

8.1 OVERVIEW

Refer to Fig. 8.1. The frequency of the voltage-controlled oscillator (VCO) when the phase detector is in the center of its range will be called the center frequency, ω_c . This can be varied by adjusting the fixed frequency-determining elements in the oscillator or perhaps some bias that is added to the tuning voltage or to the phase detector output. We will call the difference between the input frequency and the center frequency the mistuning, Ω :

$$\Omega \triangleq \omega_{\text{in}} - \omega_c. \quad (8.1)$$

The initial value of frequency error, when the signal is applied, is

$$\omega_e(0) = \Omega. \quad (8.2)$$

In general, there is a range of mistuning, called the seize or lock-in range, $|\Omega| \leq \Omega_s$, over which the ω_{out} will immediately move to the steady-state lock frequency, ω_{in} . In

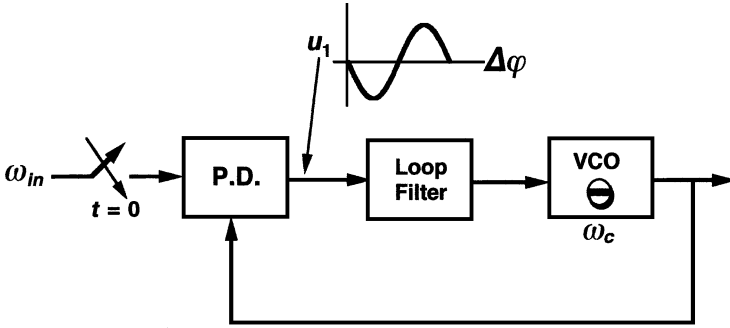


Fig. 8.1 Loop for acquisition.

other words, Ω will immediately begin decreasing toward zero when the input switch is closed.

There is a range of mistuning, called the pull-in or acquisition range, $|\Omega| \leq \Omega_{PI}$ over which the loop will eventually lock but may skip cycles, sometimes many cycles, before doing so. By skipping cycles is meant that the difference between the input and the output frequency causes the phase detector characteristic to be traversed past the phases where the voltage is minimum and maximum.

There is range of mistuning, called the hold-in or synchronization range, $|\Omega| \leq \Omega_H$, over which the loop will maintain lock but over which it will not necessarily acquire lock.

The Ω_S and Ω_{PI} ranges are guaranteed maximum values where the action described above will occur regardless of initial phase. These actions may occur at wider mistunings if the phase happens to have a fortuitous value.

These ranges are illustrated in Fig. 8.2. We will sometimes use notation such as $\pm\Omega_{PI}$ to emphasize that we are considering the deviation of ω_{in} from ω_c rather than the total range over which ω_{in} may vary, which is twice as large.

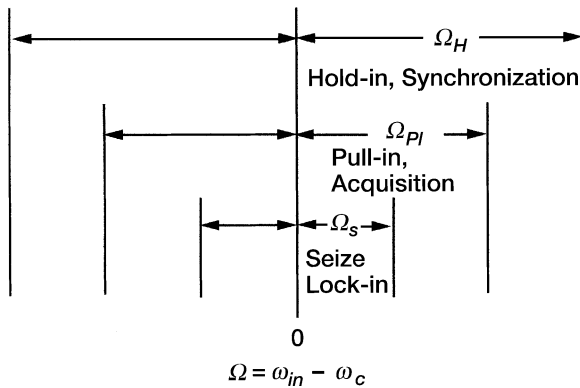


Fig. 8.2 Mistuning.

The relationship

$$\Omega_S \leq \Omega_{PI} \leq \Omega_H \quad (8.3)$$

must always hold, just because of the definitions. The equalities in this relationship apply to first-order loops and highly damped second-order loops, which act like first-order loops. First-order loops move toward steady state whenever it is possible to tune to it. The phase does not overshoot but moves directly toward its final value. In a first-order loop, the value of the phase error is instantly related to the value of the frequency; there can be only one frequency for a given phase error. In a second-order loop, due to the filter, this is not so and various total states can exist at a given frequency. Thus the second-order loop might move directly toward steady state when it is at a given frequency or it might skip cycles at the same frequency, depending on the mix of state variables under each condition—we might also say, depending on its history.

We will now describe a possible experiment with a second-order loop to illustrate the various ranges. The loop is locked with $f_{in} = 9$ MHz. The input frequency is *slowly* moved higher, past the 10-MHz center frequency, which is the frequency of the VCO when the phase detector is at midrange. At 14 MHz, lock is lost because the phase detector cannot put out enough voltage to tune the VCO to 14 MHz. (Generally we assume that the phase detector is the only nonlinear element, since its nonlinearity is inherent.) This is an edge of the total hold-in range. Now the f_{in} is slowly lowered until lock is reestablished at 12 MHz. This is an edge of the total acquisition range. When the input frequency is adjusted to 12 MHz, with the input switch open, and the switch is then closed, the loop skips cycles for a long time but finally comes into lock. But, if the frequency is a little higher when the switch is closed, no lock occurs. The experiment is repeated at lower and lower frequencies until a frequency is found, say 11 MHz, where, every time, the loop goes to steady state, after the switch is closed, without skipping cycles. This is one edge of the seize range. We can repeat the experiment on the low-frequency side and, assuming we have an ideal (linear except for the balanced-mixer phase detector) loop, the deviations from 10 MHz will be equal, for each range, on the low side to what they were on the high side. In that case we would conclude that $\Omega_H = \pm 4$ MHz, $\Omega_{PI} = \pm 2$ MHz, and $\Omega_S = \pm 1$ MHz. Or we might be more informative by saying that the hold-in range extends from 6 to 14 MHz, the pull-in range extends from 8 to 12 MHz, and the seize range extends from 9 to 11 MHz.

In contrast to the above, if the loop were first order and we tuned higher, past $f_c = 10$ MHz, the results might be as follows. At 13 MHz, for example, the loop breaks lock. As we tune slowly back, it will again acquire lock, with no cycle skipping, at 13 MHz. Thus, for this first-order loop, $\Omega_S = \Omega_{PI} = \Omega_H = 3$ MHz.

The time required to get from an initial frequency error $\omega_e = \Omega$ to $\omega_e = \Omega_S$ is the acquisition or pull-in time, T_{PI} . During this time the phase error and frequency (or, more easily observed, u_1 and u_2) might look as pictured in Fig. 8.3. The difference between ω_{in} and ω_{out} causes the phase to move continuously, leading to approximately a sinusoid at u_1 as the phasedetector characteristic is repeatedly traced out. However,

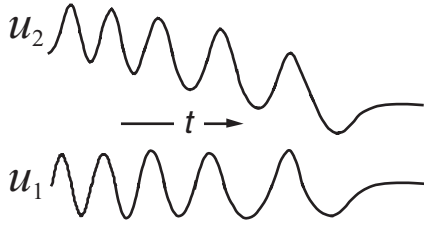


Fig. 8.3 Waveforms during acquisitions.

this is not a perfect sinusoid. Due to the small feedback that does exist, it spends slightly more time at low voltages than at high voltages, leading to a negative average value. As shown, u_2 is approximately sinusoidal, following u_1 , but the negative average value in u_1 causes the average output of the loop filter to fall. As it does so, the average output frequency comes closer to the input frequency. As this happens, $\omega_e (= \omega_{in} - \omega_{out})$ is lowered so the loop gain increases, causing more feedback and greater distortion of the sinusoid and accelerating the pull-in process.

Let us look at the same process from a slightly different perspective. Fig. 8.4 shows, on a Bode plot, some of the parameters of importance for a “high-gain” ($\omega_z \ll \omega_L$) second-order loop. [Note that, while $K < K'$ is usable for small signal analysis, in this chapter $K = K'$, $K_p = K'_p$, and parameters such as ω_L are computed with $K = K'$.] It turns out that the acquisition frequency falls approximately at the geometric mean of the hold-in range K and ω_L (i.e., $\Omega_{PI} \approx \sqrt{K\omega_L}$) while the seize frequency is approximately equal to ω_L (i.e., $\Omega_S \approx \omega_L$). When the initial mistuning is between these frequencies, the loop skips cycles while the DC content of u_1 is integrated and slowly reduces the mistuning. The process is accelerated as the seize frequency is approached and, when it is finally reached, the loop locks rapidly. During the last part of the process, when the loop is in a linear region (gains are approximately constant), the frequency changes in a manner given by the transient response curves previously studied. The appropriate response is a combination of a phase step response and a frequency step response, the combination giving the appropriate initial phase and frequency.

The developments that follow are for a sinusoidal phase detector characteristic. Somewhat improved performance should be expected (for the same K_p) when the

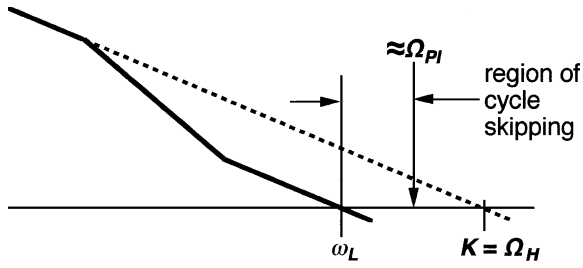


Fig. 8.4 Acquisition parameters on a Bode plot, sine PD high gain.

phase detector characteristic becomes more triangular, say as a result of near equal signal levels at the two inputs (see Section 3.1.4.2).

8.2 ACQUISITION AND LOCK IN A FIRST-ORDER LOOP

We will study the process of acquisition in a first-order loop because the relatively easy analysis will show effects that have similarities to those in higher order loops.

From Fig. 8.5, the output frequency is

$$\omega_{\text{out}} = \omega_c + K_v u_1 \quad (8.4)$$

while u_1 is given by

$$u_1 = K_p \cos[\Delta\varphi(t)]. \quad (8.5)$$

We can subtract Eq. (8.4) from ω_{in} and employ the definition of Ω in Eq. (8.1) to give

$$\omega_e = \omega_{\text{in}} - \omega_{\text{out}} = \omega_{\text{in}} - \omega_c - K_v u_1 = \Omega - K_v u_1. \quad (8.6)$$

Since $\omega_{\text{in}} - \omega_{\text{out}} = d\Delta\varphi/dt$, this can also be written

$$d\Delta\varphi/dt = \Omega - K_v u_1. \quad (8.7)$$

Thus if u_1 is greater than Ω/K_v , the value at which ω_{out} equals ω_{in} , $\Delta\varphi(t)$ will decrease. If this occurs on the positive slope of $\cos \Delta\varphi$ ($-\pi \leq \Delta\varphi \leq 0$), the decreasing $\Delta\varphi$ will cause u_1 to decrease and thus reduce the value of ω_{out} . This will continue until $\omega_{\text{out}} = \omega_{\text{in}}$ and $u_1 = \Omega/K_v$ and then cease (Fig. 8.6). The smaller the frequency error ω_e the smaller the absolute value of Eq. (8.7) and the more slowly moves $\Delta\varphi$. Thus the pull-in is eventually exponential. If this occurs on the negative slope of the phase detector characteristic ($0 \leq \Delta\varphi \leq \pi$), the decreasing $\Delta\varphi$ will lead to higher values of u_1 , which will cause the frequency error to increase and $\Delta\varphi$ to move even faster until $\Delta\varphi = 0$, at which point the process described before will occur.

There is a second equilibrium point on the opposite slope of the characteristic, but it is unstable. No error signal is generated at that point, but any deviation from that

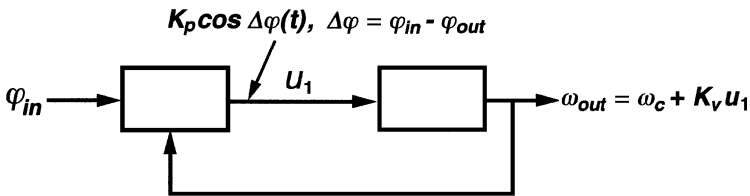


Fig. 8.5 First-order loop.

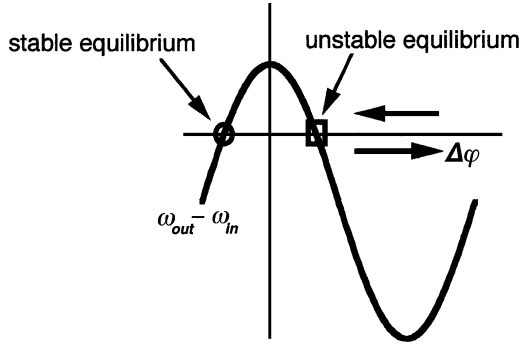


Fig. 8.6 First-order loop characteristic.

point will cause the phase to move away from it, as indicated by the arrows in Fig. 8.6, whereas any deviation from the stable point causes the phase to move back toward it. If the initial phase is near the value for unstable equilibrium, the phase error will be near zero so the phase will move very slowly initially. This is called “hang up” [Gardner, 1977] and can be a problem when the time required to attain steady state is important.

Another way to look at the process is illustrated in Fig. 8.7, which shows a phasor ox of magnitude K_p at an angle $\Delta\phi$. The horizontal projection equals u_1 . The negative of the rate of change of $\Delta\phi$ is, by Eq. (8.7), proportional to the deviation of u_1 from the line $u_1 = \Omega/K_v$, the value that, by Eq. (8.6), produces $\omega_{out} = \omega_{in}$. Thus the outside arrows in Fig. 8.7 show $\Delta\phi$ decreasing when u_1 is higher than this value and increasing otherwise; a is the point of unstable equilibrium and c is the point of stable equilibrium. If x is between b and c , u_1 will decrease continuously as c is approached. If x is between a and b , u_1 will peak at b before decreasing to its final value at c . If x is between a and b' , u_1 will reach a negative peak before settling at c . These various

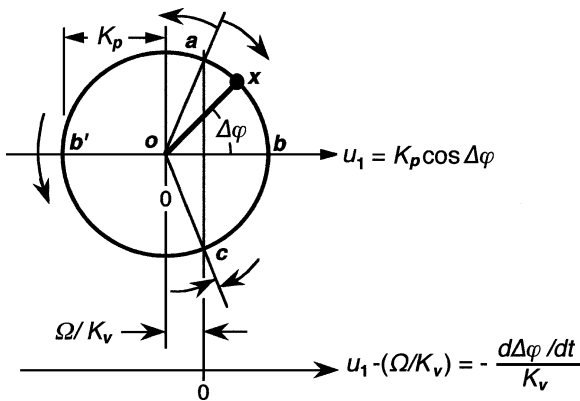


Fig. 8.7 Phasor representation of first-order loop pull-in.

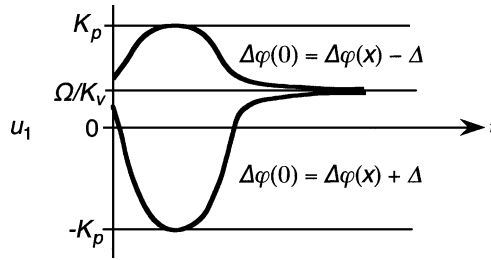


Fig. 8.8 u_1 during pull-in.

possibilities are illustrated in Fig. 8.8 where an attempt has been made to show the effect of the varying $d\Delta\varphi/dt$.

Figure 8.8 covers all the response shapes because the change in $\Delta\varphi$ is completely determined by u_1 . If the initial $\Delta\varphi$ is closer to final value than the initial condition in Fig. 8.8, we merely start further to the right on the appropriate curve. Surprisingly, the time to get from a phase near the unstable equilibrium point to a small offset from steady state is the same for both trajectories as long as the small initial and final offsets are identical, which we will now show.

8.2.1 Transient Time

Combining Eqs. (8.5) and (8.7) we obtain

$$dt = \frac{d\Delta\varphi(t)}{\Omega - K \cos \Delta\varphi(t)}. \quad (8.8)$$

This can be integrated to give the time for $\Delta\varphi(t)$ to change from an initial value $\Delta\varphi(0)$ at $t = 0$ to a value $\Delta\varphi(t)$ at time t :

$$t = \int_0^t dt = \frac{1}{\sqrt{K^2 - \Omega^2}} \ln \left| \frac{\tan[\Delta\varphi(t)/2] - \beta \tan[\Delta\varphi(0)/2] + \beta}{\tan[\Delta\varphi(t)/2] + \beta \tan[\Delta\varphi(0)/2] - \beta} \right|, \quad (8.9)$$

where

$$\beta = \sqrt{\frac{K - \Omega}{K + \Omega}}. \quad (8.10)$$

As t approaches infinity, the denominator inside the $||$ must approach zero, implying that

$$\beta = -\tan \frac{\Delta\varphi(\infty)}{2}. \quad (8.11)$$

The time t also becomes infinite when $\tan[\Delta\varphi(0)/2] = \beta$, consistent with this being a point of unstable equilibrium (hang-up). Comparing β for the two cases, we can

see that unstable equilibrium occurs when the initial phase is the negative of the final phase [i.e., $\Delta\varphi(0) = -\Delta\varphi(\infty)$], as can be seen from Fig. 8.7.

To improve our understanding of the process, let us address this question: If $\Delta\varphi(0)$ is offset by a small angle ε from the point of unstable equilibrium, how will the time required, to come within the small angular offset μ of final value, when the initial offset is positive compare to the time required when it is negative? A small offset in angle causes a small offset in the tangent related by

$$d \tan x = \frac{dx}{\cos^2 x}. \quad (8.12)$$

Thus we can write, for the point of unstable equilibrium,

$$\beta = \tan \frac{\Delta\varphi(0) + \varepsilon}{2} \approx \tan \frac{\Delta\varphi(0)}{2} + \frac{\varepsilon}{2 \cos^2[\Delta\varphi(0)/2]} \quad (8.13)$$

and, for the point of stable equilibrium,

$$-\beta = \tan \frac{\Delta\varphi(t) - \mu}{2} \approx \tan \frac{\Delta\varphi(t)}{2} - \frac{\mu}{2 \cos^2[\Delta\varphi(t)/2]}. \quad (8.14)$$

The terms $\Delta\varphi(0)$ and $\Delta\varphi(t)$ represent an initial angle slightly less than the angle of unstable equilibrium and a final angle slightly more than the angle of stable equilibrium and, thus, a transient that moves around the right side of the circle in Fig. 8.7. Substituting these into Eq. (8.9) we obtain

$$\begin{aligned} t &\approx \frac{1}{\sqrt{K^2 - \Omega^2}} \ln \left| \frac{(-2\beta)}{\mu} \frac{2\beta}{(-\varepsilon)} \frac{2 \cos^2(\Delta\varphi(t)/2)}{2 \cos^2(\Delta\varphi(0)/2)} \right| \\ &= \frac{1}{\sqrt{K^2 - \Omega^2}} \ln \left\{ \frac{16\beta^2}{\mu\varepsilon} \cos^4 \left(\frac{\Delta\varphi(0)}{2} \right) \right\}. \end{aligned} \quad (8.15)$$

Since μ and ε appear only as a product, if we repeat this process with the signs changed, to represent a locus around the left side of the circle, we obtain the same answer, which is what we were to demonstrate.

8.2.2 Acquisition

We know that, if $|\Omega| > K = K_p K_v$, the loop cannot be locked because the mistuning is outside the hold-in range. Graphically, when Ω/K_v exceeds K_p , a and c no longer exist in Fig. 8.7 because the vertical line on which they lie no longer intersects the circle. Then $d\Delta\varphi/dt$ is always positive and the phase continuously traverses the circle, slowing as it approaches the line and speeding as it moves away. Of course, similar

statements apply to mistuning in the opposite direction. We can analyze the out-of-lock waveforms to determine their shape more exactly in a manner similar to that used in the last section. Now, however, β in Eq. (8.10) will be imaginary so we define

$$\lambda = \sqrt{\frac{\Omega + K}{\Omega - K}} \tag{8.16}$$

and employ it in another form of the solution of Eq. (8.8) to obtain

$$t \approx \frac{2}{\sqrt{\Omega^2 - K^2}} \tan^{-1} \left\{ \lambda \tan \left(\frac{\Delta\varphi(t)}{2} \right) \right\} \tag{8.17}$$

or, equivalently,

$$\tan qt = \lambda \tan \left(\frac{\Delta\varphi(t)}{2} \right), \tag{8.18}$$

where

$$q = \frac{\sqrt{\Omega^2 - K^2}}{2}. \tag{8.19}$$

We can interpret this with the help of Fig. 8.9, which is drawn for $\lambda < 1$ ($\Omega < 0$).

As qt increases steadily with time, $\Delta\varphi(t)$ increases to maintain the relationship between tangents. When qt is small,¹ $\Delta\varphi(t)/2$ must increase at a rate $1/\lambda$ times faster than qt because $\tan(x) \approx x$ and the radius of the right circle is that much smaller. However, both angles reach $\pi/2$ at the same time so the rate of change of $\Delta\varphi(t)/2$ must slow as it approaches that value. This leads to the waveforms shown in Fig. 8.10. Notice the typical behavior in which more time is spent at negative values of u_1 . In

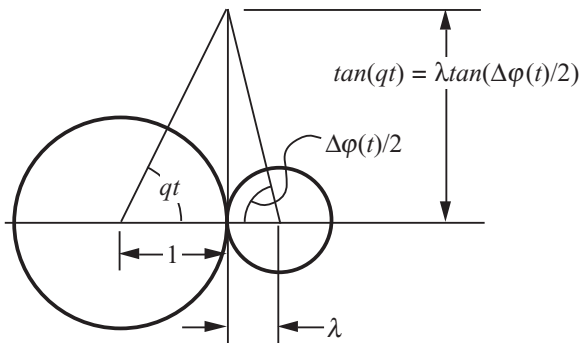


Fig. 8.9 Aid for interpreting cycle skipping waveshape.

¹ As in the typical tangent representation, when qt is outside this range, the back projection of the radius line is used to intersect the vertical line.

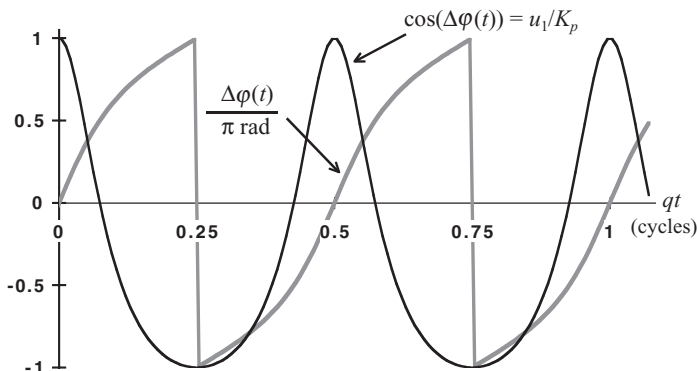


Fig. 8.10 Waveforms in first-order loop skipping cycles.

a second-order loop this tends to bring ω_{out} down to meet ω_{in} , which is lower, as implied by $\Omega < 0$. If $\Omega > 0$, the right circle would be larger than the left circle and more time would be spent at positive values to bring the VCO frequency up to the input frequency.

8.3 ACQUISITION FORMULAS FOR SECOND-ORDER LOOPS WITH SINE PHASE DETECTORS

Before developing our understanding of the acquisition further through analysis, some equations relating to acquisition in second-order loops will be presented. These have been developed through analysis or experiment and often involve approximations so it is essential to understand the limits of applicability.

All of these equations are for a sinusoidal phase detector characteristic. Results for other types [Egan, 1981, pp. 211–220; Egan, 2000, pp. 408–418] are summarized in Appendix 8A. To reiterate,

For a first-order loop,

$$\Omega_H = \Omega_{PI} = \Omega_S = K \text{ rad.} \quad (8.20)$$

For all orders,

$$\Omega_H = K \text{ rad.} \quad (8.21)$$

With an ideal integrator-and-lead filter ($\alpha = 1$), K is infinite, so the average component produced by the nonsymmetrical u_1 waveform during cycle skipping, no matter how small, will eventually be integrated to a large enough value to produce acquisition.

For an integrator-and-lead filter ($\alpha = 1$), with ideal components,

$$\Omega_H = \Omega_{PI} = \infty. \quad (8.22)$$

Even if K could be infinite, real phase detectors do not produce exactly zero average value in the absence of a signal and real op-amps do not have zero offset. There will always be some offset and it may overcome the tiny average component. In the presence of finite offset, excessive DC gain can do more harm than good. However, we defer consideration of optimization in the presence of offsets and here give equations that apply to the degree that such offsets can be neglected.

For a lag-lead filter ($0 < \alpha < 1$), with ideal components (see Fig. 8.11),

$$\Omega_{PI} \approx 2K\sqrt{x - x^2}, \quad K \gg \omega_z, \tag{8.23}$$

$$x \equiv \omega_p / (2\omega_z) \quad [\text{Greenstein, 1974}], \tag{8.24}$$

$$\Omega_{PI} \approx 2K\sqrt{x} \approx 2\sqrt{K\zeta\omega_n}, \quad K \gg \omega_z \gg \omega_p, \tag{8.25}$$

$$\Omega_S = \omega_n^2 / \omega_z = \omega_L, \quad \omega_z \ll \omega_L \quad [\text{Gardner, 2005, pp. 186–188; Kroupa, 1973, p. 177}]. \tag{8.26}$$

See Appendix 8.V for Ω_{PI} in terms of α . Under the conditions of Eqs. (8.25) and (8.26), the time to go from $\omega_e = \Omega$ to $\omega_e = \Omega_S$ is [Richman, 1954a]

$$T_{PI} = \Omega^2 / (2\zeta\omega_n^3), \quad \Omega_{PI} \gg \Omega \gg \Omega_S. \tag{8.27}$$

For a low-pass filter ($\alpha = 0$) [Rey, 1960],

$$\Omega_H \geq \Omega_{PI} \approx 3\zeta K \sqrt{\sqrt{0.423 + 1.2\zeta^4} - 1.092\zeta^2}. \tag{8.28}$$

Here the inequality indicates that the pull-in range cannot exceed the hold-in range, even if the formula gives a larger value.

The value given by Eq. (8.26) for seize frequency can be justified as follows. The loop cannot skip a whole cycle without producing the peak output from the

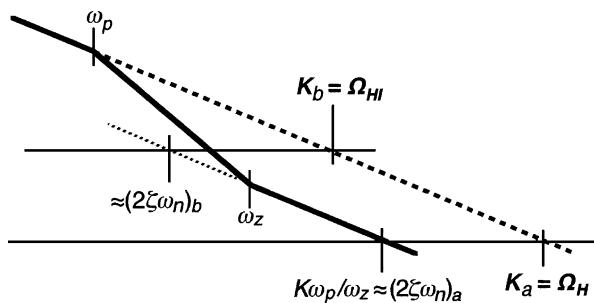


Fig. 8.11 Some parameters if $\omega_z \gg \omega_p$.

phase detector, $\pm K'_p$. If that peak output appears immediately upon connection to the reference, a worst case in which the initial phase is at the edge of the hold-in range, the resulting step will be amplified by the high-frequency gain, ω_p/ω_z , of the loop filter. This filter output will be amplified by K_v to produce an output frequency step of magnitude,

$$|\Delta\omega| = K'_p(\omega_p/\omega_z)K_v = K'\omega_p/\omega_z = \omega_n^2/\omega_z. \quad (8.29)$$

If this is bigger than the mistuning ($|\Omega| < \omega_n^2/\omega_z$), the mistuning will be overcome by the frequency step. The net frequency error will then cause the phase error to move toward lock. Therefore, locking can occur without cycle skipping with a mistuning as high as ω_n^2/ω_z .

A loop with a low-pass filter has no seize range because the filter prevents an initial phase step from being seen immediately so any mistuning can send the phase in the wrong direction initially.

Example 8.1 Acquisition Formulas Find Ω_H , Ω_S , and Ω_{PI} , for a loop with a sinusoidal phase detector and the following parameters:

$$\omega_p = 100 \text{ rad/sec}; \quad \omega_z = 500 \text{ rad/sec}; \quad K = 10,000 \text{ sec}^{-1}.$$

From (8.21),

$$\Omega_H = K = 10,000 \text{ rad/sec} = 1592 \text{ Hz.}$$

From (8.24),

$$x = 100/(2 \times 500) = 0.1.$$

From (8.25),

$$10,000 \gg 500 >? > 100$$

($> ? >$ indicates inequality satisfied weakly or questionably);

$$\Omega_{PI} \approx 2(10,000 \text{ rad/sec})(0.1)^{0.5} = 6325 \text{ rad/sec} = 1007 \text{ Hz.}$$

Equation (8.23) does not require the latter inequality needed for (8.25). For comparison, from (8.23),

$$\Omega_{PI} \approx 2(10,000 \text{ rad/sec})(0.1 - 0.01)^{0.5} = 6000 \text{ rad/sec.}$$

$$\omega_L = K\omega_p/\omega_z = 2000 \text{ rad/sec} >? > \omega_z = 500 \text{ rad/sec.}$$

From (8.26),

$$\Omega_S = \omega_L = 2000 \text{ rad/sec} = 318 \text{ Hz.}$$

8.4 APPROXIMATE PULL-IN ANALYSIS

During acquisition of lock, the spectrum of the VCO might transition through a set of states such as shown in Fig. 8.12. If the loop were slow enough for us to observe this on a spectrum analyzer, we might first see the VCO output at its initial offset from the reference signal. It would slowly move closer to the reference and, as it did, we would begin to notice two frequency-modulation (FM) sidebands. These would be caused by the beat frequency produced by the mixing of the reference and VCO output in the phase detector. Since FM sidebands are offset from the central spectral line by the modulation frequency ω_e , one of the sidebands must be at the same frequency as the reference. As the VCO frequency moves closer to the reference, the frequency of that sideband remains fixed at the reference frequency. As ω_e becomes smaller and the modulation index (and possibly the gain in the loop filter) at that frequency is therefore increased, the sidebands grow in amplitude, and more sidebands begin to appear while the VCO output closes ever more quickly on the reference. Since one sideband is always at the input frequency, it will produce a zero-frequency, or DC, signal from the phase detector. We will see that this is the signal that re-tunes the VCO to produce pull-in. In the time domain it appears as distortion of a near-sinusoidal beat note (Fig. 8.10).

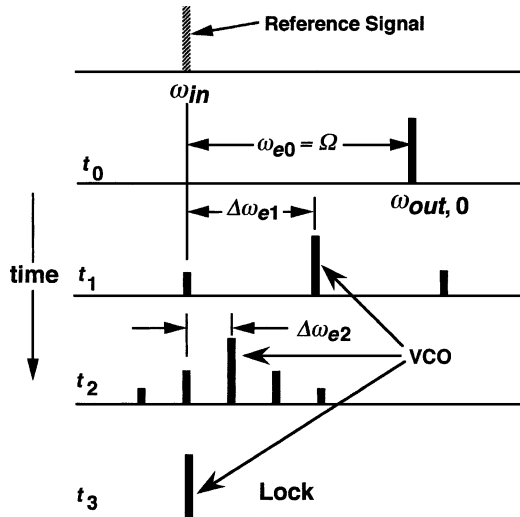


Fig. 8.12 Sequence of spectral pictures during pull-in.

More can be learned about the acquisition performance of the second-order loop through an approximate analysis that computes this bias (DC) component of u_1 . The approximation involved is the assumption, implicit in the computation of the sideband level, that ω_e is constant. That cannot really be true because we use the computed value to determine how ω_e changes with time. Thus the approximation is best when the change in bias is slow so it is poorest when ω_e approaches Ω_S . When we have developed an expression for the bias we will first consider it in the most general terms. That will permit us to predict certain performance features, such as false lock, that can occur when G is more complex than what occurs in second-order loops. Then we will again restrict our considerations to second-order loops and use the computed bias value to obtain some of the results listed in the last section.

8.4.1 Basic Equations

We will now determine the amount ω_b by which the VCO is moved, or tends to be moved, from ω_c due to the bias, which has been described above, that is developed during cycle skipping.

Define

$$\Delta\omega \triangleq \bar{\omega}_{\text{out}} - \omega_{\text{in}} = -\bar{\omega}_e, \quad (8.30)$$

where $\bar{\omega}_{\text{out}}$ is the average frequency of the VCO. Then we can write the output of the phase detector in a manner that reflects our expectations that it will consist of a distorted sinusoid at a frequency $\Delta\omega$,

$$u_1(t, \Delta\omega) = \Re_{\otimes}\{K'_p e^{j\Delta\omega t}\} + v_b + \dots \quad (8.31)$$

The use of K'_p reflects the assumption that the distortion is small so the fundamental component has the full amplitude of the waveform; v_b is the average, or bias, voltage resulting from the distortion of the phase detector (PD) output waveform. The terms that are dropped from the end of (8.31) are harmonics of $\Delta\omega$ that are not important in the interaction under consideration.

The VCO frequency will be moved from ω_c by $u_2 K_v$, which is obtained from (8.31) as

$$\delta\omega(t, \Delta\omega) \triangleq \omega_{\text{out}}(t, \Delta\omega) - \omega_c \quad (8.32a)$$

$$= K_v u_2(t, \Delta\omega) \quad (8.32b)$$

$$= K_v K_{\text{LF}} [\Re_{\otimes}\{K'_p e^{j\Delta\omega t} F(\Delta\omega)\} + v_b]. \quad (8.33)$$

Since

$$G(\Delta\omega) = K_v K_{\text{LF}} K_p F(\Delta\omega) / (-j\Delta\omega), \quad (8.34)$$

Eq. (8.33) is

$$\delta\omega(t, \Delta\omega) = [-\Re\{j\Delta\omega G(\Delta\omega)e^{j\Delta\omega t}\} + v_b K_{LF} K_v] \quad (8.35)$$

$$= \Delta\omega |G(\Delta\omega)| \sin[\Delta\omega t + \angle G(\Delta\omega)] + \omega_b, \quad (8.36)$$

where

$$\omega_b \triangleq v_b K_{LF} K_v = v_b K / K_p \quad (8.37)$$

is the frequency bias term that results from interaction of the first term with the reference and whose value we are attempting to compute.

The corresponding phase modulation is

$$\delta\varphi(t, \Delta\omega) = \int \delta\omega(t, \Delta\omega) dt = -|G(\Delta\omega)| \cos(\Delta\omega t + \angle G(\Delta\omega)) + \dots, \quad (8.38)$$

where the bias term has been dropped. We drop it here because we are analyzing the effect of the other term, but the value of the dropped bias term is represented implicitly in $\bar{\omega}_{\text{out}}$ and therefore in $\Delta\omega$ and ω_b . Thus, while we are computing ω_b , an initial value of this variable, ω_{b0} , is implied. It might seem that we are starting with the answer, but we shall see that the results of the process are nevertheless informative.

The peak phase deviation, or modulation index, can be seen from (8.38) to be

$$m = |G(\Delta\omega)|. \quad (8.39)$$

The output from the VCO is a sinusoid with average frequency $\bar{\omega}_{\text{out}}$, which is phase modulated with the deviation given by (8.39). Assuming that the deviation is small, the VCO output voltage can be written (based on FM modulation theory; see Table 21.A.1)

$$\begin{aligned} v_{\text{out}} = A \{ & \cos \bar{\omega}_{\text{out}} t - 0.5 |G(\Delta\omega)| \{ \cos [(\bar{\omega}_{\text{out}} + \Delta\omega)t + \angle G(\Delta\omega) + \pi/2] \\ & + \cos [(\bar{\omega}_{\text{out}} - \Delta\omega)t - \angle G(\Delta\omega) - \pi/2] \} \}. \end{aligned} \quad (8.40)$$

Substituting $\Delta\omega$ from (8.30) we obtain

$$\begin{aligned} v_{\text{out}} \approx A \{ & \cos \bar{\omega}_{\text{out}} t - 0.5 |G(\Delta\omega)| \{ \cos [(2\bar{\omega}_{\text{out}} - \omega_{\text{in}})t + \angle G(\Delta\omega) + \pi/2] \\ & + \cos [(\omega_{\text{in}})t - \angle G(\Delta\omega) - \pi/2] \} \}. \end{aligned} \quad (8.41)$$

Note that the second sideband (last term) here is at the input frequency, ω_{in} , and will thus produce a DC bias term when mixed in the phase detector with the input signal. The bias is given by Eq. (1.11) except that the magnitude is multiplied by the ratio of

this sideband to A ,²

$$\begin{aligned} v_b &= 0.5|G(\Delta\omega)|K'_p \cos[\angle G(\Delta\omega) + \pi/2] \\ &= 0.5|G(\Delta\omega)|K'_p \sin[\angle G(\Delta\omega)]. \end{aligned} \quad (8.42)$$

This voltage at the phase detector output will be amplified by the DC gain to the output to produce an output frequency bias of

$$\omega_b = 0.5K'F(\Delta\omega \approx 0)|G(\Delta\omega)| \sin[\angle G(\Delta\omega)] \quad (8.43)$$

$$\approx 0.5K'\Im G(\Delta\omega). \quad (8.44)$$

This is the bias signal that causes pull-in. For example, at high modulation frequencies, a loop with a lag-lead filter will have $\angle G(\Delta\omega) \approx -\pi/2[\text{sign}(\Delta\omega)]$, due to the $1/s$ term in $G(s)$, so a positive value for $\Delta\omega$ would produce $\omega_b < 0$, which would tend to cause ω_{out} to be below ω_c . If this value of ω_b were more negative than ω_{b0} , the value corresponding to the initially assumed ω_{out} , ω_{out} would tend to decrease toward ω_{in} , thus lowering $|\Delta\omega|$ and leading toward lock.

In the next section we will analyze the general behavior of loops, including possible anomalous behavior that can occur under certain circumstances, by determining a relative value for ω_b under various assumptions. This will tell us the direction of change and allow us to understand the general behavior to be expected (as was done in Section 8.2), including the effects of some undesired, but not necessarily unlikely, high-frequency poles. In the section after that we will solve for the transient values of v_b for a particular type of filter.

Example 8.2 Pull-In Bias A loop with a lag-lead filter is out of lock and skipping cycles. The VCO design is such that its center frequency is 6×10^5 rad/sec higher than the current reference frequency but, at the moment, its frequency is higher than the reference by only 10^5 rad/sec. Its open-loop gain is given by

$$G(\Delta\omega) = \frac{10^7 \text{ sec}^{-1} (1 + j(\Delta\omega/10^3 \text{ rad/sec}))}{j\Delta\omega (1 + j(\Delta\omega/1 \text{ rad/sec}))}.$$

Will the loop move toward lock or away from lock?

The bias developed, according to (8.44) is

$$\begin{aligned} \omega_b &= 0.5(10^7 \text{ sec}^{-1})\Im \left(\frac{10^7}{j10^5} \frac{1 + j(1 \times 10^5/10^3)}{1 + j(10^5/1)} \right) \\ &= -5 \times 10^5 \text{ rad/sec}. \end{aligned}$$

² Phase modulation passes through a mixer from either input to the output, so the values of the sidebands relative to the carrier that produces the normal PD output will be preserved from input to output of the PD.

Thus, a bias is developed that will tend to force the VCO frequency 5×10^5 rad/sec below its center frequency. But this is 10^5 rad/sec above the reference frequency, which is where we started. Thus there is equilibrium. If we change the initial frequency slightly we find that ω_b changes in the same direction but by an amount that is 4 times greater. Therefore we are at a point of unstable equilibrium, and the direction in which we will move depends on a small deviation from that point. Practically, it is uncertain because of the various approximations in developing the theory and inaccuracies in obtaining parameters for any real loop.

8.4.2 General Analysis

Suppose that, at some time t_0 , there is a bias voltage on the VCO tuning line that causes its frequency to be offset from ω_c by ω_{b0} . This will result in a difference frequency $\Delta\omega(0)$ [between the VCO frequency $\omega_{out}(0)$ and the reference frequency ω_{in}] that is equal to the difference between the mistuning Ω and ω_{b0} , as shown by the following equation and illustrated in Fig. 8.13:

$$\Delta\omega(0) = \omega_{out}(0) - \omega_{in} = [\omega_{out}(0) - \omega_c] - [\omega_{in} - \omega_c] = \omega_{b0} - \Omega. \tag{8.45}$$

Equation (8.45) defines the line *B* in Fig. 8.14. Given a starting value of ω_{b0} , this line shows the corresponding value of $\Delta\omega(0)$. For that $\Delta\omega(0)$ we can compute, from Eq. (8.44), a new value of ω_b that would ultimately be produced (if $\Delta\omega$ did not change). We can tell whether ω_b , and therefore ω_{out} , will increase or decrease or remain unchanged by comparing the computed value of ω_b to the starting value ω_{b0} .

The second curve in Fig. 8.14, *A*, is a plot of Eq. (8.44) for a generic loop. At very low values of $\Delta\omega$, $G(\Delta\omega) \approx -jK/\Delta\omega$. Thus ω_b is very large and negative. As $\Delta\omega$ increases, $|G(\Delta\omega)|$ drops and so does ω_b . Phase shift due to the loop filter will also decrease ω_b . At very high values of $\Delta\omega$ unintended poles, for example, those found in op-amp circuitry, will cause additional phase shift and $\Re\{G(\Delta\omega)\}$ will go through zero multiple times as $|G(\Delta\omega)|$ continues to decrease.

At any value of $\Delta\omega$, the ordinate of curve *B* gives the corresponding value of ω_{b0} while the ordinate of curve *A* gives the new value. Thus, whenever curve *A* is

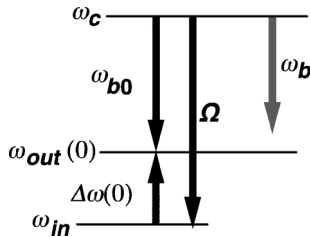


Fig. 8.13 Relationship between frequency variables.

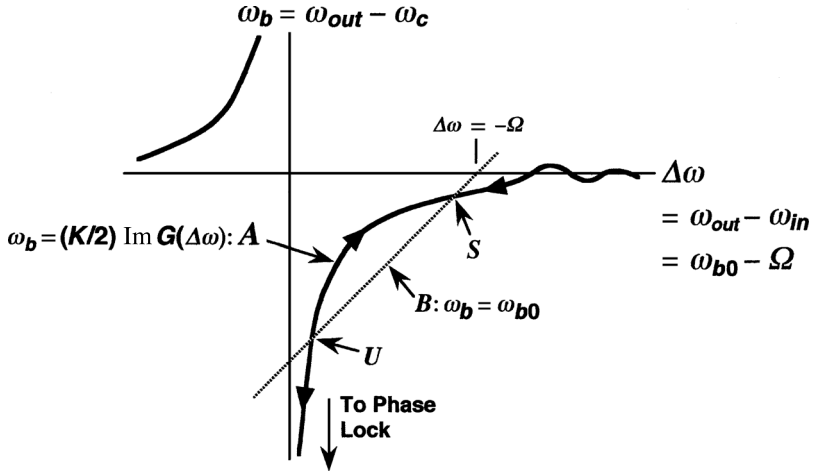


Fig. 8.14 Pull-in bias in generic PLL.

above curve B , the new value of ω_b will be greater than the original value and $\Delta\omega$ will increase. Conversely, when curve A is below curve B , $\Delta\omega$ will tend to decrease. These motions are shown by the arrows on curve A . Intersections of the two curves are equilibrium points because there the initial ω_{b0} from curve B equals the resulting ω_b from curve A . Point S is one of stable equilibrium because the arrows point toward it from both sides. Conversely, U marks a point of unstable equilibrium because the arrows point away from it. If the initial value of ω_{b0} is above U (in Fig. 8.14), the operating point will move toward S and may remain there in an unlocked condition with $\omega_{out} \neq \omega_{in}$. If it starts below U , it will move downward and $\Delta\omega$ will decrease to zero and a state of true phase lock will be acquired.

It is possible that, in the vicinity of S , while each deviation of the operating point from S causes a correction toward S , each correction is so large that it overshoots S by more than the original error. Then the operating point might oscillate about S or deviate from the vicinity of S entirely. It is not obvious how the simple theory we have developed to this point can determine these dynamics,³ but it can help us to understand the possibilities and explain observed phenomena. It is clear that, if our design includes a possible false lock point, we should be concerned. Curve A is mirrored in the graph's origin. The other part of the curve is used for $\Delta\omega < 0 \Rightarrow \omega_{out} < \omega_{in}$. This illustrates the importance of the sign of $\Delta\omega$ in Eq. (8.44).

In Fig. 8.15a, an initial offset $\Delta\omega > U$ will cause the locus to move to S at which point $\Delta\omega \approx -\Omega$, which also implies $\omega_{out} \approx \omega_c$. Thus the VCO is approximately free running. If either the oscillator center frequency is increased or the input frequency is

³ It is a simple matter to compute successive values of ω_b and to determine if successive deviations from S increase in magnitude, but this is based on a step-by-step model, as if there were a low-frequency sample-and-hold circuit in the loop. In practice, the initially assumed value will not be held until the new value is attained.

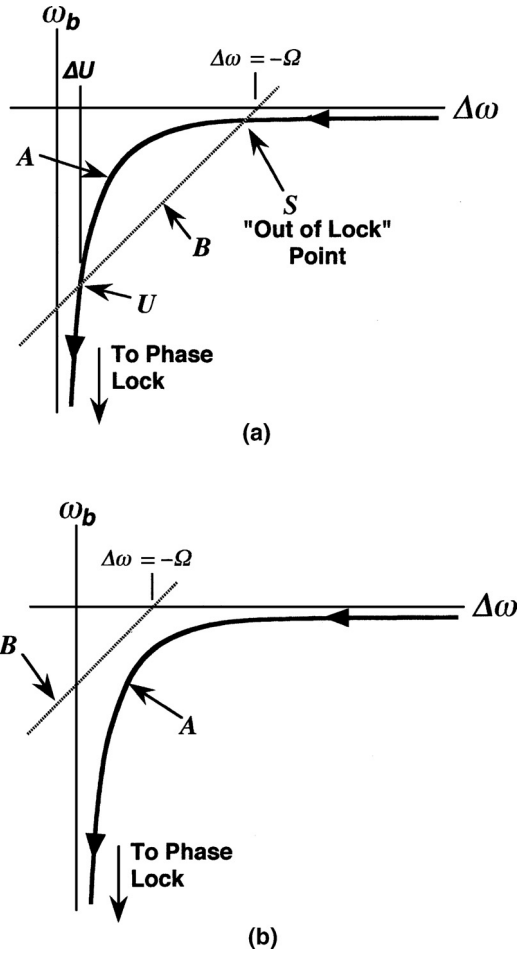


Fig. 8.15 Pull-in range: (a) beyond pull-in and (b) within pull-in.

decreased, $-\Omega = \omega_c - \omega_{in}$ will increase by the same amount, moving curve *B* to the right by that much and causing $\Delta\omega$ at point *S* to move about the same amount. Thus a change in open-loop VCO tuning will produce about the same change in frequency as would occur without the influence of the loop and, equivalently, a change in ω_{in} will have little effect upon the value of ω_{out} . For this reason, the stable operating point at *S* corresponds to an out-of-lock, or free running, condition.

If the center frequency of the oscillator is then moved lower, toward ω_{in} , or if ω_{in} moves higher toward ω_c , curve *B* moves to the left and eventually no longer intersects curve *A*, as shown in Fig. 8.15*b*. Under those conditions, curve *B* is always above curve *A*, so lock will occur from any starting point shown on curve *A*. This is an illustration of the normal pull-in process that we have discussed. However, an anomaly that we have not discussed can also be illustrated.

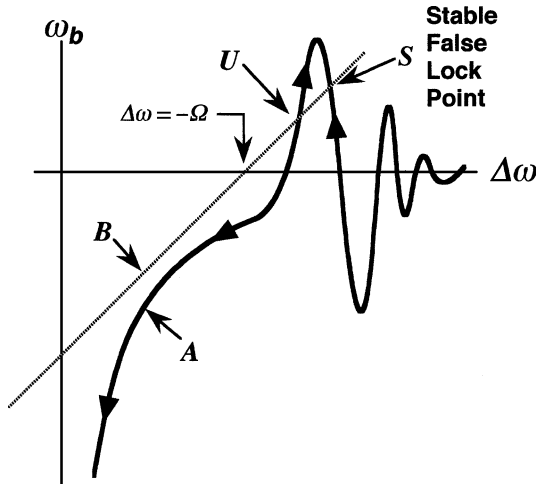


Fig. 8.16 False lock.

In Fig. 8.16, curves *A* and *B* do not cross in the normal region illustrated in Fig. 8.15, but the gain is high enough that one of the arcs, which are produced by excess phase shift at high frequencies, is intersected by curve *B*. This can cause a stable operating point at *S*. If ω_c or ω_{in} is changed, curve *B* will move correspondingly. However, because curve *A* is nearly vertical at *S*, the change in $\Delta\omega$ will be small. If the oscillator tuning is adjusted to move ω_c (and thus curve *B*), the oscillator will change frequency by only a small amount compared to the change in ω_c . If ω_{in} is changed, ω_{out} will follow it. Thus the VCO appears to be locked to the reference, but with a frequency offset, a “false lock.” One situation in which such false locks are a danger is that in which an intermediate-frequency (IF) filter is used, as shown in Fig. 8.17. Frequency-modulation sidebands may see different phase shifts in the passband of the IF filter, and that can be accentuated by multiple poles or by operating near the edge of the passband. The effect is equivalent to phase shift of the modulation waveform and therefore to phase shift in the loop.

Reducing the gain *K* will shrink curve *A* and can cause the stable point *S* to disappear, thus breaking the false lock. However, if *S* is on one of the smaller arcs,

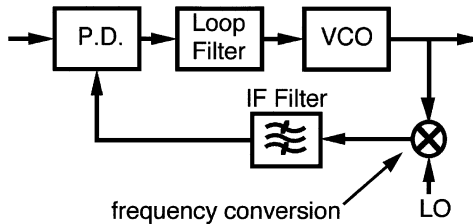


Fig. 8.17 Loop with IF filter.

reducing the gain might cause it just to change frequency offsets. Reducing the gain might also cause curve *A* to shrink to the point where the false lock is replaced by an out-of-lock condition, wherein the curves intersect on a relatively horizontal part of curve *A*. In general when the loop is not truly locked, a change in the input frequency will cause some smaller change in the VCO frequency. False locks and out-of-lock conditions describe two extremes of anomalous performance.

8.4.3 Pull-In Range

As our first attempt to gain quantitative information from the preceding analysis, we will compute the pull-in range under the conditions of Eq. (8.25), namely $K \gg \omega_z \gg \omega_p$.

Pull-in is certain if curves *A* and *B* (Fig. 8.14) do not intersect. The boundary of this condition occurs when they are tangent. That will occur at a point at which they have the same slope. Therefore, to find that point, we must find the value of $\Delta\omega$ at which the derivative of Eq. (8.44) with respect to $\Delta\omega$ is unity. Under the assumed conditions, Eq. (8.44) becomes

$$\omega_b \approx \frac{K}{2} \Im \left[\frac{K}{j\Delta\omega} \frac{\Delta\omega/\omega_z}{\Delta\omega/\omega_p} \right] = -\frac{K^2\omega_p}{2\omega_z} \frac{1}{\Delta\omega}. \quad (8.46)$$

The derivative is

$$\frac{d\omega_b}{d\Delta\omega} = \frac{K^2\omega_p}{2\omega_z} \frac{1}{\Delta\omega^2}, \quad (8.47)$$

which we set equal to 1, yielding

$$\Delta\omega_1 = K \sqrt{\frac{\omega_p}{2\omega_z}}. \quad (8.48)$$

We wish to know the value of mistuning, which is $-\Omega$, the intersection of curve *B* with the *x* axis, at which this occurs. Since curve *B* is a 45° line, the horizontal distance from *S* to $-\Omega$ is the same as the vertical distance so

$$-\Omega = \Delta\omega_1 - \omega_b(\Delta\omega_1). \quad (8.49)$$

We evaluate the last term by substituting $\Delta\omega_1$ from Eq. (8.48) into the last part of Eq. (8.46) and find that

$$-\omega_b(\Delta\omega_1) = \Delta\omega_1. \quad (8.50)$$

Thus if, in Fig. 8.15*a*, we are operating at the stable out-of-lock point *S* and we slowly

decrease the mistuning, S will reach its most extreme position on curve A , on the verge of pull-in, when the beat note $\Delta\omega$ is just half of the mistuning. At that point the last two equations indicate that the absolute value of the mistuning is

$$\Omega_{PI} = |\Omega| = 2|\Delta\omega_1| = K\sqrt{\frac{2\omega_p}{\omega_z}}. \quad (8.51)$$

Under the assumed conditions ($K \gg \omega_z \gg \omega_p$) we can write

$$\Omega_{PI} = \sqrt{\frac{2K^2\omega_p}{\omega_z}} = \sqrt{\frac{2K\omega_n^2}{\omega_z}} \approx 2\sqrt{K\zeta\omega_n}, \quad (8.52)$$

which is the same as Eq. (8.25).

8.4.4 Approximate Pull-In Time

Our next step in obtaining quantitative information from our approximate analysis is to relate the voltages at the input and output of the loop filter by Eq. (8.42), which describes how the tuning voltage causes a low-frequency term at the phase detector output, and by the filter transfer function, which describes how that term causes a tuning voltage.

By Eq. (8.42), at the input to the filter we have

$$v_b \approx 0.5K'_p \text{Im}G(\Delta\omega). \quad (8.53)$$

We assume that the frequency error is large compared to ω_z so this can be written

$$v_b \approx -0.5K'_p K\omega_p/(\omega_z \Delta\omega). \quad (8.54)$$

The filter characteristic can be written

$$v_v(s) = K_{LF} \frac{1 + s/\omega_z}{1 + s/\omega_p} v_b(s), \quad (8.55)$$

where v_v is the low-frequency tuning voltage at the output of the filter or, equivalently,

$$\left(1 + \frac{1}{\omega_p} \frac{d}{dt}\right) v_v(t) = K_{LF} \left(1 + \frac{1}{\omega_z} \frac{d}{dt}\right) v_b(t). \quad (8.56)$$

Substituting v_b from Eq. (8.54) and using $K_v v_b = \omega_b = \Delta\omega + \Omega$ we obtain

$$\left(1 + \frac{1}{\omega_p} \frac{d}{dt}\right) (\Delta\omega(t) + \Omega) = -\frac{K^2 \omega_p}{2\omega_z} \left(1 + \frac{1}{\omega_z} \frac{d}{dt}\right) \frac{1}{\Delta\omega(t)} \quad (8.57)$$

$$= -\frac{\Omega_{PI}^2}{4} \left(1 + \frac{1}{\omega_z} \frac{d}{dt}\right) \frac{1}{\Delta\omega(t)}, \quad (8.58)$$

where the last step used Eq. (8.52). Using the equality

$$\frac{d}{dt} \frac{1}{\Delta\omega(t)} = -\frac{1}{\Delta\omega(t)^2} \frac{d}{dt} \Delta\omega(t)$$

and rearranging terms, we obtain

$$\Delta\omega(t) + \Omega + \frac{\Omega_{PI}^2}{4\Delta\omega(t)} = \left(-\frac{1}{\omega_p} + \frac{\Omega_{PI}^2}{4\omega_z \Delta\omega^2(t)}\right) \frac{d}{dt} \Delta\omega(t), \quad (8.59)$$

$$dt = \frac{1}{\omega_p} \frac{-\Delta\omega(t)^2 + (\Omega_{PI}^2/4)(\omega_p/\omega_z)}{\Delta\omega(t)[\Delta\omega(t)^2 + \Omega\Delta\omega(t) + \Omega_{PI}^2/4]} d\Delta\omega(t). \quad (8.60)$$

We then integrate. We set the initial time to zero and, assuming that the VCO is initially at ω_c , set $\Delta\omega(0) = -\Omega$. We take acquisition to have been completed when $|\Delta\omega(T_{PI})|$ equals the seize frequency,⁴ $\Omega_S = \omega_n^2/\omega_z = Kr^2$ where

$$r^2 \triangleq \omega_p/\omega_z, \quad (8.61)$$

$$\begin{aligned} \int_0^{T_{PI}} dt &= \frac{\Omega_{PI}^2}{4\omega_z} \int_{-\Omega}^{-Kr^2} \frac{1}{\Delta\omega(t)[\Delta\omega(t)^2 + \Omega\Delta\omega(t) + \Omega_{PI}^2/4]} d\Delta\omega(t) \\ &\quad - \frac{1}{\omega_p} \int_{-\Omega}^{-Kr^2} \frac{\Delta\omega(t)}{\Delta\omega(t)^2 + \Omega\Delta\omega(t) + \Omega_{PI}^2/4} d\Delta\omega(t). \end{aligned} \quad (8.62)$$

These two integrals are standard forms, the solution of which give

$$T_{PI} = \frac{1}{\omega_p} \left\{ \begin{aligned} &r^2 \ln \left(\frac{K}{\Omega} r^2 \right) - \frac{1+r^2}{2} \ln \left[2 \left(r^2 - \frac{\Omega}{K} \right) + 1 \right] \\ &- \frac{(1-r^2)\Omega}{\sqrt{\Omega_{PI}^2 - \Omega^2}} \left[\tan^{-1} \left(\frac{2Kr^2 - \Omega}{\sqrt{\Omega_{PI}^2 - \Omega^2}} \right) - \tan^{-1} \left(\frac{\Omega}{\sqrt{\Omega_{PI}^2 - \Omega^2}} \right) \right] \end{aligned} \right\}. \quad (8.63)$$

This is plotted in Fig. 8.18.

⁴ The seize frequency is defined as a mistuning, that is, the difference between the input frequency and ω_c . Here we are assuming that the same value of $|\Delta\omega|$ will cause rapid pull-in even if the VCO is not at ω_c .

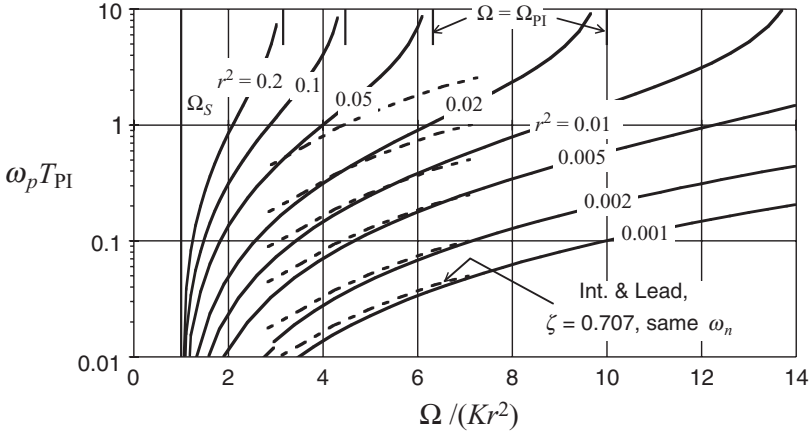


Fig. 8.18 Approximate pull-in time, $|\Delta\omega \gg \omega_z|$ Accuracy is poorest near Ω_{PI} and Ω_S .

If $\Omega \gg Kr^2$ (acquisition begins far from Ω_S),

$$T_{PI} = \frac{1}{\omega_p} \left\{ \begin{aligned} &r^2 \ln \left(\frac{K}{\Omega} r^2 \right) - \frac{1+r^2}{2} \ln \left[2 \left(r^2 - \frac{\Omega}{K} \right) + 1 \right] \\ &+ \frac{2(1-r^2)\Omega}{\sqrt{\Omega_{PI}^2 - \Omega^2}} \tan^{-1} \left(\frac{\Omega}{\sqrt{\Omega_{PI}^2 - \Omega^2}} \right) \end{aligned} \right\}. \quad (8.64)$$

When also $r^2 \ll 1$ and $K \gg \Omega$, this equation approaches Eq. (8.27). The equation can be written in normalized form as

$$\begin{aligned} \omega_p T_{PI} = &-r^2 \ln(\Delta) - \frac{1+r^2}{2} \ln[1 + 2r^2(1 - \Delta)] \\ &+ \frac{2(1-r^2)\Delta}{\sqrt{2/r^2 - \Delta^2}} \tan^{-1} \left(\frac{\Delta}{\sqrt{2/r^2 - \Delta^2}} \right), \end{aligned} \quad (8.65)$$

where

$$\Delta \equiv \Omega/(Kr^2). \quad (8.66)$$

A comparison is also shown to points obtained [Sanneman and Rowbotham, 1964]⁵ for a loop with an integrator-and-lead filter (type 2, $r^2 \Rightarrow 0$) but having the same

⁵ Figure 6 in Sanneman and Rowbotham is for $\zeta = 0.707$. The axes above are matched to the axes in their Fig. 6 using $\omega_p = \omega_n[1 - (1 - 2r^2)^{0.5}]/(2)^{0.5}$, $Kr^2 = \omega_n r^2 (2)^{0.5}/[1 - (1 - 2r^2)^{0.5}]$, which is true for $\zeta = 0.707$. Initial condition for their Fig. 6 was zero phase error.

values of ζ and ω_n . The dashed lines match the solid lines well for the smallest values of r^2 , which represent loops that approach type 2. We might also say that significant deviations occur when Ω_{PI} is approached for the lag-lead filter; this is to be expected since the type 2 loops have infinite Ω_{PI} .

8.5 PHASE-PLANE ANALYSIS

In the linear range, where phase deviations are small, linear theory applies, and we have just developed expressions applicable to the opposite extreme. If we should require accurate answers for the in-between case, we must use computer simulation, experimentation, or phase-plane analysis. (See phase-plane discussions in Gardner [2005, pp. 114–117] and Stensby [1997, pp. 121–156, 168–175].)

Figure 8.19 is a phase-plane portrait. It is constructed by computing $(d\omega_e/d\varphi_e) = F(\omega_e, \varphi_e)$, so a small segment of the locus of $\omega_e = F(\varphi_e)$ can be plotted by using the computed slope. These segments can then be connected. In other words, we can go to any point in a graph of ω_e versus φ_e and extend a line a short distance from that

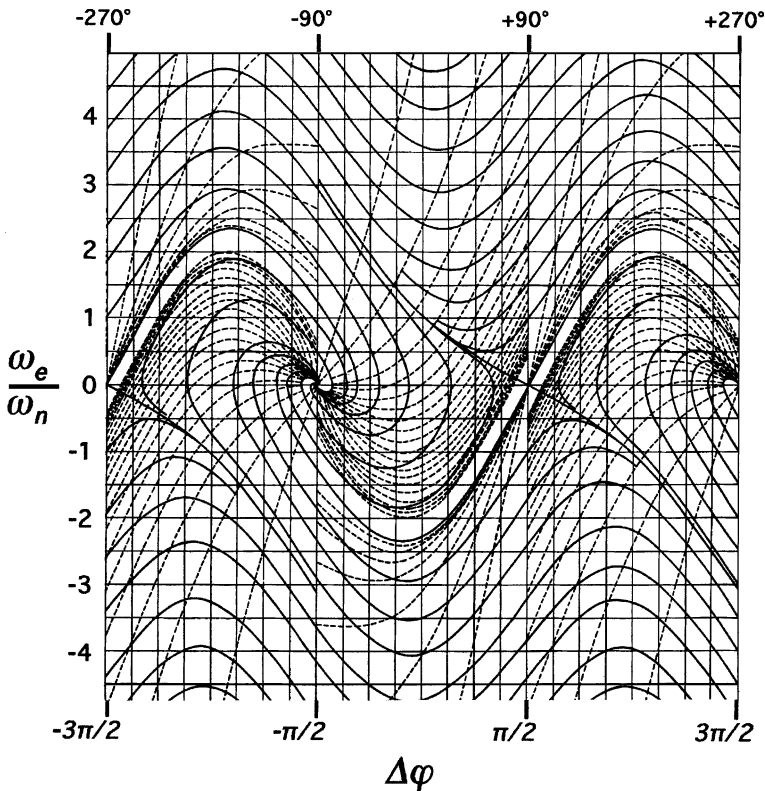


Fig. 8.19 Phase-plane plot for type 2 loop, $\zeta = 1/\sqrt{2}$. (From Sanneman and Rowbotham, 1964, Fig. 5, p. 19, ©1964 IEEE.)

point if we know the slope at that point. By doing this repeatedly, a set of loci can be created. Figure 8.19 is such a set. It repeats at increments of $\varphi_e = 2\pi$ and should be imagined to exist on a cylinder so that any vertical line is contiguous with a similar line 2π away. In the upper part of the figure, since $\omega_e = d\varphi/dt$ is positive, φ_e increases with time and the locus moves to the right. Each 2π increment in x , the $|y|$ can be seen to decrease. The decrease is slight for large $|y|$ and becomes larger for smaller $|y|$. Eventually the “separatrix,” a division between curves that do and do not reach the next $\pm\pi$ increment, is reached. Curves below the separatrix cross the $y = 0$ axis and turn back. From that point there is no cycle skipping. After the axis is crossed, since $y = dx/dt$ becomes negative, the direction of horizontal motion reverses, and the locus moves toward a steady-state value $(-\pi/2, 0)$. Depending upon ζ , it may move steadily to the final values or it may spiral in. A similar description applies if we begin in the lower part of the figure. Viterbi has drawn such plots for type 2 loops with ζ between 0.7 and 1.4 [Blanchard, 1976, pp. 256–259].

Figure 8.20 shows two possible loci superimposed on a one-cycle-wide region of Fig. 8.19. Assume $\Delta\varphi$ and ω_e happen to have values corresponding to point O' at

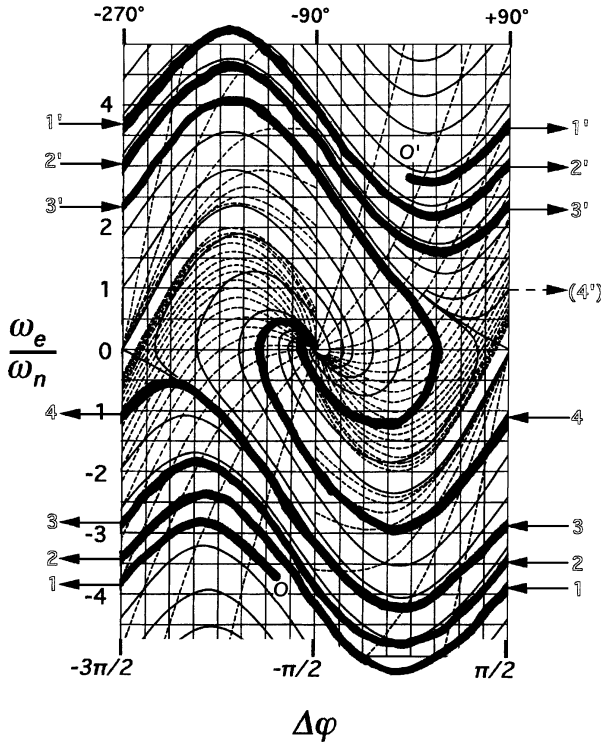


Fig. 8.20 Two loci in the phase plane. This drawing represents a cylinder connected on right and left edges. The outgoing arrow labeled n (or n') connects to the incoming arrow with the same label on the other edge. (From Sanneman and Rowbotham, 1964, Fig. 5, p. 19, ©1964 IEEE, loci added.)

$t = 0$. This may be the time when the reference is connected to the loop or might just be some instant during an acquisition sequence. (Even if t does represent the time when the reference is connected, Ω cannot be read from the graph unless the initial value of $\Delta\varphi$ happens to be $\pm\pi/2$, where the phase detector output is unchanged by the connection.)

Because $\omega_e > 0$ at O' , $\Delta\varphi$ increases with time so the locus moves to the right, guided by the curves of the graph. At $+90^\circ$ we move to the left edge of the graph (at $1'$), although this is just a way of illustrating, on a two-dimensional graph, continuous motion around a cylinder. The locus continues left to right, moving from O' to $1'$ to $2'$ to $3'$. In the process the frequency error increases above its initial value, but, at any value of $\Delta\varphi$, each locus is closer to the lock point than the previous locus. What is more, the rate at which ω_e decreases accelerates with subsequent cycles of $\Delta\varphi$ so there is a larger space between loci as time increases. After $3'$ the locus spirals into the stable lock point at $\omega_e = 0$ and $\Delta\varphi = -90^\circ$. A very slightly higher initial value of ω_e would have resulted in the locus after $3'$ being above the separatrix, and it would have gone through $4'$ before spiraling into the lock point. The spiraling nature of the approach to the lock point represents overshoot in ω_e .

A similar process occurs if we begin at point O , but now we move to the left because the negative value of ω_e represents decreasing $\Delta\varphi$. In this plot, segment 3–4 is a little below the separatrix so another cycle is skipped.

We can develop the equations needed to plot a phase-plane portrait for the lag-lead filter by beginning with one similar to Eq. (8.56), except it is written for the total filter input and output voltages rather than the low-frequency component. Relating these voltages to the corresponding frequency and phase errors, we obtain

$$\left(1 + \frac{1}{\omega_p} \frac{d}{dt}\right) \left(\frac{\Omega - \omega_e(t)}{K_v}\right) = K_{LF} \left(1 + \frac{1}{\omega_z} \frac{d}{dt}\right) K_p \cos \Delta\varphi(t), \quad (8.67)$$

$$\frac{d\omega_e(t)}{dt} + \omega_p \left(1 - \frac{K}{\omega_z} \sin \Delta\varphi(t)\right) \omega_e(t) + K\omega_p \cos \Delta\varphi(t) = \omega_p \Omega. \quad (8.68)$$

Using

$$\frac{d\omega_e(t)}{dt} = \frac{d\omega_e}{d\Delta\varphi(t)} \frac{d\Delta\varphi(t)}{dt} = \frac{d\omega_e}{d\Delta\varphi(t)} \omega_e, \quad (8.69)$$

and defining the coordinates of the plot at

$$y \equiv \frac{\omega_e}{\omega_n} \quad \text{and} \quad x = \Delta\varphi, \quad (8.70)$$

Eq. (8.68) can be written as

$$\frac{dy}{dx} = \frac{\omega_n}{\omega_z} \sin x - \frac{\omega_n}{K} + \frac{\Omega/K - \cos x}{y}. \quad (8.71)$$

If we let $K \Rightarrow \infty$, we obtain the equivalent expression for a type 2 loop,

$$\frac{dy}{dx} = 2\zeta \sin x - \frac{\cos x}{y}, \quad (8.72)$$

which is plotted in Fig. 8.19 for $\zeta = 1/\sqrt{2}$.

This figure also includes isochrones (dashed lines), which mark fixed time increments along the loci. Since $y \equiv (1/\omega_n)d\Delta\varphi/dt$, a small time increment Δt corresponds to a phase change $\Delta(\Delta\varphi) = y\omega_n\Delta t$. The isochrones in Fig. 8.19 are drawn for $\Delta t = 1/(4\omega_n)$ so, moving along a locus, the change in x between isochrones will be $\Delta(\Delta\varphi) = y/4$. The time to get from one phase and frequency to another can be determined by counting the number of isochrone increments and dividing by $4\omega_n$. It is by this method that the dashed curves in Fig. 8.18 were computed from Fig. 8.19.

8.6 PULL-OUT

In Chapter 6 we described the phase error that results from a frequency step (Fig. 6.9) but that was for a linear system. Of particular interest is the maximum frequency step ω_{po} that will not push $\Delta\varphi$ to the point where a cycle is skipped. This is inherently a nonlinear problem. Beginning at the lock point in Fig. 8.19, if the frequency is stepped (vertically, at constant $\Delta\varphi$) beyond the separatrix at $(-\pi/2, 3.2)$, a cycle will be skipped as the operating point returns to the lock point. For this type 2 loop with $\zeta = 1/\sqrt{2}$, the pull-out frequency is therefore $\omega_{po} = 3.2\omega_n$. By studying such plots, Gardner has determined, for type 2 loops, that [Gardner, 2005, p. 117]

$$\omega_{po} = 1.8 \omega_n(1 + \zeta) \quad \text{if} \quad 0.5 < \zeta < 1.4. \quad (8.73)$$

8.7 EFFECT OF OFFSETS

Here we consider how to maximize pull-in range in the presence of a phase detector or op-amp offset. Assume the general lag-lead filter and let the offset be equivalent to a phase $\sin^{-1}\varphi_{os}$, approximately φ_{os} for small values. This will produce an output offset $\omega_{os} = K'\varphi_{os}$. Presumably φ_{os} is not controllable so its sign is unknown and the loop may have to overcome both it and the mistuning that would otherwise exist. Therefore the pull-in range is effectively reduced by $|\omega_{os}|$ and the effective pull-in range is

$$\Omega_{PIos} = \Omega_{PI} - \omega_{os} = K(\sqrt{4x} - |\varphi_{os}|), \quad (8.74)$$

where x is given by (8.24).

If K is to be limited, perhaps to prevent components from saturating, then the largest value of Ω_{PIos} will occur at the highest value of x , which is 0.5 and which occurs for a first-order loop. However, the restriction of (8.25) would be violated for $x = 0.5$ and Ω_{PI} would not exceed $K(1 - |\varphi_{os}|)$.

If Ω_{PIos} is to be maximized in the presence of a fixed loop bandwidth ω_L , then we rewrite (8.74) in terms of ω_L , which is $K\omega_p/\omega_z = 2Kx$, assuming the restrictions of (8.25) apply.

$$\frac{\Omega_{PIos}}{\omega_L} = \sqrt{\frac{1}{x} - \frac{|\varphi_{os}|}{2x}}. \quad (8.75)$$

To maximize this ratio, we take the derivative of the right side with respect to x and set it equal to zero. This gives

$$x = |\varphi_{os}|^2. \quad (8.76)$$

Substituting this into (8.75), we obtain the maximum pull-in range as

$$\Omega_{PIos} \leq \frac{\omega_L}{2|\varphi_{os}|}. \quad (8.77)$$

8.8 EFFECT OF COMPONENT SATURATION

Suppose some component has a limited range that can be effectively modeled as clipping of the tuning voltage. That is, the tuning voltage, and consequently the VCO frequency, stops at some minimum or maximum value and can be moved no further. Assume that there are no other consequences of the component's saturation (e.g., strange or sluggish operation). What effect will this clipping have on pull-in?

Obviously, the reference frequency must be within the clip limits or no lock can occur. If, during pull-in, the frequency overshoots to a limit, there will be some change in the response. Suppose though that the lock frequency is within the clip limits but the VCO center frequency ω_c is not. Then the initial VCO frequency is not what would occur in a linear system with zero output from the phase detector, but corresponds rather to a clip limit. If the mistuning (in the absence of clipping) is within the seize range, the loop should lock because the high-frequency gain is adequate for the phase detector to tune the VCO to the input frequency. However, further out in the pull-in range, if the range of frequencies that would be produced during the initial cycle of ω_e is beyond the clipping limit, the VCO will not move from its initial frequency. Then there will be no feedback to develop a pull-in voltage and the loop will not lock.

8.9 HANGUP

If a loop is at the steady-state frequency but the phase is at a point of unstable equilibrium (e.g., π radians from steady-state for a type 2 loop), there will be no error signal to cause the loop to move to the stable lock point. Any small deviation from the unstable-equilibrium point can initiate acquisition of the correct steady-state phase, but the error signal will initially be small so the first part of the transient will be slow. This phenomenon is called hangup [Gardner, 1977]. An operating point on the separatrix of the phase plane leads to an unstable equilibrium point, for example, the saddle point at $(\pi/2, 0)$ in Fig. 8.19, but any slight deviation from the separatrix leads to phase lock. However, note how close are the isochrones in that region, indicating slow motion. Everlasting hangup is a zero-probability event, but finite-duration hangup has finite probability and can sometimes be a problem. One might be inclined to use a sawtooth PD, which theoretically has essentially no unstable equilibrium point of finite extent, but we will find, in Chapter 13, that noise can cause the sawtooth to lose its sharp edge.

8.10 SIMULATION OF THE NONLINEAR LOOP

The model in Fig. 8.21 represents the PLL for simulations that include the nonlinearity of the phase detector. Insertion of a sampler and zero-order hold circuit in the loop replaces the continuously changing phase-detector output by a stepped version of the same. This allows φ_e to be computed as the linear response to steps at u_{in} while each new value of u_{in} is computed from the nonlinear phase-detector response (sinusoidal in the case shown) to the previous value of φ_e . The effect of the sampling on the loop response is quantifiable and negligible if the sampling frequency is high enough. Time- and phase-plane plots of transient responses can be observed, as can the effects of a phase offset (represented by u_{off}) and acquisition equations can be verified.

We can also watch transient responses beginning with a sets of initial conditions. These can be seen to be independent of α , which is not present in the denominator (characteristic equation) of the transfer functions. See Appendix i8.M.

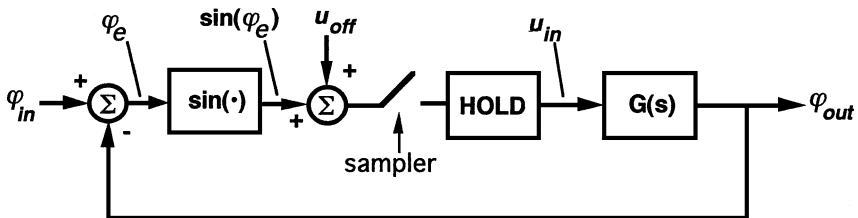


Fig. 8.21 Computation process.

8.A APPENDIX: SUMMARY OF ACQUISITION FORMULAS FOR SECOND-ORDER LOOPS

Table 8.A.1 provides a summary of acquisition formulas for second-order loops.

8.M APPENDIX: NONLINEAR SIMULATIONS

This appendix describes simulations of nonlinear behavior in PLLs and the MATLAB[®] scripts that control them. It is available from the Wiley Internet site.

8.S APPENDIX: ACQUISITION SPREADSHEET

The `ACC2` spreadsheet provides the acquisition parameters given in Section 8.3 from the values of α , ζ , ω_n , and mistuning Ω . It also gives K , ω_p , and ω_z in the process. The spreadsheet `ACCPZ` provides acquisition information based on given values of ω_p , ω_z , ω_L , and Ω . It also gives K , ω_n , ζ , and α . These are available on the Wiley Internet site.

8.V. APPENDIX: SOME VALUES IN TERMS OF α , ζ , AND ω_n

In Section 6.9 we found the open-loop transfer function in terms of the closed-loop parameters ω_n , α , and ζ . Here we find also the circuit constants ω_z , ω_p , and K and the pull-in range with a lag-lead filter in those terms

Using Eqs. (4.9) and (6.5c), we find that

$$\frac{\omega_z}{\omega_c} = \frac{1}{2\alpha\zeta} \quad (8.V.1)$$

and

$$\frac{\omega_p}{\omega_n} = 2\zeta(1 - \alpha). \quad (8.V.2)$$

Multiplying this expression by K and using Eq. (4.7), we obtain

$$\frac{K}{\omega_n} = \frac{1}{2\zeta(1 - \alpha)}. \quad (8.V.3)$$

We can now use these expressions to write Eq. (8.25) as

$$\Omega_{PI} \approx \omega_n \sqrt{\frac{2\alpha}{1 - \alpha}}, \quad 1 - \alpha \ll 1, 1/(2\zeta)^2 \quad (8.V.4)$$

[i.e., with $(1 - \alpha)$ much less than the smaller of 1 and $1/(2\zeta)^2$].

TABLE 8.A.1 Acquisition Formulas

Parameter (ranges are ± given values)	PD Type	Value of Parameter ^a	Conditions
Hold-in, Ω_H	Sine	K	
	Triangle	$\frac{\pi}{2} K$	
	Sawtooth	πK	
	PFD	$2\pi K$	
Seize (lock-in) Ω_S approximately	Sine ^b	ω_L	$\omega_z \ll \omega_L$
	Triangle ^c	$\frac{\pi}{2} \omega_L$	
	Sawtooth ^d	$\pi \omega_L$	
	PFD	Unlimited by PD	
Pull-in (acquisition) Ω_{PI} approximately	Sine	$2K\sqrt{x-x^2}; x = \frac{\omega_p}{2\omega_z}$	Lag-lead filter ^e $K \gg \omega_z$
		$3\zeta K\sqrt{\sqrt{0.423 + 1.2\zeta^4} - 1.092\zeta^2}$	Low-pass filter ^f $\Omega_{PI} < \Omega_H$
	Triangle	$\pi\sqrt{\zeta K\omega_n}$	Lag-lead filter ^g $\omega_n \ll K; \zeta < 1$
		Ω_H	Low-pass filter ^h $\zeta > 1$
	Sawtooth	$2\pi K\sqrt{\frac{\omega_p}{3\omega_z}} \approx 2\pi\sqrt{\frac{2}{3}K\zeta\omega_n}$	Lag-lead filter ^d $K \gg \omega_z > 2\omega_p$
		$\pi K \tanh \left\{ \frac{\pi}{2\sqrt{\frac{1}{\zeta^2} - 1}} \right\}$	Low-pass filter ⁱ $\zeta \leq 1$
		πK	Low-pass filter ⁱ $\zeta > 1$
	PFD	Unlimited by PD	

^aSee Egan (2005, pp. 408–418) for values under other conditions.

^bSee Gardner (1979, p. 70) and Kroupa (1973, p. 177).

^cEgan (2005, p. 411).

^dByrne (1962, p. 588).

^eSee Greenstein (1974) and Rey (1960).

^fRichman (1954a).

^gCahn (1962).

^hProtonotarios (1969).

ⁱGoldstein (1962)

The relationship of α to ζ is limited by the requirement that $\omega_p \leq \omega_z$ for the loops that we are studying. Therefore,

$$\frac{\omega_p}{\omega_z} = 4\zeta^2\alpha(1 - \alpha) \leq 1 \quad (8.V.5)$$

implies that necessarily

$$\zeta \leq \frac{1}{2\sqrt{\alpha(1 - \alpha)}}. \quad (8.V.6)$$

This is most restrictive (minimum) at $\alpha = 0.5$, where $\zeta \leq 1$. For $\alpha = 0$ or 1 there are no limits on ζ .

Equation (8.76) gives the optimum ratio of pole to zero frequency for a high-gain loop with an offset under the conditions of Eq. (8.V.4). This is

$$|\varphi_{os}|^2 = x = \frac{\omega_p}{2\omega_z} = 2\zeta^2\alpha(1 - \alpha), \quad (8.V.7)$$

which, under those conditions, corresponds to

$$\alpha = \frac{1 + \sqrt{1 - 2\left|\frac{\varphi_{os}}{\zeta}\right|^2}}{2}. \quad (8.V.8)$$

CHAPTER 9

ACQUISITION AIDS

Sometimes the pull-in range is inadequate or acquisition is too slow. An obvious solution is to increase the loop bandwidth, but other considerations, such as a filtering action that may be desired from the loop, can make that impractical. An acquisition aid often provides the solution. It can permit the loop to lock when its mistuning falls between the hold-in range and the pull-in range ($\Omega_{PI} \leq |\Omega| < \Omega_H$) or it can speed up a pull-in that might otherwise be delayed while many cycles are skipped. We will consider several acquisition aids in this chapter.

9.1 COHERENT DETECTION—LOCK INDICATOR

We can use the coherent detector as part of an acquisition aid or to warn when the loop has become unlocked. Detection refers to extraction of a signal proportional to the amplitude (or power¹) of the detected signal. Coherent detection is a method of detection that employs a second, reference, signal of identical frequency and proper phase. Coherent detection has an advantage of sensitivity over other methods, but the necessity to obtain the reference signal adds complexity.

The coherent detector (Fig 9.1) can be a balanced mixer, structurally identical to the phase detector (in which case $K'_{pd} = K'_p$). The pertinent term in Eq. (3.5) is $\cos[\varphi_1(t) - \varphi_2(t)]$, the same term that produces phase detection. However, for

¹ This is called square-law detection.

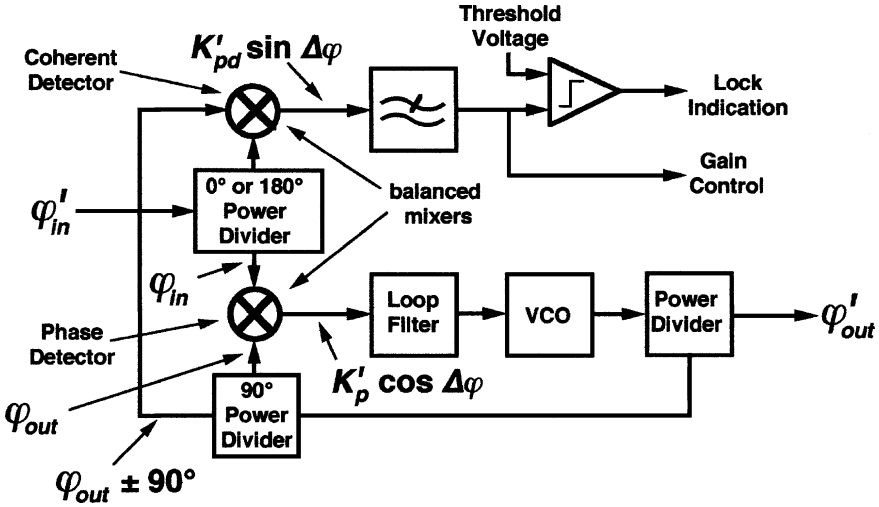


Fig. 9.1 Coherent detector with PLL.²

coherent detection, $\varphi_1(t) - \varphi_2(t)$ is approximately 0° , corresponding to maximum output, rather than the -90° required for phase detection, which corresponds to maximum slope. Because of this relationship the coherent detector in Fig. 9.1 achieves its maximum output when the phase error in the loop is zero ($\Delta\varphi = -\pi/2$). A signal to which the loop is not locked will cause an approximately sinusoidal output from the coherent detector. That can be attenuated by a low-pass filter so that only locked signals will cause sufficient output to break threshold and indicate lock.

The output of the coherent detector will be³ [Eq. (1.7)]

$$v_{CD} = K'_{pd} \cos(\Delta\varphi + \pi/2) = -K'_{pd} \sin \Delta\varphi. \tag{9.1}$$

9.1.1 During Acquisition

The output from the coherent detector is only approximately sinusoidal, just as the output from a phase detector is only approximately sinusoidal, and for the same reason. The little feedback that does exist when the loop is out of lock perturbs the sinusoid. In Chapter 8, we developed Eq. (8.42) to give the bias component from the phase detector that results from this process. We can use the same equation to give us the bias component from the coherent detector by adding the 90° phase shift that

² Power dividers split a signal into several signals, usually providing the same impedance at all ports. Different types of power dividers provide different phase relationships between signals, usually multiples of 90° .

³ The output could be positive or negative, depending on the sign of the relative phase shifts in the power dividers.

is appropriate to the latter, thus obtaining

$$|v_{bC}| \approx 0.5|G(\Delta\omega)|K'_{pd} \cos[\angle G(\Delta\omega)] \leq 0.5|G(\Delta\omega)|K'_{pd}, \tag{9.2}$$

where $\Delta\omega$ is again the error frequency during pull-in. The threshold voltage must be set high enough so v_{bC} does not exceed it when the loop is not locked. Otherwise the coherent detector might indicate phase lock during a false lock. Ideally, a loop with an integrator-and-lead or lag-lead filter will have $\angle G \approx -90^\circ$ at high frequencies so $|v_{bC}| \approx 0$ when $\Delta\omega$ is high compared to ω_z . We must control excess phase shift that leads to false lock until $|\Delta\omega| \gg \omega_L$ so v_{bC} will be small at any false lock. [We have already assumed $|G(\Delta\omega)| \ll 1$ in developing Eq. (8.42), and therefore (9.2).⁴ Under those circumstances, $|v_{bC}| \ll 0.5K_{pd}$ holds whereas $|v_{bC}| \approx K_{pd}$ is true during a normal lock if the phase error is kept low. However, there are circumstances, particularly in the presence of noise, when it is difficult or impossible to determine the presence of a false lock with the coherent detector [Gardner, 1979, p. 155].

9.1.2 During Sweep, Locked Loop

We have shown that the coherent detector can develop a voltage when the loop is not locked. We now show that its voltage during lock can be diminished by frequency sweeping. These two effects set limits for the threshold setting. By Eq. (6.18), the phase error during a linear frequency sweep settles to

$$\varphi_e(t') = \dot{\Omega} \left\{ \frac{1 - 4\zeta^2(1 - \alpha)}{\omega_n^2} + 2\zeta \frac{1 - \alpha}{\omega_n} t' \right\}, \tag{9.3}$$

where

$$\Omega \triangleq \omega_{in} - \omega_c. \tag{8.1}$$

If the input frequency is sweeping while the voltage-controlled oscillator (VCO) center frequency (not its actual instantaneous frequency) is stationary, $\dot{\omega}_c = 0$; if the input frequency is steady while the VCO center frequency is swept, perhaps by a ramp of voltage added to its tuning input, $\dot{\omega}_{in} = 0$; and, if the VCO center frequency is swept in synchronism with the input frequency, $\dot{\omega}_e = \{\dot{\omega}_{in} - \dot{\omega}_c\} = 0$. Both $\dot{\omega}_c$ and $-\dot{\omega}_{in}$ can be represented by a linear sweep of $\Delta\omega_{out}$ in Fig. 7.5.

With an integrator-and-lead filter ($\alpha = 1$), the phase error is constant at $\dot{\omega}_e/\omega_n^2$, but with a lag-lead or lag filter it increases with time. This is the error relative to steady state, when there is no sweep. The question arises, from what event are we measuring t ? When is $t = 0$? If we look at the lag-lead shown in Fig. 3.23 *f*, we see that it differs from an integrator-and-lead (Fig. 3.23 *c*) by the shunt resistor R_p , which passes some of the current that would otherwise go into C . Thus a changing phase error is required to provide this changing current as the voltage across R_p changes. However, at the moment when there is no voltage across R_p , the phase error will be the same as if R_p

⁴ Otherwise $0.5|G(\Delta\omega)|$ cannot be used as an approximation for $J_1(G\Delta\omega)$.

were absent. If we call this moment $t = 0$, then we can write Eq. (9.3) as

$$\varphi_e = \dot{\Omega} \left\{ \frac{1}{\omega_n^2} + 2\zeta \frac{1-\alpha}{\omega_n} t \right\} = \dot{\Omega} \left\{ \frac{1}{\omega_n^2} + \frac{t}{K} \right\}. \quad (9.4)$$

Here t equals t' plus a constant, and the constant is chosen such that, at $t = 0$, when the voltage across R_p is zero, the equation is the same as it would be for an integrator-and-lead filter.

If, during steady-state lock (no sweep), $\Delta\varphi = -\pi/2$, then, during sweep, it will be $\Delta\varphi = \varphi_e - \pi/2$, so Eq. (9.1) will become $v_{CD} = K'_{pd} \cos(\varphi_e)$. This is approximately equal to K'_{pd} for small φ_e . For larger errors Eq. (9.4) must be replaced by a form that accounts for the sinusoidal characteristic of the phase detector,⁵

$$\sin(\varphi_e) = \dot{\Omega} \left\{ \frac{1}{\omega_n^2} + \frac{t}{K'} \right\}, \quad (9.5)$$

and Eq. (9.1) then becomes

$$v_{CD} = K'_{pd} \cos \varphi_e = K'_{pd} \sqrt{1 - \sin^2 \varphi_e} = K'_{pd} \sqrt{1 - \left\{ \dot{\Omega} \left[\frac{1}{\omega_n^2} + \frac{t}{K'} \right] \right\}^2}. \quad (9.6)$$

Thus the output of the coherent detector decreases when ω_{in} or ω_c is swept while the loop maintains phase lock. This must be taken into account in setting the threshold voltage.

9.2 CHANGING LOOP PARAMETERS TEMPORARILY

9.2.1 Coherent Automatic Gain Control

Loop parameters can be changed to speed up acquisition. One method is to control the gain of an amplifier that precedes the circuit in Fig. 9.1 by the output voltage from the coherent detector. This arrangement is shown in Fig. 9.2. It must be done in such a manner that, when the loop is not locked, the gain is maximum. Assuming that the phase detector acts like a multiplier (or a balanced mixer in which the input signal is weak compared to the VCO output), the loop gain will then be high because K'_p is proportional to the amplitude of the input signal. When the loop pulls into lock, the output from the coherent detector contains a strong DC component that will pass through the low-pass filter and decrease the preceding amplifier's gain; K'_p will then decrease, resulting in a lower loop bandwidth. This coherently detected voltage can then be used to maintain a constant input level and thus constant phase-locked loop (PLL) parameters. We must ensure that the PLL is stable for all the possible gains as the acquisition process is taking place. Note that the total system now consists of

⁵ This assumes that zero input to the filter occurs when $\Delta\varphi = -90^\circ$.

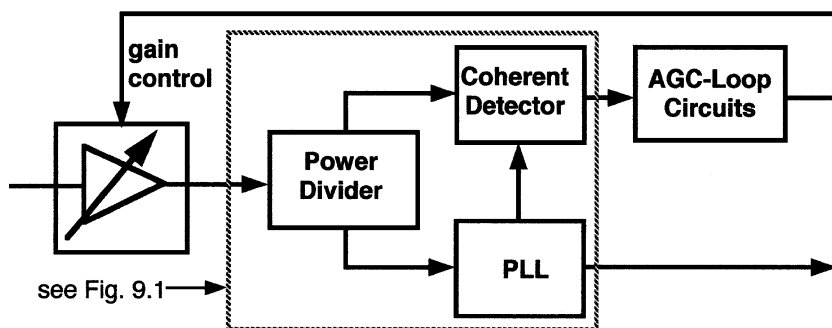


Fig. 9.2 Loop gain control by coherent AGC.

two interdependent loops, the PLL and the automatic gain control (AGC) loop, both of which must be designed for proper performance.

If the AGC loop has high DC gain, it will hold the output of the coherent detector approximately constant, and it will no longer be useful for differentiating between true and false locks. However, the signal that controls the variable-gain amplifier will indicate the strength of the signal at the amplifier's input, and it will thus be useful for false lock detection.

Whether or not coherent AGC is employed, proper false lock detection requires that the strength of the input signal be well enough known. Otherwise we cannot differentiate lock to a weak signal from a DC voltage generated by a strong signal when the loop is not locked. To make this so, it may be necessary to precede the circuit of Fig. 9.1 or 9.2 with a separate, noncoherent, AGC to establish a constant signal level. However, once again, excessive noise can defeat the system by causing the gain to be controlled by the noise rather than the signal.

9.2.2 Filter Modification

We can also modify the loop filter time constants while leaving the low-frequency gain unchanged [Rey, 1960]. During acquisition the filter can be effectively eliminated by shorting the series resistor, possibly turning the loop into a first-order loop. In Fig. 9.3a, R_{1b} might represent the series resistance of the switch, which will usually be a semiconductor device; it could be negligible or it might be purposely added to control gain in the acquisition mode, perhaps to avoid instability. Once acquisition has occurred rapidly in the wide-bandwidth loop, the switch can be opened⁶ to restore the desired loop parameters. This process must be controlled from external circuitry, perhaps a timer that is initiated when the loop center frequency is tuned to a new value or the output from a lock indicator such as the coherent detector.

Figure 9.3b shows a circuit that automatically widens the bandwidth for large changes in u_1 . Once the loop has locked, the changes will be small enough (in the

⁶ We must take care that the switch control signal does not couple excessively into the filter and cause a transient that could undo the process.

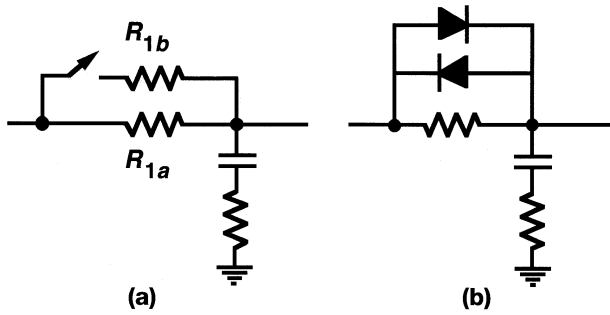


Fig. 9.3 Modification of filter for acquisition.

absence of excessive noise) that the diodes will have a very high impedance but, during pull-in, they will conduct and increase the filter pole frequency. The circuit must be carefully planned so this transition occurs at the best point in the pull-in process. There is also a danger that transients from the phase detector (especially from certain digital types) will be large enough to pass through the diodes, thus severely reducing the effectiveness of the filter.

9.2.3 Comparison of Two Types of Parameter Modifications

Figure 9.4 compares the two methods described above by showing the Bode plots of the unmodified loop and of both modified loops.

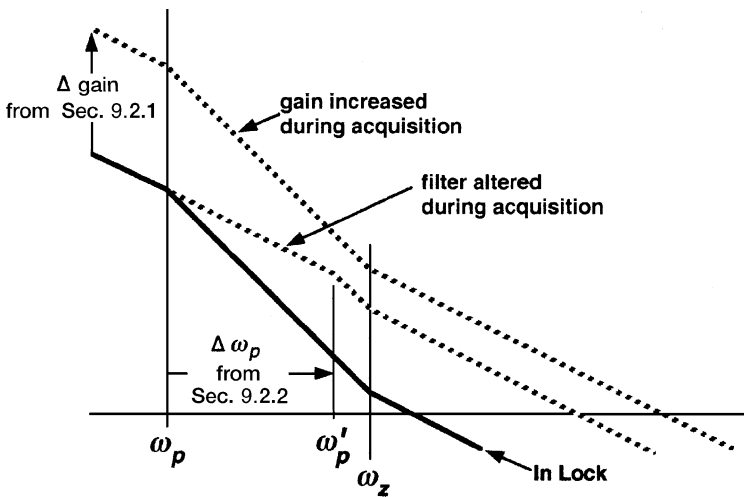


Fig. 9.4 Bode plots corresponding to loop during acquisition. These plots illustrate the modification of loop parameters based on a linearized loop although that is not an accurate representation during much of the pull-in process.

9.3 AUTOMATIC TUNING OF ω_c , FREQUENCY DISCRIMINATOR

A frequency discriminator can be used to produce a voltage that is proportional to the received frequency and that voltage can be added to u_2 to tune ω_c closer to the input frequency. The quadricorrelator [Richman, 1954b; Egan, 2000, pp. 445–448] is a particularly applicable type of discriminator. It produces a voltage that is proportional to the difference between two input frequencies. (The output of an ordinary discriminator is proportional to the frequency offset from some constant value. Thus it is not suitable for indicating the difference between two frequencies that might both be variable. It is also subject to inaccuracies due to component drifts that might change its center frequency or that of the VCO or reference.) This can be used to tune the VCO as shown in Fig. 9.5. At low-frequency errors, its output diminishes so it has little effect on the loop once lock has occurred.

The loop input and the VCO output are compared in two multipliers. One of the signals is shifted by 90° before entering one of the multipliers. As a result, the outputs from the two multipliers, which are at the difference frequency $\Delta\omega$, differ in phase by 90° . One of these is differentiated, giving it a 90° phase shift and a frequency-dependent gain. The two signals that enter the final multiplier are in phase (coherent detection), and one of them has an amplitude that is proportional to $-\Delta\omega = \omega_{in} - \omega_{out}$ so it can be added to u_2 in the loop to decrease $\Delta\omega$. Once again, the response of the PLL during pull-in depends on the combined action of two loops. This time a frequency control loop has been added.

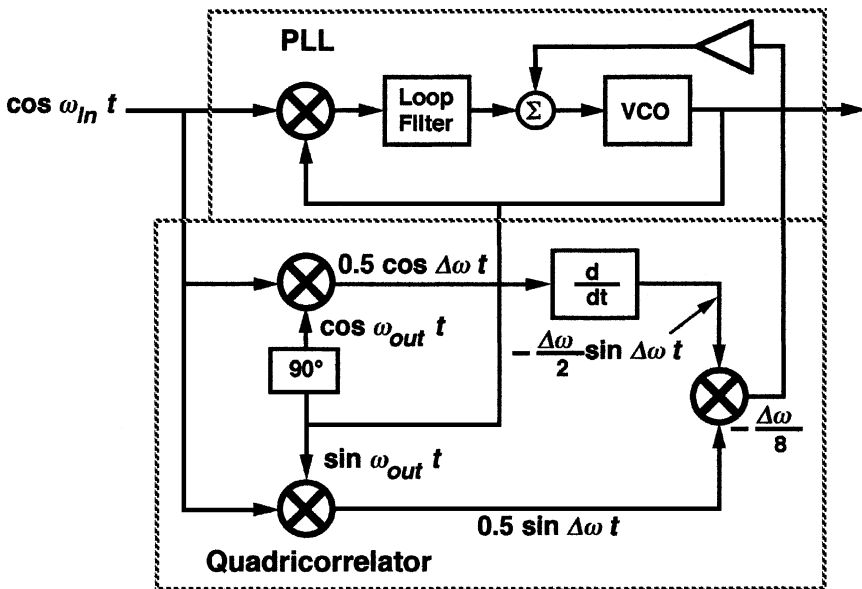


Fig. 9.5 Quadricorrelator concept.

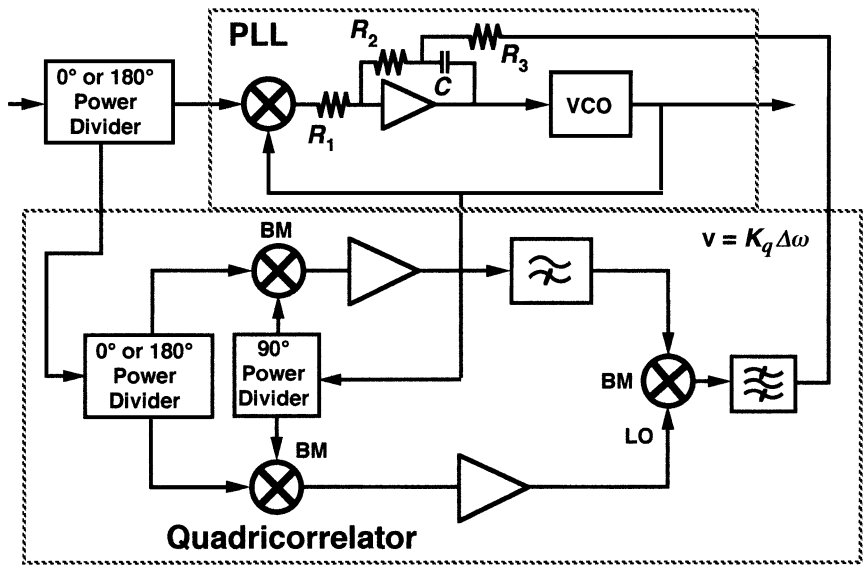


Fig. 9.6 Practical quadricecorrelator.

Figure 9.6 shows a more practical model of the quadricecorrelator. It has amplifiers and the multipliers have been replaced by balanced mixers. The differentiator has been replaced by a high-pass filter, which functions as a differentiator at frequencies that are well below its cutoff frequency ω_{co} :

$$F(\omega) = \frac{j\omega}{\omega_{co} + j\omega} \approx j \frac{\omega}{\omega_{co}} \quad \text{for } \omega \ll \omega_{co}. \quad (9.7)$$

Some options have been shown. A bandpass filter has been placed at the output. We generally assume low-pass function after balanced mixers and multipliers—it eliminates the sum frequency component—but the high-pass function can be added to decrease further the influence of the quadricecorrelator on the loop during normal in-lock operation, if desired.

The quadricecorrelator output is now shown as being added to the PLL at its loop filter [Gardner, 2005, p. 200]. The loop filter here becomes a filter, an integrator, for the quadricecorrelator loop. To see this, assume no DC output from the phase detector (PD), either because the PLL is out of lock or by superposition. Then no current flows through R_1 . Since the input to the op-amp is a virtual ground into which no current can flow, no current will then flow through R_2 either, so the junction of C and R_3 is effectively connected to the op-amp input. Since the quadricecorrelator otherwise has no pole at the origin (there is no integration to convert frequency to phase as in the PLL), the integrator causes it to change from a zero-order to a first-order loop (to the degree that this integrator is a true integrator).

9.4 ACQUISITION AIDING LOGIC

Several schemes are available that cause the phase detector to aid the acquisition process by becoming effectively a frequency discriminator under conditions where it would otherwise skip cycles (the loop being out of lock) [Egan, 2000, pp. 451–459]. The best known of these is the phase-frequency detector. It is most often used when the signals are, or have been made, digital and is not generally employed in locking to noisy signals.

Compare Fig. 9.7 with Fig. 3.1. The output at C is the same as before, and as before, it goes to 1 (positive) when triggered by the waveform at A . The difference is that, here, B resets the upper flip-flop only indirectly, although the results appear to be the same; B sets the lower flip-flop, which in turn provides a second 1 input to the AND gate (the first being C). This causes a 1 at the AND output, which then resets both flip-flops. The results are the same as before for the case shown. However, if τ should go to zero and then negative, B then preceding A , the results are considerably different. Now the output pulse appears at D and is proportional to $-\tau$. The output at D will eventually be inverted relative to that at C , possibly by entering the opposite input to the loop filter op-amp, so the pulse width at D represents negative phase relative to that at C . This phase detector has a linearly changing output (average pulse width) from τ equal to -1 reference period to $+1$ period, a range of $\pm 360^\circ$. (These linear characteristics have been described in Section 3.1.3.)

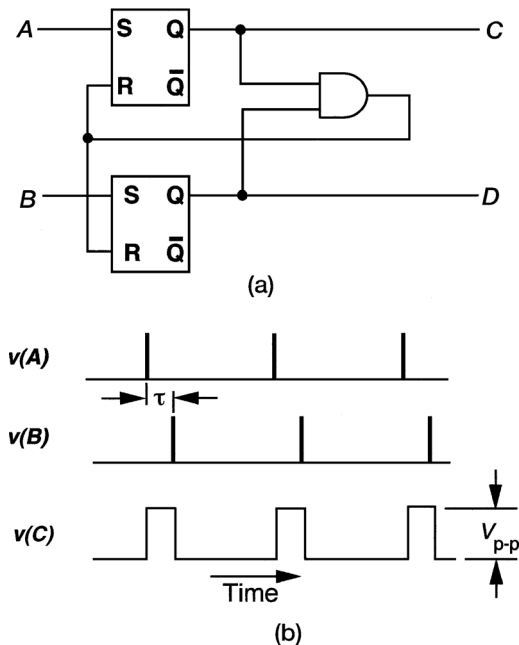


Fig. 9.7 Representation of the phase-frequency detector.

Even more important than the extended range is what happens when the loop is out of lock. If B is slightly lower in frequency than A , τ will increase until it equals a reference period. Then two A pulses will occur between B pulses and the output pulse width will drop back toward zero. Thus the characteristic will look like that in Fig. 3.1. There is a significant difference, however. The center of the lock range in Fig. 3.1 is in the middle of the sawtooth waveform, at 180° phase difference, whereas it is at the bottom of the waveform, at 0° , here. Thus, rather than generating an average value in the center of the range when out of lock, this phase detector generates an average value that corresponds to 180° above the normal lock point. In a high-gain loop the resulting voltage will drive the VCO toward the reference frequency.

If the frequency of B is higher than that of A , the same process will occur at D . Thus, when the loop is out of lock, a sawtooth will occur at C or at D , depending on which frequency is higher, and the average value produced will drive the loop toward lock.

The 180° -equivalent output occurs as the ratio of the two frequencies approaches unity and if the effect of feedback can be neglected. For higher frequency ratios, the offset becomes even greater, approaching the equivalent of 360° at large ratios. If one input should be lost, the output would become a steady value equivalent to 360° phase. In all cases the polarity will be correct to bring the frequency error toward zero.

9.5 SWEEPING ω_c , TYPE 2 LOOP

Various methods can be employed to cause ω_c to be swept past the input frequency so $\Omega \Rightarrow 0$ and lock will occur. The sweep may be discontinued when lock occurs or it may continue to be applied but be effectively overcome by the phase lock.

9.5.1 Maximum Sweep Speed, Closed-Loop Sweeping

The maximum output from the phase detector is K'_p rad. This will cause a current of K'_p rad/ R_1 into C in the filter of Fig. 3.23c, where $R_1 = 1$ for a current-source PD or is as shown in Fig. 3.25. This in turn will cause a voltage ramp of slope K'_p rad/ $(R_1 C)$ and that will produce a frequency change of slope

$$d\Omega/dt = K'_p K_v / (R_1 C) = \omega_n^2 / \text{rad}. \quad (9.8)$$

The frequency cannot ramp faster. It can move faster for a short time due to the zero in the filter—a sudden change in u_1 will be multiplied by R_2/R_1 and immediately appear at the filter output. However, the sustained ramp cannot move faster than given by Eq. (9.8). A low-pass pole, created by placing R across C , will only reduce the current into the C and therefore the slope. Even if the loop is locked it cannot follow a frequency slope greater than that given by Eq. (9.8).

By analyzing phase-plane plots, corresponding to equations such as Eq. (8.72), with $\zeta = 1/\sqrt{2}$, but with a term representing a linearly changing Ω , Viterbi [1966,

p. 58] confirmed Eq. (9.8). He also found that, as the sweep rate is lowered slightly, lock *may* occur if Ω is swept *through* zero. Whether it does depends on initial values (phase and frequency). If Ω is sweeping away from zero, the loop will not lock unless it does so immediately, without skipping a cycle (i.e., if it is almost locked to start with). But, when the sweep rate falls to

$$d\Omega/dt < \omega_n^2/(2 \text{ rad}), \tag{9.9}$$

the loop will always lock if Ω is swept through zero. The effect of much lower sweep rates is to establish a pull-in range (noninfinite) for values of Ω that correspond to sweeping away from lock,

$$\Omega_{PI} = \frac{\omega_n^2 \omega_L}{2\Omega \text{ rad}}, \quad |\Delta\omega| \text{ increasing}, \quad \dot{\Omega} \ll \frac{\omega_n^2}{2 \text{ rad}}. \tag{9.10}$$

Results for a high-gain (approaching integrator and lead) loop with $\zeta = 1/\sqrt{2}$ are summarized in Table 9.1. While this only applies precisely for $\zeta = 1/\sqrt{2}$, values of ζ that are close to $1/\sqrt{2}$ tend to be desirable for other reasons, so the information for this damping ratio is quite useful [Gardner, 2005, pp. 195–196].

9.5.2 Open-Loop Sweeping

One way to acquire lock is to sweep ω_c with the loop open, detect when $\Omega = 0$, stop the sweep at that point, and close the loop. Detection is performed by mixing the loop input and output signals, passing the mixer output through a low-pass “decision” filter, and detecting the filter output using a rectifier. See Fig. 9.8. The detected signal is then compared to a fixed threshold and the sweep is stopped when the threshold is surpassed. The bandwidth of the low-pass decision filter determines how large the frequency error can be when threshold is broken.

TABLE 9.1 Lock Probability vs. Sweep Speed, $\zeta=1/\sqrt{2}$, I & L Filter, Closed-Loop Sweeping

$R = \left \frac{(d\Omega/dt)\text{rad}}{\omega_n^2} \right $	Sweeping Away $d \Omega /dt > 0$	Sweeping Toward $d \Omega /dt < 0$
$1 < R$	No lock; cannot maintain lock	
$0.5 < R < 1$	No lock ^b	Transition between 0 and 100% probability 50% probability at $R \approx 0.92^a$
$R = 0.5$	”	100% probability of lock
$R \ll 0.5$	$\omega_{PI} \approx \omega_n^2 \omega_L / (2 \text{ rad } d\Omega/dt)$	”

^a Gardner (2005, Fig. 8.8).

^b Except if initial values are such that no cycle is skipped.

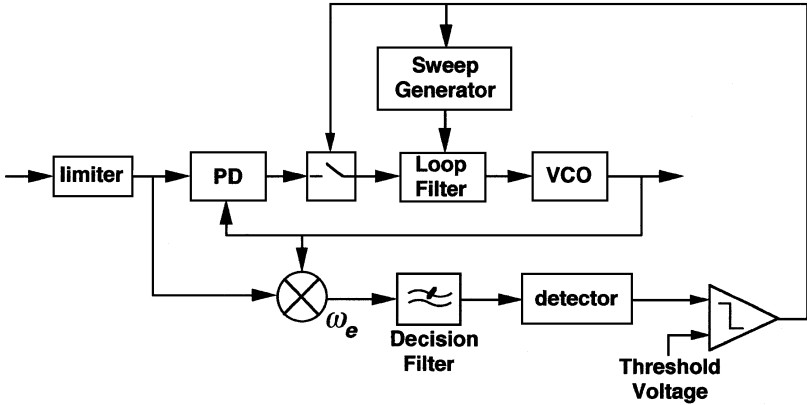


Fig. 9.8 Open-loop sweeping. When the error frequency is small enough to pass through the low-pass decision filter, the threshold is exceeded. This causes the sweep generator to stop and the loop to close.

The detector might be a half-wave rectifier as shown in Fig. 9.9a or a full-wave rectifier as in Fig. 9.9b. When the loop is being swept at a high rate, it may be that only a part of a cycle of ω_e will be seen, so the response depends on the phase that happens to occur as the signal passes through the low-pass filter [Blanchard, 1976, pp. 281–287]. Since the half-wave rectifier will have a maximum response only when the waveform is at a positive peak, whereas the full-wave rectifier has a maximum response to either a positive or a negative peak, the latter will have a greater probability of a strong response. To further increase the likelihood of a strong response, the detection circuitry can be repeated, except with one of the mixer inputs

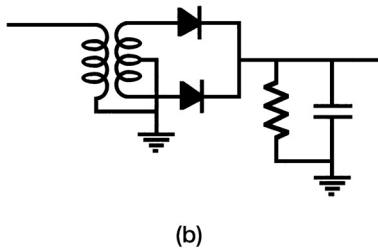
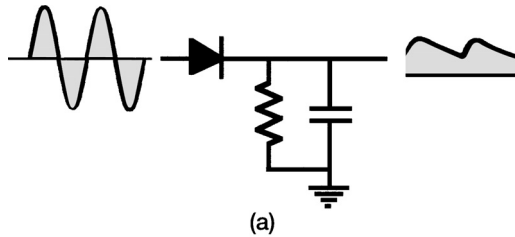


Fig. 9.9 Detectors: (a) half and (b) full wave.

shifted 90° , and the output of the second threshold circuit can be ORed with the first. Then a maximum response will exist through one of the four paths every 90° of phase, and the response will be close to maximum regardless of the phase.⁷

The narrower the decision filter, the slower must be the sweep. The time during which the difference frequency is within the pass band of the decision filter is proportional to the filter bandwidth B and inversely proportional to the sweep speed.

$$T_p \sim B/\dot{\Omega} \quad (9.11)$$

The rate at which the envelope of a pulse can rise in a filter is also proportional to the filter bandwidth, so, to reach a given relative threshold $r_T = V_T/K'_p$, the bandwidth must be proportional to the inverse of the pulse width,

$$B \sim 1/T_p \sim \dot{\Omega}/B. \quad (9.12)$$

Therefore the maximum sweep rate that will permit the detected output to rise to within r_T of its maximum possible value can be related to the filter bandwidth by a function of r_T ,

$$\dot{\Omega}_M/B^2 = c(r_T). \quad (9.13)$$

Blanchard presents⁸ experimental curves giving $c(r_T)$ for various probabilities of successful detection (stopping of the sweep) between 50% and 99%. For $0.2 < r_T < 0.8$, his curve for the ratio of maximum sweep rate to decision filter bandwidth at 99%⁹ success can be expressed approximately as

$$\dot{\Omega}_M/B^2 \approx \frac{(167)^{(1-r_T)}}{2\pi \text{ rad}}. \quad (9.14)$$

On the basis of comparison with other data given by Blanchard,¹⁰ the detector used in his experiment is probably a full-wave rectifier or a pair of quadrature rectifiers.

The question naturally arises: What is the advantage of open-loop sweeping over closed-loop sweeping, if any? We can compare them by recognizing that the bandwidth of the loop that is to be locked after the sweep stops influences both the allowed parameters in Eq. (9.13) and the speed at which it can be swept in closed loop. The wider the decision filter the greater will be the maximum frequency error when the sweep stops and the wider must be the loop if it is to acquire rapidly at that time. But

⁷ If the video filtering is not required, the detectors can be eliminated, and the threshold circuit(s) can sense the waveforms directly. The analog to the full-wave rectifier is a threshold circuit set to detect both positive and negative excursions.

⁸ See Blanchard (1976), Fig. 11.6; from unpublished document by S. Vialle, May 1969. Presumably the loop is a high-gain loop with a damping factor on the order of $1/\sqrt{2}$.

⁹ When probability of lock changes from 99 to 50%, with $\alpha = 0.5$, the value $\dot{\Omega}_M/B^2$ almost doubles.

¹⁰ Blanchard (1976), Fig. 11.5, gives probability of stopping versus sweep rate for a particular set of conditions that are also covered by Fig. 11.6. Figure 11.5 gives three curves, one for each of the configurations discussed above. The data from Fig. 11.5 fall between that for a full-wave rectifier and for two quadrature full-wave rectifiers.

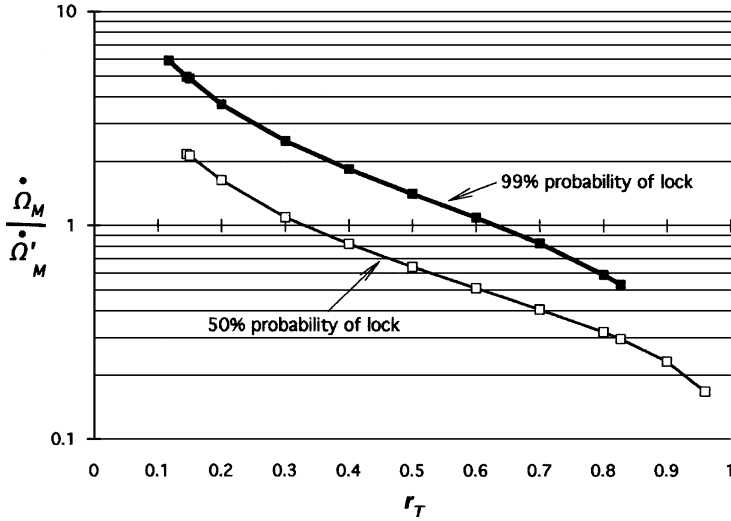


Fig. 9.10 Approximate ratio of maximum open-loop sweep rate $\dot{\Omega}_M$ to maximum closed-loop rate $\dot{\Omega}'_M$ versus relative threshold.

a wide loop can also be swept rapidly when it is closed, and that sweep rate can be compared to the open-loop rate. An approximate analysis based on such considerations is given in Appendix 9.A and results are shown in Fig. 9.10, which gives the ratio of maximum open-loop sweep rate to maximum closed-loop rate for various values of threshold r_T . From the 99% probability curve we see that there can be an advantage in open-loop sweeping at low threshold levels. Here the high sensitivity of the frequency error sensing mechanism allows the loop to lock at faster sweep rates than it could in closed loop. However, the practical limitation on low thresholds is increased susceptibility to noise. (Our analysis has not included the effects of noise.)

Example 9.1 Sweeping to Acquire The VCO center frequency in a PLL is to be swept to aid in acquiring lock. The natural frequency is 2 kHz. The threshold is set at 40% of maximum. How fast can we sweep for 99% probability of lock?

Figure 9.10 indicates that we can sweep faster at $r_T = 0.4$ with the loop open than with it closed. We can solve (9.14) for $\dot{\Omega}_M/B^2$ and substitute this into (9.20) to relate f_n to B . Equation (9.14) gives $\dot{\Omega}_M/B^2$ as 3.43. Equation (9.20) then becomes

$$f_n = 1.95B \quad \text{so } \underline{B = 6470 \text{ rad/sec} = 1030 \text{ Hz.}}$$

Since we have obtained $\dot{\Omega}_M/B^2$ and B , we can obtain

$$\dot{\Omega}_M = 3.43(6470/\text{sec})^2 = 1.44 \times 10^8 \text{ rad/sec}^2 \Rightarrow 22.9 \text{ MHz/sec.}$$

If we had used closed-loop sweeping, the rate for 100% lock probability (and

approximately the same as the 99% probability rate) would be, from (9.9),

$$\dot{\Omega}'_M = \frac{\omega_n^2}{2 \text{ rad}} \Rightarrow 12.6 \text{ MHz/sec.}$$

This is 1.8 times slower, consistent with Fig. 9.10.

9.5.3 Combined Techniques

It is also possible to use a decision circuit to terminate the sweep, as in open-loop sweeping, but while allowing the loop to remain closed. This may provide advantages from both techniques, permitting the feedback to slow the sweep as zero error is approached but still stopping it based on the detector output.

9.6 SWEEP CIRCUITS

Now that we have discussed some theory concerning the affect of sweep rate on locking, let us consider how we might generate the sweep.

9.6.1 Switched Current Source

The VCO can be caused to sweep when it is out of lock by injection of a current into the filter feedback capacitor as in Fig. 9.11. It is important that the current not pass through either of the resistors, as it would if it were injected at either end of R_1 . Otherwise there will be a step of output voltage when the current stops and the lock, once achieved, could be broken as a result. When the loop locks, even though the current continues, the loop will cancel the injected current and sweeping will stop (in this type 2 loop). There will be a steady phase error that causes I to pass through R_1 and R_2 and thus not enter C . The value of this phase offset can be easily computed from

$$\cos \Delta\varphi = IR_1/K'_p. \tag{9.15}$$

We can substitute this into the expression for the sweep rate before acquisition, obtaining

$$-d\Omega/dt = K_v I/C = K'_p \cos \Delta\varphi K_v/R_1 C = \omega_n^2 \cos \Delta\varphi, \tag{9.16}$$

which again shows that $|d\Omega/dt|$ cannot exceed ω_n^2 . The phase offset can be eliminated by stopping the current I , perhaps based on the output from a lock indicator.

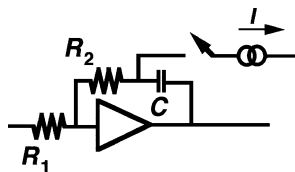


Fig. 9.11 Switched sweep current.

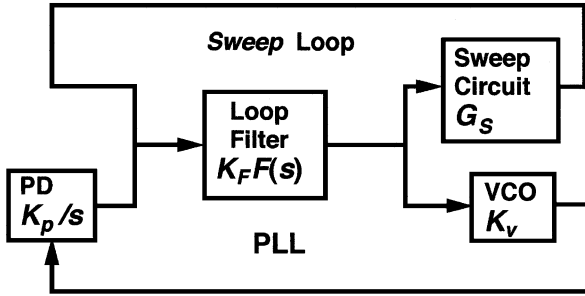


Fig. 9.12 Sweep circuit.

9.6.2 Automatic Sweep Circuit—Sinusoidal

The circuit of Fig. 9.12 can be designed to cause frequency sweeping when the PLL is out of lock. The loop filter is part of two loops, the PLL and the sweep loop. The latter is designed to be an oscillator at a frequency that is low enough that $d\Omega/dt$ will permit acquisition. When the PLL locks, it tries to cancel the input from the sweep loop. The PLL now changes the transfer function across the loop filter as seen by the sweep loop. This reduces the gain in the sweep loop so that oscillation ceases. The effect of each loop on the other should be taken into account by replacing the transfer function across the loop filter by the appropriate (from filter input to filter output) closed-loop transfer function of the other loop, assuming the other loop is locked. When the PLL is not locked, there can still be feedback [Egan, 2000, pp. 450–451] due to the false lock phenomenon previously discussed, and the sweep loop must be designed to continue sweeping under these circumstances.

9.6.3 Automatic Sweep Circuit—Nonsinusoidal

Figure 9.13 shows a modification of the sweep circuit that causes the sweep to be more linear and eliminates the influence of the sweep loop on the PLL during lock.

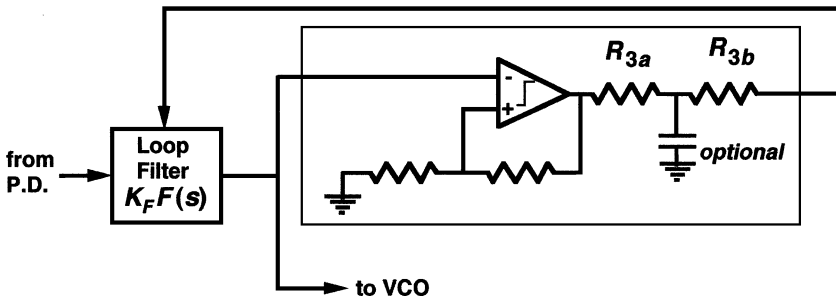


Fig. 9.13 Sweep generator with hysteresis. Hysteresis is created by the voltage divider, which gives positive feedback.

This is done by including a high-gain amplifier or comparator in the sweep loop. When the tuning voltage reaches an upper limit, it causes the output polarity of the comparator to reverse. Positive feedback provides hysteresis, preventing reversal from taking place until the lower limit is reached. The sweep may be produced by injection of the comparator output into the loop filter through R_3 , as in Fig. 9.6, for example, or the loop filter may act primarily as an amplifier for the output of an optional low-pass filter. The low-pass then acts approximately as an integrator, converting the level change from the comparator into an approximate ramp. Of course, the circuit components must be chosen to ensure sweeping between the desired limits when the loop is out of lock and to ensure that the loop is not pulled out of its hold-in range when sweep stops [Egan, 2000, pp. 448–450]. This could happen if the steady-state sweep input to the filter were too large for the phase detector to cancel.

9.A APPENDIX: MAXIMUM SWEEP RATE, OPEN-LOOP VS. CLOSED-LOOP

Suppose we want ω_L to be as great as the value of $|\Delta\omega|$ at the moment when the sweep stops so the loop will then lock rapidly. There will be a frequency error due to the time T from $\Omega = 0$ until the sweep stops. Blanchard (1976) has also observed a dispersion of $\pm B$ in the stop frequency. Thus the total error at the maximum sweep rate could be

$$\Delta\omega \lesssim \dot{\Omega}_M T + B. \quad (9.17)$$

We can set this equal to ω_L , which, for $\zeta = 1/\sqrt{2}$, equals $\sqrt{2}\omega_n$. We now make the rough approximation that the filter output is like the response to a step occurring at $\Omega = 0$. The time to rise to a relative threshold level of r_T through a single-pole low-pass filter is obtained from

$$r_T = 1 - e^{-TB}. \quad (9.18)$$

Combining these we obtain

$$T = \frac{\sqrt{2}\omega_n - B}{\dot{\Omega}_M} = \frac{\ln[1/(1 - r_T)]}{B}, \quad (9.19)$$

and we can solve this for the natural frequency of the loop that is just wide enough to acquire rapidly.

$$\omega_n = \frac{1}{\sqrt{2}} \left[\frac{\dot{\Omega}_M}{B} \ln \left(\frac{1}{1 - r_T} \right) + B \right]. \quad (9.20)$$

The maximum sweep rate for closed-loop sweeping is

$$\dot{\Omega}'_M = \lambda_m \omega_n^2, \quad (9.21)$$

where $\lambda_m \approx 0.5$ for 100% probability of lock and 0.92 for 50% [Gardner, 2005, Fig. 8.8]. Combining the last two equations, we obtain

$$\dot{\Omega}'_M = \frac{\lambda_m}{2} \left[\frac{\dot{\Omega}_M}{B} \ln \left(\frac{1}{1 - \nu_T} \right) + B \right]^2, \quad (9.22)$$

$$\frac{\dot{\Omega}'_M}{B^2} = \frac{\lambda_m}{2} \left[\frac{\dot{\Omega}_M}{B^2} \ln \left(\frac{1}{1 - \nu_T} \right) + 1 \right]^2. \quad (9.23)$$

The ratio of maximum open-loop sweeping rate to maximum closed-loop sweeping rate for the same ω_n is thus

$$r \equiv \frac{\dot{\Omega}_M}{\dot{\Omega}'_M} \approx \frac{2}{\lambda_m} \frac{\dot{\Omega}_M/B^2}{\left\{ (\dot{\Omega}_M/B^2) \ln[1/(1 - \nu_T)] + 1 \right\}^2}. \quad (9.24)$$

This ratio is plotted in Fig. 9.10 based on Blanchard's (1976) curves for Eq. (9.13).

CHAPTER 10

APPLICATIONS AND EXTENSIONS

We can apply the basic theory that we have studied to a broad range of circuits that may appear at first to be significantly different from those for which the theory was developed. The challenge is to recognize the equivalence. In this chapter we briefly discuss a number of such applications. The references provide a guide to further information.

10.1 HIGHER-ORDER LOOPS

While much of what we have studied is applicable to loops of all orders, we have concentrated on second-order loops in order to avoid the complexity resulting from an additional loop parameter. The analysis methods we have used are applicable to particular higher-order loops—that is, to loops with specified parameters—but it is difficult to discuss them in general terms, as we have done with second-order loops. However, we can extend our general discussion to an important group of third-order loops, those that are designed for maximum phase margin. That additional restriction reduces the number of parameters to be considered and permits the description of their modulation and transient performance contained in Appendix 10.A.

10.2 GENERALIZED VOLTAGE-CONTROLLED OSCILLATOR

By identifying certain collections of components as being equivalent to the voltage-controlled oscillator (VCO) in the loops that we have studied, we can use the theory

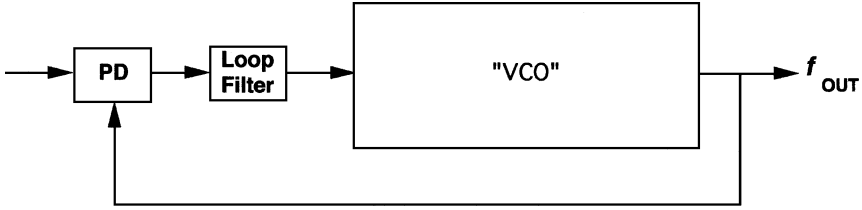


Fig. 10.1 Generalized VCO in loop.

that has been developed to treat more complex circuits. Thus the “VCO” in Fig. 10.1 can be a representation of a group of functional blocks.

10.2.1 Frequency Synthesis, Frequency Division

A phase-locked frequency synthesizer can be represented by substituting the circuit of Fig. 10.2 for the “VCO” in the generic loop. The frequency divider ($\div N$) divides the oscillator frequency (and phase) by the divide ratio N . Loop responses can be determined for the loop of Fig. 10.1, writing the tuning sensitivity of the “VCO” block as

$$K_v = K'_v / N, \tag{10.1}$$

where K'_v is the actual VCO tuning sensitivity. The usable output is taken from the output of the actual VCO since the object is to vary that frequency by changing N . We know that, in steady state, $f_{out} = f_{in}$ for the phase-locked loop (PLL) so

$$f_v = N f_{out} = N f_{in}, \tag{10.2}$$

where f_{in} is the loop’s reference frequency, of course, but the term may also be applied to a higher frequency that has been divided by a constant to provide f_{in} . The synthesizer produces, at the VCO output, frequencies that are, during steady state, multiples of f_{in} . Thus f_{in} is the minimum step size, or resolution, of the synthesizer. (Methods for overcoming this sometime severe restriction are discussed in Appendix 10.F.) Since it is usually a precise frequency derived from a crystal oscillator or other frequency

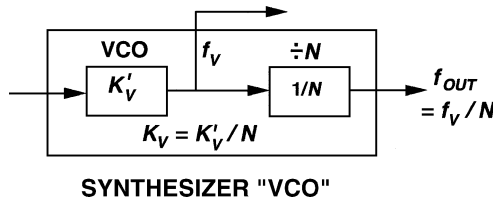


Fig. 10.2 Frequency divider in a “VCO.”

standard, the synthesized output has the same relative accuracy but can produce many frequencies rather than the single frequency available from the reference. From Eq. (10.2) can be seen that a given relative change (e.g., expressed in percentage or parts per million) in f_{in} causes the same relative change in f_v .

The loop filter usually includes a low-pass filter. In our previous studies, we have assumed a low-pass filter following the phase detector to eliminate all but the desired, difference frequency, output component. We have also assumed that its cutoff frequency is so high that it has little effect on loop response. In a synthesizer this latter assumption is not necessarily valid and the low-pass filter that eliminates undesired components at the phase detector output often must be considered part of the loop filter. We have studied the second-order loop with a low-pass filter ($\alpha = 0$) but, if a type 2 loop is employed, the additional low-pass filter implies a third-order loop. While we have also studied such higher order loops, using state-space methods and computer analysis, the detailed analysis of the class of third-order loops that is contained in Appendix 10.A is particularly useful for frequency synthesizers.

Phase-locked loop theory, such as we have studied, is used to determine the synthesizer's loop stability, its transient response, and its response to various noise sources.

10.2.1.1 Stability. Stability is evaluated using the methods in Chapter 5 with K_v given by Eq. (10.1). A complication that is sometimes significant in stability analysis, as well as in the determination of loop response in general, is due to the bandwidths of synthesizer loops often being extended so close to the reference frequency (e.g., $\omega_L = 0.1\omega_{in}$) that we can no longer consider only the low-frequency component of the phase detector's output. To permit such wide bandwidths, without excessive modulation of the output by the undesired components from the phase detector, at f_{in} and harmonics thereof, special phase detectors are employed. These minimize undesired outputs. However, the fact that information is only updated once per cycle of f_{in} affects the loop response [Egan, 2000]. While this is essentially true in all PLLs, it is not significant when the loops are narrow enough relative to the input frequency.

If the phase detector is of the sample-and-hold type, an additional phase shift of

$$\varphi_s = -\pi f_m / f_{in} \quad (10.3)$$

(corresponding to a delay of half a sample period) can be included in $G(\omega)$ to improve the accuracy of the stability analysis. This can also be used as a conservative approximation with other types of phase detectors. Such additional phase shift also causes a reduction in phase margin, leading to greater peaking in transient and frequency responses.

10.2.1.2 Transient Response. The primary transient experienced occurs when N is changed in order to change f_v . Because this is equivalent to a change in f_{in} , it can be analyzed by the methods studied in Chapter 6. If N is changed from N_0 to N_1 , the final value of the VCO frequency will be $f_{v1} = N_1 f_{in}$. Just after N is changed,

however, f_v will still be the previously synthesized frequency, $f_{v0} = N_0 f_{in}$.¹ This is also the frequency that would be produced with the final value of $N = N_1$ by a reference frequency of

$$f_{in,0} = f_{v0}/N_1 = N_0 f_{in}/N_1 \quad (10.4)$$

so the equivalent input step is

$$\Delta f_{in} = f_{in} [1 - N_0/N_1]. \quad (10.5)$$

A step that may equivalently be introduced after f_v is

$$\Delta f_v \equiv f_{v1} - f_{v0} = N_1 \Delta f_{in} = N_1 f_{in} [1 - N_0/N_1] = f_{in} [N_1 - N_0], \quad (10.6)$$

to which f_v will respond in the typical low-pass fashion of $H(s)$.

10.2.1.3 Response to Noise. Noise considerations are usually very important in synthesizer design since high-purity signals are sought. The loop imposes low-pass filtering (H) upon phase noise that enters with the reference and high-pass filtering ($1 - H$) on the open-loop noise of the VCO. Noise introduced through various loop components, such as the op-amp, is also transferred to the output. The processes are the same as we have studied in Chapter 7.

While there are many additional considerations peculiar to frequency synthesis [Egan, 2000], the theory that we have studied is basic to synthesizer design.

10.2.2 Heterodyning (Frequency Mixing)

The VCO plus a frequency mixer can also be represented as an equivalent VCO. In Fig. 10.3, the frequency before the mixer is

$$f_v = f_c + K_v u_2. \quad (10.7)$$

After the mixer the frequency is

$$f'_v = f_v \pm f_{MIX} = f'_c + K_v u_2, \quad (10.8)$$

where

$$f'_c = f_c \pm f_{MIX}. \quad (10.9)$$

Thus the mixing process modifies the center frequency of the “VCO.” If Eq. (10.9) produces a negative sign, there is also an additional inversion, or 180° phase change,

¹ More universally, this is true the instant before the change is made, but practical synthesizers have low-pass filters that do not permit a phase change to be transmitted instantly to the output.

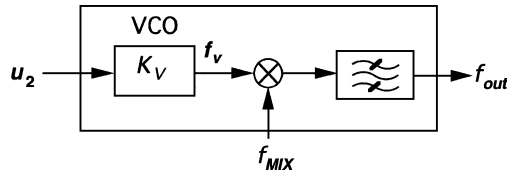


Fig. 10.3 Equivalent VCO with frequency mixing.

to be accounted for. This is easily done by changing the sign of one of the blocks in the loop. Most computations, such as response to modulation or step response, involve changes from steady state anyway, so the center frequency is not of concern. It is, of course, important from a practical standpoint.

The desired output is often f_v with the mixing process used to facilitate realization of the PLL.

If the filter following the mixer is not very wide, it must be accounted for in determining loop performance. How it affects modulation phase is of importance to loop response. For small phase deviation and with a symmetrical filter, an equivalent low-pass filter can be introduced into the loop, one that affects the modulation in a manner that produces the same effect on the modulation sidebands as would the actual bandpass filter [Egan, 2000, pp. 349–351].

10.3 LONG LOOP

A configuration called a long loop is shown in Fig. 10.4. The output frequency is multiplied by N and mixed with the input. (Recall that phase undergoes summation, multiplication, or division simultaneously with frequency.) The resulting difference frequency is phase detected against the output signal. The phases of the various signals are shown in Fig. 10.4.

An equivalent single-loop model is shown in Fig. 10.5. Note that the output of the phase detector is the same in each model. We again create an equivalent VCO

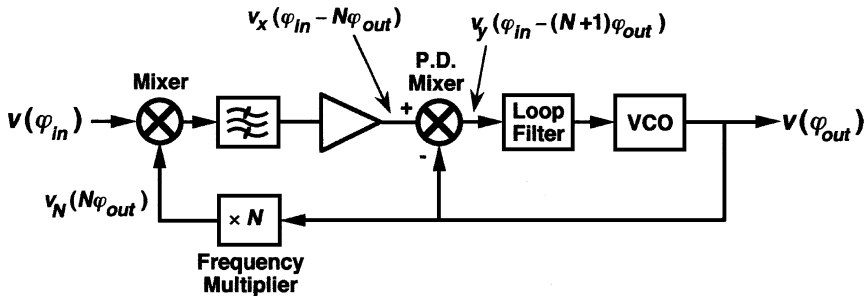


Fig. 10.4 Long loop.

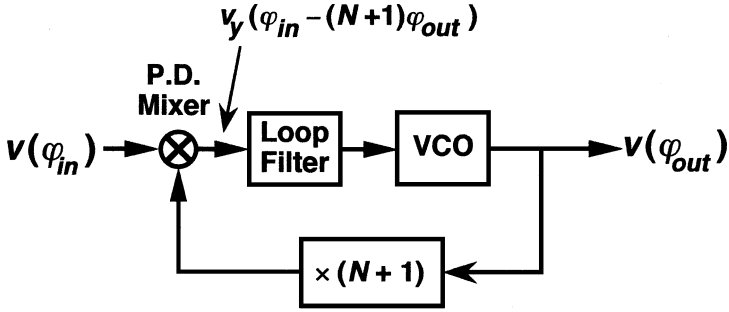


Fig. 10.5 Long loop first equivalent.

(Fig. 10.6) to substitute into our generic loop so we can see clearly how to apply our mathematical tools. Thus the long loop is analyzable by techniques that we have developed, but what are we getting for the added complexity? What we gain is the relatively narrow bandpass filter that enables us to limit the band of signals that enters the phase detector, thus rejecting unwanted energy at the input and, especially, noise. We could, of course, pass the input to any loop through a bandpass filter, but this filter can be narrower. If the input signal frequencies vary over ΔF , the passband of a filter preceding an ordinary loop would have to be ΔF wide, but the passband in a long loop can be much narrower.

From Fig. 10.4, the passband width of the filter must be

$$\Delta F \geq \Delta f_{in} - N \Delta f_{out}. \tag{10.10}$$

From Fig. 10.5, the input frequency is $N + 1$ times the output frequency, so Eq. (10.10) gives

$$\Delta F \geq \Delta f_{in} - [N/(N + 1)] \Delta f_{in} = \Delta f_{in}/(N + 1). \tag{10.11}$$

Thus the necessary width of the input filter is reduced by a factor $N + 1$ in the long loop. However, the VCO must be tuned to the correct frequency to start since the loop cannot see the input signal unless it passes through the narrow input pass band. Again, if the bandpass filter is not wide compared to the loop, the response of the modulation to the equivalent low-pass filtering must be taken into account.

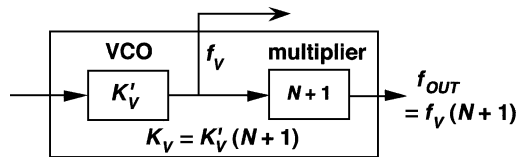


Fig. 10.6 Long loop equivalent VCO. Substitutes for the VCO and $\times(N + 1)$ in Fig. 10.5 and forms a standard loop with the PD mixer and loop filter of Fig. 10.4.

10.4 CARRIER RECOVERY

One prominent use of PLLs is in carrier recovery of phase-coded signals.

10.4.1 Biphase Costas Loop

The simplest of such signals is the biphase coded (BPSK for binary phase shift keyed) signal. Information is carried in the phase, which has two possible values, 180° apart. To decode the received signal, an unmodulated reference, operating at the same frequency, to which the phase can be compared, is needed. The problem of recovering the reference from the signal is difficult because a randomly modulated BPSK signal has no discrete energy line at the carrier frequency. The energy spectrum has the same shape as that of the individual (constant-phase) pulses of which the signal is composed with no discrete line at the carrier, or premodulation, frequency. We can imagine trying to lock a PLL to such a signal. The output from the phase detector (PD) would rapidly change polarity, as the phase changed in 180° increments with equal rapidity, producing zero average output. Some process is needed that is able to ignore the phase transitions.

Figure 10.7 shows a carrier-recovery and demodulation loop for biphase signals (BPSK), called a Costas loop. The VCO and input signals are multiplied in two places, with one of the signals shifted 90° in one place. As a result, two quadrature outputs are obtained that, when multiplied together, produce a $\sin(2\Delta\phi)$ function. That has a typical shape for a phase detector output—the slope near zero output can be used to develop u_1 ,—but the phase is doubled. That increases the slope and, what is more important, makes the function insensitive to 180° changes:

$$\sin [2(\varphi + \Theta)] = \sin [2(\varphi + \Theta + \pi)]. \tag{10.12}$$

The usual process can be used for linear analysis, with K_p taken from the shape of the $\sin(2\varphi)$ curve.

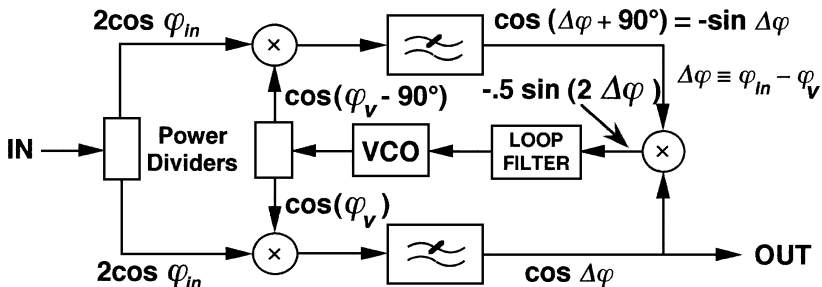


Fig. 10.7 Costas loop.

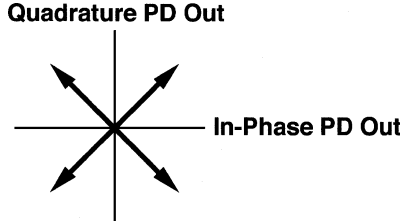


Fig. 10.8 QPSK.

Another advantage of this Costas loop is that it provides detection simultaneously with carrier recovery. The lower phase detector in Fig. 10.7 has in-phase inputs so it is a coherent detector, producing a maximum-amplitude positive or negative output.

10.4.2 N-Phase Costas Loop

In addition to two-phase signals, four- and eight-phase signals are also common. A four-phase (QPSK for quadrature phase shift keyed) signal is illustrated in Fig. 10.8. More complex versions of the Costas loop [Lindsey and Simon, 1973, pp. 74–75; Ziemer and Peterson, 1985; Gardner, 1979, Chapter 11] employing four or eight multiplications of the VCO by the input, rather than the two shown above, can be used to derive a carrier from these. One of these circuits is described in Appendix i10.B.

10.4.3 Multiply and Divide

Another method for carrier recovery from M -phase signals involves the multiplication of the signal frequency by M , as illustrated in Fig. 10.9. The allowed phases of the coded signal are

$$\varphi_{in} = \varphi + i2\pi/M, \tag{10.13}$$

where $0 \leq i \leq M - 1$. After multiplication by M , the phases are

$$\varphi_M = M\varphi_{in} = M\varphi + i2\pi \Rightarrow M\varphi, \tag{10.14}$$

regardless of which of the allowed states (i) exists. This must be filtered because, in practice, the signals lose amplitude, sometimes going through zero, as they transition



Fig. 10.9 Carrier recovery by multiplication with a PLL providing filtering.

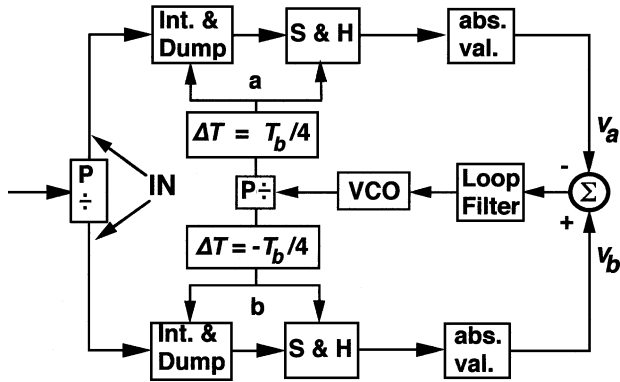


Fig. 10.10 Early-late gate bit synchronizer.

from one state to another. A narrow filter is needed to filter out this amplitude modulation, and possibly accompanying noise, and the PLL is often used for this purpose.

10.5 DATA SYNCHRONIZATION

In addition to the need for carrier recovery, recovery of the clock that synchronizes the data is also important. This process too involves various circuits to which basic PLL principles can be applied. [Gardner, 1979, Chapter 11; Spilker, 1977]. We will discuss some of them briefly.

10.5.1 Early-Late Gate Bit Synchronizer

An early-late gate bit synchronizer is shown in Fig. 10.10. Waveforms are shown in Fig. 10.11.

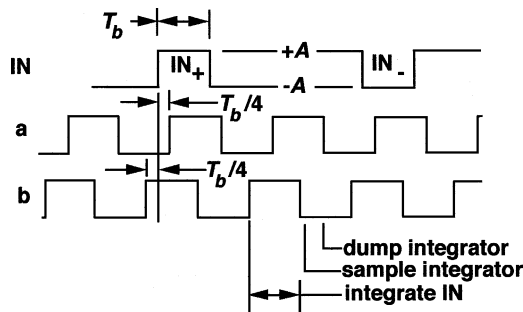


Fig. 10.11 Bit synchronizer waveforms.

The VCO output is split, and one path is delayed by one quarter of the bit period while the other is advanced by that much. The time-shifted pulses are used to control integrate-and-dump circuits. Integration of the input pulse train begins at the beginning of the delayed or advanced VCO pulse and stops at the end of that pulse. The integrator output is then sampled and held, after which the integrator is dumped, remaining in that state until the start of the next time-shifted pulse. The sample-and-hold (S & H) outputs are converted to their absolute values and then subtracted from each other. The values of the individual voltages, v_a and v_b , are shown in Fig 10.12 as a function of the delay of the incoming pulse relative the position shown in Fig. 10.11. The resulting average voltage into the loop filter equals $v_a - v_b$, which is plotted in Fig. 10.12. Either a positive or a negative pulse will produce an output of AT_b when shifted from the position in which they are shown in Fig. 10.11 by $T_b/4$. However, when there is no transition (no change in bit value), $v_a - v_b = 0$. Since there is a 50% probability of a transition, the average voltage into the loop filter is half of the value produced after a transition. Thus the loop filter output goes from $-0.5AT_b$ to $+0.5AT_b$ with a triangular characteristic, giving $K_p = 2AT_b/\text{cycle}$.

The PLL theory we have studied, especially as it pertains to triangular PD characteristic, applies to this loop. The quarter bit period delay and advance could be obtained by decoding a four-stage shift frequency divider driven by the VCO operating at four times the data rate.

10.5.2 Synchronizing to a Pseudorandom Bit Sequence

A pseudorandom bit sequence of length M has an autocorrelation function such as shown in Fig. 10.13. If the incoming signal is composed of repetitions of this sequence, we can synchronize to it using a method that has similarities to that which was employed for the simple code in the previous section. We do this by multiplying

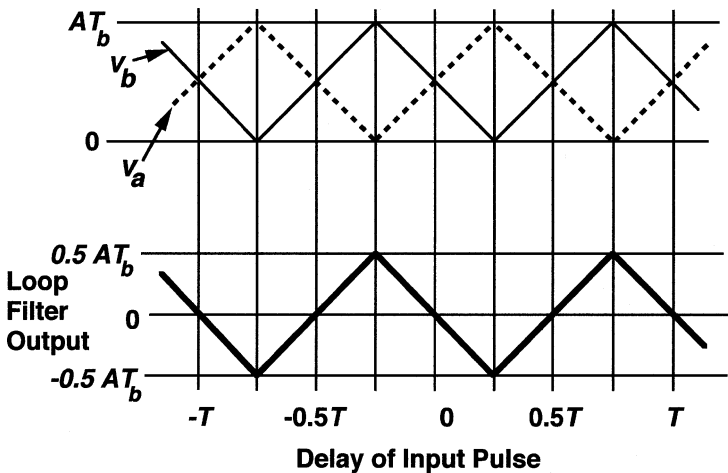


Fig. 10.12 PD characteristic and its components.

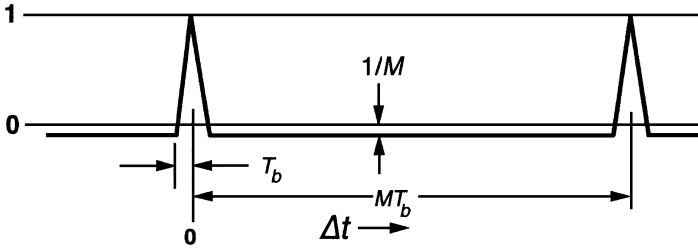


Fig. 10.13 Autocorrelation function of pseudorandom sequence.

the incoming code sequence by the identical sequence synchronized by the VCO and by the same sequence shifted by two bits and then subtracting the results. This is illustrated in Fig. 10.14.

In this “delay lock loop,” [Spilker, 1963; Spilker and Magill, 1961] the bit streams entering the multiplier have zero average value (e.g., ± 1). The circuit could also be realized using ExNOR gates and binary (0 and 1) signals. The average value of v_a varies with the time offset between the incoming code and the code generator with the same characteristic as the autocorrelation function. So does the average of v_b , but its characteristic is offset from the characteristic of v_a by two bit periods (smaller offsets are also used). These two characteristics are superimposed in Fig. 10.15, where v_b has also been given the minus sign it acquires at the summer. The resulting phase detector characteristic is shown in Fig. 10.16.

These diagrams have been normalized so 1 represents the voltage that appears at the loop filter output due to fully correlated (in phase) codes in one of the two correlators. The resulting phase detector gain (still so normalized) is

$$K_p = \frac{(M + 1/M) MT_b}{T_b \text{ cycle}} = \frac{M + 1}{\text{cycle}}. \tag{10.15}$$

Except in the narrow region between $\pm 2T_b$, the phase detector has zero output and there is no feedback. We then depend upon the difference between the incoming bit rate and the VCO frequency to move the operating point toward the region where the PD has gain. The bias voltage might be used as part of a strategy to ensure acquisition of lock in an acceptable time.

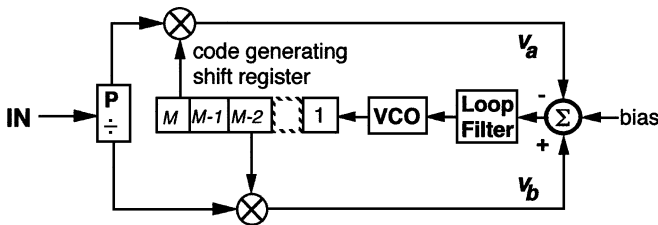


Fig. 10.14 Code synchronizing loop.

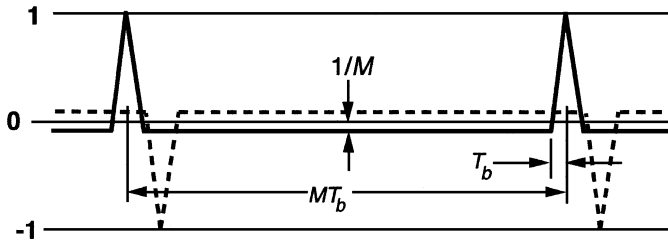


Fig. 10.15 Characteristics of v_a and $-v_b$ (normalized).

10.5.3 Delay-and-Multiply Synchronizer

Another method manipulates the data stream to create a spectral line at the data rate and then uses the PLL as a narrow tracking filter. This method is discussed in Appendix 10.C.

10.6 CLOCK AND DATA TIMING CONTROL

In Section 10.5 we studied the use of PLLs to recover the data clock from the data. Here we study the use of phase-locked loops (PLLs) and delay-locked loops (DLLs) to synchronize the phase of one clock to another and the use of a DLL to adjust data timing relative to a reference clock. Then we will see how a PLL can be combined with such a DLL, sharing its circuitry, to provide that reference clock by recovering it from the data.

10.6.1 Phase-Locked Loops

Phase-locked loops can be used to synchronize various copies of a system clock to a master clock in order to reduce skew (time offset) between clocks [Jeong et al., 1987]. Figure 10.17 shows a PLL where the signal at f_{out} is to be time (phase) synchronized with the signal at f_{ref} . The times T shown on the diagram refer to waveform

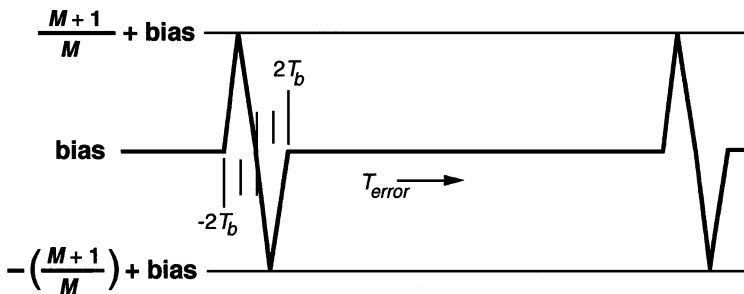


Fig. 10.16 Phase discriminator characteristic (normalized).

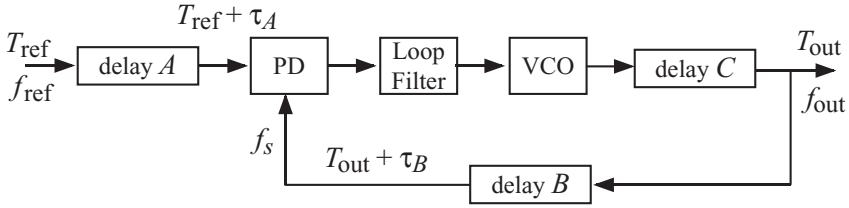


Fig. 10.17 PLL used to eliminate clock skew.

transitions—for example, positive-going edges of rectangular waveforms—that are to occur simultaneously in the synchronized clocks. Assume that the phase detector is of the charge-pump type (Section 3.1.3), which is common in such applications, and that it responds to those transitions. Then zero PD output will occur when the input waveforms are in phase—that is, synchronized such that those effective transitions occur simultaneously in the two PD inputs. If the loop is of type 2, zero phase error will occur in steady state and, as a result,

$$T_{\text{ref}} + \tau_A = T_{\text{out}} + \tau_B. \tag{10.16}$$

Thus, the synchronized clock at f_{out} will produce a transition at

$$T_{\text{out}} = T_{\text{ref}} + \tau_A - \tau_B, \tag{10.17}$$

and, if $\tau_A = \tau_B$, T_{out} will equal T_{ref} and the clock at f_{out} will be time synchronized with that at f_{ref} .

The delays, A–C, result both from propagation delays in the signal paths and from delays in amplifiers that are needed to raise the clocks to adequate power levels, as illustrated in Fig. 10.18 for Delay C. To accomplish $\tau_A = \tau_B$, the phase detector would be placed effectively at a midpoint between the VCO and the reference input, although that might not occur at a physical midpoint if there are different amplifying elements in the two path. Delay from phase detector to VCO can be included in Delay

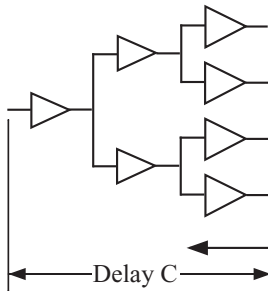


Fig. 10.18 An amplifier tree, a possible source of Delay C.

C. While Delays B and C are within the loop, they usually have no significant effect on its operation. A delay that might be a significant part of a cycle at f_{out} will usually be a small part of a cycle at the modulation frequencies of importance in the loop, and information does not have to pass rapidly within the loop in order to synchronize f_{out} with f_{ref} . See Section 5.3.5 concerning the representation of delays in stability analysis.

By including a frequency divider in the feedback path, an output can be generated at a multiple of the reference frequency. This changes the loop into a frequency synthesizer (Section 10.2.1). It is also possible to synthesize multi-phase clocks by logically combining the outputs from the clock at differing delays within Delay B or C or within the VCO [Young et al., 1992].

10.6.2 Delay-Locked Loops

10.6.2.1 For Clock Synchronization. Another circuit used to synchronize clocks is shown in Fig. 10.19 [Johnson and Hudson, 1988]. This circuit has been called a PLL although it has no VCO and is basically of Type 0, characteristics that do not apply to what are traditionally termed phase-locked loops. It seems to fall better into the category that has been termed delay-locked loops (DLLs). It is a control loop that adjusts a voltage-variable delay line to maintain minimum time (phase) error relative to a reference clock. In this case the object is to synchronize two clocks, both of which have been delayed, rather than to remove the delay in one clock so it matches an undelayed version. The mathematical block diagram is shown in Fig. 10.20a. Here $K_{pt} = K_p f_{ref}$ is the ratio of voltage change to time change and K_B is the ratio of change in τ_B to a change in the voltage that controls that delay.

Figure 10.20b is equivalent but in terms of phase rather than time. Given a fixed clock frequency, one implies the other:

$$\varphi = \tau f_{clock} \tag{10.18}$$

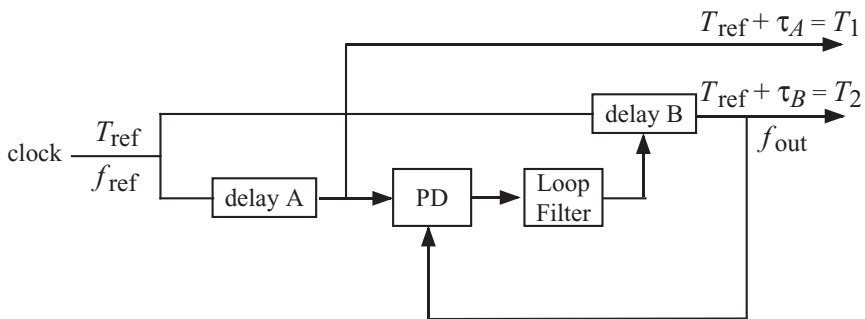


Fig. 10.19 Delay-control-loop.

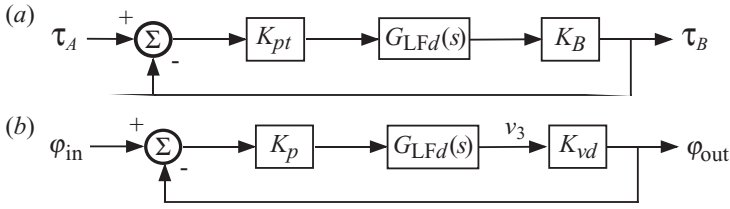


Fig. 10.20 Mathematical block diagrams for the delay-control-loop, in terms of (a) time (b) phase.

The delay-line gain is here given by

$$K_{vd} = d\phi_{out}/dv_3. \tag{10.19}$$

This loop has no inherent pole at the origin because there is no necessary conversion from frequency to phase, as is the case with a PLL. If $G_{LFd}(s) = K_{LF}$, this is a type 0 loop. Any change in either delay will be only partially compensated since a phase error must be developed to change the output from Delay B. If we include an integrator in the loop filter,

$$G_{LFd}(s) = K_{LF}F_d(s) = K_{LF}F(s)/s, \tag{10.20}$$

the diagram becomes as shown in Fig. 10.21. This has the same basic form as Fig. 4.1 from ϕ_{in} to ϕ_{out} and represents a loop of type 1 or higher. Therefore, the linear analyses of Chapters 2 through 7 apply).

10.6.2.2 For Data Synchronization. Delay-locked loops are also used to adjust the timing of data relative to a reference clock. Figure 10.22 shows such a circuit.

The phase detector here measures the offset between data transitions and the reference clock. Since transitions do not occur every clock cycle, this will be a special type of phase detector that responds to data transitions, such as the PD described in the bit synchronizer in Section 10.5.1 and those described by Lee (1992, Section IV-A). While the phase of the data is being adjusted here, we could as well consider that we are adjusting the underlying data clock, whose phase we must discern from the data. We can represent the circuit in Fig. 10.22 by the mathematical block diagram in Fig. 10.23.

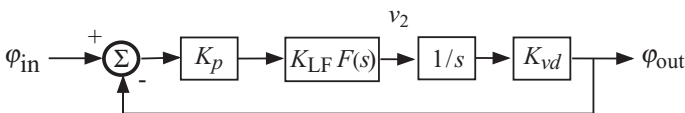


Fig. 10.21 Delay control loop with integrator.

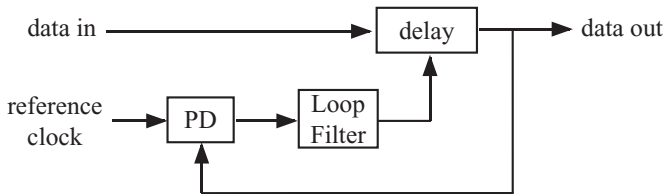


Fig. 10.22 Delay-locked loop for data synchronization.

Since we are interested in the response to changes in the input data and not to changes in the data clock, which we assume to be constant, we can set the latter to zero (change) and remove the summer, leaving Fig. 10.24.

Now, if we include an integrator in $G_{LFd}(s)$, as we did for the previous loop, we again have a loop as shown in Fig. 2.4 from φ_{in} to φ_{out} , with $G_{LFd}(s)$ given by Eq. 10.20, and we can again apply Chapters 3 through 7.

10.6.3 Combined Loops

It is possible to combine this DLL with a PLL that extracts the reference clock from the data. That configuration is analyzed in Appendix i10.T.

10.7 ALL-DIGITAL PHASE-LOCKED LOOP (ADPLL)

Another loop that is analyzable by the theory that we have developed is the all-digital PLL (ADPLL). Sometimes an ordinary PLL that uses a phase detector that is driven by, and provides, logic-level signals is called a digital PLL. However, we refer here to a loop that is truly digital. In fact, it has no continuously variable signal. State variables that were voltages or currents in analog loops are numbers in the ADPLL. Of course, the numbers are represented by voltages in a binary form, but the variables of interest are the represented numbers. Many different ADPLLs are discussed in the literature [Lindsey and Chie, 1981]. We will discuss two in order to show how the general principles of PLLs can be applied to ADPLLs. The many possible implementations of ADPLLs and various special considerations applicable to them provide opportunities for additional study that is beyond the scope of this book.

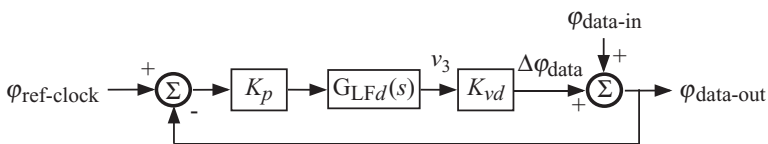


Fig. 10.23 Mathematical diagram for a delay-locked loop performing data synchronization.

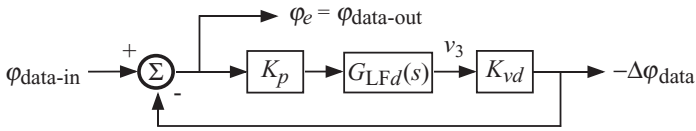


Fig. 10.24 A rearrangement of Fig. 10.23.

10.7.1 Basic Digital Implementation

10.7.1.1 The Loop. First consider the block diagram of a simple ADPLL shown in Fig. 10.25. It is labeled “too simple” for reasons we will discover, but it will serve as an introduction to the equivalencies that can be drawn between the digital blocks and those which we have studied as well as to some new potential problems.

All of the signals are numbers, represented electronically. The input is a binary clock of frequency f_{in} , and the output is to be a number that, in the manner of phase-locked loops, has an average repetition rate of f_{in} but follows the input with a closeness that depends on the loop parameters. The VCO has been replaced by an output accumulator (OA). Its output is a number that advances each clock cycle by an amount equal to its input N_2 . Each time the OA reaches its capacity N_v , it recycles to 0 ($N_v - 1$ being the highest number that can be attained). Thus one output cycle is represented by N_v and the output phase of the OA is

$$\varphi_{out} = (N_{out}/N_v) \text{ cycles.} \tag{10.21}$$

The output frequency is

$$f_{out} = \Delta\varphi_{out}/\Delta t = (N_2/N_v) f_{clock} \tag{10.22}$$

since the output is incremented by N_2 each cycle ($T = 1c/f_{clock}$) of the OA clock. If we consider N_2 to be the tuning signal, we can obtain the OA gain constant from (10.22) as

$$K_v = df_{out}/dN_2 = f_{clock}/N_v. \tag{10.23}$$

The register stores the value of N_{out} , representing the loop’s output phase, each cycle of the input signal. Thus the stored number represents the output phase at the last

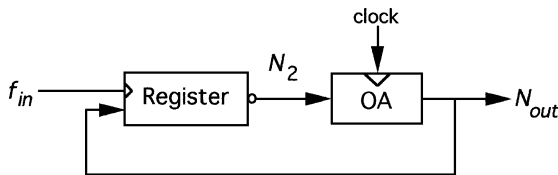


Fig. 10.25 Too simple ADPLL.

transition of f_{in} , and this can be considered the phase difference between the output and input signals. The register thus functions as a phase detector (and zero-order hold), but we need a phase inversion. One simple way to do this is to make N_2 the complement of the sampled value of N_{out} (represented by the small circle at the output of the register). During the n th input cycle (following the n th input clock transition), we have

$$N_2 = -N_{out} (\varphi_{in} = 2\pi n). \quad (10.24)$$

The phase detector gain constant is

$$K_p = -\frac{\partial N_2}{\partial \varphi_{out}} = -\frac{\partial N_2}{\partial N_{out}} \frac{\partial N_{out}}{\partial \varphi_{out}} = -(-1) \frac{N_v}{1 \text{ cycle}} = N_v \text{ cycle}^{-1}. \quad (10.25)$$

The forward gain constant is

$$K = K_p K_v = f_{\text{clock}} \text{ cycle}^{-1}. \quad (10.26)$$

We have the essential parts of a PLL, an OA whose output phase advances at a rate proportional to the control signal that is the output from a phase detector/register. The mathematical block diagram is given by Fig. 2.4 with $K_{LF} = 1$ and φ_{out} , K_p and K_v defined above. For this first-order loop, $u_1 = u_2 = N_2$.

Let us consider what the output looks like as the frequency changes; it will be affected by its discrete or sampled nature. For small values of N_2 , the output frequency will advance with increasing N_2 , beginning with $f_{out} = f_{\text{clock}}/N_v$ when $N_2 = 1$. For small N_2 it will be a relatively smoothly increasing sawtooth value. As N_2 increases, the number of output changes between overflows will eventually become small so the output will no longer resemble a smoothly increasing phase. At $N_2 = N_v/2$, the output will change by half of the capacity of the OA each input period and will appear to be a square wave. At higher values of N_2 , the output frequency will be negative, as demonstrated by the slope of the phase change with time, and decrease until, at the maximum value of $N_2 = N_v - 1$, it will appear to be again at a frequency of f_{clock}/N_v , but with a negative sign. These appearances are due to the fact that the advancing value of N_{out} is sampled at a frequency of f_{clock} .

10.7.1.2 Sampling and Stability. There are two sampling processes occurring in the simple loop, one in the OA at f_{clock} and one in the register at f_{in} . To make the OA operate like a VCO that has approximately a continuous phase change at its output, it is necessary that

$$f_{\text{clock}} \gg f_{in}, \quad (10.27)$$

since f_{in} , is the frequency of the OA's output during lock. If this inequality is not met, the output will appear to be a few steps rather than a ramp of phase. Suppose f_{in} should have a discrete phase change of $\Delta\varphi$ between two of the transitions that clock

the register. At the first input sample after this step, a frequency change will occur at N_{out} equal to

$$\Delta f_{\text{out}} = K \Delta \varphi = f_{\text{clock}} \Delta \varphi \text{ cycle}^{-1}. \quad (10.28)$$

This results in a phase change at the next input sample, which occurs 1 cycle/ f_{in} later, of

$$\Delta \varphi' = -\Delta f_{\text{out}} \text{ cycle}/f_{\text{in}} = -\Delta \varphi f_{\text{clock}}/f_{\text{in}}. \quad (10.29)$$

So the phase error grows by $-f_{\text{clock}}/f_{\text{in}}$ each input sample and, according to (10.27), this is a substantial increase each time. This “first-order loop” is unstable due to the sampling process. It repeatedly overshoots the final value by an amount much larger than the previous error, causing a growing oscillation. To solve this dilemma, let us divide N_2 by 2^q by shifting bits between the register and the OA, connecting 2^i from the register to 2^{i-q} at the OA input. Now we have reduced the gain by 2^q , but we have also reduced the maximum value of N_2 by the same amount so now

$$N_2 \leq (N_v - 1)2^{-q} \quad (10.30)$$

so the highest frequency we can accommodate will be

$$f_{\text{in}} = f_{\text{out}} \leq f_{\text{clock}} \frac{N_v - 1}{N_v} 2^{-q} \approx 2^{-q} f_{\text{clock}}. \quad (10.31)$$

The ratio of phase change magnitudes between adjacent samples that could be obtained from (10.29) is now, due to our gain change of 2^{-q} ,

$$-\frac{\Delta \varphi'}{\Delta \varphi} = 2^{-q} \frac{f_{\text{clock}}}{f_{\text{in}}} > 1, \quad (10.32)$$

where unity is approached at maximum f_{in} . Unity is desired in Eq. (10.32), since the output phase will then follow the input step in one input sample period. Instability occurs when $-\Delta \varphi'/\Delta \varphi = 2$, since the magnitude of the error in $\Delta \varphi$ will then increase each input sample. Thus we are constrained to operate in a narrow range of frequencies, depending on the value of f_{clock} and q . We could alleviate this restriction by obtaining N_2 by adding a constant to the PD output so the value of N_2 would not be tied so directly to f_{in} . Then Eqs. (10.30) and (10.31) would not hold, and we could lower the gain by operating at higher input frequencies than allowed by (10.31). With all of these changes, the block diagram would look like Fig. 10.26.

We can see these stability problems more succinctly from the z -transform representation of the loop. Again we will assume Eq. (10.27) so we can simplify analysis by ignoring the faster sampling in the OA. The mathematical block diagram with the sampler shown is Fig. 10.27, where the blocks are represented in terms of Laplace

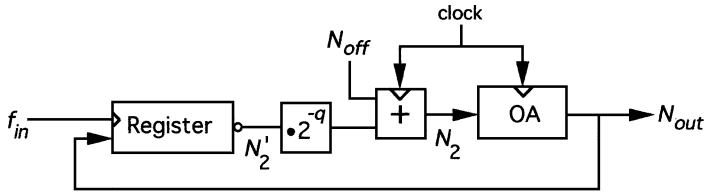


Fig. 10.26 Workable simple ADPLL.

transforms. Now, however, Eq. (10.25) has changed to

$$K_p = N_v 2^{-q} \text{ cycle}^{-1} \tag{10.33}$$

so the forward transfer function is

$$G(s) = \frac{1 - e^{-Ts}}{s^2} f_{\text{clock}} 2^{-q} \text{ cycle}^{-1}, \tag{10.34}$$

in Laplace terms, and the corresponding z -transform is

$$G(z) = 2^{-q} \frac{f_{\text{clock}}}{\text{cycle}} \frac{(1 - z^{-1})Tz^{-1}}{(1 - z^{-1})^2} = 2^{-q} \frac{f_{\text{clock}}}{f_{\text{in}}} \frac{1}{z - 1}. \tag{10.35}$$

Given the position of the sampler, the z -transform of the closed-loop response is

$$H(z) = \frac{G(z)}{1 + G(z)} = \frac{C}{z - 1 + C}, \tag{10.36}$$

where

$$C = 2^{-q} \frac{f_{\text{clock}}}{f_{\text{in}}}. \tag{10.37}$$

The open- and closed-loop poles for this loop are shown on the z -plane in Fig. 10.28. The closed-loop pole locus begins at the open-loop pole when $C = 0$ and moves along the real axis as C increases. The best response will be in the center of the unit circle, where $C = 1$ and oscillation occurs when the locus reaches the unit circle and enters the region of instability outside of it, at $C = 2$. This corresponds to what we

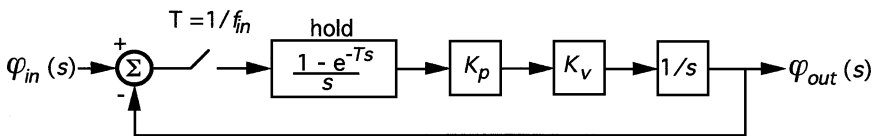


Fig. 10.27 Mathematical diagram of ADPLL in Fig. 10.26.

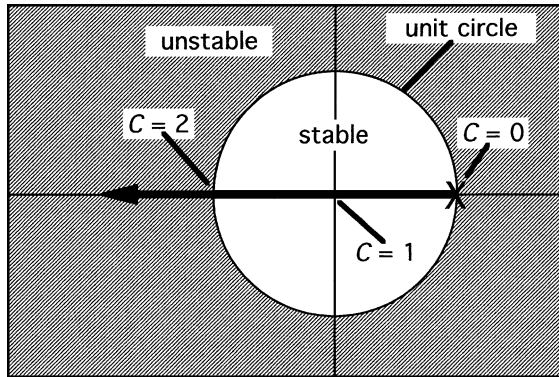


Fig. 10.28 Z-plane representation of the loop.

ascertained using the simple analysis above. Without an offset added to the PD output, we were restricted to operating where $C > 1$, in the left half plane only.

10.7.1.3 Choice of Values

Example 10.1 First-Order ADPLL, A Particular Implementation Given: $N_v = 2^{12} = 4096$; $f_{\text{clock}} = 8.192 \text{ MHz}$; $f_{\text{in}} = 400 \text{ kHz}$. Find: (a) a good value for q ; (b) N_2 ; (c) N_{off} if required; (d) ω_L ; (e) OA output range.

- a. We would like to have $C = 1$. Since $f_{\text{clock}}/f_{\text{in}} \approx 21$, we consider $q = 4$ and 5 for $2^{-q} = \frac{1}{16}$ and $\frac{1}{32}$, respectively. If we use the former, $C = 1.31$, which is not far from optimum.
- b. From Eq. (10.22),

$$f_{\text{in}} = f_{\text{out}} = 4 \times 10^5 \text{ Hz} = (N_2/4096)8.192 \times 10^6 \text{ Hz} = N_2 \times 2 \text{ kHz},$$

$$N_2 = 200,$$

$$N'_2 = 16N_2 = 3200.$$

- c. No offset is required since $N'_2 = 3200 < N_v$.
- d. For this first-order loop,

$$\omega_L = K = 2^{-4} f_{\text{clock}} \text{ cycle}^{-1} = 5.12 \times 10^5 \text{ sec}^{-1} \Rightarrow f_L = 81.49 \text{ kHz}.$$

- e. Since N_2 can vary from 0 to $\text{INT}(4095/16) = 255$, Eq. (10.17) also shows that the possible range of f_{out} is from 0 to

$$f_{\text{out, max}} = (255/4096)8.192 \text{ MHz} = 510 \text{ kHz}.$$

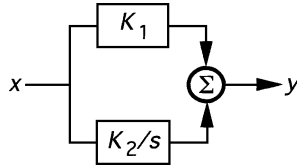


Fig. 10.29 Digital integrator- and -lead filter.

Also note in the example that N_2 can be constant if f_{in} is a power-of-two multiple of the step size, 2 kHz, but, for other frequencies, it will cycle between two numbers such that the average output frequency, which cycles between two adjacent multiples of 2 kHz, equals the input frequency. This is an example of jitter or ripple, an effect that is due to quantization, the fact that the signals are quantized rather than continuously variable. False locks, wherein the input and output frequencies are related by a ratio of small numbers, are another unwelcome possibility in sampled systems. [Egan, 2000, pp. 426–434]. The program described in Appendix i10.M.2 should be helpful for exploring some of these effects.

10.7.1.4 Higher-Order Loops We have shown a simple loop with no loop filter, but loop filters can be implemented digitally. Figure 10.29 shows a simple integrator-and-lead filter. The $1/s$ integration block can be realized with an accumulator. The values of K_1 and K_2 can be adjusted by how the digits are shifted in the two legs. We can write the resulting transfer function as

$$\frac{y}{x}(s) = K_1 + \frac{K_2}{s} \tag{10.38}$$

$$= \frac{K_2(1 + sK_1/K_2)}{s} \tag{10.39}$$

This is the same general form as Eq. (3.20) with $\omega_z = K_2/K_1$.

With this filter we should ensure that a number in the middle of the output range of the PD is interpreted as zero, lower numbers being considered negative. This will allow steady state, where the accumulator output is fixed, to occur in the middle of the PD range. Such DC shifts do not affect mathematical blocks but can be necessary or advisable in various implementations.

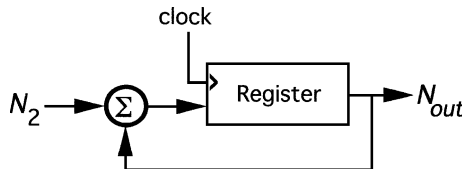


Fig. 10.30 Output accumulator.

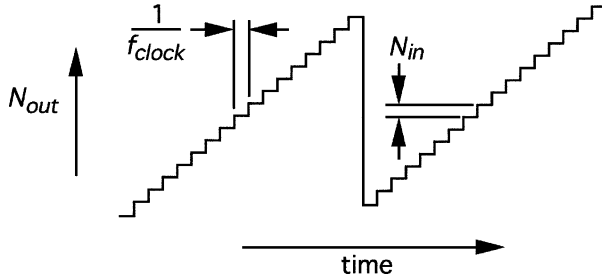


Fig. 10.31 OA output.

10.7.2 OA, NCO, DDS

The OA is shown in more detail in Fig. 10.30. The control input, N_2 , is added (numerically) to N_{out} . Each clock cycle the sum is entered into the register. (We assume that delays or some other mechanism prevents the register output from changing before the new input has been completely entered into the register. This is the kind of practical detail that will not be addressed here.) Figure 10.31 shows how N_{out} advances each clock cycle until it recycles when the register’s capacity has been exceeded. The OA is an accumulator, a combination of memory and adder, that we have designated OA to differentiate it from accumulators that may be used elsewhere in the ADPLL.

Adding a read-only memory (ROM) to the OA, as in Fig. 10.32, produces a numerically controlled oscillator (NCO).² The ROM produces the sine (or cosine) of the phase represented by the output of the OA, which provides the address to the ROM. The output N'_{out} is now a numerical representation of a sinusoid.

Adding a digital-to-analog (DA) converter and a low-pass filter produces a direct digital synthesizer (DDS). The DA converts the numerical sine wave to a stepped voltage representation of the sine wave, and the low-pass filter rejects all but the fundamental component, so we finally arrive at a sine voltage V_{out} such as would be produced by a VCO.

The ADPLL with the OA output N_{out} fed back to the PD performs like an analog loop with a sawtooth phase detector whereas a ADPLL in which the NCO output N'_{out} is fed back performs more like an analog loop with a sinusoidal phase detector. Their linear small signal representations are the same, but their large signal performances differ. In some cases the sampling type of PD may not be appropriate (e.g., when the input is not a simple square wave but a sinusoid, to which we must lock, plus noise or other signals). We can achieve phase detection that is similar to analog multiplication (see Section 3.1.4.3) by multiplying N'_{out} by a digital representation of an analog input signal and passing the result through a digital low-pass filter. We also could use V_{out} with an analog phase detector and convert the phase detector’s output to a digital signal using an AD, thus producing a hybrid digital/analog PLL.

We might use the transition of the most significant bit (MSB) of N_{out} or N'_{out} as a clock edge. (This suggests a different set of PLLs in which phase detection

² NCO is a common term whereas OA is not.

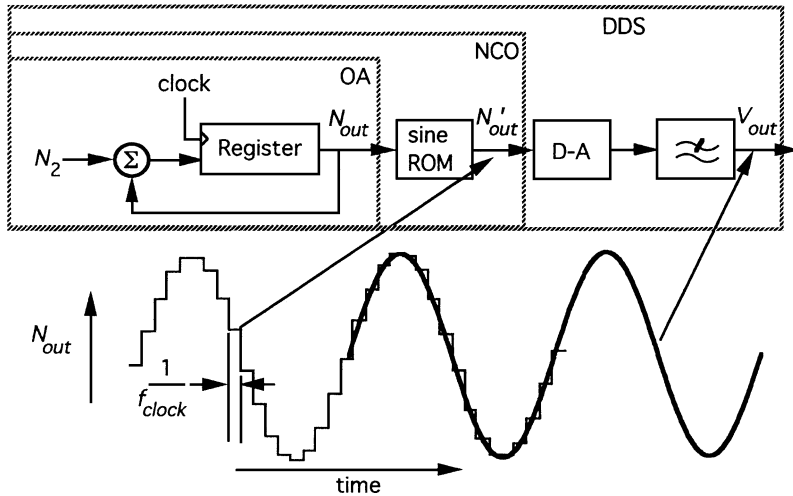


Fig. 10.32 OA, NCO, DDS, and waveforms.

is accomplished by sampling an appropriate input waveform at a time determined by the output waveform.) However, it has an inherent jitter, with a total range of one clock cycle. To understand this peculiarity of the sampled nature of the digital representation, let us compare the transitions of the most significant bit to a clock edge that is obtained by processing V_{out} through a limiter to produce a rectangular clock waveform, $V'_{out} = \text{sign}(V_{out})$.

If $N_2 = 2^i$, where i is a whole number, each cycle of N_{out} and N'_{out} will be the same as each other cycle, and there will be no jitter. If we add a small number ϵ to N_2 for one clock cycle only, it will change all subsequent values of N_{out} by ϵ , but it will not necessarily affect the transition time of the MSB. If it does, it will shift it by $(1/f_{clock}) \gg 1/\epsilon$. Transition of the MSB indicates that N_v has been reached sometime in the previous sample period, but the information as to just when is contained in other bits. However, while the MSB transition moves by zero or $1/f_{clock}$, the transition time of V'_{out} will change by ϵ/N_v cycles, as is appropriate, because V_{out} is affected by all of the bits in N'_{out} .

If $N_2 \neq 2^i$, all cycles will not be identical. Some cycles of the most significant bits of N_{out} and N'_{out} will be one sample period longer than others (this is jitter or “ripple”). The cycles of V_{out} and V'_{out} , however, will all be of the same duration.

10.7.3 Implementing an ADPLL by Pulse Addition and Removal

A second ADPLL implementation is shown in Fig. 10.33. This diagram corresponds to an available integrated circuit (IC).³ We will study it to improve our ability to recognize the correspondence between the fundamentals we have studied and structures that look

³ All the blocks in Fig. 10.25 are available in the TI SN74LS297 IC excepting the $\div N$, which must be supplied externally. A flip-flop phase detector is also included. See the data sheet but notice some difference in symbols from what we use here.

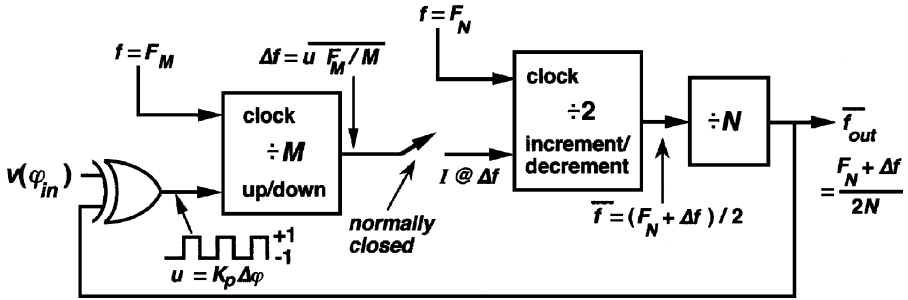


Fig. 10.33 ADPLL using pulse removal.

much different at first observation. Additional description and analysis, including for a second-order loop, is in Appendix 10.D. More information can be obtained from the IC data sheet and from an application note [Troha, 1994].

The VCO function is provided by removing pulses from, or adding pulses to, a fixed-frequency clock and dividing the result by a large number so the jitter is relatively small. (Adjacent cycles are approximately of the same duration because a small number of pulses, relative to the total, have been removed.)

We begin our examination at the switch, which is open for purposes of discussion only. The signal to the right of the switch we call I . This is a series of pulses with value 1 or -1 and an average net frequency of Δf (number of positive pulses less number of negative pulses divided by time). The subsequent $\div 2$ circuit divides the clock at frequency F_N by 2 but, when it receives an increment (+1) or decrement (-1) command at I , it adds one clock edge or subtracts one, respectively. Thus the average frequency from the $\div 2$ circuit is

$$\bar{f} = \frac{F_N + \Delta f}{2}. \tag{10.40}$$

The output frequency is this value divided by N :

$$\bar{f}_{out} = \frac{F_N + \Delta f}{2N} \tag{10.41}$$

10.7.3.1 Transfer Function The phase detector we are using is an Exclusive-OR gate (as in Section 3.1.2); we will call its output u and define it as true = 1 or false = -1 . When its inputs are exactly in phase, its output is -1 . As the relative phase of its inputs shifts, its average output becomes more positive until it equals $+1$ when the inputs are exactly out of phase. During the half-cycle change in the relative phase of its inputs, its average output changes proportionately from -1 to $+1$ and could be represented by the triangle characteristic in Fig. 3.2 with those limits. Thus the phase detector gain constant is

$$K_p = 2/(0.5 \text{ cycle}) = 4/\text{cycle}. \tag{10.42}$$

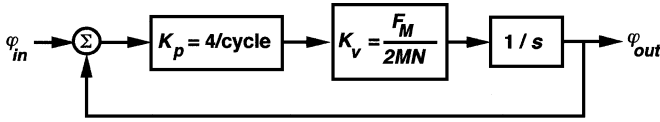


Fig. 10.34 ADPLL, mathematical block diagram.

When the PD output is +1 the $\div M$ circuit counts up; when it is -1 the circuit counts down. The output frequency from the $\div M$ therefore has an average value of

$$\Delta f = u F_M / M. \tag{10.43}$$

The gain of the VCO block can be obtained from Eq. (10.41) and (10.43) as

$$K_v = \frac{\overline{df_{out}}}{du} = \frac{F_M}{2MN}. \tag{10.44}$$

A block diagram using these constants is shown in Fig. 10.34. The $1/s$ block, as usual, converts frequency to phase.

Figure 10.34 represents a first-order loop with unity-gain bandwidth of

$$\omega_L = K = [2F_M / (MN)] \text{ rad/cycle}. \tag{10.45}$$

Transfer functions are given by Eqs. (2.25b) and (2.26b). Transient responses are as given in Table 2.1.

10.7.3.2 Tuning Range The hold-in range can be seen from Fig. 10.33 and Eqs. (10.41) and (10.43) to be

$$F_H = \pm F_M / (2MN), \tag{10.46}$$

the limits corresponding to +1 or -1 from the phase detector.

From Eq. (10.41) the center frequency can be seen to be

$$f_c = F_N / (2N). \tag{10.47}$$

Thus both the center frequency and the bandwidth can be controlled independently by means of external clocks (F_M and F_N) or by setting divider ratios (M and N).

10.8 SUMMARY

There are many other circuits and software functions to which these principles can be applied. Some types of PLLs require further considerations (e.g., control of jitter) that are peculiar to those types, but this chapter has illustrated how the theory that we have studied can be used to obtain a basic understanding of diverse circuits that employ the common properties of phase-locked loops.

10.A APPENDIX: EXACT ANALYSIS OF A SPECIAL-CASE THIRD-ORDER LOOP

In this appendix we will analyze a special case of a third-order loop that is particularly valuable when a type 2 loop is desired, but an additional significant high-frequency pole is also needed. As suggested in Section 10.2.1, this is often the case for frequency synthesizers. Two possible loop filters for this third-order loop are shown in Fig. 10.A.1.

The transfer function for either filter is

$$-K_{LF}F(s) = \frac{1}{R_1 C_1} \frac{1 + \frac{s}{\omega_z}}{s \left(1 + \frac{s}{\omega_p}\right)}. \quad (10.A.1)$$

For Fig. 10.A.1a, the zero and pole are at

$$\omega_z = \frac{1}{R_2 C_1} \quad (10.A.2)$$

and

$$\omega_p = \frac{1}{R_{p2} C_{p2}} \quad (10.A.3)$$

while for Fig. 10.A.1b, they are at

$$\omega_z = \frac{1}{R_2(C_1 + C_2)} \quad (10.A.4)$$

and

$$\omega_p = \frac{1}{R_2 C_2}. \quad (10.A.5)$$

A loop in which unity open-loop gain occurs midway between the filter zero and the second (finite) pole on a log plot is of special significance. As will be shown, it gives

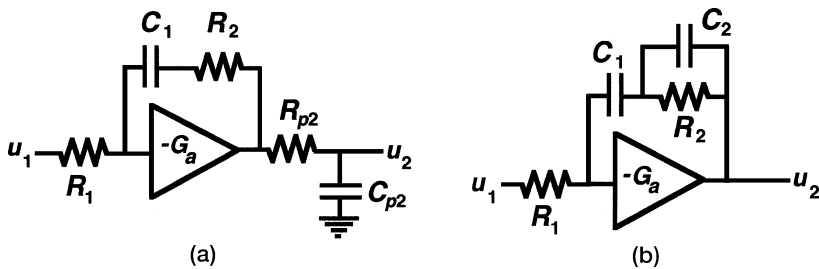


Fig. 10.A.1 Two possible loop filters for a third-order loop.

maximum phase margin for any given ratio of pole to zero frequency. Of course, we could increase the phase margin by moving the pole frequency higher or the zero frequency lower. However, there are good reasons to not do either. Increasing the pole frequency decreases the rejection of undesired outputs from the phase detector; that is particularly important in frequency synthesizers. Decreasing the zero frequency can have several undesired effects.

To lower the zero frequency, a larger capacitor is required in the loop filter and that can present a problem if space is limited. In addition, lowering the zero frequency, while holding ω_L constant, decreases the low-frequency gain of the loop and thus the suppression of VCO noise at those frequencies (see Section 12.2). We also can expect slower response at lower zero frequencies. We can see this trend in the type-2 loop that results from ignoring the pole at ω_p . For example, from Fig. 6.6, the $\zeta = 4$ curve reaches a relative error of $10^{-3.4}$ at $t = 30/\omega_n$ whereas the $\zeta = 2$ curve reaches the same level at $t = 20.2/\omega_n$. Using Eqs. (6.14) and (4.16), we find that, for approximately the same ω_L , an ω_z decrease of 4 has produced a time increase of 3. Thus we would often like to have the zero occur as high in frequency as possible. Therefore a good compromise can often be obtained by placing unity open-loop gain at the point of maximum phase margin, midway (on a log scale) between the zero frequency and the second pole frequency, as shown in Fig. 10.A.2.

We set

$$\frac{\omega_p}{\omega_L} = \frac{\omega_L}{\omega_z} = r, \quad (10.A.6)$$

where r^2 was defined in Eq. (8.61). Figure 10.A.3 shows phase margin versus r . Zero phase margin occurs at $r = 1$, where the zero and pole are coincident, since they then cancel leaving a pure integrator. Fortunately, we can obtain the transient response for this loop with reasonable ease.

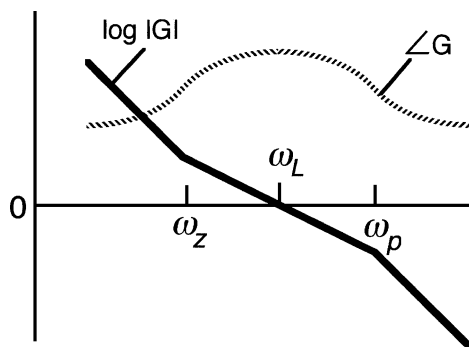


Fig. 10.A.2 Gain and phase of a type 2, third-order loop.

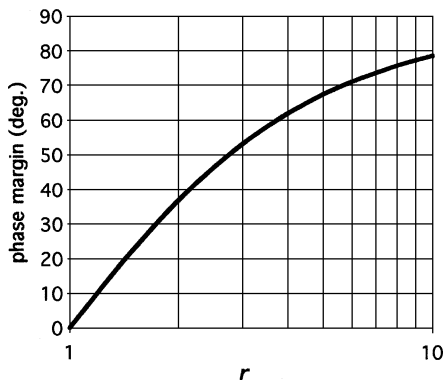


Fig. 10.A.3 Phase margin vs. r .

10.A.1 Loop Response

The open-loop transfer function

$$G(s) = \frac{\omega_L^2}{r} \frac{1 + \frac{s}{\omega_L/r}}{s^2 [1 + s/(\omega_L r)]} = \omega_L^2 r \frac{s + \omega_L/r}{s^2 (s + \omega_L r)} \tag{10.A.7}$$

has two poles at zero frequency and a pole and a zero equally spaced, on a log scale, about the unity-gain frequency ω_L , as desired. The corresponding closed-loop transfer function is

$$H(s) = \frac{G(s)}{1 + G(s)} = \omega_L^2 r \frac{s + \omega_L/r}{s^3 + s^2 \omega_L r + s \omega_L^2 r + \omega_L^3} \tag{10.A.8}$$

We can factor $(s + \omega_L)$ out of the denominator,⁴ giving

$$\frac{G(s)}{1 + G(s)} = \omega_L^2 r \frac{s + \omega_L/r}{(s + \omega_L) [s^2 + \omega_L(r - 1)s + \omega_L^2]}. \tag{10.A.9}$$

This can be written as a partial-fraction expansion [Truxal, 1955, p. 14],

$$\frac{G(s)}{1 + G(s)} = \frac{a}{s + \omega_L} + f(s), \tag{10.A.10}$$

⁴ This is shown by Encinas (1993), pp. 30–41.

where a and $f(s)$ are to be determined. We can solve for a using a common method wherein we multiply both sides by $(s + \omega_L)$ and then set s equal to $-\omega_L$, giving

$$a = \omega_L^2 r \frac{-\omega_L + \omega_L/r}{\omega_L^2 + \omega_L^2(1-r) + \omega_L^2} = \omega_L \frac{r-1}{r-3}. \quad (10.A.11)$$

Therefore, (10.A.10) can be written

$$\frac{G(s)}{1+G(s)} = \omega_L \frac{r-1}{r-3} \frac{1}{s+\omega_L} + f(s). \quad (10.A.12)$$

By equating (10.A.12) to (10.A.9), we can obtain

$$f(s) = \frac{\omega_L^2 r (s + \omega_L/r) - \omega(r-1/r-3) [s^2 + \omega_L(r-1)s + \omega_L^2]}{(s + \omega_L) [s^2 + \omega_L(r-1)s + \omega_L^2]} \quad (10.A.13)$$

$$= -\frac{\omega_L}{r-3} \frac{(r-1)s^2 + (r+1)\omega_L s + 2\omega_L^2}{(s + \omega_L) [s^2 + \omega_L(r-1)s + \omega_L^2]} \quad (10.A.14)$$

$$= -\frac{\omega_L}{r-3} \frac{(r-1)s + 2\omega_L}{s^2 + \omega_L(r-1)s + \omega_L^2}. \quad (10.A.15)$$

This part of the response is second order with standard parameters

$$\omega_n = \omega_L \triangleq \omega_{n2} \quad (10.A.16)$$

and

$$\zeta = (r-1)/2 \triangleq \zeta_2. \quad (10.A.17)$$

We add the subscript 2 to emphasize that these parameters apply only to the part of the response represented by $f(s)$. The overall response [as in Eq. (10.A.8)] is third-order and therefore such second-order parameters do not apply to it. After inserting (10.A.15) into (10.A.12), the closed-loop transfer function can be written in terms of these parameters as

$$H(s) = \frac{G(s)}{1+G(s)} = \frac{\omega_{n2}}{\zeta_2 - 1} \left[\frac{\zeta_2}{s + \omega_{n2}} - \frac{\zeta_2 s + \omega_{n2}}{s^2 + \omega_{n2}(2\zeta_2)s + \omega_{n2}^2} \right]. \quad (10.A.18)$$

We write this as the sum of two responses,

$$H(s) = \frac{\zeta_2}{\zeta_2 - 1} H_1(s) - \frac{1}{\zeta_2 - 1} H_2(s), \quad (10.A.19)$$

where

$$H_1(s) = \frac{\omega_{n2}}{s + \omega_{n2}} \quad (10.A.20)$$

is the response of a first-order loop (or of a low-pass filter—see Section 2.2.1 or Table 2.1) and

$$H_2(s) = \omega_{n2} \frac{\zeta_2 s + \omega_{n2}}{s^2 + \omega_{n2}(2\zeta_2)s + \omega_{n2}^2} \quad (10.A.21)$$

is the response of a second-order loop with $\alpha = 0.5$ [see Eq. (6.4)]. The responses, for various types of inputs (e.g., step, ramp), to this third-order loop can thus be obtained by combining the responses for the two parts, which are available in Table 2.1 and in Chapter 6.

The corresponding error response is

$$1 - H(s) = \frac{1}{1 + G(s)} = \frac{\zeta_2}{\zeta_2 - 1} [1 - H_1(s)] - \frac{1}{\zeta_2 - 1} [1 - H_2(s)], \quad (10.A.22)$$

where

$$1 - H_1(s) = \frac{s}{s + \omega_{n2}} \quad (10.A.23)$$

and

$$1 - H_2(s) = \frac{s(s + \omega_{n2}\zeta_2)}{s^2 + \omega_{n2}(s\zeta_2) + \omega_{n2}^2}. \quad (10.A.24)$$

This is the error response of a first-order loop [or a high-pass filter—see Eq. (2.12a)] plus the error response of a second-order loop with $\alpha = 0.5$. Again, this information is available in Table 2.1 and in Chapter 6. Some of the third-order responses that are more likely to be of value in synthesizer design are provided below.

10.A.2 Final Values

Combining Eq. (10.A.22) to (10.A.24), we can obtain

$$1 - H(s) = \frac{s^2 [s + \omega_{n2}(2\zeta_2 + 1)]}{(s + \omega_{n2})(s^2 + 2\zeta_2\omega_{n2}s + \omega_{n2}^2)} \quad (10.A.25)$$

with a limit for small s of

$$\lim_{s \rightarrow 0} [1 - H(s)] = \frac{2\zeta_2 + 1}{\omega_{n2}^2} s^2 = \frac{r}{\omega_{n2}^2} s^2. \quad (10.A.26)$$

By the final value theorem, when the loop is driven by $1/s^n$, the error response approaches a final value of

$$\lim_{s \rightarrow 0} \frac{s[1 - H(s)]}{s^n} = \frac{r}{\omega_{n2}^2} s^{3-n}. \quad (10.A.27)$$

This indicates that the phase error resulting from a phase step ($n = 1$) or a frequency step ($n = 2$) will settle to zero, and the phase error will not reach steady state for $n > 3$. A frequency ramp ($n = 3$) will produce a steady-state phase error of r/ω_L^2 which is also $1/(\omega_z \omega_L)$, the same phase error that occurs in a second-order type 2 loop with $\omega_z \leq \omega_L$ (as is true here). This is not surprising since one way to realize the third-order loop is to follow the integrator-and-lead of Fig. 3.23c with a passive low-pass filter, as in Fig. 10.A.1. The output of a low-pass filter driven by a ramp is another ramp with the same slope, offset from the input by a constant voltage.

10.A.3 Triple Roots

For $\zeta_2 = 1$ ($r = 3$), Eq. (10.A.25) becomes

$$[1 - H(s)]_{\zeta=1} = \frac{s^2(s + 3\omega_n)}{(s + \omega_n)^3}. \quad (10.A.28)$$

The step response can be obtained by partial fraction expansion:

$$\frac{1}{s}[1 - H(s)]_{\zeta=1} = \frac{s(s + 3\omega_n)}{(s + \omega_n)^3} = \frac{1}{(s + \omega_n)} + \frac{\omega_n}{(s + \omega_n)^2} - \frac{2\omega_n^2}{(s + \omega_n)^3}. \quad (10.A.29)$$

From this the time response to a unit step can be written in terms of normalized time,

$$x \equiv \omega_n t, \quad (10.A.30)$$

as

$$Y_{u, \text{error}}(x) = e^{-x}(1 + x - x^2). \quad (10.A.31)$$

10.A.4 Step Response

The step response corresponding to Eq. (10.A.24) is given by Eq. (6.9) or (6.11) with $\alpha = 0.5$. [Eq. (6.10) corresponds to the triple-root case that is covered by Eq. (10.A.31)]. With this we combine, as in Eq. (10.A.22), the high-pass response

corresponding to (10.A.23), to obtain

$$Y_{u, \text{error}}(t) = \frac{\zeta_2}{\zeta_2 - 1} e^{-\omega_L t} - \frac{1}{\zeta_2 - 1} e^{\zeta_2 \omega_L t} \left[\cosh \left(\omega_L t \sqrt{\zeta_2^2 - 1} \right) \right] \quad (10.A.32a)$$

$$= \frac{r-1}{r-3} e^{-\omega_L t} - \frac{2}{r-3} e^{-[(r-1)/2]\omega_L t} \left[\cosh \left(\omega_L t \sqrt{\left(\frac{r-1}{2}\right)^2 - 1} \right) \right] \quad (10.A.32b)$$

for $\zeta_2 > 1 (r > 3)$, or

$$Y_{u, \text{error}}(t) = \frac{\zeta_2}{\zeta_2 - 1} e^{-\omega_L t} - \frac{1}{\zeta_2 - 1} e^{-\zeta_2 \omega_L t} \left[\cos \left(\omega_L t \sqrt{1 - \zeta_2^2} \right) \right] \quad (10.A.33a)$$

$$= \frac{r-1}{r-3} e^{-\omega_L t} - \frac{2}{r-3} e^{-[(r-1)/2]\omega_L t} \left[\cos \left(\omega_L t \sqrt{1 - \left(\frac{r-1}{2}\right)^2} \right) \right] \quad (10.A.33b)$$

for $\zeta_2 < 1 (r > 3)$.

These step responses are shown in Fig. 10.A.4. The envelope for large times is given in Fig. 10.A.5. Figure 10.A.6 gives the error response to a ramp, which can be used to compute phase error during a phase ramp, that is, a frequency step. The ramp response was obtained by numerically integrating the step response.

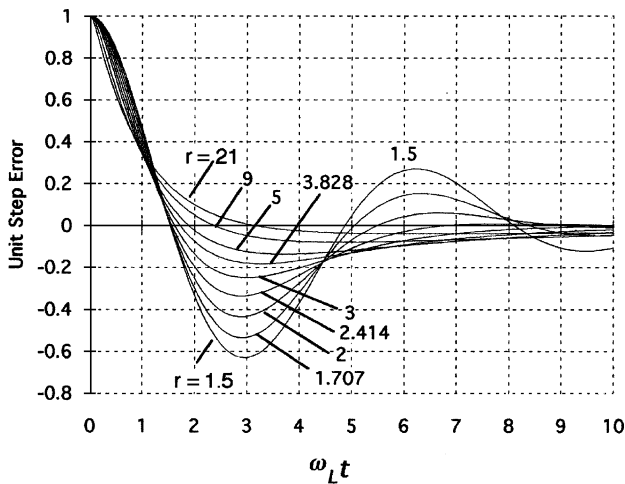


Fig. 10.A.4 Error response to unit step, third-order loop with ω_z, ω_p symmetrical about ω_L .

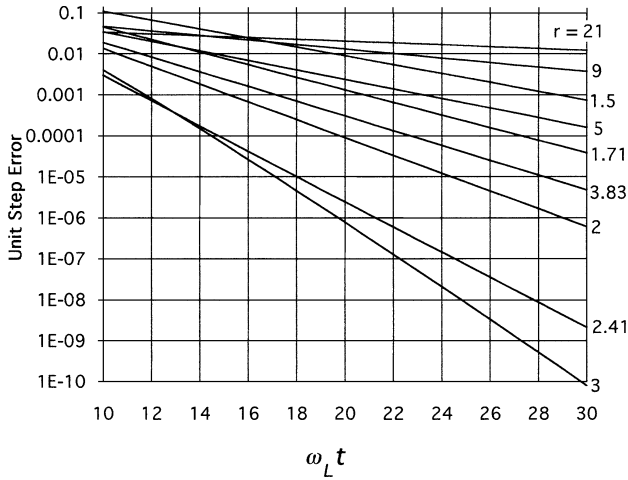


Fig. 10.A.5 Log of the envelope of the error, third-order loop with ω_z, ω_p symmetrical about ω_L .

Example 10.A.1

Problem: What is the frequency error $40\mu\text{sec}$ after a commanded frequency change of 2 MHz if the loop has unity gain at 10 kHz and a loop filter that is an integrator plus a zero at 4142 Hz and a pole at 24.14 kHz? What is the error at $400\mu\text{sec}$? Use the method of combining first- and second-order responses and compare results to

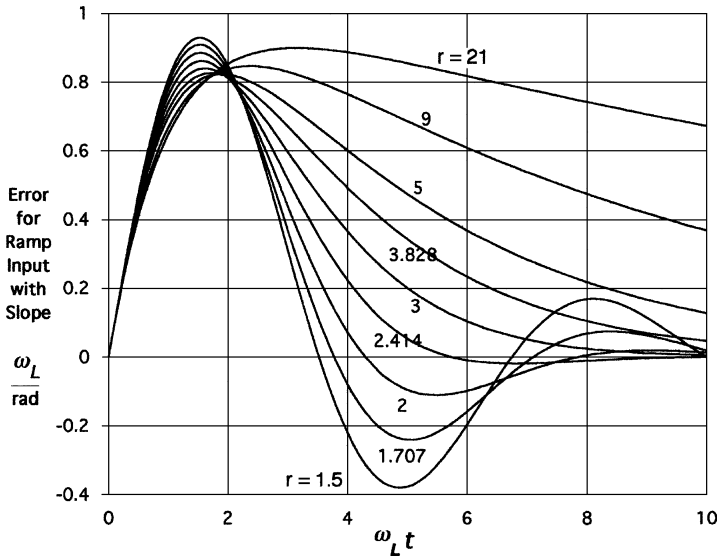


Fig. 10.A.6 Error response to a ramp with slope ω_L/rad , third-order loop with ω_z, ω_p symmetrical about ω_L .

those given by Fig. 10.A.4 or 10.A.5. (This will verify the equivalence and demonstrate how other third-order responses, for example, to parabolic inputs, could be obtained.)

Solution: The ratio of f_p to f_L is $r = 2.414$, which is also the ratio of f_L to f_z (which is, of course, the way we designed the loop). From Eq. (10.A.16), $f_{n2} = f_L = 10$ kHz. From (10.A.17), $\zeta_2 = 0.707$. The error response from Eq. (10.A.22) is

$$[1 - H(s)] = 2 \text{ MHz}\{-2.413[1 - H_1(s)] + 3.413[1 - H_2(s)]\},$$

so the related time functions are

$$f_{\text{error}}(t) = 2 \text{ MHz}\{-2.413 f_{\text{error},1}(t) + 3.413 f_{\text{error},2}(t)\}.$$

$1 - H_1(s)$ from Eq. (10.A.23) is the impulse response of a high-pass filter. At

$$\omega_n t = 2\pi 10^4 (40 \times 10^{-6}) = 2.51$$

this first-order part of the response to a 2-MHz step is

$$f_{\text{error},1} = 2 \text{ MHz}\{-2.413 \exp(-2.51)\} = -0.395 \text{ MHz}.$$

The second-order part of the response can be found from the upper and lower curves in Fig. 6.3 at $\omega_n t = 2.51$ (but we will use corresponding equations for better accuracy). The error response to a unit step for $\alpha = 1$ and $\zeta = 0.707$ is 0.0132 and for $\alpha = 0$ is -0.200 so the error response, for $\alpha = 0.5$, is the average of these or -0.034 . The second part of the response thus equals

$$f_{\text{error},2} = 2 \text{ MHz}\{3.413(-0.034)\} = -0.235 \text{ MHz}$$

at $40 \mu\text{sec}$. The sum of the two parts is

$$f_{\text{error}} = -0.395 \text{ MHz} - 0.235 \text{ MHz} = -0.63 \text{ MHz}.$$

Figure 10.A.4 gives an error for a unit step of -0.314 at $\omega_{n2}t = 2.51$. Multiplying by the 2-MHz excitation we again get -0.63 MHz.

If we repeat the process at $400 \mu\text{sec}$, the first-order part of the error after a unit step is -3×10^{-11} . From Fig. 6.5, the envelope of the error response to a unit step at $\omega_n t = 25.1$, when $\zeta = 0.707$ and $\alpha = 0.5$, is $10^{-7.7} = 1.96 \times 10^{-8}$. Multiplying by 3.413 we obtain 6.7×10^{-8} , which is obviously much larger than the first-order part and essentially equals the total error envelop. This is also the answer obtained from Fig. 10.A.5. Multiplying by 2 MHz, we obtain an error

$$f_{\text{error}}(400\mu \text{ sec}) \leq 0.13 \text{ Hz}.$$

Example 10.A.2

Problem: What is the peak phase error in a frequency synthesizer that steps from 120 to 150 MHz if $r = 2.4$, unity open-loop gain occurs at 5 kHz, and the reference frequency is 100 kHz [assuming Eq.(10.A.6)]?

Solution: Refer to Figs. 10.1 and 10.2. During the transient, $N = 150 \text{ MHz}/100 \text{ kHz} = 1500$ so the frequency step at the output of the “VCO” is $\Delta f = d\varphi/dt = (150 \text{ MHz} - 120 \text{ MHz})/1500 = 20 \text{ kHz} = 2 \times 10^4 \text{ cycles/sec}$.

By Fig. 10.A.6, a ramp with slope ω_L/rad will produce a peak phase error of 0.84 for $r = 2.4$. For this loop, $\omega_L = 5 \times 10^3 \text{ Hz}(2\pi \text{ rad/cycle}) = 31,400 \text{ rad/sec}$. Thus a phase ramp of $\omega_L/\text{rad} = 31,400/\text{sec}$ produces a peak phase error of 0.84. We have a step of $2 \times 10^4 \text{ cycles/sec}$, so it will produce a peak phase error of

$$\varphi_{\text{peak}} = 0.84 (2 \times 10^4 \text{ cycles/sec}) / (31,400/\text{sec}) = 0.54 \text{ cycle}.$$

This is larger than the linear range of a sinusoidal PD but is within the range of some types of PDs that are typically used in frequency synthesizers.

10.A.5 Modulation Response

As in Chapter 7, we will write the frequency response in terms of a normalized variable. Here we will use

$$\Omega_L \triangleq \omega/\omega_L \quad (10.A.34)$$

and the parameter for the individual curves will be r rather than ζ . Equation (10.A.7) can be written in these terms as

$$G(\Omega_L) = -\frac{j\Omega_L r + 1}{\Omega_L^2(j\Omega_L + r)}, \quad (10.A.35)$$

leading to expressions for gain and phase:

$$|G(\Omega_L)|^2 = \frac{\Omega_L^2 r^2 + 1}{\Omega_L^4(\Omega_L^2 + r^2)}, \quad (10.A.36)$$

$$\angle G(\Omega_L) = \tan^{-1}(\Omega_L r) - \tan^{-1}(\Omega_L/r) - 180^\circ. \quad (10.A.37)$$

These are shown in Fig. 10.A.7, where the optimum nature of this function, in terms of phase margin, is apparent.

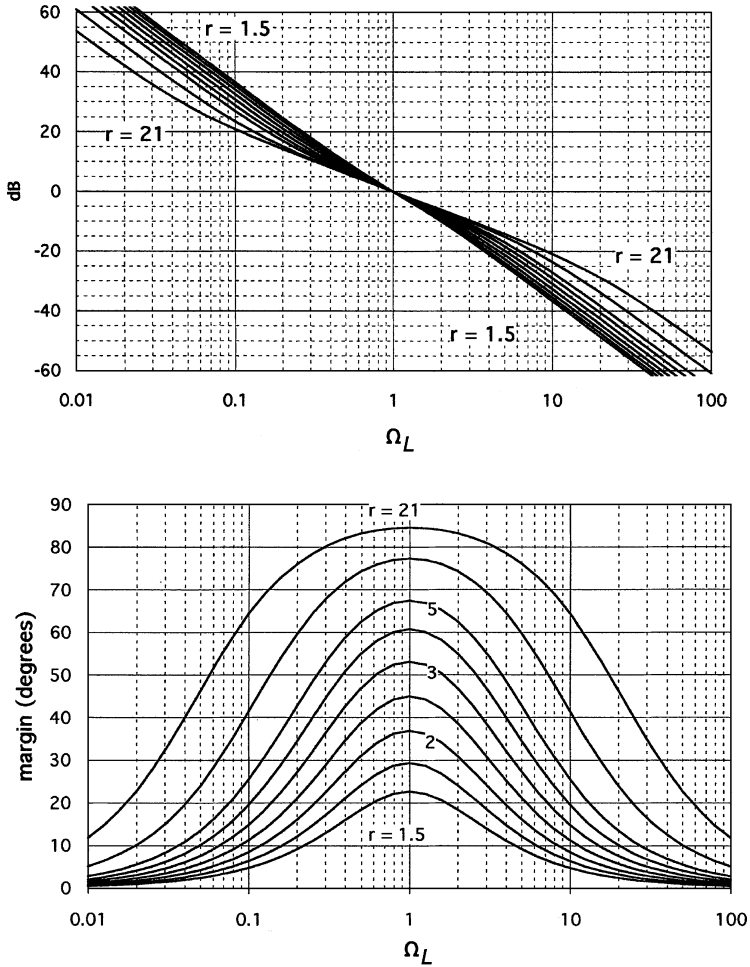


Fig. 10.A.7 Open-loop gain and phase for a third-order loop with ω_z, ω_p symmetrical about ω_L . $r = \omega_p/\omega_L = 1.5, 1.707, 2, 2.414, 3, 3.828, 5, 9, \text{ and } 21$.

Similarly, the closed-loop responses can be developed from Eq. (10.A.8):

$$H(\Omega_L) = \frac{j\Omega_L r + 1}{-j\Omega_L^3 - \Omega_L^2 r + j\Omega_L r + 1}, \tag{10.A.38}$$

$$|H(\Omega_L)|^2 = \frac{(\Omega_L r)^2 + 1}{\Omega_L^2(\Omega_L^2 - r)^2 + (\Omega_L^2 r - 1)^2}, \tag{10.A.39}$$

$$\angle H(\Omega_L) = \tan^{-1}(\Omega_L r) - \tan^{-1} \left[\frac{\Omega_L(\Omega_L^2 - r)}{\Omega_L^2 r - 1} \right]. \tag{10.A.40}$$

These relationships are shown in Fig. 10.A.8.

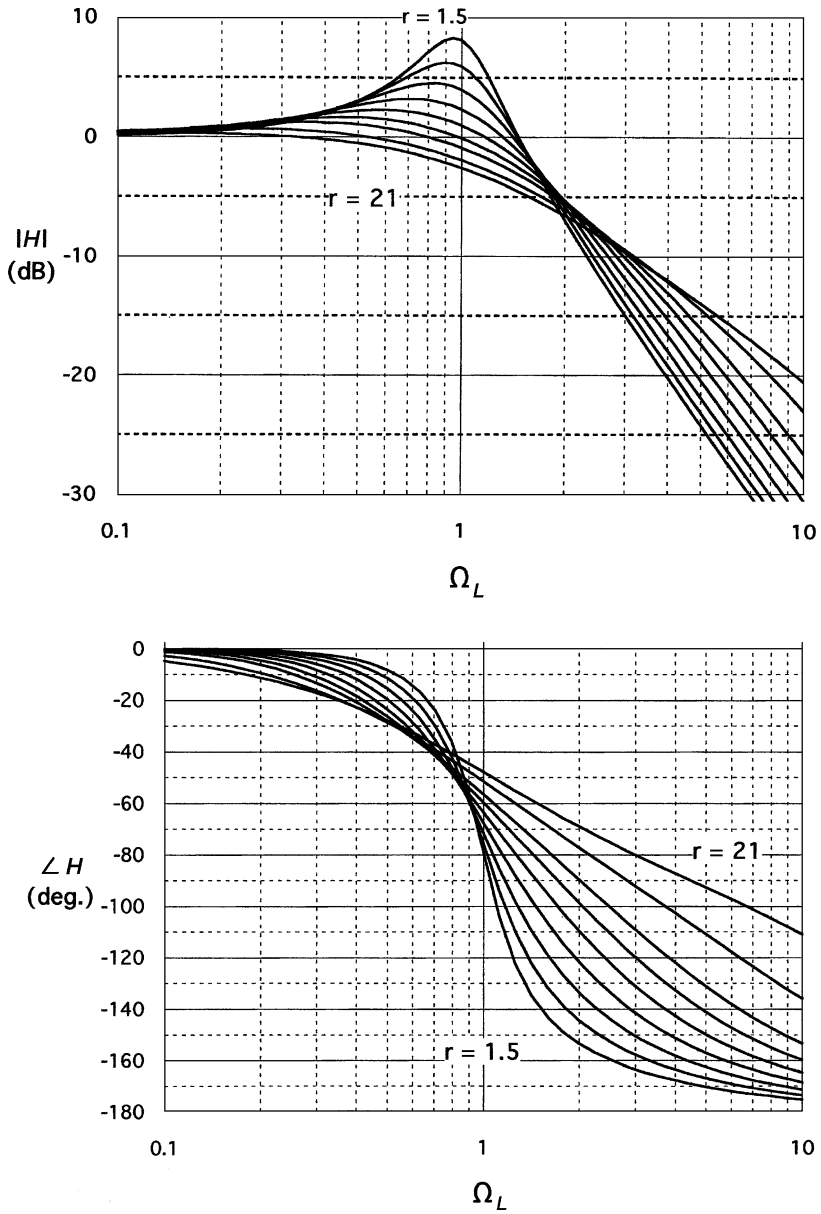


Fig. 10.A.8 Magnitude and phase of $H(\Omega)$ for a third-order loop with ω_z, ω_p symmetrical about ω_L . $r = \omega_p/\omega_L = 1.5, 1.707, 2, 2.414, 3, 3.828, 5, 9,$ and 21 .

The magnitude and phase of the error response can be obtained from Eq. (10.A.38):

$$1 - H(\Omega_L) = \frac{j\Omega_L^3 + \Omega_L^2 r}{+j\Omega_L^3 + \Omega_L^2 r - j\Omega_L r - 1}, \quad (10.A.41)$$

$$|1 - H(\Omega_L)|^2 = \frac{\Omega_L^4(\Omega_L^2 + r^2)}{\Omega_L^2(\Omega_L^2 - r)^2 + (\Omega_L^2 r - 1)^2}, \quad (10.A.42)$$

$$\angle[1 - H(\Omega_L)] = \tan^{-1}\left(\frac{\Omega_L}{r}\right) - \tan^{-1}\left[\frac{\Omega_L(\Omega_L^2 - r)}{\Omega_L^2 r - 1}\right]. \quad (10.A.43)$$

In the second term in Eqs. (10.A.40) and (10.A.43), the ambiguity in the quadrant of the \tan^{-1} must be solved so the value of the \tan^{-1} is always increasing.

We can easily show, by comparing Eqs. (10.A.39) and (10.A.42) that

$$|1 - H(\Omega_L)|^2 = \left|H\left(\frac{1}{\Omega_L}\right)\right|^2. \quad (10.A.44)$$

Thus we can use Fig. 10.A.8 to give the error also by taking the value from a frequency equal to the reciprocal of the frequency of interest. (For example, for $r = 2$, $\Omega_L = 0.5$, Fig. 10.A.8 shows $|H| \approx 2.7$ dB. This will also be the value of $|1 - H|$ at $\Omega_L = 2$.)

Similarly, we obtain

$$\angle H(\Omega_L) = -\angle[1 - H(1/\Omega_L)]. \quad (10.A.45)$$

Thus we can obtain the phase of $1 - H(\Omega_L)$ by reading the phase of $H(1/\Omega_L)$ from Fig. 10.A.8.

i10.B APPENDIX: COSTAS LOOP FOR N PHASES

This appendix is available from the Wiley Internet site.

i10.C APPENDIX: SYMBOL CLOCK RECOVERY

This appendix is available from the Wiley Internet site.

i10.D APPENDIX: ADPLL BY PULSE ADDITION AND REMOVAL, ADDITIONAL MATERIAL

This appendix is available from the Wiley Internet site.

10.F APPENDIX: FRACTIONAL- N AND SIGMA-DELTA

As long as a frequency synthesizer's minimum step size equals its reference frequency f_{in} , we may be forced to use a reference frequency that is too low, one that restricts the maximum loop bandwidth or causes an excessive divide ratio N (some noise sources are effectively multiplied by N). The fractional- N synthesizer provides a solution to this dilemma. In it, N need not be an integer, so the step size can be smaller than f_{in} . If we can change from $N = 100.123$ to $N = 100.124$, for example, the step size becomes $f_{in}/1000$. Here the VCO is effectively divided by a fractional number. This is achieved by alternating between integer divide ratios that are just smaller and just greater than the desired fractional number. For example, dividing by 1000, except by 1001 every tenth time, produces a fractional divide number of 1000.1, on the average.

If the only advantage sought is a small value of N , the loop might be made narrow enough to filter out the frequency variations that naturally result from changing N periodically. However, if fractional- N is being used to achieve wide loop bandwidth, another way must be found to counter the resulting frequency modulation. Some synthesizers use the known sequence of N -values, with a digital-to-analog converter, to create a waveform that is added to the phase-detector output to cancel the undesired modulation. This does not change the average frequency, only the deviations from it. However, there is a practical limit to the accuracy of this process and, therefore, to the suppression of the resulting FM sidebands.

More recently, designers have resorted to sigma-delta modulators [Schreier and Temes, 2005] to create patterns of N -values that cause the resulting FM to be noiselike. (These N -values typically extend beyond the two nearest integer values, which would be used for fractional- N .) Moreover, the sigma-delta modulator shapes the noise so it appears predominantly at higher modulation frequencies where it can be more effectively filtered by the loop. The goal is to reduce the phase modulation and resulting spectral sidebands to a level that is acceptable for the application. One reason that the significantly increased complexity can be acceptable is that its implementation is completely digital. [De Muer and Steyaert (2003), pp. 168–220]

i10.M APPENDIX: MATLAB[®] SIMULATIONS

This appendix describes the use of MATLAB to simulate higher-order loops, producing responses such as those in Figs. 10.A.4, 10.A.6, 10.A.7, and 10.A.8, and

to obtain closed-loop responses given open-loop parameters. Simulation of the ADPLL in Fig. 10.26 is also discussed. It is available from the Wiley Internet site.

i10.T APPENDIX: COMBINED PLL AND DLL

This appendix is available from the Wiley Internet site at ftp://ftp.wiley.com/public/sci_tech_med/phase_lock.

PART 2

PHASE LOCK IN NOISE

CHAPTER 11

PHASE MODULATION BY NOISE

We have previously studied how the phase-locked loop (PLL) responds to sinusoidal phase modulation of the input signal. This chapter will again consider the response of the PLL to phase modulation but, this time, the modulation will be noiselike.

11.1 REPRESENTATION OF NOISE MODULATION

Figure 11.1a shows a noise waveform as a function of time. The random variable, y , could represent voltage v , current i , phase φ , or frequency f . Variable y is described by its probability density, $\text{Prob}(y)$, which is a Gaussian function of y with zero mean (otherwise the mean is removed from y and considered to be a second variable) and with a variance, or mean square value, σ_y^2 (Fig. 11.1b). If $y = v$ or if $y = i$, and a 1- Ω system is assumed for convenience, σ_y^2 is power. Mean square phase σ_y^2 and mean square frequency σ_y^2 can also be called power by analogy, as we shall see.

If the noise waveform passes through a narrow filter (Fig. 11.2), the output waveform, as viewed on an oscilloscope, will appear to be a sine wave at the frequency of the filter passband. The narrower the filter, the smaller will be the spread, or fuzziness, of the waveform. Also, the narrower the filter, the more slowly will be the amplitude and phase change. This is because changes in the time domain imply modulation sidebands in the frequency domain. These are separated from the spectral center by the modulation frequency, and a narrow filter allows only a small spread of modulation sidebands and so inhibits high-frequency variations of amplitude or phase (frequency).

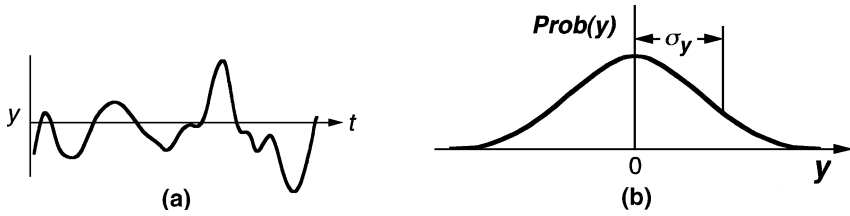


Fig. 11.1 Noise: (a) as a function of time and (b) probability density.

As the bandwidth is narrowed, the amplitude and phase of the signal in that band change ever more slowly. Thus, for any given duration of observation, a filter bandwidth can be chosen that is narrow enough that the amplitude and phase, both of which are unknown except by their statistics, remain essentially constant. Thus, the random waveform can be represented by a continuum of adjacent passbands (e.g., ideal rectangular contiguous passbands), which are sufficiently narrow that the power in each can be considered to represent a single sine wave. Since these sine waves have differing frequencies and unknown phase, we do not know how they combine. We know only that their mean square values, “powers,” add; that is, values of σ^2 add, as is usual for random variables. This is an important point. We cannot add volts, amperes, radians, or hertz. We can add volts squared, amperes squared, radians squared, and hertz squared. We refer to these squared variables by the generalized name “powers.”

If the center frequency of the filter in Fig. 11.2 is moved or scanned, as in a spectrum analyzer, the average power coming through the filter might be a function of the filter center frequency such as is illustrated in Fig. 11.3. If adjacent filters pass the same mean power, we expect twice as much power from a filter that covers the total band of two individual filters, three times the power from a filter that covers three times the band, and so forth. Thus, over any frequency range where the power is independent of the frequency, we can express the power in a filter bandwidth δf at frequency f' as

$$P_y(f') = \delta f S'_y(f'), \tag{11.1}$$

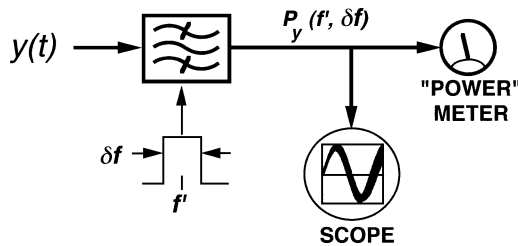


Fig. 11.2 Noise through a narrow filter.

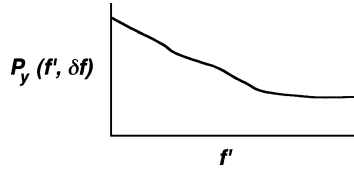


Fig. 11.3 Frequency distribution of noise power in bandwidth δf vs. f .

where $S'_y(f')$ is a constant. But, unless $S'_y(f')$ is flat in the region of interest, its value will change¹ somewhat as δf changes. However, the relative change will be smaller as δf becomes smaller [barring discontinuities in $S'_y(f')$]. Thus, in the limit where δf becomes differential, $S'_y(f')$ becomes a constant $S_y(f')$ and is called the power spectral density of y at frequency f' :

$$S_y(f') \equiv \lim_{\delta f \rightarrow 0} \frac{P_y(f')}{\delta f} = \lim_{\delta f \rightarrow 0} \frac{\mathbf{E}[y^2(f' - \delta f/2 < f < f' + \delta f/2)]}{\delta f}. \tag{11.2}$$

Here $\mathbf{E}[y^2(\cdot)]$ is the expected value of y^2 in the frequency band indicated. Then the power, or mean square value, of y , $\mathbf{E}(y^2) = \sigma_y^2$, over any frequency range can be obtained by summing the differential contributions:

$$\mathbf{E}(y^2)|_{f_1}^{f_2} = \int_{f_1}^{f_2} S_y(f) df. \tag{11.3}$$

If $y = \varphi$, then Eq. (11.3) gives the variance of phase represented by the phase power spectra density² (PPSD) $S_\varphi(f_m)$ over a range of modulation frequencies from $f_m = f_1$ to $f_m = f_2$,

$$\sigma_\varphi^2|_{f_1}^{f_2} = \mathbf{E}(\varphi^2)|_{f_1}^{f_2} = \int_{f_1}^{f_2} S_\varphi(f_m) df_m. \tag{11.4}$$

We can describe frequency modulation by a frequency power spectral density in the same manner as is done for phase modulation. The relationship between frequency deviation at a frequency f_m and phase modulation at the same frequency [see Eq. (7.7)] leads directly to the relationship between the two densities at the same modulation frequency:

$$\frac{S_f(f_m)}{S_\varphi(f_m)} = \frac{S_f(f_m)\delta f_m}{S_\varphi(f_m)\delta f_m} = \frac{[\Delta f(f_m)]^2}{[\Delta \varphi(f_m)]^2} = f_m^2. \tag{11.5a}$$

¹ When $S_y \sim f$ there is no error due to large δf if S_y is taken in the middle of δf .

² The power spectral density $S_y(f)$ is also the Fourier transform of the autocorrelation function of $y(t)$.

We obtain the variance of frequency deviation in a manner similar to Eq. (11.4) and relate it to S_φ by Eq. (11.5a):

$$\sigma_f^2|_{f_1}^{f_2} = \mathbf{E}(f^2)|_{f_1}^{f_2} = \int_{f_1}^{f_2} S_f(f_m) df_m = \int_{f_1}^{f_2} f_m^2 S_\varphi(f_m) df_m. \quad (11.6a)$$

Similarly, in different units,

$$\frac{S_\omega(f_m)}{S_\varphi(f_m)} = \frac{S_\omega(\omega_m)}{S_\varphi(\omega_m)} = \omega_m^2. \quad (11.5b)$$

and

$$\sigma_\omega^2|_{f_1}^{f_2} = \int_{f_1}^{f_2} \omega_m^2 S_\varphi(f_m) df_m = \frac{\text{cycle}}{2\pi} \int_{\omega_1}^{\omega_2} \omega_m^2 S_\varphi(\omega_m) d\omega_m. \quad (11.6b)$$

Note that S_φ is in radians squared per hertz (rad^2/Hz) in each of the last four equations, which can be seen from consideration of the units.

11.2 PROCESSING OF NOISE MODULATION BY THE PHASE-LOCKED LOOP

The output and input phase can be related by Eq. (4.3) written for $s = j\omega$:

$$\varphi_{\text{out}}(\omega) = H(\omega)\varphi_{\text{in}}(\omega). \quad (11.7)$$

We write the mean squared value of the output phase as

$$\sigma^2 [\varphi_{\text{out}}(\omega)] = \sigma^2 [H(\omega)\varphi_{\text{in}}(\omega)] = |H(\omega)|^2 \sigma^2 [\varphi_{\text{in}}(\omega)]. \quad (11.8)$$

Here $|H(\omega)|^2$ can be brought outside because it is not a function of time, the variable over which the squared value is being averaged.

Since noise modulation can be represented as modulation by an infinite set of sine waves, each at its own frequency f' and having power $S(f')df$, the response of the loop to each of these differential oscillations is the same as its response to a discrete sinusoidal modulation at the same frequency. Thus, in response to an input phase power spectral density $S_{\varphi,\text{in}}(f')$, the output phase power spectral density will be

$$S_{\varphi,\text{out}}(f') = S_{\varphi,\text{in}}(f')|H(f')|^2. \quad (11.9)$$

Similarly, the phase error $S_{\varphi,e}(f')$ will be described by a phase power spectral density

$$S_{\varphi,e}(f') = S_{\varphi,\text{in}}(f')|1 - H(f')|^2. \quad (11.10)$$

Figure 11.4 illustrates the reaction of the loop to a limited band of white phase noise at the input according to Eqs. (11.9) and (11.10). It is important that σ_e^2 , the integral of $S_{\varphi,e}(f_m)$, be small enough to permit linearity to be assumed at the phase detector (e.g., $\sigma_e^2 \ll 1 \text{ rad}^2$).

Of course, as in the case of discrete signals, the same relationships hold for frequency modulation:

$$S_{f,\text{out}}(f') = S_{f,\text{in}}(f')|H(f')|^2, \quad (11.11)$$

$$S_{f,e}(f') = S_{f,\text{in}}(f')|1 - H(f')|^2. \quad (11.12)$$

11.3 PHASE AND FREQUENCY VARIANCE

The phase power, or variance, of the output, or of the error, in response to an input phase modulation, or to modulation originating at the voltage-controlled oscillator (VCO), is obtained by integrating the corresponding values of S_φ over the frequency range of interest. Often this range covers all frequencies, limited by the response of the circuit or the fall-off of the input modulation. Since both $S(f)$ and $H(f)$ are often

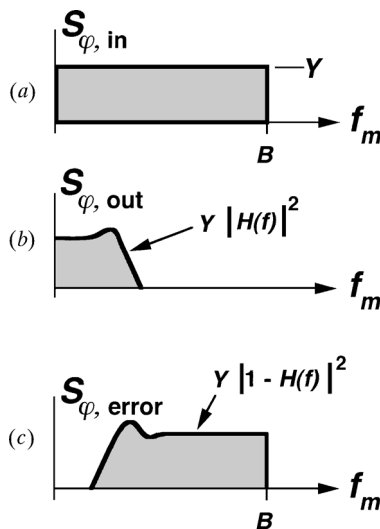


Fig. 11.4 Phase power spectral densities, output at (b) and error at (c) in response to input at (a).

represented well by straight lines on a log–log plot, it is helpful to have formulas for integrating under such curves.

The integral from f_1 to f_2 under a curve that has value $S_y(f_1)$ at f_1 and a slope such that $S_y(f) = Kf^b$ is³

$$P_y|_{f_1}^{f_2} = \frac{f_1 S_y(f_1)}{b + 1} \left[\left(\frac{f_2}{f_1} \right)^{b+1} - 1 \right] = \frac{f_2 S_y(f_2)}{b + 1} \left[1 - \left(\frac{f_1}{f_2} \right)^{b+1} \right], \quad (11.13)$$

unless $b = -1$. If $b = -1$,

$$P_y|_{f_1}^{f_2} = f_1 S_y(f_1) \ln(f_2/f_1) = f_2 S_y(f_2) \ln(f_2/f_1). \quad (11.14)$$

We can use these equations to determine the error that occurs if we simply multiply a frequency range by the density in the center of the range, rather than using a differential width and integrating. Figure 11.5 shows how wide the frequency range Δf can be, normalized to the center value f , before a 1 dB error occurs for various slopes b . It implies that it is most important to keep Δf small when the curves are steep. At slopes of $b = 0$ or 1 there is no error. The errors cause high estimates except for slopes between 0 and 1 but the errors there are smaller than 0.3 dB, even if the region extends from 0 to $2f$.

11.4 TYPICAL OSCILLATOR SPECTRUMS

Figure 11.6 shows typical shapes of phase power spectral density in oscillators. Here S_ϕ is plotted in decibels relative to 1 rad² per hertz (dBr/Hz). At high modulation

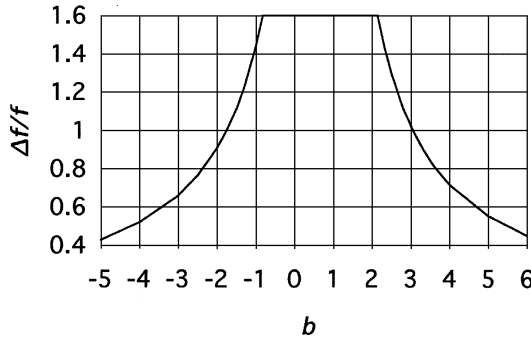


Fig. 11.5 Relative averaging range that causes 1-dB error in power. This is plotted against the slope of the density curve.

³This equation is represented by a straight line of slope b on a log–log plot,

$$\log[S_y(f)] = \log(K) + b \log(f)$$

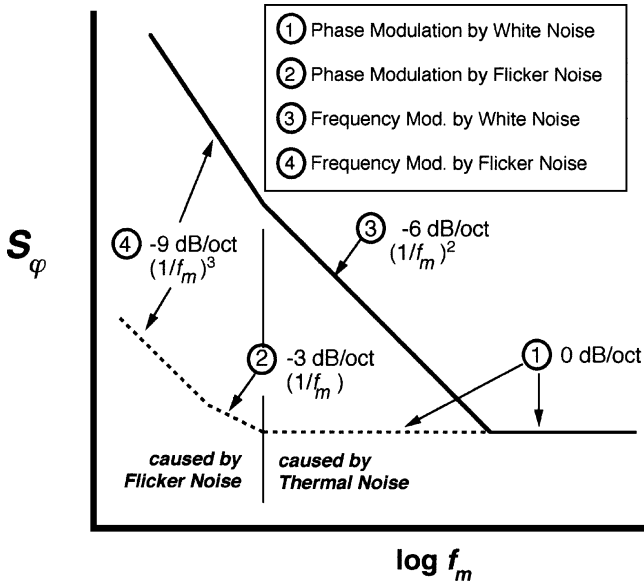


Fig. 11.6 Typical shapes of oscillator phase power spectral densities.

frequencies (region 1 in Fig. 11.6), S_φ is flat. This is white phase noise and results from amplification of the active device’s input noise by its gain and noise factor. At lower frequencies, flicker noise predominates (region 2) and S_φ takes on a slope of -3 dB/octave ($= -10$ dB/decade, $b = -1$). Below a frequency that is related to the oscillator’s frequency f_{out} and the loaded Q , namely $f_m = f_{out}/(2Q)$, the spectrum rises, taking on an additional -6 dB/octave (-20 dB/decade, $b = -2$), for a net of -9 dB/octave (-30 dB/decade, $b = -3$) (region 4). This profile (regions 1, 2, and 4) belongs primarily to high- Q oscillators such as crystal oscillators. For most oscillators, $f_m = f_{out}/(2Q)$ occurs at a frequency where the phase noise is still white. As a result, a -6 dB/octave region (3) is created between regions 1 and 4.

Since $S_\varphi = S_f/f_m^2$, white frequency noise (flat S_f) causes an S_φ that falls at -6 dB/octave due to the $1/f_m^2$ term. That is why the -6 dB/octave region is so named—and similarly for the -9 dB/octave region.

Example 11.1 Integrating Noise Density

- a. What is the variance of phase σ_φ^2 between 1 and 10 kHz if the phase power spectral density S_φ is 2×10^{-5} rad²/Hz at 100 Hz and falls at -30 dB/decade throughout the region?

At the low-frequency end of the region, S_φ is 30 dB lower than at 100 Hz, that is 2×10^{-8} rad²/Hz. We use Eq. (11.13), placing the variables into the equation in

the order shown there, to get

$$\begin{aligned} \sigma_\varphi^2 \Big|_{10^3 \text{ Hz}}^{10^4 \text{ Hz}} &= \frac{(1000 \text{ Hz})(2 \times 10^{-8} \text{ rad}^2/\text{Hz})}{-3 + 1} \left[\left(\frac{10^4 \text{ Hz}}{10^3 \text{ Hz}} \right)^{-3+1} - 1 \right] \\ &= 0.99 \times 10^{-5} \text{ rad}^2. \end{aligned}$$

b. What is the variance of frequency?

Using Eq. (11.5a),

$$\begin{aligned} S_f(1 \text{ kHz}) &= S_\varphi(1 \text{ kHz})(10^3 \text{ Hz})^2 \\ &= (2 \times 10^{-8} \text{ rad}^2/\text{Hz})(10^6 \text{ Hz}^2) = 0.02 \text{ Hz}^2/\text{Hz}. \end{aligned}$$

Since $S_f = S_\varphi f_m^2$ and since S_φ has a slope of -30 dB/decade ($b = -3$), S_f has a slope of -10 dB/decade ($b = -1$). Therefore we use Eq. (11.14) to obtain

$$\sigma_f^2 \Big|_{10^3 \text{ Hz}}^{10^4 \text{ Hz}} = (1000 \text{ Hz})(0.02 \text{ Hz}^2/\text{Hz}) \ln \left(\frac{10^4 \text{ Hz}}{10^3 \text{ Hz}} \right) = 46 \text{ Hz}^2.$$

11.5 LIMITS ON THE NOISE SPECTRUM—INFINITE VARIANCES

We can rewrite $S_y(f_1)$ in terms of the $S_y(f_x)$ at some fixed frequency f_x :

$$S_y(f_1) = S_y(f_x) \left(\frac{f_1}{f_x} \right)^b. \tag{11.15}$$

This allows us to write Eq. (11.13) in terms of a fixed density level $S_y(f_x)$ as the starting frequency f_1 changes.

$$P_y \Big|_{f_1}^{f_2} = \frac{f_1 S_y(f_x)(f_1/f_x)^b}{b + 1} \left[\left(\frac{f_2}{f_1} \right)^{b+1} - 1 \right] = \frac{S_y(f_x)}{f_x^b(b + 1)} [f_2^{b+1} - f_1^{b+1}]. \tag{11.16}$$

From this we see that, if integration begins at $f_1 = 0$, b must be greater than -1 to maintain a finite integral. At the limit where $b = -1$, $f_x S_x$ is constant (since $S_x = K/f_x$). Thus Eq. (11.14) shows that P_y is the same in each octave or each decade or in any other set of ranges for which $f_2 = K f_1$. Therefore, as one integrates a curve of this slope, beginning at lower and lower starting frequencies, the integral increases uniformly for each decade, for example, of decrease in f_1 . For curves with steeper negative slopes the increase under such a process is more rapid. So how is one to interpret these seemingly possible infinities?

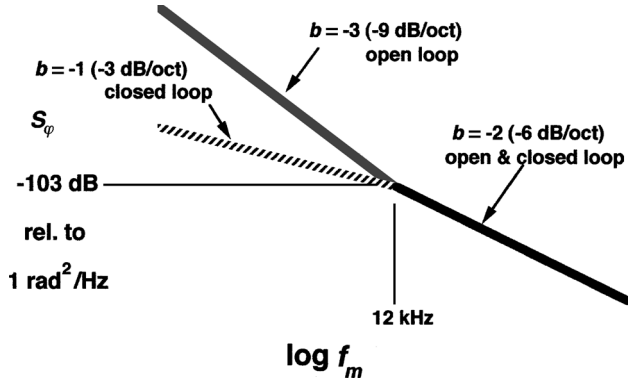


Fig. 11.7 Phase power spectral density for a 240- to 352-MHz oscillator at 296 MHz.

If an infinite answer is obtained to a practical question, the problem has probably not been formulated correctly or the answer has not been interpreted properly. For example, Fig. 11.7 shows the open-loop phase power spectral density for a certain (real) oscillator at 296 MHz.

Suppose that a simple PLL is created using this VCO with a unity-gain frequency f_L at the same frequency where the open-loop S_φ bends, about 12 kHz (just to make the problem simpler). Then, with the loop closed, the loop gain will cause the slope to change from -9 dB/octave to -3 dB/octave (effective high-pass) to the left of 12 kHz (approximately). But Eq. (11.14) will give an infinite value for $\sigma_{\varphi e}$ if we integrate to zero frequency. We might think that the slope must flatten at some low frequency, but modulation frequencies with periods as long as a year have been measured [Allan, 1966] so the continuation of the noise slope to such low frequencies had better not cause a problem in a practical application because it does happen—a true lower limit for this slope is not known. Suppose then that we compute the frequency $f_x = c/T_x$ below which this noise must terminate in order that $\sigma_{\varphi e}$ be no greater than 1 rad^2 . By Eq. (11.13) the contribution from frequencies above 12 kHz (to infinity) is $6 \times 10^{-7} \text{ rad}^2$ and, by Eq. (11.14), the contribution from frequency f_x to 12 kHz, which we set equal to 1 rad^2 is

$$6 \times 10^{-7} \ln(12 \text{ kHz } T_x/c) = 1, \tag{11.17}$$

so

$$T_x = 8.3 \times 10^{-5} \text{ sec} \exp(1/6 \times 10^{-7}) = 10^{7.2 \times 10^5} \text{ sec} \approx 10^{7.2 \times 10^5} \text{ years.} \tag{11.18}$$

This is 100...00 years, where the number of zeros is 720,000. Changes in phase that occur with periods this long will have a greatly attenuated effect during the period when we will be concerned about any oscillator.

The real question of interest is not what is the variance of phase error but what is the variance over some period, perhaps a year or ten years or maybe a thousand years?

The mathematical formulation of *that* problem will include a factor that attenuates the effect of $S_\varphi(f_m)$ for $f_m \ll c/T_x$. The actual expression for the variance of phase during a period T_x/c is [Egan, 2000, p. 504]

$$\sigma_\varphi^2 = \int_0^{f_{\max}} \left[1 - \left(\frac{\sin \pi f_m T_x/c}{\pi f_m T_x/c} \right)^2 \right] S_\varphi df_m. \quad (11.19)$$

For $\omega_m \ll c/T_x$, the integrand here becomes approximately $S_\varphi[\pi f_m T_x/c]^4/3$, which we obtain by taking the first two terms in the series for a sine. Since $[\pi f_m T_x/c]^4$ has a +12 dB/octave-of- f_m slope, the net slope for the expression is +9 dB/octave at low frequencies and thus the contribution for $f_m < c/T_x$ is small and the integral is finite.

Questions related to the effects of phase noise statistics over finite time periods are addressed in Egan (2000), pp. 499–506, and Egan (2003), pp. 269–271.

11.6 POWER SPECTRUM

The spectrum of a cosine that is sinusoidally phase modulated with a peak phase deviation of m at a frequency f_m can be written (see Appendix 21A)

$$\begin{aligned} \cos(\omega_c t + m \sin \omega_m t) \\ = J_0(m) \cos \omega_c t + J_1(m) [\cos(\omega_c + \omega_m)t - \cos(\omega_c - \omega_m)t] \\ + J_2(m) [\cos(\omega_c + 2\omega_m)t - \cos(\omega_c - 2\omega_m)t] + \dots \end{aligned} \quad (11.20)$$

11.6.1 Spectrum for Small m

For a small enough modulation index m (peak phase deviation in radians), we have $J_0(m) \approx 1$, $J_1(m) \approx m/2$, and we can ignore the higher order Bessel functions, $J_i(m)$ with $i > 1$. More specifically, for $m \leq 0.1$: $J_0(m)$ is within 0.25% of 1; $J_1(m)$ is within 0.12% of $m/2$, and $J_2(m) < 0.024 J_1(m)$. Therefore, for small phase deviations, the spectrum of the modulated cosine becomes

$$\cos(\omega_c t + m \sin \omega_m t) \approx \cos \omega_c t + \frac{m}{2} [\cos(\omega_c + \omega_m)t - \cos(\omega_c - \omega_m)t], \quad (11.21)$$

as illustrated in Fig. 11.8. The carrier power is undiminished, 0.5, corresponding to a sinusoidal amplitude of 1. (We assume a 1- Ω system in converting between voltage and power.) The sidebands' powers are lower by $(m/2)^2$, but this is written in terms of the root mean square (rms) value, $\sigma = m/\sqrt{2}$. We call the spectrum one-sided because all the power is shown at positive frequencies. In a two-sided spectrum, such

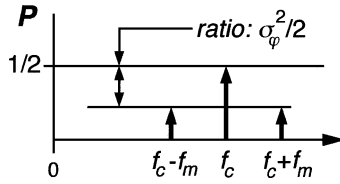


Fig. 11.8 One-sided power spectrum of PM sinusoid, small modulation index.

as is used in Fourier transform theory, half of the power is at negative frequencies. The ratio of sideband to carrier level is the same in either case.

If the phase modulation, rather than being discrete, should represent S_φ in a narrow bandwidth δf , then the relative sideband amplitudes would also be

$$L_\varphi \delta f = \sigma_\varphi^2 / 2 = S_\varphi \delta f / 2, \tag{11.22}$$

so

$$L_\varphi = S_\varphi / 2. \tag{11.23}$$

The φ subscript just indicates that L is due to phase noise.

11.6.2 Single-Sideband Density

Here σ_φ^2 is the mean square phase deviation in radians squared and L is the single-sideband (SSB) power spectral density relative to the carrier, a dimensionless ratio that is often expressed in decibels relative to the carrier per hertz. It is termed single sideband because it gives the relative power of each sideband separately. As can be seen from Eq. (11.23), the relative power in the sum of both sidebands equals S_φ .

In decibels, Eq. (11.23) can be expressed as

$$L_\varphi |_{\text{dBc/Hz}} = S_\varphi |_{\text{dBBr/Hz}} - 3\text{dB}. \tag{11.24}$$

One would measure $S_\varphi(f_m)$ by observing the output of a phase detector, as in Fig. 11.9, while L would be observed directly on the spectrum analyzer. The displays that would be observed in each case are illustrated in Fig. 11.10.

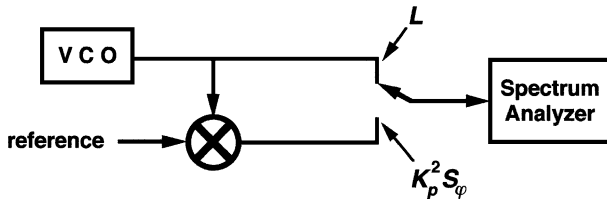


Fig. 11.9 Measuring L and S_φ .

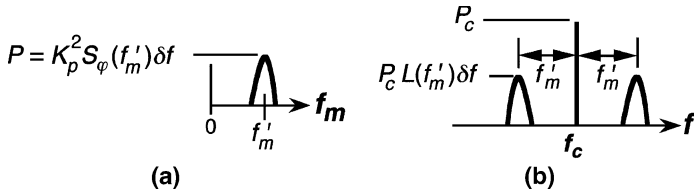


Fig. 11.10 Spectrum analyzer display of phase power density and power spectral density; δf is the analysis bandwidth in the spectrum analyzer.

As long as the analysis bandwidth of the spectrum analyzer is narrow enough that S_ϕ and L are nearly constant across that bandwidth, we can obtain an accurate value for S_ϕ and L by dividing the observed power by δf and, in the case of S_ϕ , also by K_p^2 . Thus a decrease of the analysis bandwidth by a decade will cause the displayed density to drop by 10 dB, as illustrated in Fig. 11.11. Note, however, that the proper bandwidth to use is the “noise bandwidth.” This is the bandwidth that gives the proper ratio between the density and the displayed level. If the analyzer had an ideal rectangular filter, the noise bandwidth would be the same as the -3 dB bandwidth, but, with real filters, there is generally some difference between the two bandwidths. Noise bandwidth will be discussed in some detail in Chapter 14. Other corrections may be required also. Analog spectrum analyzers typically read noise power about 2.5 dB low relative to the power of a sine wave [Hewlett-Packard, 1989]. If a spectrum analyzer can display density, that mode should be used both for convenience and because it can include appropriate corrections. Otherwise, consult the instrument’s manual.

Example 11.2 Spectral Display of Noise Sidebands If Fig. 11.11 represents an expansion of part of Fig. 11.10b and if the discrete carrier power is displayed at a level of 0 dB (without saturating the analyzer), then the peak value of L displayed in Fig. 11.11 is

$$L = 10^{-4}/100 \text{ Hz} = 10^{-5}/10 \text{ Hz} = 10^{-6}/\text{Hz} \tag{11.25}$$

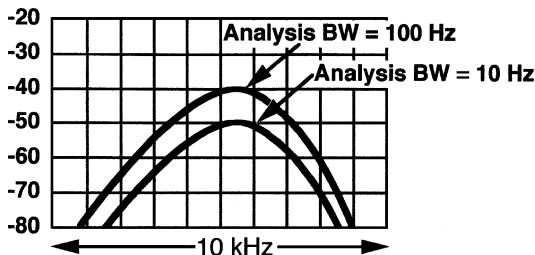


Fig. 11.11 Spectrum analyzer display of noise power at two different analysis bandwidths.

or, in terms of dB,

$$L|_{\text{dB}} = -40 \text{ dBc/Hz} - 20 \text{ dB} = -50 \text{ dBc/Hz} - 10 \text{ dB} = -60 \text{ dBc/Hz}. \quad (11.26)$$

Here we have expressed the density in terms of various bandwidths to indicate what happens when the analysis bandwidth changes. Unlike the effect on noise, changing the analyzer's analysis bandwidth has no effect on the display of the discrete carrier level.⁴

Example 11.3 Spectral Display of Phase Noise If Fig. 11.11 represents an expansion of Fig. 11.10a and if 0 dB is equivalent to 0.01 V rms and the phase detector gain is $K_p = 0.1 \text{ V/rad}$, then 0 dB represents 0.1 rad rms [i.e., $0.01 \text{ V}/(0.1 \text{ V/rad})$]. With a 100-Hz analysis bandwidth, 0 dB represents a noise density of

$$S_\varphi = (0.1 \text{ rad})^2/100 \text{ Hz} = 10^{-4} \text{ rad}^2/\text{Hz} \Rightarrow -40 \text{ dBBr/Hz}. \quad (11.27)$$

Thus the peak of -40 dB would represent a density of

$$S_\varphi \Rightarrow -40 \text{ dBBr/Hz} - 40 \text{ dB} = -80 \text{ dBBr/Hz}. \quad (11.28)$$

Repeating this process for the 10-Hz analysis bandwidth will produce the same density.

11.6.3 When is the Modulation Small?

We showed that the approximation of Eq. (11.23) is valid if $m \ll 1 \text{ rad}$. For the noise power spectrum of Fig. 11.10 the equivalent restriction would be $\sigma_\varphi^2 \ll 1 \text{ rad}^2$. If that restriction does not hold, then L will not have a shape so similar to that of S_φ because the carrier at f_c will diminish, $S_\varphi(f_m)$ will cause sidebands of L not only at $f_c \pm f_m$ but at $f_c \pm if_m$, for various i , and the magnitudes of the first sidebands will not be linearly proportional to m . Moreover, each sideband will be modulated by components that produced the others. We might be tempted to think that, since S_φ is the density in a differential bandwidth, it is small enough; but further reflection should convince us that dividing the spectrum into smaller frequency segments will not make a fundamental difference. The more segments we have, the more multiple sidebands are produced to add to each other. If σ_φ in Fig. 11.10a is larger than a radian, we can expect to see multiple sidebands on each side of the carrier in Fig. 11.10b.

The criterion is more difficult to apply in the case of typical oscillator spectrums, such as shown in Fig. 11.6, because we have shown that σ_φ^2 grows without known

⁴ Actually, even "discrete" signals have finite widths, but we speak of them as being discrete when their power is concentrated in a bandwidth small enough to be considered zero for the problem at hand. The level of a discrete will not change as long as practically all of its power is contained within the analysis bandwidths used.

limit as the lower frequency of integration decreases. The criterion that we will apply to such spectrums is [Shoaf et al., 1973, p. 43]

$$\int_{f_1}^{\infty} S_{\varphi}(f_m) df_m \ll 1 \text{ rad}^2, \tag{11.29}$$

where f_1 is the lowest modulation frequency at which we employ the small modulation index approximation.

We can rationalize this as follows [Egan, 2000, pp.105–106]. Consider modulation at frequencies much lower than f_1 , where the phase variance is appreciable, as belonging to the carrier. This will broaden the line width of the carrier, which is being modulated by the frequencies that are greater than f_1 , and the broadening will result in smearing of the spectrum. However, the smearing will not be noticeable for $f_m > f_1$ because S_{φ} changes little across frequency increments that are small compared to f_1 at offsets that are greater than f_1 . For example, if f_1 is 1 kHz, modulation frequencies much lower than this will broaden the “carrier,” perhaps spreading power over 1 Hz or so. Phase modulation at a 1-kHz rate will produce sidebands on the carrier at offsets of ± 1 kHz. If the carrier is 1 Hz wide, then the ± 1 -kHz sidebands will actually spread over a few hertz at ± 1 kHz. Usually the magnitude of S_{φ} does not change much over a few hertz in the vicinity of 1 kHz, so the smearing of this power over a small frequency increment, along with similar smearing from adjacent frequencies, will not make a noticeable difference.

When the criterion of Eq. (11.29) is not met, we can expect differences in the shape of $S_{\varphi}(f'_m)$ and L at offsets from the carrier of f'_m . As S_{φ} climbs indefinitely at lower frequencies, leading to unbounded σ_{φ}^2 , the power spectrum will remain finite near the spectral center.

11.6.4 Modification of Spectral Shape at Higher Modulation Index

The phase power spectral density of Fig. 11.12a results in a power spectrum, shown at (b), which is given by Eq. (11.23) if the deviation is small enough for Eq. (11.29)

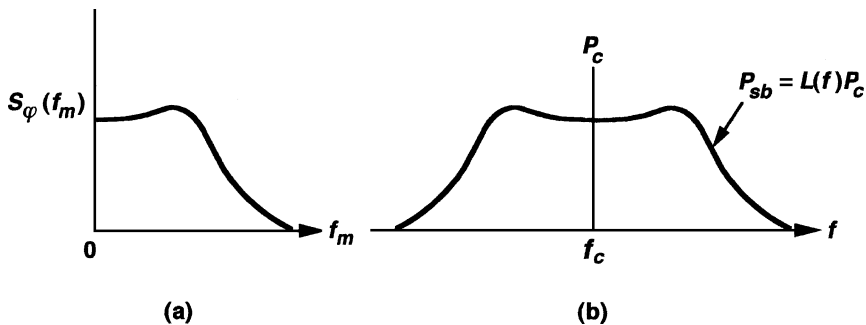


Fig. 11.12 (a) Phase power spectral density and (b) resulting sideband power spectrum.

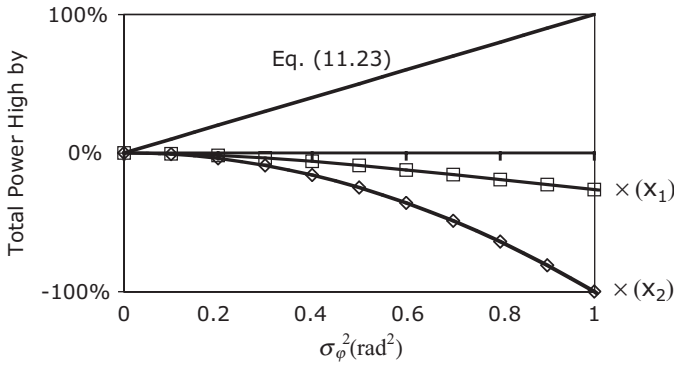


Fig. 11.13 Change in total spectral power versus phase variance with several spectral representations. The power should remain unchanged.

to be satisfied. When the deviation increases to the point where that restriction is not well satisfied, we can approximate the effect on the spectrum by multiplying the entire spectrum, central line and sidebands, by $x_1 = \exp(-\sigma_{\phi, \text{out}}^2) \approx x_2 = (1 - \sigma_{\phi, \text{out}}^2)$ [Blanchard, 1976, pp. 168–171]. A derivation for x_2 is given in Appendix 11.H. These changes just improve the representation but are still depend on small m . One indication of their effectiveness is whether the total power of the spectrum, which is not changed by the phase modulation, remains closer to the true value with the modified spectrums (Fig. 11.13). On this basis, multiplication by

$$x_1 = \exp(-\sigma_{\phi, \text{out}}^2) \tag{11.30}$$

gives the best estimate.

11.6.5 Script \mathcal{L}

The symbol L is not commonly used. Rather, \mathcal{L} has been in common use for single-sideband power spectral density. Unfortunately, carelessness in usage has caused it to be often taken as equal to $S_\phi/2$, regardless of whether this was justified by Eq. (11.29). Therefore, \mathcal{L} was redefined by IEEE Standard 1139 [Allan et al., 1988] to be simply equal to $S_\phi/2$. As a result of its prior usage and its revised usage, \mathcal{L} is somewhat ambiguous, meaning perhaps $S_\phi/2$ and perhaps L . (Of course, in most cases of interest, they will be the same.) Therefore, we will not use \mathcal{L} in this text, except where Eq. (11.29) has been met so that $\mathcal{L} = L = S_\phi/2$.

11.7 FREQUENCY MULTIPLICATION AND DIVISION

We briefly considered devices that multiply or divide frequency in Chapter 10 (Fig. 10.9). Multiplying frequency by M automatically multiplies frequency, and

phase, deviation by M . Thus it increases S_φ , S_φ , σ_φ^2 , and so forth by a factor of M^2 ($20\log_{10}M$). Conversely, frequency division by N decreases these variables by $20\log_{10} N$.

11.8 OTHER REPRESENTATIONS

Figure 11.14a is identical to Fig. 11.4b. It shows phase power spectral density S_φ in radians-squared/hertz or dBr/Hz versus modulation frequency f_m . Figure 11.14b shows the single-sideband spectral density L , usually given in decibels relative to carrier per hertz. This differs from the representation in Fig. 11.10b in three ways (not counting the fact that they represent different spectral shapes). First, only one sideband is shown—there is no loss of information, however, because of symmetry. Second, no powers are given, only the level of the sideband density relative to the carrier power or total power [which are approximately the same under the restriction of Eq. (11.29)]. Third, the abscissa shows offsets from spectral center rather than actual frequencies. While Fig. 11.10b represents what one would observe on a spectrum analyzer, the information is usually plotted as shown here. In both Fig. 11.14a and Fig. 11.14b the horizontal axis is often logarithmic (of course 0 cannot be shown when it is). At Fig. 11.14c is shown the two-sided phase power spectral density. This uses negative frequencies, as in Fourier transforms, so the power is distributed over twice the frequency range, causing the density to be only half as great as at Fig. 11.14a. This has the advantage of causing Fig. 11.14c to look very much like a spectrum analyzer display (Fig. 11.10b) with the carrier level set at 0 dB except that the analyzer's display would be centered on the carrier frequency rather than on zero. The density level then also is the same as at Fig. 11.14b, rather than being twice

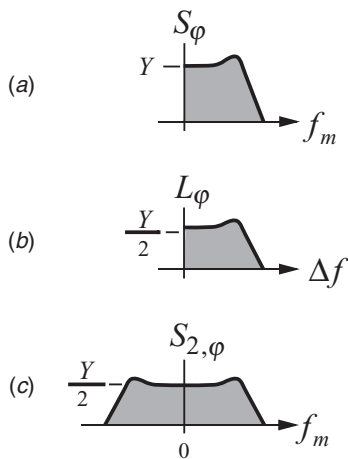


Fig. 11.14 Other Representations.

as great, as at Fig. 11.14a. However, this uniformity does not seem to be worth the potential confusion. If the two-sided representation were used also for Fig. 11.14b, the similarity would disappear again so it depends on using two different representations, one for the power spectrum and another for the modulation density, at the same time. We will generally avoid the use of the two-sided spectrum.

i11.H SHAPE OF OUTPUT SPECTRUM

This appendix is available from the Wiley Internet site.

11.S APPENDIX: SPREADSHEETS FOR INTEGRATING DENSITIES

Many practical noise density plots can be approximated by straight-line segments on semilog paper (decibels vs. log frequency; refer to Sections 11.3 and 11.4). Enter the coordinates at the ends of these segments into the spreadsheet IntPDens and it will give the total integrated power. Companion spreadsheets give specifically the phase deviation or frequency deviation when the segments described represent L_ϕ [Eq. (11.29) assumed].

IntPDens is an Excel spreadsheet that integrates relative (to carrier) power density to give relative power, using Eqs. (11.13) and (11.14). The rms value of the integrated density is presented for one sideband and for both identical sidebands. We could also enter absolute power density in decibels relative to 1 W/hertz (dBW/Hz) and thus obtain the answers relative to 1 W plus rms values in volts in a 1-Ω system.

IntPhNs also converts the density to the phase deviation it represents. Root mean square phase deviation is given in radians and degrees. The spreadsheet is shown in Fig. 11.S.1. The modulation frequency and SSB density corresponding to the vertices of the segments are entered in columns B and C over the sample data shown there.

	A	B	C	D	E	F	G	H
1	INTEGRATED PHASE DENSITY							
2	Enter SSB density representing phase noise to obtain phase deviation.							
3	ENTER DATA		Do not enter data below. Copy last line OK.					
4	Mod.	SSB		Integrated				
5	Freq.	density		Segment	sum	RMS Phase		
6	Hz	dBc/Hz	slope	rad^2	rad^2	rad	degrees	
7	1.00E+2	-89						
8	1.00E+4	-109	-1	1.16E-06	1.16E-06	0.00108	0.0617	
9	1.00E+5	-139	-3	1.246E-07	1.284E-06	0.00113	0.06493	
10	2.00E+6	-140	-0.08	4.06E-08	1.325E-06	0.00115	0.06595	
11								

Fig. 11.S.1 IntPhNs.

	A	B	C	D	E	F	G	H	I
1	INTEGRATED FREQUENCY DENSITY								
2	Enter SSB density representing phase noise to obtain frequency deviation.								
3	ENTER DATA		Do not enter data below. Copy last line OK.						
4	Mod.	SSB	Freq. Dens.		Integrated	sum		RMS Frequency	
5	Freq.	density	Hz^2/Hz		Segment	Hz^2	Hz	rad/sec	
6	Hz	dBc/Hz	in dB	slope	Hz^2				
7	1.00E+2	-89	-46						
8	1.00E+4	-109	-26	1	0.0006294	0.0006294	0.03548	0.22292	
9	1.00E+5	-139	-36	-1	2.899E-07	0.0006297	0.03549	0.22298	
10	2.00E+6	-140	-10.9794	1.923	2.736E-06	0.0006324	0.03556	0.22346	
11									

Fig. 11.S.2 IntFNs.

The bottom line is to be copied as often as required, depending on the number of segments.

Operation of the spreadsheet can be understood from the formulas contained in typical cells. Cell D9 obtains the value of b , using $= (C9 - C8) / \text{LOG}(B9 / B8)$. Cell E9 chooses either Eq. (11.13) or (11.14) to give the integral, depending on how close b is to -1 :

$$\begin{aligned}
 &= \text{IF}(\text{OR}(D9 < -1.0001, D9 > -0.9999), \\
 &B8 * 10^{(C8 / 10) / (1 + D9)} * ((B9 / B8)^{(1 + D9) - 1}), \\
 &B8 * 10^{(C8 / 10) * \text{LN}(B9 / B8)} * 2
 \end{aligned}$$

The factor of 2 at the end comes from Eq. (11.23) and is required to convert from L_ϕ to S_ϕ . Cell F9 contains $= E9 + F8$ to sum the integral under the current segment with those from previous segments. Cells G9 and H9 contain $= \text{SQRT}(F9)$ and $= G9 * 180 / \text{PI}()$, respectively.

IntFNs converts the single-sideband relative power density to frequency density and then integrates to give mean square frequency deviation. Root mean square deviations in radians/second and hertz are also given. The spreadsheet is shown in Fig. 11.S.2. It is similar to IntPhNs but it contains one more column to convert SSB density L_ϕ to frequency power spectral density S_f before the slope is taken. This follows Eq. (11.5a) and is represented in cell D9 by $= C9 + 3 + 10 * \text{LOG}(B9)$. Here +3 serves the same purpose as the *2 did previously, to double the density to represent phase power spectral density rather than SSB density, but we add 3 because we are still working in decibels at this point. Since the doubling takes place prior to integration here, cell F9 does not contain the *2 that the equivalent cell E9 did before.

CHAPTER 12

RESPONSE TO PHASE NOISE

The loop processes phase noise just as it processes discrete (single-frequency) phase modulation. We studied this concept in Chapter 7. Here we will apply those same responses to the power spectral density S , which was defined in Chapter 11.

12.1 PROCESSING OF REFERENCE PHASE NOISE

Figure 12.1 indicates where the input phase power spectral density $S_{\varphi, \text{in}}$ and the output and error responses to it are located relative to loop components. This is a special case of Fig. 7.5 wherein the phases are modulation density profiles rather than being at a discrete modulation frequency. The responses in Chapter 7 apply equally well to S_{φ} as they do to φ . However, since the modulation phase of the noise is unknown, the phase of $H(\Omega)$ (e.g., in Fig. 7.7) is of no value and, based on Eqs. (11.9) and (11.10), squared values of $|H(\omega)|$ and $|1-H(\omega)|$ are of interest. When the latter are given in decibels, the squaring process is automatic since

$$|H(\omega)|_{\text{dB}} = 20 \log_{10} |H(\omega)| = 10 \log_{10} |H(\omega)|^2 = |H(\omega)|_{\text{dB}}^2. \quad (12.1)$$

Thus, for example, Fig. 7.6, 7.8, or 7.9 could be applied to determine $S_{\varphi, e}$ or $S_{\varphi, \text{out}}$. The value of $S_{\varphi, \text{in}}$, in decibels relative to 1 rad²/hertz (dBr/Hz) would be added to the appropriate response in dB to determine the output in dBr/Hz.

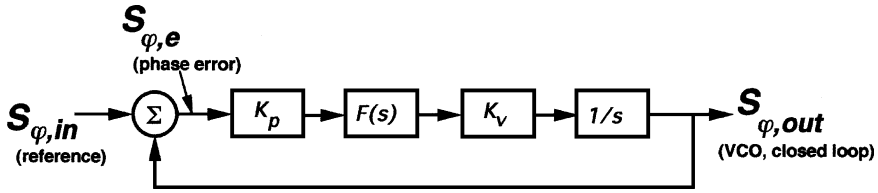


Fig. 12.1 Phase noise on the reference and responses.

The value of $S_{\phi,out}$ is important in many applications. Often the purpose of the phase-locked loop (PLL) is to improve $S_{\phi,out}$ relative to $S_{\phi,in}$. One reason that $S_{\phi,e}$ is important is that it must be small if the assumption of linearity of the loop is to be valid. For example, with a sinusoidal phase detector, the integrated $S_{\phi,e}$ should be small compared to $(\pi/4 \text{ rad})^2$.

Example 12.1 Loop Response to Reference Phase Noise, Standard Plots

Given: $S_{\phi,in} = -90 \text{ dBBr/Hz}$; $\alpha = 0$; $\omega_m/\omega_n = 0.7$; $\zeta = 1$.

Find: $S_{\phi,e}$ and $S_{\phi,out}$.

From Fig. 7.8, $|H(\alpha = 0; \omega_m/\omega_n = 0.7; \zeta = 1)| \Rightarrow -3 \text{ dB}$, so

$$S_{\phi,out} = -90 \text{ dBBr/Hz} - 3\text{dB} = -93 \text{ dBBr/Hz}.$$

From Fig. 7.9, $|[1 - H](\alpha = 0; \omega_m/\omega_n = 0.7; \zeta = 1)| \Rightarrow 0 \text{ dB}$, so

$$S_{\phi,e} = S_{\phi,in} = -90 \text{ dBBr/Hz}.$$

By the same figure, $S_{\phi,e}$ would be -99.7 dBBr/Hz if α were to equal 1.

As long as the conditions of Eq. (11.29) are met, the same graphs can be used in the same manner to determine $L_{\phi,e}$ or $L_{\phi,out}$ from $L_{\phi,in}$, since L differs from S by a constant under those conditions.

The same information can be obtained approximately from the Bode gain plot. This is often more convenient and is especially important where the loop is higher than second order, in which case the plots in Chapter 7 would not apply. Figure 12.2 shows a tangential approximation to typical crystal oscillator noise as $S_{\phi,in}$ and a log plot of $|G(\omega_m)|$ for a loop.

Since

$$S_{\phi,out}(\omega_m) \approx S_{\phi,in}(\omega_m) \quad \text{for } \omega_m \ll \omega_L$$

and

$$S_{\phi,out}(\omega_m) \approx S_{\phi,in}(\omega_m) |G(\omega_m)|^2 \quad \text{for } \omega_m \gg \omega_L$$

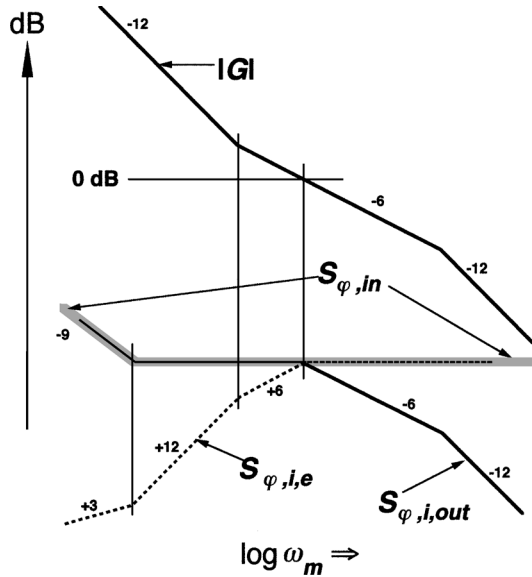


Fig. 12.2 Effect of loop on input phase noise. Subscripts *i* indicate responses to noise at loop input (to PD). Numbers are slopes in decibels per octave.

[see Sections 4.1 and 7.4.1], $S_{\varphi,out}(\omega_m)$ can be roughly determined from this information. As shown by $S_{\varphi,i,out}$ in Fig. 12.2, $S_{\varphi,out}$ follows $S_{\varphi,in}$ until the unity-gain point and, beyond that, the decreasing values of $|G|_{dB}$ are added to $S_{\varphi,in}|_{dB}$.

Similarly, since $S_{\varphi,e}(\omega_m) \approx S_{\varphi,in}(\omega_m)$ for $\omega_m \gg \omega_L$ and $S_{\varphi,e}(\omega_m) \approx S_{\varphi,in}(\omega_m)/|G(\omega_m)|^2$ for $\omega_m \ll \omega_L$ [see Section 7.4.2], $S_{\varphi,e}(\omega_m)$ follows $S_{\varphi,in}$ for ω_m greater than ω_L but is reduced by $|G|$ at lower frequencies. This is shown as $S_{\varphi,i,e}$, in Fig. 12.2.

Practically, the response plots are often drawn largely by modifying slopes of $S_{\varphi,in}$ based on the slopes of $|G|$. Overshoot or attenuation near the unity-gain point might be estimated from the response curves for second-order loops, using Fig. 5.16 to convert computed phase shift to equivalent ζ .

Example 12.2 Loop Response to Reference Phase Noise, Bode Plot In Fig. 12.2, 100-Hz modulation frequency occurs between the filter zero and unity gain. Unity gain is at 500 Hz. The phase power spectral density (PPSD) of the reference input at 100 Hz is 10^{-6} rad²/Hz. What are the PPSDs for the voltage-controlled oscillator (VCO) output and for the error at the phase detector (PD) at $f_m = 100$ Hz?

The loop voltage gain at 100 Hz is approximately $500 \text{ Hz}/100 \text{ Hz} = 5$. The phase error at the phase detector will equal the input phase suppressed by this amount to $10^{-6} \text{ rad}^2/\text{Hz}(\frac{1}{5})^2 = 4 \times 10^{-8} \text{ rad}^2/\text{Hz}$. The output phase will essentially equal $S_{\varphi,in}$, $10^{-6} \text{ rad}^2/\text{Hz}$.

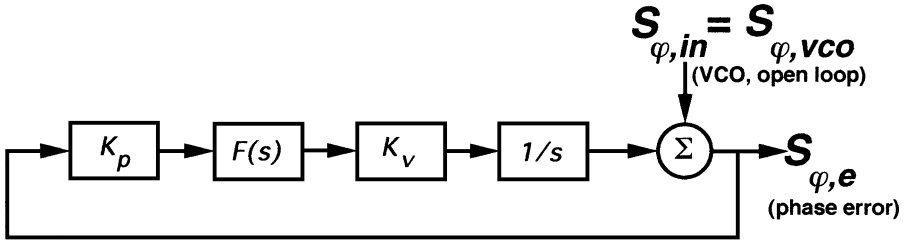


Fig. 12.3 Phase noise on the VCO and responses.

12.2 PROCESSING OF VCO PHASE NOISE

Figure 12.3 applies to phase noise originating in the loop’s VCO. In this case $S_{\varphi, in}$ is the phase noise that would appear at the output of the VCO if the loop were open. With the loop closed, what is actually seen at the output of the VCO is $S_{\varphi, e}$, the combination of $S_{\varphi, in}$ and the phase variations that result from feedback arriving on the tuning voltage. And $S_{\varphi, e}$ is also the error seen by the phase detector, so it must again be maintained small for linearity.

Figure 12.4 shows a tangential approximation to typical VCO noise $S_{\varphi, vco}$ and the same gain plot as in Fig. 12.2. The phase noise at the loop output $S_{\varphi, v, e}$ equals $S_{\varphi, vco}$ reduced by the loop gain at low frequencies, where the gain is significant.

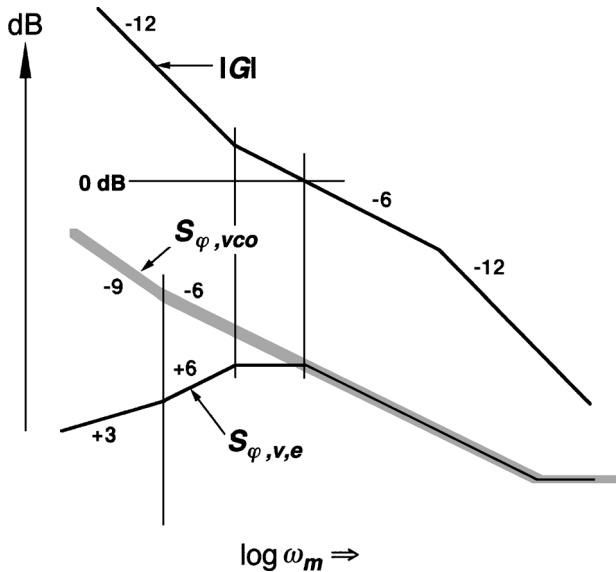


Fig. 12.4 Effect of loop on VCO phase noise. Subscripts v indicate responses to noise at VCO input. Numbers are slopes in decibels per octave.

Example 12.3 Loop Response to VCO Phase Noise What would be the answers in Example 12.2 if the phase noise originated in the VCO?

The numbers would be the same but the correspondence between the general loop diagram and the system is different. Now the loop (VCO) output would be the error signal, 4×10^{-8} rad²/Hz, which remains after the loop has canceled most of the original VCO noise. This is also the phase error at the PD, as it was before. In Example 12.1, the phase detector error was the difference between the noisy reference input and the VCO output phase, which closely tracked the input at low ω_m . In this example, the difference signal is generated at the VCO and passed to the PD. Since the other PD input is noiseless, the phase detector error has the same magnitude as the phase noise from the VCO.

Both Figs. 12.2 and 12.4 show $S_{\phi,e}$ as having a positive slope at the lowest frequencies. This is due to the integrator in a type 2 loop but, if there is no integrator, or at sufficiently low frequencies where the integrator runs out of gain, the slope will become negative. Then the considerations of Section 11.5 apply; however, the resulting, theoretically infinite, variance may be, practically, inconsequential.

12.3 HARMFUL EFFECT OF PHASE NOISE IN RADIO RECEIVERS

As an example of why we may wish to minimize phase noise, consider the application of the local oscillator (LO) in a radio receiver. Its job is to convert the frequency of the received signal to a new frequency. To do this, it enters a mixer along with the received signal and the mixer's output, the intermediate frequency (IF), is at the desired frequency sum or difference. Just as the frequencies combine, so do the corresponding phases. Therefore, any phase variation in the LO will be added to the phase of the signal. Figure 12.5 depicts at *A* the spectrum analyzer display of a strong signal that has been converted by a noisy LO. A weak signal is shown at *B* and it is masked by the phase noise sidebands on the strong signal. If a PLL can reduce the phase noise on the LO so it is as shown at *C*, then the weak signal will be discernible above the sidebands.

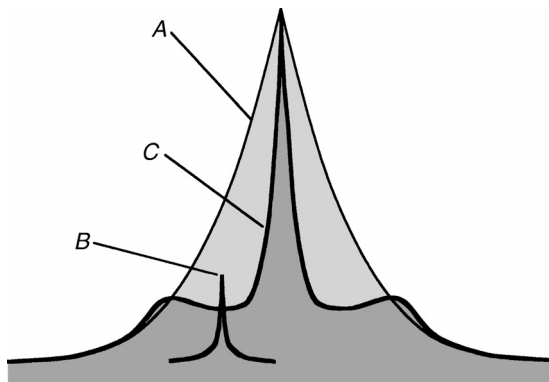


Fig. 12.5 Spectral display illustrating desensitization due to phase noise.

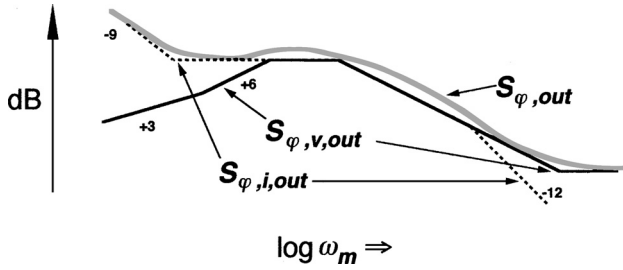


Fig. 12.6 Total phase noise at the output.

on the strong signal. The process by which sidebands on a strong signal increase noise in the vicinity of another signal is one form of receiver “desensitization,” sometimes called “reciprocal mixing.”

Example 12.4 Desensitization How small a signal can be seen above the noise if the signal is separated from a second, -10 dBm signal by 20 kHz and the relative sideband density L_{ϕ} of the LO is -80 dBc/Hz at $f_m = 20$ kHz? The detection Bandwidth is 1 kHz.

In 1 kHz, -80 dBc/Hz produces a noise power of $[(10^{-8}/\text{Hz})(1000 \text{ Hz}) =] 10^{-5}$ or, in decibels, $[-80 \text{ dBc} + 10 \text{ dB} \log_{10}(1000) =] -50$ dBc. Thus the noise produced by the -10 dBm signal will be at $-10 \text{ dBm} - 50 \text{ dBc} = -60$ dBm and that is the weakest signal level that can be seen (if we require 0 dB signal-to-noise ratio).

12.4 SUPERPOSITION

Since noise powers add, the total phase noise can be a combination of phase noise from the reference and phase noise from the VCO (as well as from other sources in the loop). Figure 12.6 shows the closed-loop output noise due to both reference and VCO. These are taken from Figs. 12.2 and 12.4. It also shows their sum, $S_{\phi, out}$.

Figure 12.7 shows the total phase noise at the PD output. There is a lack of symmetry between Figs. 12.6 and 12.7. The former shows the sum of an error response

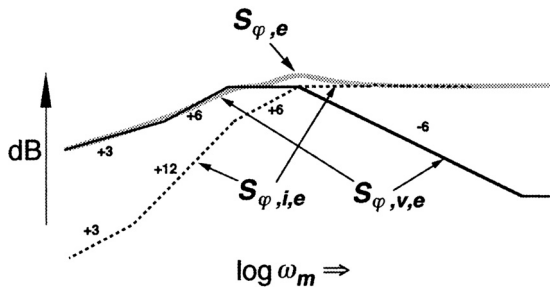


Fig. 12.7 Total phase noise at the phase detector output.

and an output response. The latter shows the sum of two error responses. To understand why this should be so, study the position of the noise inputs and outputs using Fig. 7.5.

12.5 OPTIMUM LOOP WITH BOTH INPUT AND VCO NOISE

The value of modulation frequency at which the open-loop phase noise densities of the reference oscillator and the VCO cross is often nearly the optimum loop bandwidth ω_L for minimizing closed-loop output phase noise. At lower frequencies the VCO noise, which is the higher of the two at low frequencies, is attenuated by the loop whereas at higher frequencies the reference noise, which is the higher at high frequencies, is attenuated. The gain in Figs. 12.2 and 12.4 was chosen to produce this “optimum” bandwidth with the results shown in Fig. 12.6. To obtain a more accurate optimum for minimum phase variance, one should write the total integrated noise, using equations of the type given in Section 11.5, and solve for a minimum or plot the sum against ω_L .¹

Figure 12.8 shows the result with higher gain. The higher gain reduces low-frequency noise due to the VCO but only a small improvement is possible because the $S_{\varphi, in}$ is not affected at low frequencies. However, high-frequency $S_{\varphi, in}$ that had been attenuated is now passed to the output of the loop. Continued gain increases will further increase high-frequency noise without greatly reducing low-frequency noise.

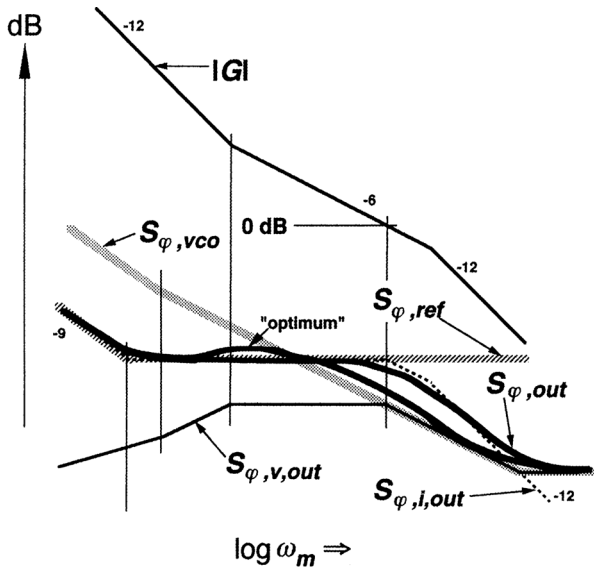


Fig. 12.8 Output phase noise with gain higher than optimum. “Optimum” from Fig. 12.6 is also shown.

¹ Even that is not quite optimum because of the straight-line approximations. A truly accurate optimum would require the actual response equations to be used.

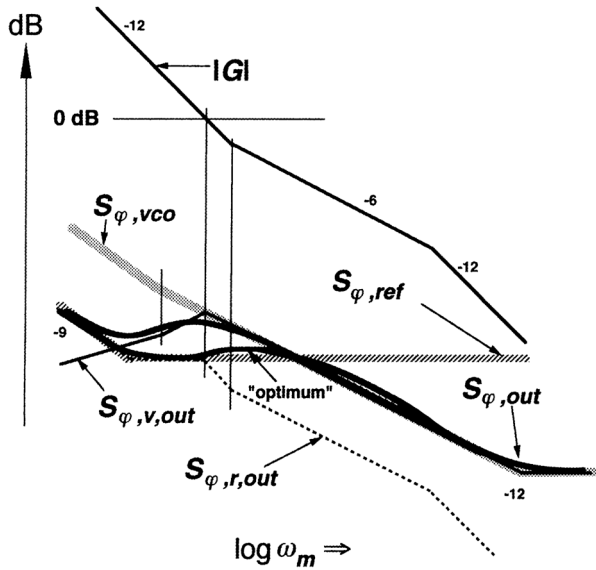


Fig. 12.9 Output phase noise with gain lower than optimum. “Optimum” from Fig. 12.6 is also shown.

Figure 12.9 shows the opposite effect when gain is too low. Here the amount of low-frequency VCO noise that passes through increases, but there is little improvement in high-frequency noise compared to the “optimum.”

Example 12.5 Optimum Loop Bandwidth Estimate the loop bandwidth that will minimize total phase variance for modulation frequencies between 1 and 10^6 Hz if the reference phase noise is flat at -130 dB/Hz and the VCO PPSD is -40 dB/Hz at 10 Hz and has a slope of -30 dB/decade.

We estimate the optimum to occur where the PPSD curves cross. The VCO noise must drop 90 dB from its level at 10 Hz to reach the flat reference noise; 90 dB at 30 dB/decade implies 3 decades. Therefore the curves cross at $f_m = 10^4$ Hz.

To check the estimate, we plot the phase variance, using Eqs. (11.15) and (11.16). Figure 12.10 shows the contributions from the reference phase noise in curve 1 and from the VCO phase noise in curves 2 and 3. Curve 2 is for a first-order loop and curve 3 is for a type 2 loop with the zero at one-third of the unity-gain bandwidth. Without a zero (and the corresponding low-frequency gain), the VCO’s PPSD continues to climb at lower frequencies because its -30 dB/decade slope cannot be overcome by the loops -20 dB/decade gain slope. If it were not for the restriction that we are integrating noise only above 1 Hz, the integral of the closed-loop VCO noise below f_L would be infinite. When a zero is added, the increased low-frequency gain causes the closed-loop PPSD to fall at frequencies below the zero frequency, and we are no longer dependent on the lower limit of integration to a significant degree. Even if the zero frequency is moved up to the unity-gain bandwidth, curves 1 and 3 will cross at a frequency that is about 70% of the crossing point shown.

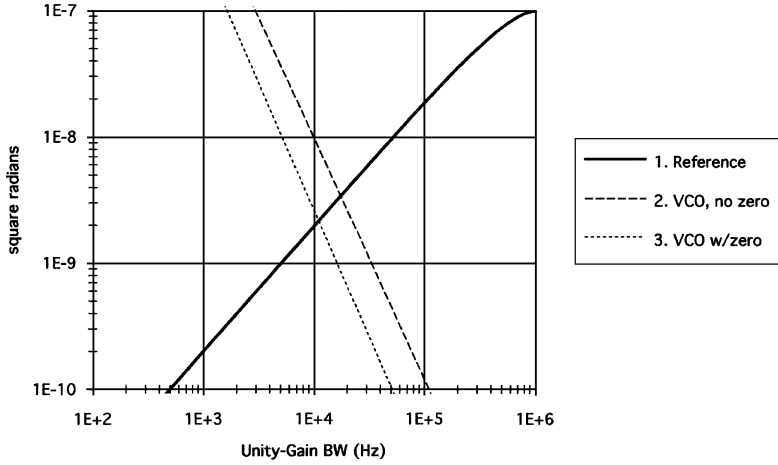


Fig. 12.10 Integrated phase noise, reference and VCO. This is the variance of the closed-loop VCO phase. The integral is from 1 Hz to 1 MHz.

The total phase noise corresponding to Fig. 12.10, curves 2 and 3, is shown in Fig. 12.11, curves 4 and 5 respectively. Our estimate was fairly accurate for the type 2 loop (curve 5). Also shown, at curve 6, is the total noise when the -30 dB/decade slope of the VCO noise is changed to -20 dB/decade, still crossing the reference noise at 10 kHz. We see that the minimum phase variance occurs very close to the estimated point for this case.

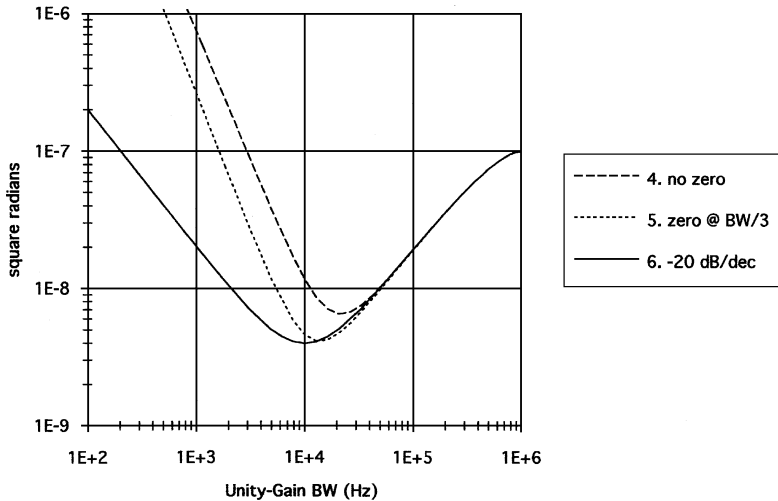


Fig. 12.11 Total phase variance. The sum of reference and VCO noises at the output of the closed loop are shown. One curve is for a first-order loop and the second is for a type 2 loop, both with -90 dBc/Hz open-loop VCO PSD. Also shown is the variance with -20 dB/decade phase noise.

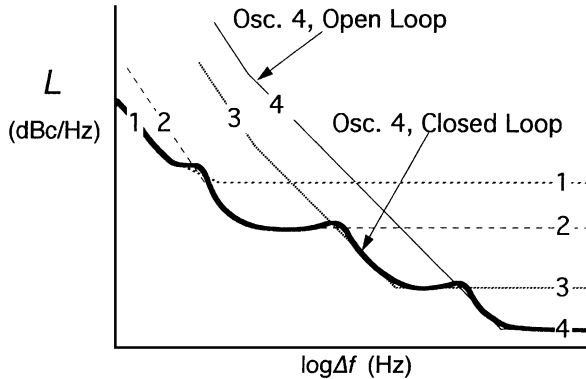


Fig. 12.12 Multiple loops.

12.6 MULTIPLE LOOPS

Sometimes more than one loop is used to minimize phase noise. Figure 12.12 shows an example in which four loops are employed. Oscillator 4 might be a microwave oscillator that is locked to oscillator 3, a surface-acoustic-wave (SAW) oscillator whose frequency has been multiplied to the desired output frequency. The higher Q of the SAW oscillator gives it better phase noise, but the multiplication process increases the noise floor (as well as other phase noise), which is due to thermal noise, so that it exceeds the noise floor of oscillator 4. The loop has an optimum bandwidth near the point at which the noise levels cross. Low-frequency noise modulation is further improved by locking the SAW oscillator to a frequency-multiplied crystal oscillator in a similar manner and with similar results. The crystal oscillator may then be locked to an atomic standard that has superior close-in performance. For each of the three loops, the reference has better close-in purity than the VCO because it is of a type that has higher effective Q . The VCO, on the other hand, has better far-out purity (at least after frequency multiplication).

12.7 EFFECTS OF NOISE INJECTED ELSEWHERE

There are many noise sources other than the VCO and the reference. For example, the op-amp in the loop filter may produce enough noise to dominate in certain regions of the output spectrum. In such cases, use the procedures of Section 7.4. While they are described in terms of discrete modulations, the extension to density is straightforward. Remember to use the square of the transfer gains, however, since S is a power density.

Example 12.6 Component Noise What is the contribution to the noise of the PLL output, at modulation frequencies well above the loop bandwidth, due to the 4-k Ω resistor in Fig. 12.13.

Thermal noise density is -174 dBm/Hz = -204 dBW/Hz. This is what the resistor would deliver to a matched load, in which case half of the voltage appears across the

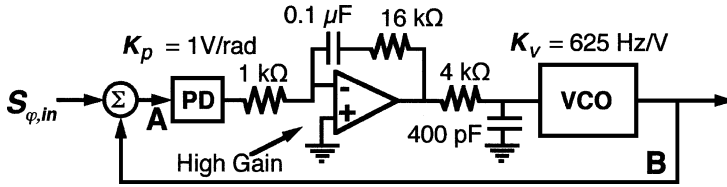


Fig. 12.13 Loop with component noise.

load. Therefore, the mean square noise voltage density \tilde{v}_T^2 of its equivalent internal noise generator is obtained from the expression for power, \tilde{v}^2/R , as

$$\begin{aligned}
 (\tilde{v}_T/2)^2 &= RN_T = (4000 \Omega) 10^{-20.4} \text{ W/Hz}, \\
 \tilde{v}_T^2 &= 6.37 \times 10^{-17} \text{ V}^2/\text{Hz},
 \end{aligned}$$

where \tilde{v}_T is the open-circuit rms thermal noise voltage. Multiplying by $K_v^2 = (625 \text{ Hz/V})^2$, we obtain $2.5 \times 10^{-11} \text{ Hz}^2/\text{Hz}$. At high frequencies the voltage is reduced, due to the low-pass filter, by the impedance ratio, $1/(R\omega_m C) = 99,471 \text{ Hz}/f_m$. The phase power spectral density is therefore

$$S_\varphi = S_f/f_m^2 = (2.5 \times 10^{-11} \text{ Hz})(99,471 \text{ Hz})^2/f_m^4 = 0.25 \text{ rad}^2 \text{ Hz}^3/f_m^4.$$

At frequencies well above the loop bandwidth, $1 - H \approx 1$, so this is the value of S_φ at the output. For example, the single-sideband level would be -163 dBc/Hz , corresponding to $S_\varphi = 10^{-16} \text{ rad}^2/\text{Hz}$, at $f_m = 7071 \text{ Hz}$, if that is much greater than the loop bandwidth (with which we will not concern ourselves for this example).

Some resistors have troublesome flicker ($1/f$) noise also, but it is proportional to the direct current (DC) voltage across the resistor. Since the steady-state value of that voltage is zero, $1/f$ noise is probably small, especially at higher frequencies.

12.8 MEASURING PHASE NOISE

There are three basic ways to measure phase noise [Egan, 2000, pp. 507–512]:

- We can measure it directly, using a phase detector.
- We can measure frequency deviation, using a frequency discriminator, and convert S_f to S_φ using Eq. (11.5).
- We can observe the noise sidebands directly using a spectrum analyzer or other receiver, and, if we know the sidebands are due to phase modulation (PM) [rather than amplitude modulation (AM), which we will consider in Chapter 13], use Eq. (11.23) to convert L_φ to S_φ .

In the first two methods the signal is demodulated before the noise is measured, whereas in the third method the spectrum of the signal is observed directly without

demodulation. Direct observation using a spectrum analyzer is often the most convenient method, but it is often less sensitive than the phase detector. That is, with the phase detector, lower noise levels can be measured before the measurement is limited by the noise of the instrumentation.

The discriminator sensitivity has a different dependence on modulation frequency than do the other two devices. As modulation frequency increases, the output from a discriminator increases relative to the output from the others, as is evident from Eq. (11.5) and Fig. 12.14. With the spectrums shown in Fig. 11.6, a phase detector will produce a voltage that increases at lower frequencies where phase noise climbs. Thus, the output from the phase detector will grow relative to the noise of the amplifier or detector that follows it. The discriminator output will be flat in region 3 but will increase with frequency in region 1, so its output will tend to overcome subsequent instrument noise at higher values of f_m . However, noise that precedes the discriminator is partially seen as phase noise by the discriminator. Therefore, it will tend to be the limitation on sensitivity at high values of f_m , climbing along with the noise being measured.

Although the spectrum analyzer usually serves as the final measurement device in all of these methods, we give its name to the method where it is not preceded by a demodulator but rather measures the signal directly. While the similarity of response shapes for the phase detector and spectrum analyzer methods might lead us to expect identical sensitivities, spectrum analyzer performance is limited, in the latter method, by the presence of the strong un-demodulated signal and by requirements placed on it to cover a broad frequency range.

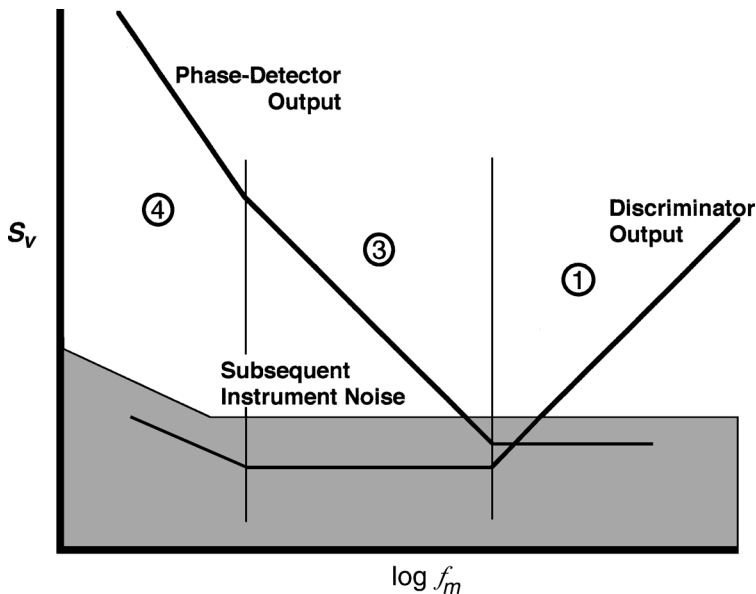


Fig. 12.14 Voltage power spectral density from phase noise measurements in the noise regions of Fig. 11.6. The various curves will shift depending on details of the noise and the measurements.

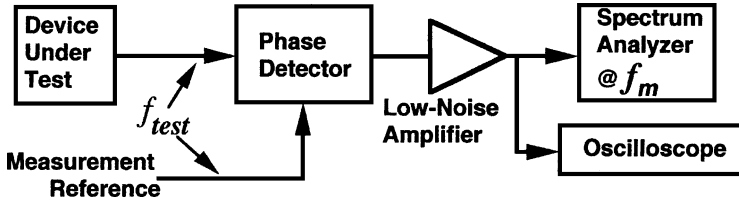


Fig. 12.15 Phase noise measurement with a phase detector.

12.8.1 Using a Phase Detector

The test setup is shown in Fig. 12.15. The reference must be maintained at the proper phase relationship to the signal being tested. Ideally the reference is much cleaner than the device under test (DUT). The DUT will often be an oscillator or a PLL. The spectrum analyzer is tuned to the noise modulation frequency. The noise figure of the low-noise amplifier, and, to some degree, the spectrum analyzer, along with K_p of the test phase detector, set a minimum measurable level for S_ϕ .

The power spectral density S_p , read on the analyzer at f_m , is converted to $S_\phi(f_m)$ using the PD gain constant,

$$S_\phi(f_m) = \frac{S_v(f_m)}{K_p^2} \tag{12.2}$$

Here $S_v = R_0 S_p$, where R_0 is the impedance of the measurement system.²

Note that when variables such as S_p , S_v , and K_p are used in an expression, they must all generally be referenced to the same location in the circuit, for example, the output of the phase detector or the input to the analyzer.

If the effective noise density of the amplifier plus analyzer, in the vicinity of the signal frequency, equals N_0 , the resulting measured output will be the same as is produced by a PPSD of $S_{\phi n} = R_0 N_0 / K_p^2$. Levels of S_ϕ lower than this measurement noise floor cannot be measured accurately. Having considered that, the limitations that we will consider in the rest of this section will be due to additional noise from other sources.

12.8.1.1 Calibration

Measuring K_p Directly. We can ascertain the PD gain constant K_p by observing, on the oscilloscope, the slope of the PD characteristic at the relative phase (e.g., 90° for a balanced mixer) to be maintained during the test. We can do that by changing the phase of one of the inputs to the PD with a phase shifter or line stretcher or by offsetting the reference frequency slightly from the frequency of the DUT and observing the characteristic as the phase changes. We must then maintain the average (DC) voltage during test at a value that will give that measured slope.

²This is simply an expression of the relationship: power = (rms voltage)²/(resistance).

When the PD Characteristic is Sinusoidal. If the PD characteristic is truly sinusoidal (note Fig. 3.13), K_p can be measured using the spectrum analyzer. When the reference frequency is offset slightly from the frequency of the DUT, the root mean square (rms) voltage from the PD will be, by Eq. (1.10),

$$\tilde{v}_{\text{test}} = \frac{K'_p \text{ rad}}{\sqrt{2}}. \quad (12.3)$$

If we are to make the measurement of S_φ at the highest value of K_p , which is K'_p , we can combine the last two equations to obtain

$$S_\varphi(f_m) = \frac{S_v(f_m)}{2\tilde{v}_{\text{test}}^2}. \quad (12.4)$$

The level that we observe during the measurement will be $\sigma_v^2 = S_v B_n$, where B_n is the effective noise bandwidth (to be defined exactly in Chapter 14) of the analyzer. Thus we can write Eq. (12.4) as

$$S_\varphi(f_m) = \left[\frac{\sigma_v^2(f_m)}{\tilde{v}_{\text{test}}^2} \right] \frac{1}{2B_n} \quad (12.5)$$

or, in decibels,

$$\begin{aligned} S_\varphi(f_m)|_{\text{dBr/Hz}} &= [\sigma_v(f_m)]_{\text{dBV}} - 20 \text{ dBV} \log(\tilde{v}_{\text{test}}/V) \\ &\quad - 10 \text{ dB} \log(B_n/\text{Hz}) - 3 \text{ dB}. \end{aligned} \quad (12.6)$$

The ratio of the level observed on the analyzer during measurement to the level observed during calibration, when the signals were offset by f_m , is enclosed in brackets in the last two equations. Thus we set the level read during calibration as a reference and compare measured noise levels to that reference. Then we divide by $2B_n$ to get density.

Example 12.7 Calibration With a Sinusoidal Characteristic In Fig. 12.15, the reference signal is offset 10 kHz from the output frequency of the DUT. Because the reference is weak compared to the other PD input, the resulting PD output is a good sinusoid at 10 kHz with 0 V average (DC) value. The spectrum analyzer is adjusted so the resulting 10 kHz line is at the top line on the display. Then the reference signal is set to the same frequency as the DUT and adjusted for 0 V DC output. The level of the noise at 10 kHz is 65 dB below the top line. The analyzer's noise bandwidth is 30 Hz. What is S_φ (10 kHz)?

Using Eq. (12.6),

$$S_\varphi(f_m) = -65 \text{ dB} - 10 \text{ dB} \log(30) - 3 \text{ dB} = -83 \text{ dBr/Hz}.$$

By Modulation. We can also calibrate by phase modulating one of the inputs by a known amount. In the absence of a calibrated phase modulator, we can determine the phase deviation by observing the sideband level of the modulated signal using a second spectrum analyzer (refer to Section 11.6).

Example 12.8 Calibration Using Modulation We phase modulate one of the two signals to be compared and observe the response on the spectrum analyzer. The modulation frequency is 20 kHz and a single line is produced at that frequency on the analyzer in Fig. 12.15. The gain of the analyzer is adjusted to place this line at 0 dB on the display. The spectrum of the modulated signal, as observed directly on a second analyzer, consists of a carrier line and two sidebands, 40 dB smaller and offset 20 kHz on either side. The modulation is removed and the noise level is observed, on the first analyzer, to be at -80 dB. The analysis bandwidth is 100 Hz. What is $S_\varphi(20 \text{ kHz})$?

Based on Eq. (11.21) and Fig. 11.8, the variance of the modulated signal during calibration is

$$\sigma_\varphi^2 = 2 \times 10^{-40 \text{ dB}/10 \text{ dB}} \text{ rad}^2 = 0.0002 \text{ rad}^2.$$

In a 100-Hz measurement bandwidth this is equivalent to a density of

$$S_\varphi = 0.0002 \text{ rad}^2/100 \text{ Hz} = 2 \times 10^{-6} \text{ rad}^2/\text{Hz},$$

or -57 dBBr/Hz . Since the measured noise is 80 dB weaker, its level is

$$S_\varphi(10 \text{ kHz}) = -57 \text{ dBBr/Hz} - 80 \text{ dB} = -137 \text{ dBBr/Hz}.$$

Calibration can be performed as a function of f_m if necessary.

12.8.1.2 Obtaining a Measurement Reference. Sometimes very stable and accurate sources (e.g., atomic standards) can be compared without locking. Absent that, however, how do we maintain a constant phase relationship between the DUT and the reference? There are several methods.

Comparing Output Phase to Reference Phase. If the DUT is a PLL, we can use the PLL's reference as the measurement reference (Fig. 12.16). Phase noise of the reference will appear at the PLL output multiplied by H . Since the PD subtracts the reference phase from this transformed reference phase, the net result is multiplication of the reference phase by $(H - 1)$, leading to a measured phase of

$$\varphi_m = \varphi_{\text{out},1} - \varphi_{\text{ref}} \quad (12.7)$$

$$= \varphi_{\text{res},1} + \varphi_{\text{ref}}(H - 1), \quad (12.8)$$

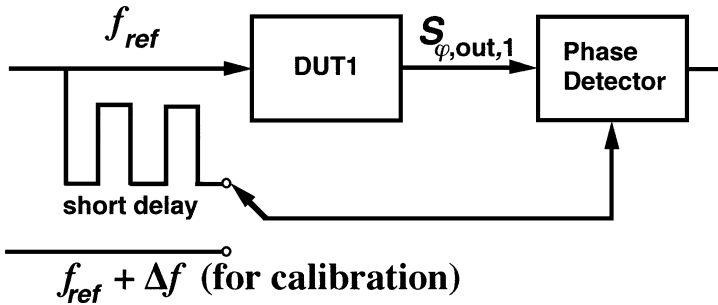


Fig. 12.16 Measurement referred to the DUT’s reference. The short delay is used to obtain the correct relative phase at the PD.

where $\varphi_{res,1}$ is the residual phase noise, defined as the phase noise with a perfect reference (i.e., when $\varphi_{ref} = 0$). The corresponding PPSDs are

$$S_{\varphi,m} = S_{\varphi,out,1} + S_{\varphi,ref} \tag{12.9}$$

$$= S_{\varphi,res,1} + S_{\varphi,ref}|1 - H|^2. \tag{12.10}$$

The result can be considered a measurement of $S_{\varphi,res}$ corrupted by a measurement noise, $S_{\varphi,ref}|1 - H|^2$, which is small if $f_m \ll f_L$, since H is then approximately 1.

Comparing Output Phase to a Locked Source. Another way to obtain a measurement reference is to lock a clean oscillator to the DUT with a bandwidth narrow compared to the f_m of interest so the locked oscillator does not track the noise of interest. In Fig. 12.17 the measurement point (output) is at the error node of the phase-locked oscillator (PLO) so the phase measured there is

$$\varphi_m = \varphi_{res,PLO} + \varphi_{out,1}(1 - H) \tag{12.11}$$

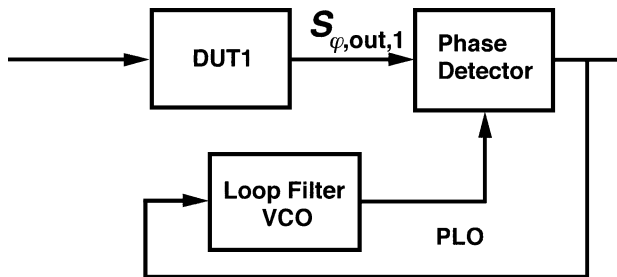


Fig. 12.17 Measurement referred to a PLO.

and the PPSD is

$$S_{\varphi,m} = S_{\varphi,\text{res,PLO}} + S_{\varphi,\text{out},1} |1 - H|^2. \tag{12.12}$$

If $S_{\varphi,\text{res,PLO}}(f_m)$ is small and $|1 - H(f_m)|^2 \approx 1$, then $S_{\varphi,m} \approx S_{\varphi,\text{out},1}$. We can extend the measurements to lower frequencies by measuring $[1 - H(f_m)]$ and using the measured value to cancel its effect mathematically.

Comparing Output Phases of Similar Sources. When a reference that is clean compared to the DUT cannot be obtained, perhaps because the DUT is very clean, two “identical” DUTs are sometimes compared (Fig. 12.18). For example, we might lock two PLLs of identical design to the same reference and compare their output phases. The worst-case noise for each is the measured noise, but often each is assumed to be contributing half of the noise power. If three DUTs having similar noise levels are compared two at a time, the three resulting measurements provide enough information to solve for the noise level of each DUT.

The phase from DUT1 is

$$\varphi_{\text{out},1} = \varphi_{\text{res},1} + \varphi_{\text{ref}} H_1 \tag{12.13}$$

and similarly for $\varphi_{\text{out},2}$. The difference between the two is

$$\varphi_m = \varphi_{\text{out},1} - \varphi_{\text{out},2} \tag{12.14}$$

$$= \varphi_{\text{res},1} - \varphi_{\text{res},2} + \varphi_{\text{ref}}(H_1 - H_2). \tag{12.15}$$

The corresponding PPSD is

$$S_{\varphi,m} = S_{\varphi,\text{res},1} + S_{\varphi,\text{res},2} + S_{\varphi,\text{ref}} |H_1 - H_2|^2. \tag{12.16}$$

Most of the phase noise of the common reference is either attenuated, at $f_m \gg f_L$, where H is small, or canceled, at $f_m \ll f_L$, where H_1 and H_2 are both very close to

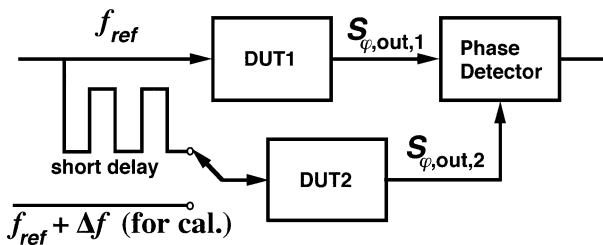


Fig. 12.18 Comparing two DUTs. The short delay is used to obtain the correct relative phase at the PD.

one. At $f_m \approx f_L$, however, attenuation is small and good cancellation depends on a good match between H_1 and H_2 .

Example 12.9 Measurement with a Phase Detector We are to measure the phase noise of a high-spectral-purity phase-locked oscillator using the setup of Fig. 12.15. The output frequency is a multiple of the reference frequency (see Section 10.2.1), so the PLO and its reference cannot be measured against each other directly. We decide to compare two “identical” PLOs, as in Fig. 12.18. Initially we replace DUT2 by a signal generator set to the same amplitude³ as DUT2 but offset in frequency by 2 kHz. The output of the balanced-mixer PD, as observed on the oscilloscope, is shown in Fig. 12.19 where the slope at the intended operating point is

$$K_p = 2.5 \text{ V}/0.2 \text{ cycle} = 12.5 \text{ V/cycle}.$$

We then connect PLO2 as DUT2. We connect both DUTs to the same PLL reference and adjust the lengths of the coaxial cables to obtain the required 90° phase shift between the inputs to the test PD. The DC output from the PD is 0 V, corresponding to the operating point where K_p was measured. The spectrum analyzer indicates a power of -55 dBm at 2 kHz when its analysis bandwidth is 10 Hz, which is also approximately its noise bandwidth, so the indicated power density is

$$S_p(2 \text{ kHz}) = \frac{10^{-55/10} \text{ mW}}{10 \text{ Hz}} = 3 \times 10^{-7} \text{ mW/Hz} = 3 \times 10^{-10} \text{ W/Hz}.$$

In a $50\text{-}\Omega$ system, this corresponds to a voltage power spectral density of

$$S_v(2 \text{ kHz}) = 50\Omega S_p(2 \text{ kHz}) = 1.5 \times 10^{-8} \text{ V}^2/\text{Hz}.$$

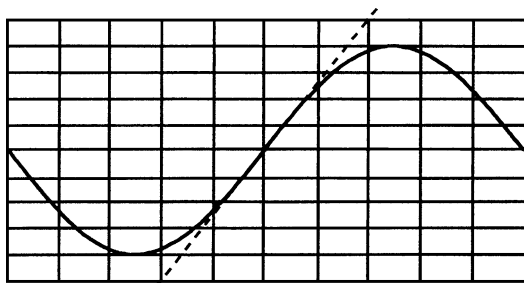


Fig. 12.19 PD calibration output at 0.5 V/division.

³ This is equivalent to calibration by offsetting the reference to one of the PLOs, as in Fig. 12.18, as long as the PLO's amplitude is maintained constant.

From Eq. (12.2), this implies a PPSD of

$$\begin{aligned}
 S_\varphi(2 \text{ kHz}) &= \frac{1.5 \times 10^{-8} \text{ V}^2/\text{Hz}}{(12.5 \text{ V/cycle})^2} = 9.6 \times 10^{-11} \frac{\text{cycle}^2}{\text{Hz}} \left(\frac{2\pi \text{ rad}}{\text{cycle}} \right)^2 \\
 &= 3.8 \times 10^{-9} \text{ rad}^2/\text{Hz}.
 \end{aligned}$$

This corresponds to -84 dBc/Hz . If we attribute half of the noise to each PLO, each has a PPSD at 2 kHz of $1.9 \times 10^{-9} \text{ rad}^2/\text{Hz}$ or -87 dBc/Hz .

We attenuate the signal from DUT2 until the noise level on the spectrum analyzer stops decreasing. This occurs at -90 dBm , so we know that the noise due to the amplifier and analyzer is at least 35 dB below the measured level and, therefore, makes no significant contribution to measurement inaccuracy. Assuming the last term in Eq. (12.16) is small, the noise we have measured is primarily the “residual” noise of the two PLOs.

12.8.2 Using a Frequency Discriminator

One advantage of the use of a frequency discriminator is that it does not require a reference. In a sense it provides its own phase reference. We can see this from a particular implementation of the discriminator (Fig. 12.20) in which the output from the DUT is split and both parts are sent to the phase detector but one path is delayed by T relative to the other. If $T \ll 1/\omega_m$, the phase detector sees a phase difference of $\varphi = \omega T$. This relationship implies

$$S'_\varphi = S_\omega T^2, \tag{12.17}$$

where S'_φ is the PPSD measured at the PD and S_ω is the actual frequency PSD. Combining with Eq. (11.5), we obtain the PPSD S_φ of the DUT in terms of S'_φ as

$$S_\varphi = S_\omega/\omega_m^2 = S'_\varphi/(\omega_m T)^2. \tag{12.18}$$

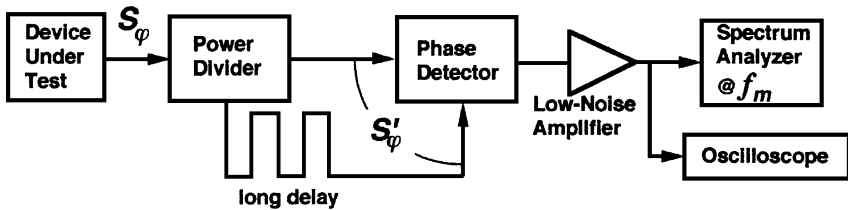


Fig. 12.20 Measurement of S_φ with a frequency discriminator.

The maximum measurable deviation is limited by the assumption that K_p remains constant. Therefore, the integrated noise should not be so high that the instantaneous operating point moves from the linear region in Fig. 12.19. If this restriction is met, the sensitivity, by Eq. (12.18), increases with T , and thus with the differential cable length. While $S'_\varphi > S_\varphi$ suggests greater sensitivity with a discriminator than with a phase detector (alone), for S'_φ to exceed S_φ at small values of ω_m , the delay cable must be quite long (roughly 1000 ft at 100 kHz, proportionally longer at lower frequencies). As the delay increases, so too do losses, and additional amplifiers may be necessary along the cable to prevent thermal noise from overwhelming the noise being measured.

12.8.3 Using a Spectrum Analyzer or Receiver

The spectrum analyzer is a swept radio receiver plus a display that indicates the received power versus frequency. Curves *A* and *B* in Fig. 12.5 illustrate signals displayed on a spectrum analyzer. It can be used to measure directly the relative sideband amplitudes, from which can be obtained S_φ , by Eq. (11.23) as long as the condition of Eq. (11.29) is met. The spectral purity of the analyzer may be compromised due to other requirements (such as the necessity to be able to sweep over a wide frequency range), so a discrete, single-frequency-at-a-time version may be more sensitive. Of course, a low noise amplifier can be used to improve the sensitivity, but there may then be a problem of overdriving the analyzer so the problem becomes one of dynamic range, sensitivity in the presence of a strong signal. [With the phase detector and discriminator, strong signals are eliminated by cancellation (adjusting for 0 V DC output) and simple filtering, low-passing and possibly high-passing, at the PD output.] We can enhance the dynamic range by using a narrow-band rejection filter (notch filter) at the operating frequency of the DUT to remove most of the power from the center of the spectrum, leaving the sidebands to be measured. Of course, it is necessary then to determine what the total power would have been had it reached the analyzer with the same attenuation as the sidebands, but this is not a fundamentally difficult problem. The spectrum analyzer or receiver, enhanced by filtering when necessary, is the usual choice for measuring the noise floor far from spectral center.

CHAPTER 13

REPRESENTATION OF ADDITIVE NOISE

We have studied, in the two previous chapters, how the loop responds to phase noise. However, many problems require us to determine how the loop acts in the presence of additive noise. Additive noise is noise that accompanies a signal rather than modulation on that signal. So we can apply, to this problem, what we have already learned, we will determine how to represent additive noise as equivalent phase noise. We will see that half of the noise power can be treated as if it were the result of phase modulation.

13.1 GENERAL

As before, we will treat noise as a collection of very narrow-band signals characterized by root mean square (rms) values \tilde{v} and having unknown phase. Figure 13.1 illustrates how we can represent a voltage of any phase as two voltages in mutual phase quadrature, a real component v_r and an imaginary component v_i . The total power (1Ω assumed) is $\tilde{v}_T^2 = \tilde{v}_r^2 + \tilde{v}_i^2$. Since the phase is random, the power is as likely to be in the real component as in the imaginary component, so each has a mean value of half of the total:

$$P_i = P_r = P_T/2. \quad (13.1)$$

We now wish to represent the real and imaginary noise components as sidebands on some central carrier signal so we can relate these sidebands to modulation of that

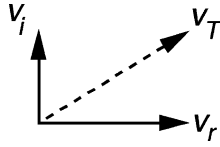


Fig. 13.1 Voltage phasor decomposed into real and imaginary parts.

signal. We will be able to interpret some of the modulation as amplitude modulation (AM) and the remainder as phase modulation (PM). The former has only a secondary effect (K_p may increase with amplitude), and we already know how the loop responds to PM. Since phase and amplitude modulation are recognizable by the even or odd sign relationship (see Appendix 21.A) between sidebands (which are offset from the carrier by the modulation frequency), we first consider how to decompose the noise sidebands into even and odd pairs.

Figure 13.2 shows how we can do this. The larger sideband is equivalent to the sum of two new sidebands, which have the same sign. The smaller, symmetrically located, sideband is equivalent to the difference between the same two new sidebands. Thus we can decompose the original pair into an even (same sign) and an odd (different signs) pair. (In fact, we could do this one noise sideband at a time, rather than in pairs, as shown in Appendix i13.A.) As before, each component represents the rms voltage corresponding to the power in a narrow or differential bandwidth.

Again, due to randomness, the likelihood of the even pair is the same as that of the odd pair. Thus the sideband power at any pair of frequencies, located at $f_c \pm \Delta f$, can be equated to eight equally likely, and therefore equally powerful, components. Four are real and four are imaginary. For each of these two types, two form an even pair and two form an odd pair. For each of these pairs, one component is located at $f_c + \Delta f$ and the other is located symmetrically about f_c at $f_c - \Delta f$. We will only treat noise for which $0 < (f_c \pm \Delta f) < 2f_c$ [Blanchard 1976, p. 140]. If the noise power density is symmetrical about the carrier (as it will be with white noise) the total noise power in two differential-width sidebands located at $f_c \pm \Delta f$ is $2N_0 df$, where N_0 is the one-sided noise power density. This is divided equally among the eight equally likely components so each will have power $N_0(f)df/4$. Thus the total power of the four components at the frequency of each single sideband equals the power of the sideband, $N_0(f) df$.

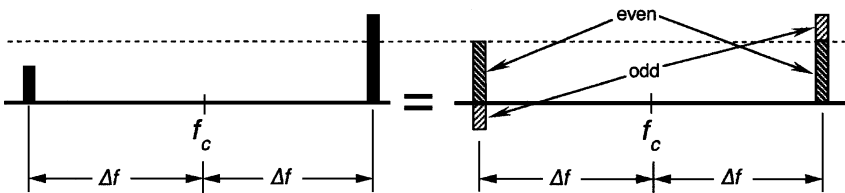


Fig. 13.2 Decomposition of a pair of sidebands into the sum of an even pair and an odd pair.

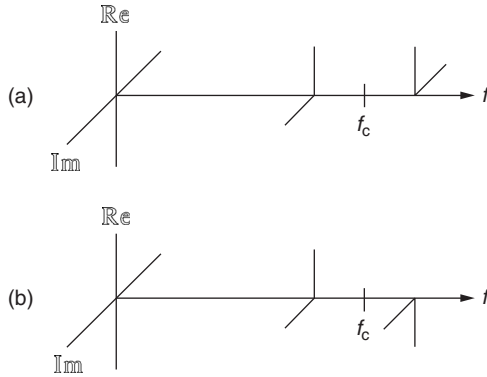


Fig. 13.3 Decomposition into eight sidebands. Only the positive half of the Fourier spectrum is shown.

The eight components are shown in Fig. 13.3.¹ The set at Fig. 13.3a corresponds to AM sidebands on a real (cosine) carrier at f_c . The real pair is produced by a cosine modulation and the imaginary pair by a sine modulation. However, since AM has no first-order effect on a locked loop, the sidebands shown at Fig. 13.3b, which represent half of the total additive noise power, are of primary importance.

13.2 PHASE MODULATION ON THE SIGNAL

Figure 13.3b represents phase modulation (PM) or, equivalently, frequency modulation (FM) sidebands on a real (cosine) carrier as long as $m \ll 1$ rad. The representation is restricted to small m because there would otherwise be multiple sidebands and a reduction in the carrier amplitude. When the noise is true phase modulation, the phase error in the loop must be small if the loop is to be approximately linear. The input phase modulation can be quite strong as long as the VCO tracks it to keep the phase error small. Here, however, even if the phase error is kept small, the input phase deviation must also be kept small for the noise to be accurately modeled as PM.

The real pair in Fig. 13.3b corresponds to sine modulation and the imaginary pair to cosine modulation. If the noise power is symmetrical about the signal at f_c , the total power of the two PM components at each sideband frequency equals $N_0(f_c \pm \Delta f) df/2$. Since $L(\Delta f) df$ is the ratio of the sideband power at $+\Delta f$ or $-\Delta f$ to the carrier (signal) power P_c , the single-sideband power density representing phase noise L_φ is

$$L_\varphi(\Delta f) df = L(\Delta f) df/2 = N_0(f_c \pm \Delta f)df/(2P_c), \tag{13.2}$$

$$L_\varphi(\Delta f) = L(\Delta f)/2 = N_0(f_c \pm \Delta f)/(2P_c). \tag{13.3}$$

¹ This can be considered to be the positive half of the Fourier spectrum or to be an analytic signal representation. In the latter case the spectral line at ω represents $\exp(j\omega t)$, the real part of which is the signal.

Based on Eq. (11.23), and under condition that Eq. (11.29) is met, we also have the phase power spectral density in terms of the ratio of additive noise density to signal power,

$$S_\varphi(f_m) = N_0(f_c \pm f_m)/P_c. \quad (13.4)$$

Thus, whereas for phase noise we have $S_\varphi(f_m) = 2L_\varphi(\Delta f)$, when the noise is additive it is also true that $S_\varphi(f_m) = L(\Delta f)$.

Example 13.1 PPSD from Additive Noise Power Density A signal at a frequency of 10 MHz and power 6 dBm is embedded in noise at a density of -120 dBm/Hz and extending ± 1 MHz from the signal (at which point it is attenuated rapidly by a filter). This is equivalent to a single-sideband density of

$$\begin{aligned} L(\Delta f < 1 \text{ MHz}) &= N_0(10 \text{ MHz} \pm \Delta f)/P_c = (10^{-12} \text{ mW/Hz})/(10^{0.6} \text{ mW}) \\ &= 2.5 \cdot 10^{-13} / \text{Hz} \\ &\Rightarrow -120 \text{ dBm/Hz} - 6 \text{ dBm} = -126 \text{ dBc/Hz}. \end{aligned}$$

The equivalent phase power spectral density is the same except for units,

$$S_\varphi(f_m < 1 \text{ MHz}) = -126 \text{ dBr/Hz}.$$

The requirement for validity, Eq. (11.29), is, for this case,

$$\begin{aligned} \int_0^{1 \text{ MHz}} 10^{-12.6} \text{ rad}^2/\text{Hz} df &= 10^6 \text{ Hz} \cdot 10^{-12.6} \text{ rad}^2/\text{Hz} \\ &= 10^{-6.6} \text{ rad}^2 = 2.5 \times 10^{-7} \text{ rad}^2 \ll 1 \text{ rad}^2. \end{aligned}$$

Because of the natural thermal noise floor and because of the equivalence between additive noise and phase noise that was shown above, all oscillator spectrums eventually drop to a flat region, such as those that can be seen in Figs. 11.6 and 12.4.

Even if m does not meet the assumed restriction of smallness, additive noise still produces equivalent phase noise. The zero crossings of the waveform still vary in a noiselike manner. As m becomes larger, however, the theory described above becomes less justifiable as a means of computing the value of the equivalent S_φ from a given L_φ . In such cases we can use a second representation of the spectrum in Fig. 13.3b. We will see results that are the same as if we had simply ignored the limitations on m , but we will have a better understanding of the process.

13.3 MULTIPLICATIVE MODULATION ON QUADRATURE CARRIERS

In this second representation, the components in Fig. 13.3*b* are represented as AM sidebands on an imaginary (sine) carrier. As in Fig. 13.3*a*, the sidebands in the same plane as the carrier are even while those in the quadrature plane are odd. And, as before, the former represent cosine modulation and the latter sine modulation. This time, however, suppressed-carrier AM is employed so no carrier signal is required. This is multiplicative modulation. These sidebands exist independently of a desired signal.

Using trigonometric identities, we can determine the spectral components produced by multiplicative modulation. We will represent noise voltages as \tilde{v}_{mab} , where a and b are r or i for real or imaginary, and represent the phase (cosine or sine) of the carrier and of the modulating signal, respectively. Modulation of a cosine can be represented as

$$\begin{aligned} & \sqrt{2}\{\tilde{v}_{mrr} \cos \omega_m t + \tilde{v}_{mri} \sin \omega_m t\} \cos \omega_c t \\ &= \frac{\tilde{v}_{mrr}}{\sqrt{2}} [\cos(\omega_c + \omega_m)t + \cos(\omega_c - \omega_m)t] \\ & \quad + \frac{\tilde{v}_{mri}}{\sqrt{2}} [\sin(\omega_c + \omega_m)t - \sin(\omega_c - \omega_m)t]. \end{aligned} \quad (13.5)$$

We can identify the right side with the components in Fig. 13.3*a*. Similarly, the components in Fig. 13.3*b* can be represented as

$$\begin{aligned} & \sqrt{2}\{\tilde{v}_{mir} \cos \omega_m t + \tilde{v}_{mii} \sin \omega_m t\} \sin \omega_c t \\ &= \frac{\tilde{v}_{mir}}{\sqrt{2}} [\sin(\omega_c + \omega_m)t + \sin(\omega_c - \omega_m)t] \\ & \quad + \frac{\tilde{v}_{mii}}{\sqrt{2}} [\cos(\omega_c + \omega_m)t - \cos(\omega_c - \omega_m)t]. \end{aligned} \quad (13.6)$$

The power in each component [e.g., $(\tilde{v}_{mii}/\sqrt{2}) \cos(\omega_c + \omega_m)t$] on the right side of these equations is $\tilde{v}_{mab}^2/4$ (where $\tilde{v}_{mab} = \tilde{v}_{mii}, \tilde{v}_{mir}$, etc.). But we have seen that each component has power $N_0 df/4$ so

$$N_0(f_c \pm f_m) df = \tilde{v}_{mab}^2. \quad (13.7)$$

This means that the total modulation power at frequency ω_m that is multiplying $\sin \omega_c t$ is

$$\langle [\sqrt{2}\{\tilde{v}_{mir} \cos \omega_m t + \tilde{v}_{mii} \sin \omega_m t\}]^2 \rangle = \langle \tilde{v}_{mir}^2 \rangle + \langle \tilde{v}_{mii}^2 \rangle = 2N_0(f_c \pm f_m) df, \quad (13.8)$$

where $\langle \rangle$ indicates mean value, and an average has here been taken over the period of the sinusoids.

By the same process we can show that the modulation power multiplying $\cos \omega_c t$ has the same expected value. Thus we can write the density of the noise modulation power N_m on either carrier in terms of the noise sideband power density N_0 as

$$N_m(f_m) = 2N_0(f_c \pm f_m). \quad (13.9)$$

13.4 NOISE AT THE PHASE DETECTOR OUTPUT

Now we should answer an important question: What voltage does this noise produce at the output of the phase detector? Assume a sinusoidal phase detector characteristic and that the voltage-controlled oscillator (VCO) is in quadrature with the signal (otherwise there will be a contribution from AM). Then, by the first representation, the phase modulation $S_\varphi(f_m) df_m$ will produce noise power

$$\langle u_{1n}^2 \rangle = (K'_p)^2 S_\varphi(f_m) df_m = (K'_p)^2 N_0(f_c \pm f_m) df / P_c. \quad (13.10)$$

Let us compare this to the result obtained using the second representation of additive noise. For convenience, we choose the time origin such that the sinusoidal carrier is in phase² with the VCO and assume that the mixer acts like a multiplier.

We represent the modulation voltage due to noise in a narrow bandwidth as

$$v_m(t) = (2N_m df)^{0.5} \sin(\omega_m t + \theta). \quad (13.11)$$

Here θ is an unknown phase and the modulation power is N_m , as before. The PD output due to noise is proportional to the product of the noise voltage, represented by the modulated sine carrier, and the VCO voltage, both at frequency $\omega = \omega_c$. This is

$$u_{1n} = \eta_p [v_m(t) \sin \omega t] [V_v \sin \omega t], \quad (13.12)$$

where V_v is the VCO signal amplitude and η_p indicates the phase detector's efficiency as a multiplier. Combining the last two equations, the noise power at the PD output is

$$\langle u_{1n}^2 \rangle = \eta_p^2 \langle [(2N_m df)^{0.5} \sin(\omega_m t + \theta) (\sin \omega t V_v \sin \omega t)]^2 \rangle. \quad (13.13)$$

The first sine is the noise modulation. The other two are the noise carrier and the VCO. The product of those two is first averaged to eliminate the second harmonics, a process that occurs at the PD output. This leaves a DC component of 0.5. Then the mean squared value of the modulation is taken to obtain

$$\langle u_{1n}^2 \rangle = \eta_p^2 N_m df V_v^2 / 4 \quad (13.14)$$

$$= \eta_p^2 N_0(f_c \pm f_m) df V_v^2 / 2. \quad (13.15)$$

² We do this for simplicity. The same results can be obtained by assuming an arbitrary angle but the expressions are longer.

To put this in more useful form we must obtain a value for η_p . We do that by writing the PD output voltage due to the desired signal and relating η_p to its signal strength. The output produced by the VCO and in-phase signal has an average value (i.e., removing the high-frequency terms as usual)

$$\langle u_{1s} \rangle = \eta_p \langle V_v \sin \omega t (2P_c)^{0.5} \sin \omega t \rangle = \eta_p V_v (2P_c)^{0.5} / 2, \quad (13.16)$$

where V_v is the VCO amplitude. Note that this is independent of the amount of noise. Since the voltage from a sinusoidal phase detector, when the signals are in phase, is K'_p radians, $\langle u_{1s} \rangle$ is equal to K'_p radians so

$$(K'_p)^2 = (\eta_p / \text{rad})^2 V_v^2 P_c / 2 \quad (13.17)$$

or

$$\eta_p^2 = 2(K'_p \text{ radians})^2 / (V_v^2 P_c). \quad (13.18)$$

Substituting this value into (13.15) gives the mean square noise voltage as

$$\langle u_{1n}^2 \rangle = (K'_p \text{ radians})^2 N_0(f_c \pm f_m) df / P_c. \quad (13.19)$$

Notice that we have obtained the same expression for $\langle u_{1n}^2 \rangle$ using either representation [compare to Eq. (13.10)].

The assumption that the phase detector operates, relative to the signal, as a multiplier with some efficiency factor η_p is fulfilled if the phase detector is a balanced mixer used in the normal fashion for mixers, with a relatively powerful local oscillator (LO) and total input at the signal port that is small enough that amplitudes are preserved in the mixing process (perhaps 10 dB or more below the LO power).

13.5 RESTRICTIONS ON THE NOISE MODELS

Here we discuss some conceptual limitations on the models that we have developed.

We have indicated that the PM model cannot be used when the additive noise is too great. The multiplicative (AM) representation, on the other hand, does not have this restriction. Equation (13.19) does not depend upon $\int (N_0 / P_c) df$ being small.

Figure 13.4 shows u_{1n} entering the loop after the phase detector. Since we have formulated all of our previous theory for PM, there is an advantage to interpreting u_{1n} in those terms also. It is the voltage that would be produced, in the linearized model, by a phase deviation of

$$n'(t) \equiv u_{1n}(t) / K_p \quad (13.20)$$

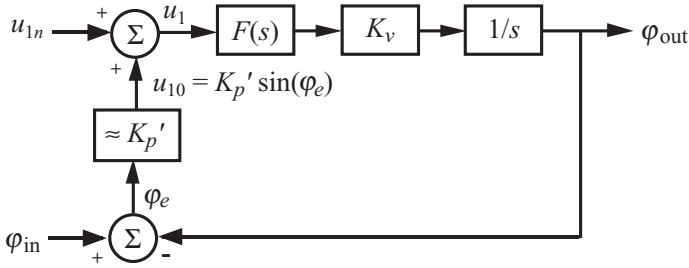


Fig. 13.4 Equivalent noise source entering the loop.

Use of K_p to express the ratio of u_1 to phase is a linear approximation of the sinusoidal relationship, but Eq. (13.20) is accurate, for whatever value of K_p is used, regardless of the accuracy of that linear approximation. Equation (13.19) was developed without reference to a sinusoidal characteristic and without reference to the desired input signal. K'_p and P_c were introduced in Eq. (13.19) only as a means of replacing the “new” variable η_p with variables that we are already using. If, for example, the VCO should deviate by Θ from quadrature with the reference, K_p will be reduced by $\cos \Theta$ [see Eq. (1.8)] and, as a result, n' in Eq. (13.20) will increase, since u_{1n} is fixed.

In using the multiplicative representation, we have assumed a constant VCO phase that is always aligned with the sine noise carrier, or, at least, we have not provided for the effects of phase variation. Thus, in Fig. 13.5, $\varphi_{out} = 0$, causing u_{1n} to equal u_{ns} , the noise associated with the sine carrier in Eq. (13.6). If the VCO phase should not be aligned with that carrier, the noise voltage produced by the corresponding noise source at the output of the phase detector would be attenuated, multiplied by $\cos \varphi_{out}$. [The last product in Eq. (13.13) would be $\langle \sin \omega t V_v \sin(\omega t + \varphi_{out}) \rangle = 0.5 V_v \cos \varphi_{out}$ if φ_{out} had not been 0.] However, at any fixed value of φ_{out} , the total noise power would not have decreased because the contribution from the noise modulation on the cosine carrier would replace the reduction of noise from the sine carrier.³ The two

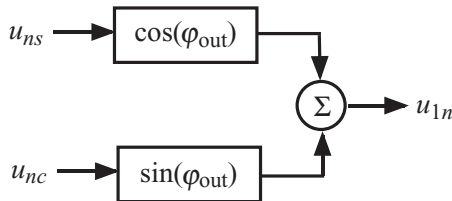


Fig. 13.5 Two noise sources adding to produce total noise.

³ The same is true for our first model, PM on a signal. If the VCO phase is $\Delta\theta$ from quadrature, $\langle u_1^2 \rangle$ in Eq. (13.11) will be reduced by a factor $\cos^2 \Delta\theta$. However, the equivalent AM component would be $\langle u_1^2 \rangle$ from Eq. (13.11) multiplied by $\sin^2 \Delta\theta$ so the total value of the two components would be independent of $\Delta\theta$.

noise sources in Fig. 13.5 each still produce random noise after multiplication by the constant sine or cosine value and the total power at u_{1n} is the same as the power of each of the individual noise sources.

However, we are still assuming a constant VCO phase and we know that the phase of the VCO will change—that is why we are studying the effects of noise—and that produces a kind of feedback that is embodied in the noise source of Fig. 13.5.

If the additive noise has unlimited bandwidth, it will be uncorrelated to φ_{out} . By the time the loop responds to u_{1n} , u_{1n} will have a new value that is independent of the value that had influenced φ_{out} and multiplication by its sine or cosine will not affect the random nature of the result.

At the other extreme, if the additive noise is band-limited to frequencies that are close to the input signal, frequencies producing equivalent phase noise to which the loop can respond, φ_{out} will be correlated to the noise. The value of the noise voltage at any instant will not be independent of a just previous value that has influenced the current value of φ_{out} .

The degree of correlation between values of a noise $n(t)$ at two times separated by T can be expressed by the autocorrelation function [Bracewell, 1965].

$$\frac{R(T)}{R(0)} = \frac{\int_0^\infty u(t)u(t+T)dt}{\int_0^\infty u(t)u(t)dt} \quad (13.21)$$

which is here normalized to its maximum value at $T = 0$. Since $R(T)$ is the Fourier transform of the power spectral density, for a rectangular band of noise, one of constant density from $-\omega_{\text{max}}$ to $+\omega_{\text{max}}$, this would be

$$\frac{R(T)}{R(0)} = \frac{\sin(\omega_{\text{max}}T)}{\omega_{\text{max}}T} \quad (13.22)$$

$$\leq \frac{1}{\omega_{\text{max}}T}, \quad (13.23)$$

so the correlation is small if

$$\omega_{\text{max}}T \gg 1. \quad (13.24)$$

We would expect a similar requirement for band shapes that are only approximately rectangular. The time required for a loop to respond to input changes is on the order of

$$T \approx 1/\omega_L. \quad (13.25)$$

Therefore, combining (13.24) and (13.25) we see that the output phase will have little correlation with the input noise if

$$f_{\text{max}} \gg f_L. \quad (13.26)$$

Otherwise, the simplifying assumption that allowed the use of a single noise source will not be valid and dual noise sources, according to the model of Fig. 13.5, become essential for accurate simulation of additive noise.

None of this is of any concern if $\sigma_{\varphi, \text{out}}^2 \ll 1 \text{ rad}^2$, since we can then use $\varphi_{\text{out}} \approx 0$ in Fig. 13.5, effectively giving a single source, or use the PM model, which produces the same results, for that portion of the noise that is close to the signal and which, therefore, most influences φ_{out} . However, the accuracy of a single-source model is doubtful when $\sigma_{\varphi, \text{out}}^2$ is significant and the noise is narrow band. We can use the more complex two-source model of Fig. 13.5 in simulations, but it is too complex for some theoretical developments, which, therefore, apply strictly only to very broadband input noise.

Exactly how wide the noise band must be to be considered broadband, in the sense that a wider bandwidth would make no difference in the loop output, will be easier to ascertain when we have obtained some results with bandlimited noise. Since both of our noise models employ modulation that causes the noise to be symmetrical about the signal frequency f_c , the represented noise cannot extend past $2f_c$. Therefore, in order for the loop bandwidth to be narrow compared to the bandwidth of the noise, the loop must have a small relative, or percentage, bandwidth.

Example 13.2 u_1 Due to Additive Noise A -30-dBm signal is embedded in white noise of density -100 dBm/Hz . A loop with a 1-kHz unity-gain bandwidth is locked to the signal. What is the noise voltage density at the output of the phase detector in the vicinity of 10 kHz if $K_p = 0.1 \text{ V/rad}$?

From Eq. (13.4), $S_\varphi = -100 \text{ dBm/Hz} + 30 \text{ dBm} = -70 \text{ dBc/Hz} \Rightarrow -70 \text{ dBBr/Hz}$. In a 1-kHz bandwidth, this is $\sigma_\varphi^2 = (10^{-7} \text{ rad}^2/\text{Hz})(1000 \text{ Hz}) = 10^{-4} \text{ rad}^2$, which is much less than 1 rad^2 . Therefore the representations just discussed apply. The loop will not track noise appreciably at 10 times its bandwidth so the error phase power spectral density (PPSD) at 10 kHz is the same as the input PPSD, and the 10-kHz voltage power spectral density at the PD output is

$$S_{u1} = S_\varphi K_p^2 = (10^{-7} \text{ rad}^2/\text{Hz})(0.1 \text{ V/rad})^2 = 10^{-9} \text{ V}^2/\text{Hz}.$$

13.6 DOES THE LOOP LOCK TO THE ADDITIVE NOISE?

When the signal is truly phase modulated, the VCO attempts to follow the modulated signal, and voltage at the phase detector output indicates its lack of success in doing so. The same is not true with additive noise. Even though we may represent the additive noise as an equivalent phase noise for computational purposes, the error signal required to maintain lock is generated by phase detection between the (unmodulated) signal and the VCO output.

We can discover one difference between the responses with true phase modulation and those with equivalent phase modulation by considering what happens when the angle between the VCO and signal moves away from quadrature (in an uncorrelated fashion, i.e., not in a locked loop). With true phase modulation, the detected signal

decreases due to the loss of sensitivity of the phase detector at other phase angles. Not so with the equivalent noise. The magnitude of u_{1n} is unaffected by the phase angle.

However, within the bandwidth of a locked loop, u_{1n} will be reduced (see Fig. 11.4c) because the loop will track out the *equivalent* modulation. In this case, phase detection between the signal and the modulated VCO output, which has variance $\sigma_{\varphi n}^2$, produces a voltage that tends to cancel the voltage produced by detection between the VCO output and the additive noise. Note, however, that $\sigma_{\varphi n}^2$ is also the variance of the phase error and, while u_{1n} may be greatly reduced, that does not imply that $\sigma_{\varphi n}^2$ has been reduced. It is the existence of this noise-induced phase modulation at the loop output that causes u_{1n} to be reduced but which also produces a noiselike phase error. Within the loop bandwidth, u_1 is a direct measure of neither the additive noise nor $\sigma_{\varphi n}^2$ but, rather, of a combination of the two.

If the signal and accompanying additive noise pass through a limiter that removes amplitude variations, however, the AM component is stripped and the signal at the limiter output is a true phase-modulated signal. Figure 13.6 illustrates how maximum sensitivity to changes in phase is attained when the VCO is in quadrature to the signal in the presence of additive noise, but, once the signal is limited, maximum sensitivity occurs when the VCO maintains quadrature to the resulting phase-modulated signal. This supports our contention that the phase error should be considered to be between the VCO and a phase-modulated signal (not just its carrier), but, in the presence of additive noise, it is between the VCO and the actual signal (not the equivalent phase-modulated signal). When the VCO is at an angle that gives best sensitivity to signal changes, it is most likely to respond to those changes and maintain lock. Thus, in the presence of additive noise, the VCO tends to lock to, and maintain quadrature with, the signal and, in the case of phase noise, it tends to lock to, and maintain quadrature with, the modulated signal, not with its unmodulated carrier.

Let us summarize the answer to the question posed in the title of this section.

- With a signal that is phase modulated by noise, the loop locks to the total resultant noise-modulated signal.
- With a signal in additive noise, the loop locks to the signal, not to the equivalent phase-modulated signal, in the sense that it is the error between the signal and the VCO output that must be minimized for proper tracking.
- With a signal in additive noise, the loop acts as though it locks to the equivalent phase-modulated signal in the sense that it tends to reduce the noise at the phase detector output as if that were true.

13.7 OTHER TYPES OF PHASE DETECTORS IN THE PRESENCE OF NOISE

With our first representation of additive noise, in terms of PM, we need give no unusual consideration to the type of phase detector employed. However, when the additive noise is large such that we represent it as multiplicative modulation, there is some variation in its influence as the phase detector type changes [Blanchard,

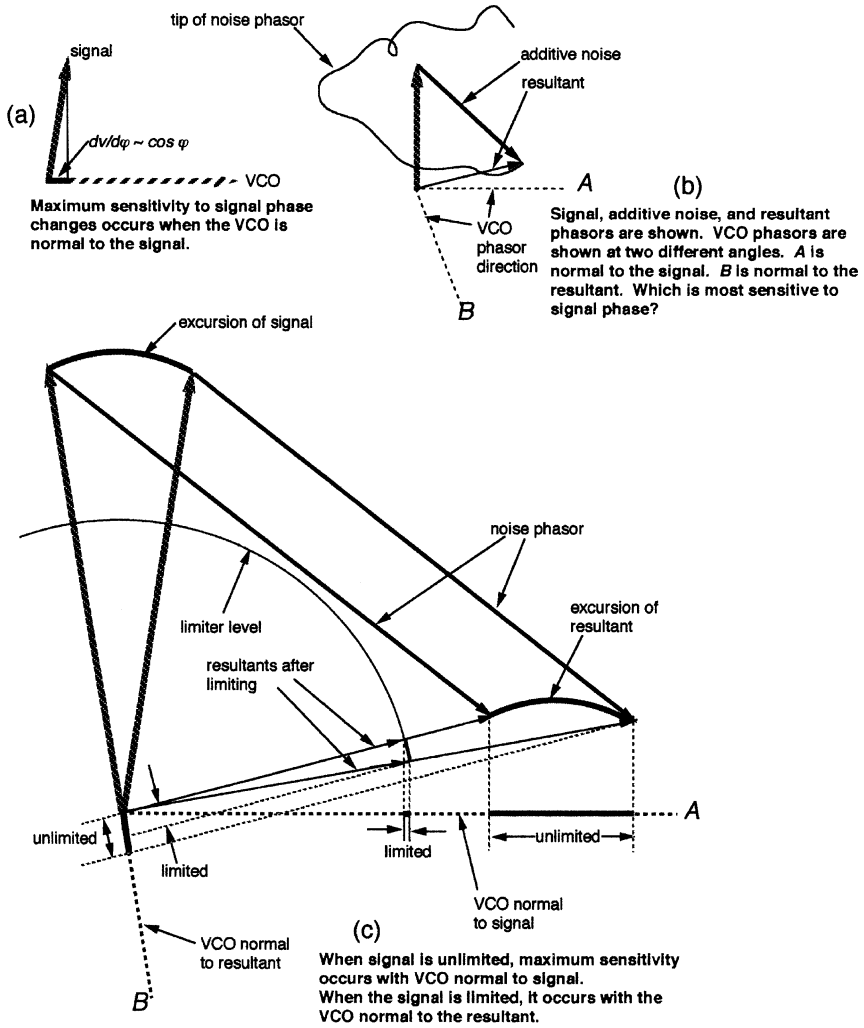


Fig. 13.6 Tracking the signals in additive noise. In this figure the noise phasor is very large for purposes of illustration. Limiting would not be maintained for much of the time with such large noise.

1976, pp. 145–152]. The development above was for a phase detector that acts like a multiplier. As noted, this applies to balanced mixers with relatively weak signals and strong LOs, even square-wave LOs that switch the diode bridge in a manner similar to strong sinusoids.

13.7.1 Triangular Characteristic

If the input and VCO signals are both square waves, we have shown that a triangular characteristic is produced. If the input signal is a square wave in broadband noise,

and we assume an infinite bandwidth in order to pass all the harmonics necessary to maintain a square signal, the equivalent phase noise, relative to that from the sinusoidal PD, will be 4 dB greater for the same signal power. (This is reduced to 1 dB difference, if the square wave has the same peak amplitude as the sine wave in the sinusoidal PD. The 3-dB difference is because a square wave has 3 dB more power than a sine wave with the same amplitude.) The added noise comes from the mixing of LO harmonics with noise that exists at frequencies that are far from the signal and that are assumed to be filtered out when the signal is sinusoidal but that must be passed to maintain the edges of a square wave.

If the signal is squared by passing it through a limiter that follows an input filter, thus allowing the input signal to be filtered, there is no effect as long as the signal-to-noise ratio (S/N) remains high. At the other extreme, where S/N approaches zero, a small degradation is seen in the low-frequency components of u_{1n} (0.3 dB for noise passed through a single-section bandpass input filter; 0.7 dB for an ideal rectangular filter) [Davenport and Root, 1965, Chapter 13, as given in Blanchard, 1976, p. 152]. The low-frequency components tend to be those that the loop follows because they are within its bandwidth.

13.7.2 Sawtooth Characteristic

This type of phase detector triggers on zero crossings (see Section 3.1.1). For purposes of analysis, the signal is assumed to pass through a hard limiter and then a differentiator before triggering the phase detector. As with the triangular characteristic, the performance is similar to that of a sinusoidal detector (mixer) at high S/N while a deterioration is seen at low S/N. By S/N = 0 dB, the peak of the phase detector output characteristic has dropped to about half of its noise-free value and the characteristic has begun to look sinusoidal. The low-frequency S/N deteriorates about 2.9 dB as S/N approaches zero.

13.8 MODIFIED PHASE DETECTOR CHARACTERISTIC WITH PHASE NOISE

Pouzet (1972) has shown that the triangular, sawtooth, and bang-bang (two-value) phase detector characteristics all approach sinusoidal shapes and shrink in amplitude with increasing noise as does the sinusoidal amplitude of a limiter followed by a multiplier. Wolaver (1991) has given a technique for computing the modification of phase detector characteristics in the presence of phase noise that applies, in a particularly uncomplicated manner, to the sinusoidal characteristic. We will consider that first. Recall that additive noise did not affect the PD characteristic of an ideal multiplier [Eq. (13.16)] but here we are considering phase noise. The average output from the PD, in response to the loop phase error φ_e (between reference signal⁴ and VCO) and

⁴ In the case of true phase modulation that is not followed by the VCO, the reference signal is the modulated signal, and φ_n is the difference between its phase and that of the unmodulated VCO. In the case of additive noise, the reference signal is the noiseless signal, and φ_n is the difference between its phase and that of the noisy VCO.

in the presence of phase noise φ_n , is obtained as the average of the output voltage $v(\varphi_e + \varphi_n)$, weighted by the probability density $p(\varphi_n)$ of the random variable φ_n ,

$$\bar{u}_1 = \int_{-\infty}^{\infty} p(\varphi_n) v(\varphi_e + \varphi_n) d\varphi_n. \tag{13.27}$$

Assuming that probability is an even function of φ_e , we can substitute $p(-\varphi_e)$ for $p(\varphi_e)$, causing (13.27) to become the convolution of the probability density with the phase detector characteristic,

$$\overline{u_1(\varphi_e)} = p(\varphi_n) * v(\varphi_n). \tag{13.28}$$

The Fourier transform of this equation is

$$\overline{u_1(x_e)} = p(x) \times v(x). \tag{13.29}$$

Thus the Fourier transform of the average PD characteristic in noise is the product of the transforms of the phase probability density function and the PD characteristic without noise.

$\sin(\varphi)$ is represented by a pair of impulses at $x = \pm 1$ in the Fourier domain,⁵ so the amplitude of the sinusoid will be multiplied by the transform of the probability density function there. Assuming a Gaussian distribution, the transform of the distribution will also be Gaussian with variance as shown in Fig. 13.7. Thus the amplitude of the sinusoidal characteristic will be multiplied by the value of a Gaussian at $x = \pm 1$,

$$\overline{K_p} = K_p \exp\left(-\frac{x^2}{2\sigma_x^2}\right)\Big|_{x=\pm 1} = K_p \exp\left(-\frac{1}{2\sigma_x^2}\right) = K_p \exp\left(-\frac{\sigma_\varphi^2}{2}\right), \tag{13.30}$$

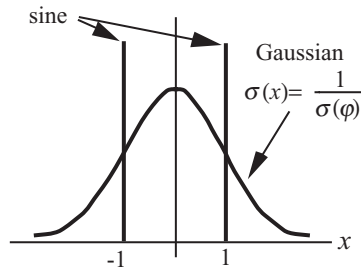


Fig. 13.7 Transforms of PD characteristic and phase probability density function.

⁵ We need not relate this Fourier variable to any quantity defined for the loop. It is merely a variable in the transform domain, which is being employed because it facilitates calculations.

where σ_φ is the rms phase deviation and σ_x is the corresponding standard deviation in the transform domain.

We can also see, from this development, why the triangular and sawtooth PD characteristics approach sinusoidal shape in the presence of noise. The transform of those characteristics, $v(x)$, contains many harmonics (at $x = \pm n$). When $\sigma(\varphi)$ is small, the Gaussian in the transform domain $\{\exp[-n^2/(2\sigma_x^2)]$ at harmonic $n\}$ will be very broad and will have little effect on the significant harmonics. Therefore, in Eq. (13.29), $\overline{u_1(x_e)} \approx v(x)$, the value without noise. However, as noise increases and $\sigma(\varphi)$ becomes wider, $\sigma(x)$ shrinks, and so, therefore, does the width of the Gaussian. As it does so, it attenuates the harmonics in the PD characteristic but has least effect on the fundamental at $x = \pm 1$. Thus the transform-domain representation in Eq. (13.29) approaches that of a sinusoidal characteristic and, therefore, so does the phase-domain characteristic.

Example 13.3 Phase Detector Characteristic Change by Noise A PLL is being used to demodulate a phase-modulated signal. How much will the demodulated output be reduced if additive noise causes the VCO to have an rms phase variation of 0.5 rad? (Assume the strength of the input signal remains fixed.)

The variance is

$$\sigma_\varphi^2 = (0.5 \text{ rad})^2 = 0.25 \text{ rad}^2.$$

By Eq. (13.30), the average PD sensitivity will be multiplied by $\exp(-0.25/2)$. The resulting gain change will be

$$20 \text{ dB } \log_{10}[\exp(-0.125)] = \underline{\underline{-1.1 \text{ dB}}}$$

By applying (13.29) to the fundamental and harmonics of the triangular phase detector characteristic, we can find how the shape of the characteristic changes in the presence of phase noise. The effect on the peak values of both the sinusoidal and triangular characteristics is shown by the curves in Fig. 13.8. Note how the peak value of the triangular characteristic approaches that of the sinusoidal characteristic (even on a relative basis) as both decrease with increasing phase noise. The three data points are from Blanchard (1976), but these represent changes in the triangular characteristics in the presence of additive noise where the abscissa is the equivalent phase noise according to (13.4). In the former case (the curves) one of the two inputs to the phase detector is phase modulated by noise and, in the latter, it is accompanied by additive noise as it passes through a limiter in the process of turning into a square wave. Blanchard (1976) and Stensby (1997) give curves that show changes with additive noise for both triangular and sawtooth characteristics.

By the same process we can determine how the slope of the PD characteristic in the center of the operating range, K'_p , is modified. The results are shown in Fig. 13.9.

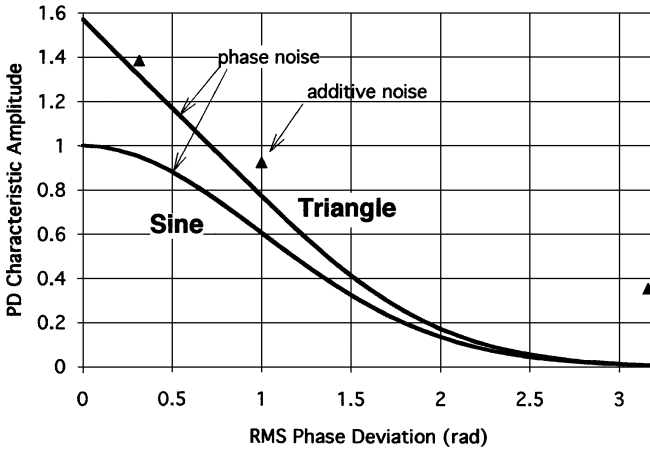


Fig. 13.8 Modification of the amplitude of the characteristics of phase detector in the presence of noise.

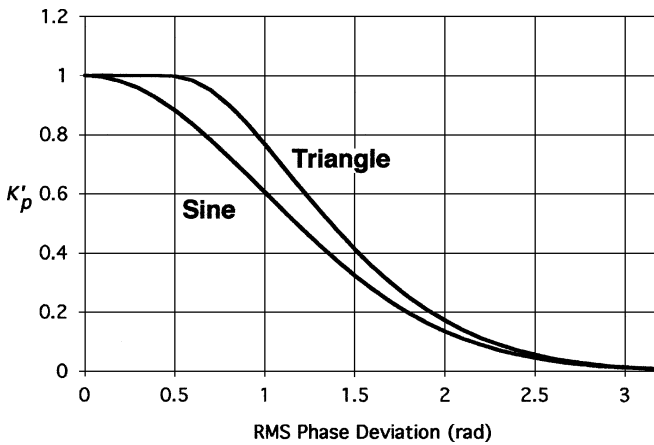


Fig. 13.9 Modification of the slopes of the characteristics of phase detectors in the presence of phase noise.

i13.A APPENDIX: DECOMPOSITION OF A SINGLE SIDEBAND

This appendix, which is available from the Wiley Internet site, shows how a single sideband can be decomposed into two pairs of modulation sidebands.

CHAPTER 14

LOOP RESPONSE TO ADDITIVE NOISE

In this section we will study the response of the loop to additive noise, using the representations of additive noise developed in the last chapter. Since the analysis will be carried out through an equivalent phase modulation, we could have studied this material before the last chapter. However, we will be considering flat noise spectrums that are characteristic of additive noise, so it is more appropriate to have studied the equivalence first. We will also be able to develop our skills in applying the equivalence. See Example 14.1.

14.1 NOISE BANDWIDTH

The variance, or mean square value, of the phase deviation of the voltage-controlled oscillator (VCO), $\sigma_{\varphi 0}^2$, is the integral of $S_{\varphi, \text{out}}$ over all frequency. We saw in Chapter 12 how phase modulation on the reference appeared in a filtered form on the VCO of the locked loop. We saw in Chapter 11 how this $S_{\varphi}(f_m)$ can be integrated to obtain σ_{φ}^2 . If a flat noise spectrum $S_{\varphi, \text{in}}(f_m) = S_{\text{flat}}$ on the reference is processed by the loop, the phase power spectral density at the output is

$$S_{\varphi, \text{out}}(f_m) = S_{\text{flat}} |H(f_m)|^2. \quad (14.1)$$

The output phase variance is

$$\sigma_{\varphi, \text{out}}^2 = \int_0^{\infty} S_{\varphi, \text{out}}(f_m) df_m \triangleq B_n S_{\text{flat}}, \quad (14.2)$$

where B_n is the one-sided¹ noise bandwidth. It is an equivalent bandwidth such that an ideal rectangular filter that passes modulation from $f_m = 0$ to $f_m = B_n$ will produce the actual output phase variance $\sigma_{\varphi, \text{out}}^2$.

Comparing the last two equations we see

$$B_n = \int_0^\infty |H(f_m)|^2 df_m. \tag{14.3}$$

The concept of noise bandwidth has application considerably wider than the field of phase-locked loops (PLLs). Figure 14.1 illustrates the concept in general. Some level G' in a frequency response characteristic, usually the highest, is taken as nominal. The noise bandwidth is then the bandwidth B_n of a rectangular filter that has that nominal power gain G' within its passband and zero gain outside the passband and passes the same total power as does the actual filter when they are both excited by a uniformly distributed power density. In general, the right side of Eq. (14.3) equals $G'B_n$ but, for the particular case of a PLL, the maximum gain G' is unity.

If S_{flat} represents additive noise, Eqs. (13.4) and (14.2) can be combined to give

$$\sigma_{\varphi, \text{out}}^2 = B_n N_0 / P_c. \tag{14.4}$$

For a first-order loop,

$$B_n = \int_0^\infty \left| \frac{K/j\omega}{1 + K/j\omega} \right|^2 df_m = \int_0^\infty \frac{K}{K + j\omega} \frac{K}{K - j\omega} df_m \tag{14.5}$$

$$= \frac{\text{cycle}}{2\pi} \int_0^\infty \frac{K^2}{K^2 + \omega^2} d\omega = \frac{K}{2\pi} \left[\tan^{-1} \frac{\omega}{K} \right]_0^\infty \text{ cycles} \tag{14.6}$$

$$= \frac{K}{4} \text{ cycles}. \tag{14.7}$$

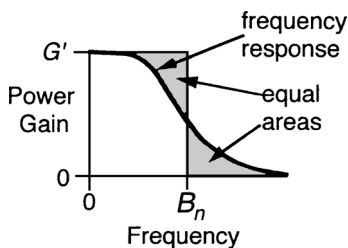
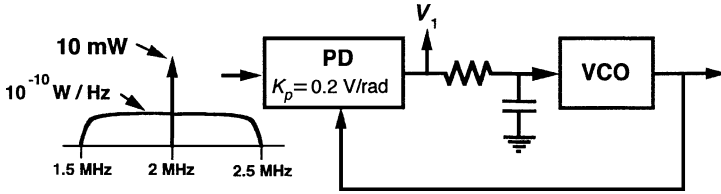


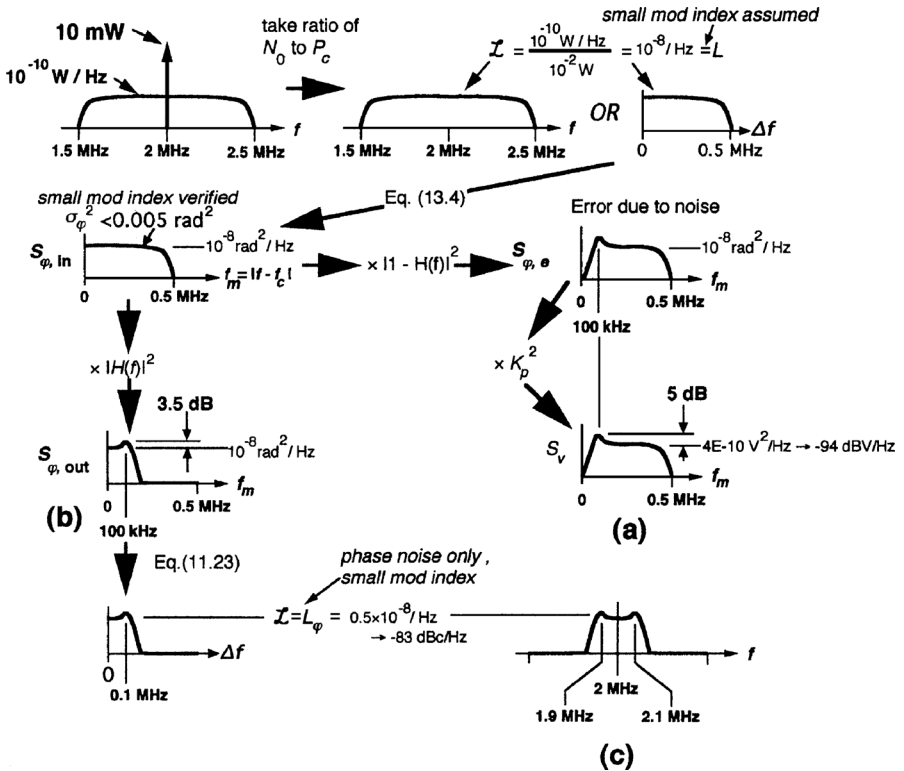
Fig. 14.1 Noise bandwidth.

¹ One-sided means that only positive frequencies are considered.

Example 14.1 Loop Responses to Additive Noise The noise shown below accompanies the 10-mW carrier into the loop shown. The loop has $\omega_n = 2\pi \times 10^5$ rad/sec and $\zeta = 1/(2\sqrt{2})$.



Sketch: (a) The power spectrum at V_1 ; (b) The phase power spectral density S_ϕ at the VCO output; and (c) The sideband density (as one might see on a spectrum analyzer) at the VCO output.



Note that, since K has units of reciprocal seconds (sec^{-1}), the units of B_n are hertz. Since the 3-dB bandwidth of a first-order loop is

$$\omega_L = 2\pi f_L \text{ rad/cycle} = K, \tag{14.8}$$

B_n can be written in terms of loop bandwidth as

$$B_n = \frac{\pi}{2} f_L. \quad (14.9)$$

Thus the equivalent noise bandwidth extends beyond the 3-dB bandwidth by almost 60% for the single-pole response of the first-order loop.

For the second-order loop we integrate $|H(\omega)|^2$ from Eq. (6.4a) to get

$$B_n = \frac{\omega_n}{4} \left[\frac{1}{2\zeta} + 2\zeta\alpha^2 \right] \frac{\text{cycle}}{\text{rad}}. \quad (14.10)$$

The integration is described in Appendix 14.A. Surprisingly, Eq. (14.10) reduces to (14.7) for $\alpha = 0$. In other words, the addition of a low-pass loop filter has no effect on the noise bandwidth regardless of pole frequency; all of the responses in Fig. 7.8 for $\alpha = 0$, when taken as power gains, have the same area under them. This is illustrated in Fig. 14.2.

We might consider B_n to be the width of a filter on the demodulated signal. For example, we might pass the reference through a phase detector and then place a rectangular filter of width B_n on its output. This would be equivalent to filtering the phase variation with a rectangular filter, if we had a device that could do that, prior to demodulation. Alternately, we could put a rectangular filter on the reference signal before demodulation. It would extend from $f_c - B_n$ to $f_c + B_n$, and thus have a width of $2B_n$, and filter the noise that would produce phase modulation extending from 0 to B_n (assuming small m). These processes are illustrated in Fig. 14.3.

14.2 SIGNAL-TO-NOISE RATIO IN THE LOOP BANDWIDTH

The power of that half of additive input noise to which the loop responds and that passes through the effective filter is

$$P_n = 2B_n(N_0/2) = B_n N_0. \quad (14.11)$$

From this we can write the signal-to-noise (S/N) ratio as

$$S/N = P_c/P_n = P_c/(B_n N_0). \quad (14.12)$$

From Eq. (14.4) this is

$$S/N = 1/\sigma_{\varphi, \text{out}}^2 \quad (14.13)$$

and is sometimes called signal-to-noise ratio in the loop bandwidth or “in the loop.”

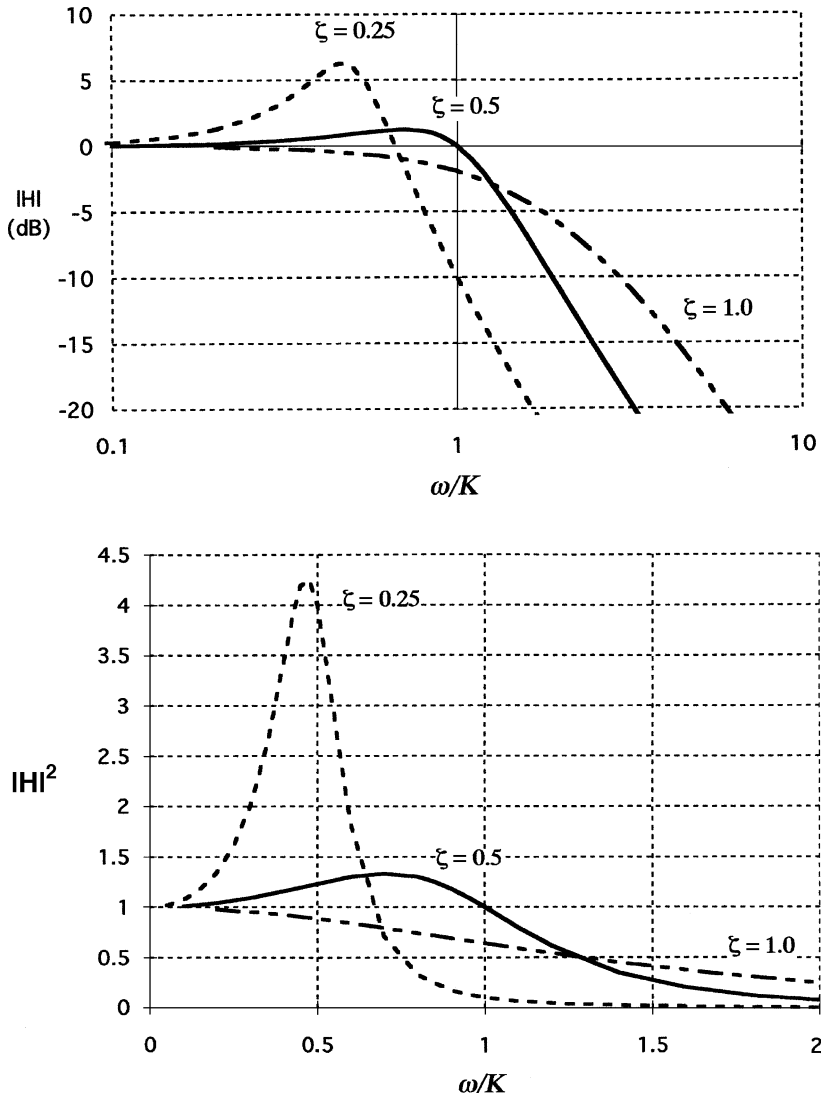


Fig. 14.2 Squared magnitude of $H(\omega/K)$ for $\alpha = 0, \zeta < 1$. These curves, which are also shown in Fig. 7.8, are drawn here to illustrate how the power under them can be the same for all $\omega_p [= K/(2\zeta)^2]$ if K is the same for each.

14.3 LOOP OPTIMIZATION IN THE PRESENCE OF NOISE

Jaffe and Rechlin (1955) have developed a procedure for optimizing loop parameters in the presence of noise. It is an optimization in the sense that the noise is minimized under a constraint on the loop error when the loop is following certain input transients (step, ramp, parabola). In this section we will give the forms of the optimum transfer

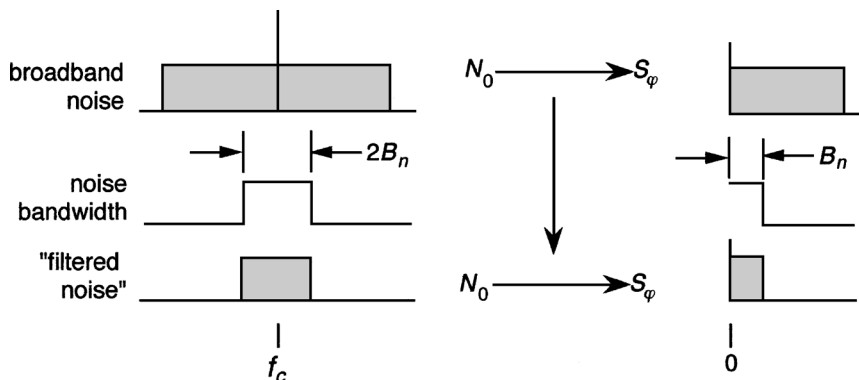


Fig. 14.3 Noise bandwidth, input and modulation equivalents.

functions, which will tell us the general nature of the optimum loops, for each of the three kinds of inputs that they treated. For each kind we will then compute or indicate the parameters that show the tradeoff between response accuracy and output noise.

A theoretical basis, which should aid in extension of the theory to other kinds of inputs, is developed in Appendix i14.B.

14.3.1 The Problem

Optimize the loop transfer function for operation in the presence of noise to give simultaneously (a) *a constrained phase error in the response to the input* and (b) *minimum steady-state output phase noise, subject to that constraint.*

Ideally we would like to minimize both the noise and the error, but the optimum filter for noise would have zero bandwidth and the optimum filter for following the input would have infinite bandwidth. So we accept a certain error in the exactness with which we follow the input and discover the filter that will give the least noise while meeting that constraint.

14.3.2 Measures to Be Used

1. Steady-state noise:

$$\sigma_{\varphi, \text{out}}^2 = B_n N_0 / P_c. \tag{14.4}$$

2. Integrated square error of the output $\varphi_{\text{out}}(t)$ relative to the input $\varphi_{\text{in}}(t)$ in the absence of noise: Let

$$\varepsilon(t) = \varphi_{\text{out}}(t) - \varphi_{\text{in}}(t), \tag{14.14}$$

as illustrated in Fig. 14.4. Then define E ,

$$E^2 \equiv \int_0^\infty |\varepsilon(t)|^2 dt = \int_{-\infty}^\infty |\varepsilon(f)|^2 df, \tag{14.15}$$

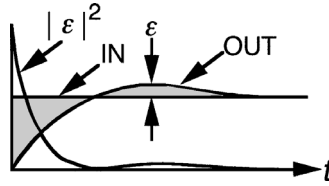


Fig. 14.4 Error in following a step input.

where $\varepsilon(f)$ is the Fourier transform of $\varepsilon(t)$. The last equation is true by Rayleigh’s theorem if $|\varepsilon(t < 0)| \equiv 0$; The value of E^2 is a measure of the inaccuracy with which the output has followed the input.

14.3.3 Optimum Loop for a Phase Step Input

For a phase step, $\Phi(s) = \theta/s$, the optimum loop transfer function is

$$H(s)_{\text{opt}} = \frac{K}{s + K}. \tag{14.16}$$

This form implies a first-order loop [see Eq. (2.25b)] with open loop $G = K/s$.

The integrated error is

$$\begin{aligned} E^2 &= \int_{-\infty}^{\infty} |\Phi(f)|^2 |1 - H(f)|^2 df \\ &= \frac{\theta^2}{2\pi} \int_{-\infty}^{\infty} \frac{1}{\omega^2} \frac{\omega^2}{\omega^2 + K^2} d\omega = \frac{\theta^2}{2\pi} \left[\frac{1}{K} \tan^{-1} \left(\frac{\omega}{K} \right) \right]_{-\infty}^{\infty} = \frac{\theta^2}{2K}. \end{aligned} \tag{14.17}$$

Therefore, we can write

$$K = \frac{1}{2} \frac{\theta^2}{E^2}. \tag{14.18}$$

Thus the loop gain can be chosen to give an acceptable ratio between the phase step θ and the error E in the response to it.

Using Eq. (14.7), the output variance equals

$$\sigma_{\varphi, \text{out}}^2 = \frac{N_0}{P_c} B_n = \frac{KN_0}{4P_c} \text{ cycle} = \frac{N_0}{8P_c} \frac{\theta^2}{E^2} \text{ cycle}. \tag{14.19}$$

(With θ^2 in rad^2 , E^2 in $\text{rad}^2\text{-sec}$, and N_0/P_c in rad^2/Hz , $\sigma_{\varphi, \text{out}}^2$ will be in rad^2 and K will be in sec^{-1} without unit conversions.) As we would expect, this shows that a larger tolerated error in response permits a smaller output steady-state phase variance. In practice, one would trade off E^2 against $\sigma_{\varphi, \text{out}}^2$. Then, having chosen these values,

one would find K from Eq. (14.18) to get the minimum $\sigma_{\varphi, \text{out}}$ and the accepted value of E^2 .

14.3.4 Optimum Loop for a Frequency Step Input

For the phase ramp (frequency step) input $\Delta\omega/s^2$, the optimum filter is a second-order loop with an integrator-and-lead filter and $\zeta = 1/\sqrt{2}$,

$$H(s) = \omega_n^2 \frac{1 + (\sqrt{2}/\omega_n)s}{s^2 + \sqrt{2}\omega_n s + \omega_n^2}. \quad (14.20)$$

Following a procedure similar to that for the phase step, Jaffe and Rechtin (1955) give the expressions that enable us to determine the optimum natural frequency as a function of the ratio of step size to allowed squared error and to determine the variance, due to noise, at the output of the optimum loop,

$$\sigma_{\varphi, \text{out}}^2 = \frac{3}{4\sqrt{2}} \frac{\omega_n N_0}{P_c} \text{ rad-cycle} \quad (14.21)$$

and

$$E^2 = \frac{(\Delta\omega)^2}{2\sqrt{2}\omega_n^3} \text{ rad}^3. \quad (14.22)$$

We can rearrange this expression to give the natural frequency,

$$\left(\frac{\sqrt{2}\omega_n}{\text{rad}} \right)^3 = \frac{\Delta\omega^2}{E^2}, \quad (14.23)$$

and combine (14.21) and (14.23) to show the tradeoff between error and noise,

$$\sigma_{\varphi, \text{out}}^2 = \frac{3}{8} \frac{N_0}{P_c} \left(\frac{\Delta\omega^2}{E^2} \right)^{1/3} \text{ rad}^2\text{-cycle}. \quad (14.24)$$

With $\Delta\omega$ and ω_n in radians/second and other units as given above, unit conversions are again unnecessary.

14.3.5 Optimum Loop for a Frequency Ramp Input

For a frequency ramp input, which is a phase parabola having a slope of qt , the optimum loop is of third order. The filter transfer function is defined by

$$KF(s) = 2\omega_3 \frac{[s + (\omega_3 + j\omega_3)/2][s + (\omega_3 - j\omega_3)/2]}{s^2} \quad (14.25)$$

This has a complex zero pair and two integrators. The resulting closed-loop transfer function is

$$H(s) = \frac{2x^2 + 2x + 1}{x^3 + 2x^2 + 2x + 1}, \tag{14.26}$$

where

$$x = s/\omega_3. \tag{14.27}$$

The closed-loop frequency response of this third-order loop is plotted in Fig. 14.5.²

From Jaffe and Rechlin's (1955) results we can also derive

$$\sigma_{\varphi, \text{out}}^2 = \frac{5}{12} \frac{N_0}{P_c} \omega_3 \text{ rad-cycle} \tag{14.28}$$

and

$$E^2 = \frac{(q \text{ sec})^2}{3} \left(\frac{\text{rad}}{\omega_3} \right)^3, \tag{14.29}$$

where q is the slope of the frequency ramp.

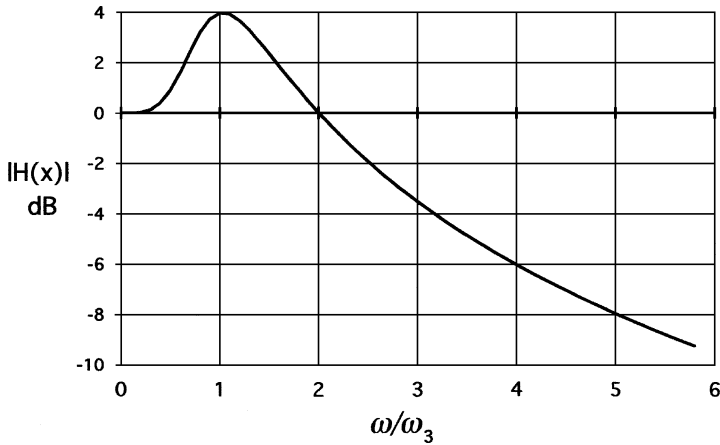


Fig. 14.5 Response of the optimum loop for frequency ramp.

² Blanchard (1976, p. 166) suggests approximating the filter with two identical integrator-and-lead filters in series. He obtains a response that has almost the 4-dB peak of Fig. 14.5 but is about 25% higher in frequency. He briefly describes an exact filter for which he references Gupta and Solem (1965) for discussion. While we will not discuss it further here, it does not appear that realization should be especially difficult. However, one must consider the stability of the active filter during design.

i14.A APPENDIX: INTEGRATION OF EQ. (6.4a)

This appendix is available from the Wiley Internet site.

i14.B APPENDIX: LOOP OPTIMIZATION IN THE PRESENCE OF NOISE

This appendix, which provides a theoretical basis for Section 14.3, is available from the Wiley Internet site at ftp://ftp.wiley.com/public/sci_tech_med/phase_lock.

CHAPTER 15

PHASE-LOCKED LOOP AS A DEMODULATOR

The effects of noise often determine whether a particular design is usable. In this chapter we will study the loop as a demodulator, much as we did in Chapter 7, but this time we will describe the effects of noise. We will also consider the effects of noise in the carrier recovery circuits of Chapter 10.

15.1 PHASE DEMODULATION

A phase demodulator is pictured in Fig. 15.1. The signal passes through a filter, which limits the noise bandwidth, and enters the phase-locked loop (PLL) as its reference. Because the loop is designed to be too narrow to follow the modulation, the output of the phase detector is the demodulated signal. Also, because the loop does not follow the modulation, m must be small to prevent nonlinear operation.

15.1.1 Noise

The power spectrums of the phase-modulated input signal and additive noise are shown in Fig. 15.2. We assume that the signal is in the center of the input filter passband. The corresponding input phase, φ_{in} , consists of signal and noise, which are described by the signal phase power spectral density $S_{\varphi_s}(\omega_m)$ and the noise phase power spectral density $S_{\varphi_n}(\omega_m)$. We can multiply each by $|1 - H(\omega_m)|^2$ and then by K_p^2 and sum them to give the power spectrum S_{u1} at \mathbf{x} , which is illustrated by Fig. 15.3.

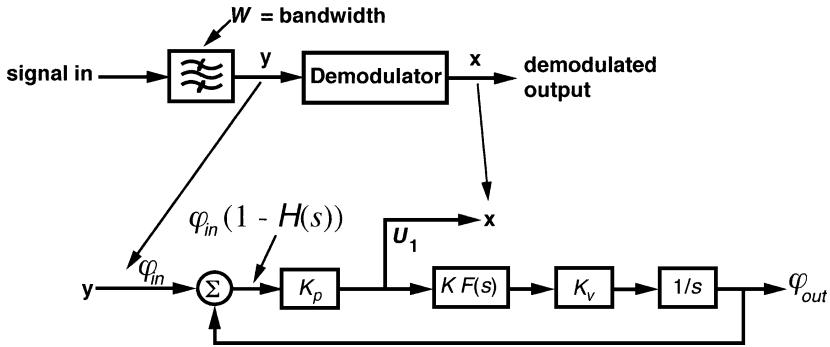


Fig. 15.1 Phase demodulator.

The signal spectrums in Fig. 15.2 look like those in Fig. 15.3 because the modulation index m is small and, therefore, Eq. (11.23) holds. Since m is small and the input phase deviation appears, essentially unaltered, as the phase error, u_1 will be proportional to phase. At x the power spectral density of the demodulated signal is

$$S_{s1} = S_{\phi_s} K_p^2 |1 - H(s)|^2, \tag{15.1}$$

and the noise power spectral density is

$$S_{n1} = S_{\phi_n1} K_p^2 |1 - H(s)|^2 = (N_0/P) K_p^2 |1 - H(s)|^2, \tag{15.2}$$

where P is the signal power. We can find the demodulated signal-to-noise ratio (S/N) by integrating the power spectral densities of the signal and noise. These are restricted by the input filter to a noise bandwidth $B_n = W/2$, where W is the equivalent input [radio frequency (RF)] bandwidth. The expression for the demodulated signal power S is

$$S = K_p^2 \int_0^{W/2} |1 - H(f_m)|^2 S_{\phi_s}(f_m) df_m = K_p^2 \sigma_{\phi_s}^2, \tag{15.3}$$

where $\sigma_{\phi_s}^2$ is the mean square phase deviation of the filtered signal. If $|1 - H(f_m)| \approx 1$ for all f_m of significance in the signal, $\sigma_{\phi_s}^2$ is the mean square phase deviation of the

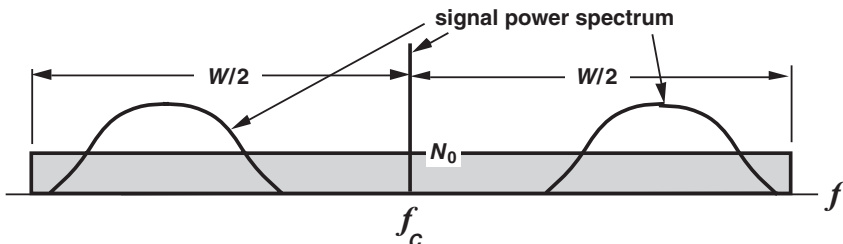


Fig. 15.2 Input power spectrums.

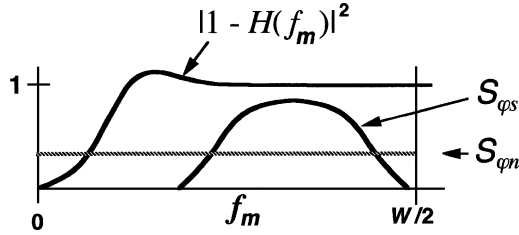


Fig. 15.3 Phase power spectrums and loop response.

original signal. The expression for the demodulated noise power is

$$N = K_p^2 \int_0^{W/2} |1 - H(f_m)|^2 S_{\phi_n} df_m = K_p^2 \left(\frac{N_0}{P} \right) \int_0^{W/2} |1 - H(f_m)|^2 df_m. \quad (15.4)$$

If $|1 - H(f_m)| \approx 1$ for $f_m < W/2$, which implies that the loop bandwidth is small compared to $W/2$ (i.e., $f_L \ll W/2$), this becomes

$$N \approx K_p^2 \left(\frac{N_0}{P} \right) \left(\frac{W}{2} \right), \quad (15.5)$$

and the signal-to-noise ratio becomes

$$\frac{S}{N} \approx \frac{\sigma_{\phi_s}^2}{(N_0/P)(W/2)} = \frac{2P\sigma_{\phi_s}^2}{WN_0}. \quad (15.6)$$

If $W/2$ exceeds the signal bandwidth (i.e., $W > 2f_m$ for all f_m), the S/N can be further improved by low passing the output at some video bandwidth, $B_v < W/2$. Then $W/2$ can be replaced by B_v in the expressions above.

The same results could be obtained with a traditional (standard) frequency discriminator followed by an integrator, producing a demodulated signal

$$v_d = K_d \int \delta f(\omega_m, t) dt = K_d \delta_\phi(\omega_m, t), \quad (15.7)$$

where K_d is the detector sensitivity and δf is the frequency deviation. This is not restricted to small modulation index, as is the PLL, but is less sensitive in the presence of noise [Blanchard, 1976].

15.1.2 Distortion of the Demodulated Signal

The clearest way to see how distortion can occur in the demodulated signal is to consider the phase detector as an ideal multiplier, multiplying the modulated input signal by the steady VCO signal. The spectral description of a signal modulated by a

sinusoid is given by

$$v_{\text{in}} = A \left\langle \begin{array}{l} \text{control signal} \quad \text{demodulated phase} \\ \frac{J_0(m) \sin \varphi_c(t) + J_1(m) \{\cos[\varphi_c(t) + \varphi_m(t)] + \cos[\varphi_c(t) - \varphi_m(t)]\}}{-J_2(m) \{\sin[\varphi_c(t) + 2\varphi_m(t)] + \sin[\varphi_c(t) - 2\varphi_m(t)]\}} \\ \frac{-J_3(m) \{\cos[\varphi_c(t) + 3\varphi_m(t)] + \cos[\varphi_c(t) - 3\varphi_m(t)]\} + \dots}{\text{distortion in demodulated phase}} \end{array} \right\rangle. \quad (15.8)$$

Here $\varphi_c(t)$ and $\varphi_m(t)$ include ωt and a fixed phase θ . The parts of the input signal are marked in Eq. (15.8), to indicate their use after multiplication by the VCO signal, $B \cos \omega_c t$. This is shown in detail in Fig. 15.4. The first term produces a constant; this is the control signal that keeps the loop locked. The second term produces the desired sinusoidal signal. The remaining terms produce distortion. This occurs because of the fundamental nature of the demodulation process with a balanced-mixer phase detector, that is, multiplication.

The even harmonics of ω_m (with coefficients J_2, J_4 , etc.) appear in the product after multiplication by $B \cos \omega_c t$ only to the degree that the phase relationship between the signal and VCO are not ideal, that is $\theta_c \neq 0$. This is analogous to the zero low-frequency output from a phase detector when one input is a sine and the other is ideally a cosine.

To eliminate the distortion, the components that produce it must be filtered out of the input. Thus the input filter would have to pass signals offset from the carrier by the highest modulation frequency, $\omega_{m, \text{max}}$, but reject signals offset by $3\omega_{m, \text{min}}$ or, in case the signal and VCO are out of quadrature, $2\omega_{m, \text{min}}$. Since modulation frequencies often cover many octaves, this is not necessarily possible. Moreover, such a filter might be very difficult to realize due to the narrow relative bandwidth or due to drift of the carrier frequency. The distortion can be reduced to an arbitrary level by keeping m small since the higher order Bessel functions drop more quickly as m drops.

Extra distortion terms also appear when the signal is modulated by multiple frequencies. These occur at frequencies that are offset from the carrier by the sums and differences of the modulation frequencies and the harmonics thereof. Such components do not represent distortion in the phase modulation but cause distortion as a natural result of the use of multiplicative demodulation, which treats the spectral components separately.

15.1.3 Demodulation with a Linear Phase Detector Characteristic

A phase detector with a sawtooth or triangular characteristic can be analyzed most efficiently in terms of its response to the modulation, since that is linear. If we try to analyze the sinusoidal phase detector in terms of its nonlinear phase response or if we try to analyze the triangular or sawtooth phase detector in terms of multiplication by an equivalent series of sinusoidal harmonics, we will get the same answers. We will just be less efficient and more prone to confusion.

Because u_1 is linearly proportional to ϕ_i , no distortion is produced (within the range of the phase detector, of course). These types of phase detectors do not need

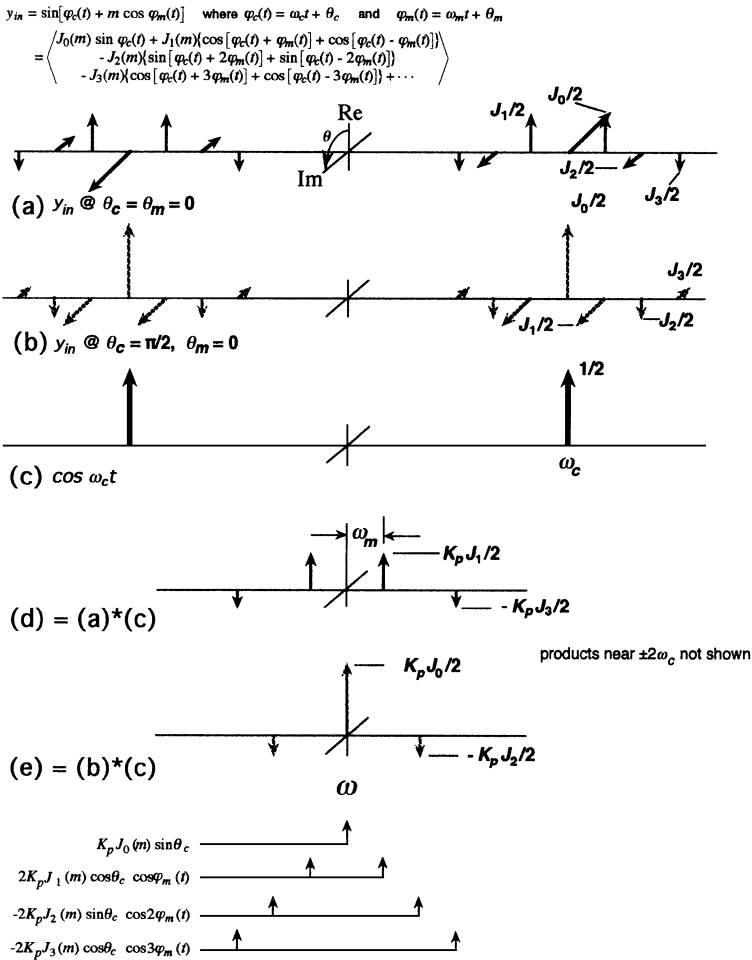


Fig. 15.4 Multiplication of a sinusoidally modulated signal by a cosine at the same carrier frequency. (* represents corresponding convolution in the frequency domain.)

a carrier to lock to either. Figure 15.5 shows possible operating points for the four phases in a QPSK signal (Fig. 10.8).

Since these states are generally equiprobable, the signal has no spectral line at the carrier frequency. Nevertheless, it is obvious from Fig. 15.5a that the average voltage will change as Φ_i , the average value of ϕ_{in} , changes relative to Φ_o . When Φ_e changes, the average voltage will change by the same amount as does the voltage at each state. However, the linear range is reduced by $M = 4$, as illustrated in Fig. 15.5b.

15.2 FREQUENCY DEMODULATION, BANDWIDTH SET BY A FILTER

A PLL used as a frequency demodulator is pictured in Fig. 15.6. It is in many ways analogous, or complementary, to the phase demodulator. We might expect the signal

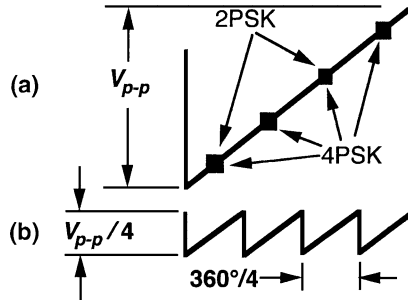


Fig. 15.5 Operating points for (a) QPSK signal (voltage versus phase) and (b) equivalent phase detector characteristic.

spectrum to differ from that of Fig. 15.2 somewhat because constant frequency deviation implies greater phase deviation at lower modulation frequencies. For small modulation index, this would cause the signal power to climb near the carrier (or center) frequency. However, this loop is designed to follow the modulation so the restriction to small modulation index is relieved. The phase error can remain small even though the input phase deviates by many radians. At sufficiently high values of m the input spectrum approaches one in which the power is spread evenly between $f_c - \Delta f$ and $f_c + \Delta f$. The demodulated signal spectrum would look like the original information spectrum, perhaps like S_{ω_s} in Fig. 15.3. Generally the power of the demodulated signal will be given by an expression similar to Eq. (15.3),

$$S = \sigma_{\omega_s}^2 / K_v^2, \tag{15.9}$$

where $\sigma_{\omega_s}^2$ is the frequency variance of the signal, the mean square frequency deviation.

Other features in Fig. 15.3 will also differ for frequency modulation (FM), however. The loop response will be $H(\omega_m)$, and thus low-pass in nature, and the power created by additive white input noise will not be flat.

Combining Eqs. (11.5b) and (13.4), we obtain the frequency power spectral density due to additive noise as

$$S_{\omega_m}(\omega_m) = \omega_m^2 N_0 / P \tag{15.10}$$

so the demodulated spectrums appear as in Fig. 15.7.

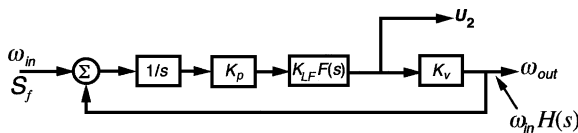


Fig. 15.6 Frequency demodulator.

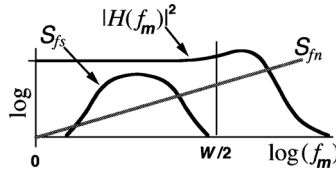


Fig. 15.7 Frequency power spectrums and loop response.

For the case where the loop is wide enough that $H(f_m) \approx 1$ for $f_m < W/2$, signal-to-noise ratio can be computed as it was for the phase demodulator. The power in the demodulated signal is

$$S = \frac{1}{K_v^2} \int_0^{W/2} |H(f_m)|^2 S_{\omega_s}(f_m) df_m = \frac{\sigma_{\omega_s}^2}{K_v^2}, \quad (15.11)$$

where S_{ω_s} is the frequency power spectral density of the signal and $\sigma_{\omega_s}^2$ is its mean square frequency deviation. The detected noise power is

$$N = \frac{1}{K_v^2} \int_0^{W/2} |H(f_m)|^2 S_{fn}(f_m) df_m \quad (15.12a)$$

$$= \frac{1}{K_v^2} \int_0^{W/2} |H(f_m)|^2 f_m^2 S_{\varphi n} df_m \quad (15.12b)$$

$$\approx \frac{N_0}{K_v^2 P} \int_0^{W/2} |H(f_m)|^2 f_m^2 df_m. \quad (15.13)$$

A Note on Units Equation (15.13) may use mixed units. We have attempted to use radian units for uniformity, but convention sometimes makes this awkward. Noise bandwidths are often in Hz and densities are usually given on a per-hertz basis. We could have used $S_{\omega n}$ in Eq. (15.12a), as we did in Eq. (15.11), but Eq. (15.13) would then contain ω_m explicitly, even though the variable of integration is f_m . Some change would have to be made before integrating, even if integration were being done graphically, so we use S_{fn} and f_m . There will be no difficulty in execution as long as we carry units and treat them properly. The integral will have units of Hz^3 . This will cancel Hz^{-1} in N_0 and Hz^{-2} in $1/K_v^2$ if K_v is given in Hz/V . If K_v is in $(\text{rad}/\text{sec})/\text{V}$, as we have been trying to maintain it, the result must be multiplied by $(2\pi \text{ rad}/\text{cycle})^2$, which is, of course, equal to one. We could insert that factor in the equation, but it does not seem worth the increased complexity and might promote the thought that we can depend on answers that are obtained without consideration of units. Note also that, while Eq. (15.9) was written in radian units, it holds equally for Hz units and will be used in that form in some S/N equations to follow.

If $H(f_m) \approx 1$ for $f_m < W/2$ but ≈ 0 otherwise, this is

$$N \approx \frac{N_0}{K_v^2 P} \left(\frac{1}{3}\right) \left(\frac{W}{2}\right)^3 = \frac{N_0 W^3}{24 K_v^2 P}, \quad (15.14)$$

and, using also (15.11),

$$\frac{S}{N} \approx \frac{24 P \sigma_{fs}^2}{W^3 N_0}. \quad (15.15)$$

Again we can provide video filtering on the demodulated signal, and the bandwidth of the noise will then be limited to the video bandwidth B_v instead of to $W/2$, giving

$$\frac{S}{N} \approx \frac{3 P \sigma_{fs}^2}{B_v^3 N_0}. \quad (15.16)$$

For sinusoidal modulation with amplitude Δf , $^1 \sigma_{fs}^2 = \Delta f^2/2$, and the above expression can be written

$$\frac{S}{N} \approx \frac{3 P \Delta f^2}{2 B_v^3 N_0}. \quad (15.17)$$

Since we do not want to admit any more noise than necessary, we set B_v equal to the maximum modulation frequency F_m , giving

$$\frac{S}{N} \approx \frac{3 P \Delta f^2}{2 F_m^3 N_0} = \frac{3}{2} \frac{\Delta f^2}{F_m^2} \left[\frac{P}{F_m N_0} \right]. \quad (15.18)$$

The expression in brackets is the S/N for a coherently detected 100% AM signal of modulation bandwidth F_m and carrier power P .² The preceding factor, $1.5 (\Delta f/F_m)^2$, is called the FM improvement factor. To take advantage of this factor, $\Delta f/F_m$ must be large and the price that is paid is additional bandwidth. According to Carson's rule, the required input bandwidth for FM is

$$W \geq 2(F_m + \Delta f), \quad (15.19)$$

whereas only $W \geq 2F_m$ is required for narrow-band FM.

¹ For any sine wave, the root mean square (rms) value is the amplitude divided by $\sqrt{2}$ so the mean square value is this squared.

² This assumes a rectangular filter to pass the signal band. Even so, the transmitted power of a 100% amplitude modulation (AM) signal is $1.5P$ because there is power in the sidebands in addition to that in the carrier, as compared to P for the FM signal. On the other hand, a suppressed carrier AM signal can have a transmitted power approaching that of the sidebands alone, $P/2$.

The loop that we have been discussing acts like a standard discriminator that multiplies the frequency deviation by a gain factor, $K_d = 1/K_v$ and employs a noise-limiting filter. We shall next consider the use of the PLL to provide not only conversion from frequency to voltage but also filtering.

15.3 FREQUENCY DISCRIMINATOR, FIRST-ORDER LOOP

If the first-order loop acts as the noise filter, the noise in Eq. (15.13) becomes

$$N = \frac{N_0}{K_v^2 P} \int_0^{W/2} \frac{f_m^2}{1 + (2\pi f_m / (K \text{ cycle}))^2} df_m \quad (15.20)$$

$$= \left(\frac{K \text{ cycle}}{2\pi} \right)^2 \frac{N_0}{K_v^2 P} \left[f_m - \frac{K \text{ cycle}}{2\pi} \tan^{-1} \left(\frac{2\pi}{K \text{ cycle}} f_m \right) \right]_0^{+W/2} \quad (15.21)$$

$$= (K \text{ rad})^2 \frac{N_0}{K_v^2 P} \left[\frac{W}{2} - \frac{K \text{ cycle}}{2\pi} \tan^{-1} \left(\frac{\pi}{K \text{ cycle}} W \right) \right]. \quad (15.22)$$

If the loop bandwidth is wide compared to $W/2$, the first terms of the small-argument expansion, $\tan^{-1} x \approx x - x^3/3$, can be used. The first term in the expanded product equals $W/2$, which cancels the preceding term leaving

$$N = \left(\frac{K \text{ cycle}}{2\pi} \right)^2 \frac{N_0}{K_v^2 P} \left[\frac{K \text{ cycle}}{2\pi} \left(\frac{\pi}{K \text{ cycle}} W \right)^3 \frac{1}{3} \right] = \frac{N_0 W^3}{24 K_v^2 P}. \quad (15.23)$$

This, not surprisingly, is the same as Eq. (15.14).

At the other extreme, let

$$K \text{ cycle} \ll \pi W \quad (15.24)$$

or, equivalently, $f_{-3 \text{ dB}} \ll W/2$, where $f_{-3 \text{ dB}}$ is the cutoff frequency of the first-order loop. Then $\tan^{-1}(\cdot)$ approaches $\pi/2$ and Eq. (15.22) becomes

$$N \approx K^2 \frac{N_0}{K_v^2 P} \left[\frac{W}{2} - \frac{K \text{ cycle}}{4} \right] \approx K^2 \frac{N_0 W}{2 K_v^2 P} \quad (15.25)$$

$$= \omega_{-3 \text{ dB}}^2 \frac{N_0 W}{2 K_v^2 P}. \quad (15.26)$$

Note that the filtering action is not very good here because the noise power still increases in direct proportion to the bandwidth of the input filter and could become extremely large. This is not surprising when we consider that $H(f_m)$ for a first-order loop never falls faster than f_m^2 , which is the rate at which S_{f_m} increases [see Eq. (15.10) and Fig. 15.7].

Note that we cannot also make use of the wide input bandwidth to get a higher FM improvement factor because the peak phase error could exceed the linear range, which we will now demonstrate. Set the loop bandwidth f_L equal to the maximum modulation frequency F_m . At this frequency, $\omega_L = K$, the phase error is related to the input phase by Eq. (2.26b), $\varphi_e = \varphi_{in}/(1 - j)$. Combining this with Eq. (7.7), the peak phase deviation can be written in terms of the input peak frequency deviation as

$$\Delta\varphi_e = \frac{\Delta f}{f_L\sqrt{2}} = \frac{\Delta f}{F_m\sqrt{2}} \quad (15.27)$$

If we restrict the phase deviation by $\Delta\varphi_e < \pi/4$ for linearity, (15.27) requires $\Delta f < 1.1F_m$, and the FM improvement factor is limited to

$$1.5(\Delta f/F_m)^2 = 1.8. \quad (15.28)$$

To summarize, while a first-order loop can be used as a frequency discriminator, it should not be depended on for noise filtering.

We have seen (Section 7.8.1) that a response similar to that of the first-order loop is available at the phase detector output u_1 when the loop has a low-pass filter. The low-frequency demodulation sensitivity is the same as for the first-order loop but, as illustrated by Fig. 7.17 for $\alpha = 0$, a choice of response shape is available; $\zeta = 0.7$ offers a particularly flat response. This response has the same problems relative to noise filtering, however, since it also falls at only -6 dB/octave at high frequencies. The noise at u_1 is given by Eq. (15.4) and, for a loop narrow compared to $W/2$, by Eq. (15.5). But this is the same as Eq. (15.25) when $K = K_p K_v$ so S/N is the same as with a first-order loop [Blanchard, 1976, p. 182].

15.4 FREQUENCY DISCRIMINATOR, SECOND-ORDER LOOP

If a second-order loop has a response similar to a first-order loop, we can expect similar results when using it as a frequency demodulator. In order to produce better noise filtering, the loop response must fall faster than -6 dB/octave at high frequencies. This implies that the loop filter must be low pass. (We showed in Section 7.8.1 that this response can also be obtained across the capacitor of a passive lag-lead loop filter.) We will find an expression for the demodulated noise in that case. As usual, from these results we can estimate the performance of higher order loops having similarly shaped responses.

We use Eq. (15.13) again and obtain the expression for $|H(f_m)|^2 f_m^2$ for $\alpha = 0$ (low-pass filter) by manipulating Eq. (6.4). That will enable us to obtain the integral in Eq. (15.12) in terms of B_n , which we have already found. From Eq. (6.4) we can obtain

$$|H(\omega)|^2 = \omega_n^4 \frac{1 + \omega^2(2\alpha\zeta/\omega_n)^2}{(\omega_n^2 - \omega^2)^2 + (2\zeta\omega_n\omega)^2}. \quad (15.29)$$

From this equation alone we can write the algebraic relationship

$$|\omega_m H(f_m)_{\alpha=0}|^2 = \left(\frac{\omega_n}{2\zeta}\right)^2 \left[|H(f_m)_{\alpha=1}|^2 - |H(f_m)_{\alpha=0}|^2 \right]. \quad (15.30)$$

Substituting this into Eq. (15.12b), we obtain

$$N = \left(\frac{N_0}{K_v^2 P}\right) \left(\frac{\omega_n}{2\zeta}\right)^2 \left\{ \int_0^{W/2} |H(f_m)_{\alpha=1}|^2 df_m - \int_0^{W/2} |H(f_m)_{\alpha=0}|^2 df_m \right\}. \quad (15.31)$$

As W approaches infinity, the integrals become noise bandwidths, B_n . Substituting their values from Eq. (14.10), we obtain the noise power at u_2 ,

$$N = \left(\frac{N_0}{K_v^2 P}\right) \left(\frac{\omega_n}{2\zeta}\right)^2 \frac{2\zeta \omega_n \text{ cycle}}{4 \text{ rad}} = \frac{1}{8} \left(\frac{N_0}{K_v^2 P}\right) \left(\frac{\omega_n^3}{\zeta} \text{ cycle}\right). \quad (15.32)$$

This response does provide effective (but not necessarily optimum) filtering, justifying our ignoring W when it is much wider than the effective filtering of the loop. We have not obtained an expression for the general lag-lead or integrator-and-lead filter, but Fig. 7.8 illustrates the benefits of the response that we have treated. The noise filtering action is improved by the presence of filter roll off at high f_m and a two-pole Butterworth response is available for $\zeta = 0.707$.

15.5 EXPECTED PHASE ERROR

We will now show that, in the absence of noise, a frequency demodulation loop that is made wider than the input half bandwidth $W/2$ should remain locked. By implication, a loop that is used also for filtering, that is, one that is made as wide as the maximum modulation frequency F_m but narrower than $W/2$, may not. Carson's rule states that the bandwidth necessary for a frequency-modulated (FM) signal is

$$W \geq 2[\Delta f + F_m], \quad (15.33)$$

where Δf is the peak frequency deviation and F_m is the highest modulation frequency. If the bandwidth of the PLL is made at least $W/2$ wide, the unity-gain frequency will be approximately

$$f_L \geq [\Delta f + F_m], \quad (15.34)$$

and the peak phase error will be

$$\varphi_e = \frac{\Delta f}{f_m}(1 - H(f_m)) \approx \frac{\Delta f}{f_m G(f_m)}. \quad (15.35)$$

Here the approximation is possible because the modulation frequencies are within the loop bandwidth where $(1 - H)$ can be approximated as $1/G$. Since $G(f_m)$ will be falling at least as fast as $1/f_m$, the value of φ_e will be no larger than what occurs at the unity-gain frequency f_L , so Eqs. (15.34) and (15.35) imply

$$\varphi_e \leq \Delta f/f_L < 1 \text{ rad}, \quad (15.36)$$

where the last inequality comes from Eq. (15.34). Therefore, even with no margin relative to $W/2$, operation is approximately linear.

15.6 SUMMARY OF FREQUENCY DISCRIMINATOR S/N

The mean square signal at the voltage-controlled oscillator (VCO) tuning line is σ_{fs}^2/K_v^2 and the noise is given by the various equations developed above. Since those noise expressions contain $N_0/(K_v^2 P)$, the signal ratio can be conveniently written

$$\frac{S}{N} = \gamma_d \frac{P \sigma_{fs}^2}{N_0}, \quad (15.37)$$

where γ_d is given in Table 15.1. The equations giving the noise are also indicated in the table. The equations have been written to make it easy to obtain the proper units for γ_d , which are, as can be seen from Eq. (15.37), Hz^{-3} . The voltage at the point where the demodulated signal is taken is v_D .

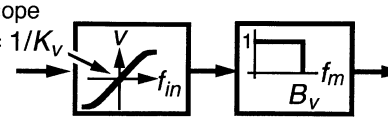
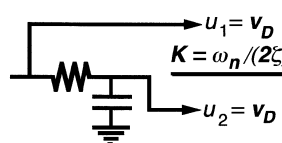
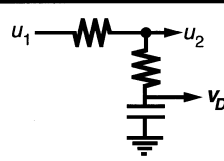
The first entry in Table 15.1 applies to the standard discriminator followed by a video filter. The first entry under “Loops” applies to any loop that does not supply filtering but depends on a preceding filter (W wide) or a subsequent video filter (B_v wide).

The remaining entries are for the condition wherein the loop provides significant filtering ($B_n \ll W/2$).

The second value of γ_d is given for the first-order loop and for a particular second-order loop when v_D is taken at the input of the loop filter, rather than at its output. The third value of γ_d applies to the second-order loop with a low-pass filter when v_D is taken from the loop-filter output and also to the more general second-order loop when v_D is taken across the capacitor in the passive loop filter. The equivalence of these two voltages was established in Section 7.8.1.

Example 15.1 S/N in a Frequency Discriminator A 1-mW signal is accompanied by a broad floor of additive noise of power density 10^{-6} W/Hz. The frequency

TABLE 15.1 Frequency Discriminator Characteristics

DISCRIMINATOR TYPE CONVENTIONAL + rectangular filter	$\gamma_d = \left(\frac{S}{N}\right) / \left(\frac{P\sigma_{fs}^2}{N_0}\right)$	CONDITIONS
slope $= 1/K_v$ 	$\frac{3}{B^3}$ where $B = W/2$ or B_v , whichever is much narrower (15.15), (15.16), (15.23)	
LOOPS		
Any wide loop	$u_2 = v_D$	$B_n \gg W/2$
First Order $u_1 = u_2 = v_D$ $\left\{ \gamma = \frac{2}{W\omega_{-3\text{ dB}}^2} \right\}$ see (15.26), (15.25)	$\frac{8\pi^2}{(K \text{ cycle})^2 W}$	
Second Order With low pass $(\alpha = 0)$  see (15.32)		$B_n \ll W/2$
with lag-lead, output across the shunt capacitor (see Section 7.8.1) $(0 < \alpha < 1)$ 	$\frac{4\zeta}{\pi f_n^3}$	

v_D is the discriminator output.

demodulated signal is obtained from the VCO tuning line of a loop with a low-pass filter. The loop has $K_v = 1 \text{ MHz/V}$, $\omega_n = 3000 \text{ rad/sec}$, $\zeta = 0.5$. The signal has an rms frequency deviation of 50 kHz. Find the signal-to-noise ratio of the demodulated signal ($\omega_m \ll \omega_L$).

From Eq. (15.32),

$$\begin{aligned}
 N &= \frac{1}{8} \left(\frac{10^{-6} \text{ W/Hz}}{(10^6 \text{ Hz/V})^2 10^{-3} \text{ W}} \right) \left(\frac{(3 \times 10^{-3} / \text{sec})^3}{0.5} \text{ cycle} \right) \left(\frac{\text{cycle}}{2\pi} \right)^2 \\
 &= 1.7 \times 10^{-7} \text{ V}^2.
 \end{aligned}$$

The signal power is obtained from (15.9) as

$$S = (5 \times 10^4 \text{ Hz})^2 / (10^6 \text{ Hz/V})^2 = 2.5 \times 10^{-3} \text{ V}^2.$$

The S/N is given by the ratio as 1.5×10^4 or 41.7 dB.

We could also use Eq. (15.37) with γ_d given by Table 15.1 as $4\zeta / (\pi f_n^3)$,

$$\gamma = \frac{4(0.5)}{\pi [(3000 \text{ rad/sec})(\text{cycle}/2\pi \text{ rad})]^3} = \frac{5.85 \times 10^{-9}}{\text{Hz}^3}.$$

This would be multiplied by

$$\frac{P\sigma_{fs}^2}{N_0} = \frac{10^{-3} \text{ W}(5 \times 10^4 \text{ Hz})^2}{10^{-6} \text{ W/Hz}} = 2.5 \times 10^{12} \text{ Hz}^3$$

from Eq. (15.30), giving $S/N = 1.5\text{E}4$, as before.

15.7 STANDARD DISCRIMINATOR AND CLICK NOISE

The expression for S/N in a standard discriminator (with video filtering dominant) was given by Eq. (15.18). This can be written in decibels as

$$\left[\frac{S}{N} \right]_{\text{dB}} = \left[\frac{S_i}{N_i} \right]_{\text{dB}} + 20 \log_{10} m + 1.8 \text{ dB}. \tag{15.38}$$

However, this does not hold at low S/N , as illustrated in Fig. 15.8. This is because of pulse noise, or “clicks.”

When listening to the output of a standard frequency discriminator (a circuit that produces voltage change proportional to frequency change) in a low S/N

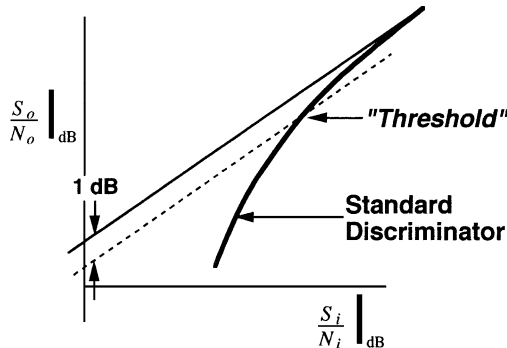


Fig. 15.8 S/N for standard discriminator.

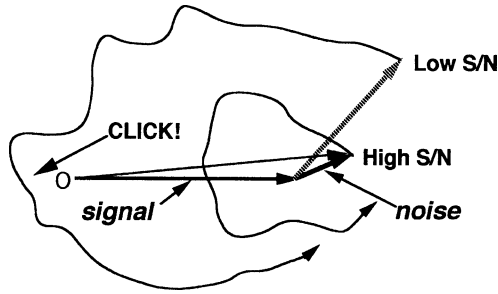


Fig. 15.9 Phasor representation of click noise.

environment, one can hear clicking sounds caused by impulsive noise, which results from sudden one-cycle phase transitions.

Figure 15.9 shows a phasor representation of a signal plus noise. At high S/N the noise phasor rotates about the tip of the signal phasor producing the path shown, and the deviation of the phase of the resultant from the signal's phase produces phase noise and, simultaneously, frequency noise. At low enough S/N , however, the trajectory of the noise vector sometimes circles the origin (O). When this happens the phase changes by 2π rad, causing a severe perturbation. Figure 15.10a illustrates the phase noise with steps corresponding to 2π rad changes when the origin is circled. Figure 15.10b shows the derivative of the phase noise, which is the frequency noise. The assumption of a narrow video filter permits these transitions to be approximated as phase steps and corresponding frequency impulses. The implied requirement that B_v be narrow compared to $W/2$, and a similar requirement on B_n of the PLL, is somewhat relieved by the rapidity of these nonlinear transitions.

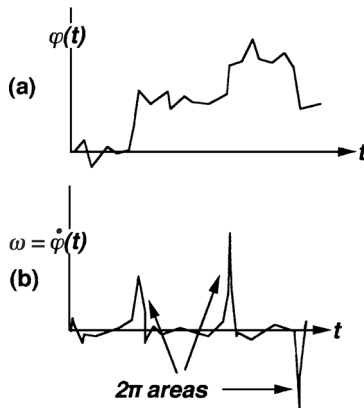


Fig. 15.10 (a) Phase and (b) frequency, $\omega = \dot{\phi}$, during clicks.

The effective frequency power spectral density due to the clicks is approximately³ (see Appendix 15.C)

$$S_{\omega} = 2\bar{F} \quad (15.39)$$

where \bar{F} is the rate of clicks. We will use units of Hz to represent clicks per second. If the signal is centered in the input filter and unmodulated,

$$\bar{F} = \frac{\gamma}{2\pi} \operatorname{erfc}\sqrt{\rho}, \quad (15.40)$$

where

$$\gamma^2 = \frac{\int_{-\infty}^{\infty} \Omega_o^2 |F(\Omega_o)|^2 d\Omega_o}{\int_{-\infty}^{\infty} |F(\Omega_o)|^2 d\Omega_o} \quad (15.41)$$

and $\operatorname{erfc}\sqrt{\rho}$ is plotted in Appendix 15.E.1. Here Ω is the offset from the center of the RF filter $F(\Omega)$ and

$$\operatorname{erfc}(x) \triangleq \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-u^2} du \quad (15.42)$$

is the complementary error function.

For a rectangular filter with RF bandwidth W ,

$$\begin{aligned} \gamma^2 &= \frac{\int_{-W/2}^{W/2} \Omega_o^2 d\Omega_o}{\int_{-W/2}^{W/2} d\Omega_o} = \frac{\frac{2}{3}(W/2)^2}{W} = \frac{W^2}{12} \\ \gamma &= \frac{W}{2\sqrt{3}} \left(\frac{2\pi \text{ rad}}{\text{cycle}} \right) = \frac{\pi}{\sqrt{3}} \frac{W}{\text{cycle}}. \end{aligned} \quad (15.43)$$

Combining this with Eqs. (15.39) and (15.40), we obtain the frequency power spectral density as

$$S_{\omega} = \left(\frac{W}{\sqrt{3} \text{ Hz}} \operatorname{erfc}\sqrt{\rho} \right) \frac{\text{Hz}^2}{\text{Hz}} = \frac{W}{\sqrt{3}} \operatorname{erfc}\sqrt{\rho}. \quad (15.44)$$

Some factors not included here are an increase in the number of clicks due to modulation (we might have guessed this from Fig. 15.9) and that clicks decrease the signal's apparent frequency deviation, causing a further deterioration at low S/N .

³ See Schwartz et al. (1966, pp. 144–154) with references to Rice (1963).

Adding this noise to what we included in Eq. (15.18) (based on the assumption of rectangular filters), we obtain

$$\frac{S}{N} = \frac{\frac{1}{2}\Delta f^2}{\frac{1}{3}(N_0/P)F_m^3 + S_\omega F_m} \quad (15.45)$$

$$= \frac{\frac{1}{2}\Delta f^2}{\frac{1}{3}(N_0/P)F_m^3 + (WF_m/\sqrt{3})\operatorname{erfc}\sqrt{\rho}}. \quad (15.46)$$

The numerator is the signal power; the first term in the denominator is the filtered input noise; and the second term is noise due to clicks. For comparison to S/N without clicks, we can rearrange this such that the numerator is Eq. (15.17):

$$\frac{S}{N} = \frac{(3P\Delta f^2)/(2N_0F_m^3)}{1 + \sqrt{3}(W^2/F_m^2)\rho \operatorname{erfc}\sqrt{\rho}}. \quad (15.47)$$

15.8 CLICKS WITH A PLL

In a PLL, if the transition is fast enough, the loop will not follow the phase change during the clicks. The phase error will skip a cycle, producing a relatively small phase perturbation at the VCO and resulting in a bipolar frequency transient that contains little energy compared to a frequency transient that has a one-cycle area. This is illustrated in Fig. 15.11. A limiter can be harmful here because it can prevent a desirable drop in input level during a click.

The effects of noise that is severe enough to cause cycle skipping properly belong in the subsequent chapters. However, we will borrow the results from there so we can treat, in one place, the effects of noise in a loop being used for frequency demodulation.

Based on an estimation that the times of the clicks are governed by a Poisson probability law (Blanchard, 1976, p. 340), the frequency power spectral density will be

$$S_\omega = 2 \bar{F}, \quad (15.48)$$

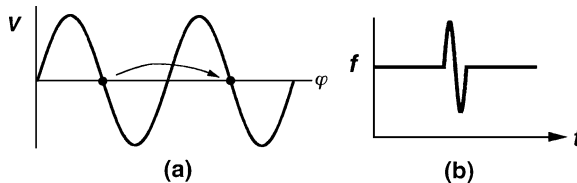


Fig. 15.11 Fast phase step of 1 cycle may cause the phase error in the PLL to skip a cycle (a), resulting in a low-energy frequency transient (b) at the PLL output.

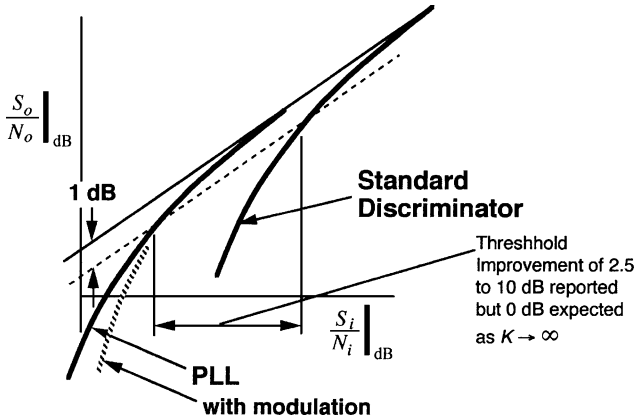


Fig. 15.12 S/N for a discriminator PLL.

where \bar{F} is the mean frequency of cycle skipping in hertz. Substituting this into Eq. (15.45), we obtain (still assuming a sharp postdetection filter)

$$\frac{S}{N} = \frac{\frac{1}{2} \Delta f^2}{\frac{1}{3} (N_0/P) F_m^3 + 2 \bar{F} F_M}. \tag{15.49}$$

We can use the results of Chapter 20 to give us \bar{F} . Then the threshold for a PLL demodulator, based on this theory, is better than for a standard discriminator, as shown in Fig. 15.12 [Gupta et al., 1968]. While modulation makes \bar{F} worse, as shown in the figure, it does not much affect the threshold. However, Rice (1963) has indicated a degradation of about 2 dB in one case. Although conditions and results vary widely the potential for improvement with a PLL is evident.

15.9 NOISE IN A CARRIER RECOVERY LOOP

If a carrier recovery loop (for example, Fig. 10.9) receives the same phase noise $S_{\varphi, \text{in}}$, as does the demodulator that it supports, that part of the phase noise that the carrier recovery loop follows will not corrupt the demodulated signal. However, the same cannot be said for additive noise.

In the presence of phase noise φ_n we can rewrite Eq. (10.14) as

$$\varphi_M = M\varphi_{\text{in}} = M(\varphi + \varphi_n) + i2\pi \Rightarrow M(\varphi + \varphi_n). \tag{15.50}$$

At the output of the carrier recovery PLL, the phase deviation due to noise at modulation frequency f_m , will be

$$\varphi'_n(f_m) = MH_{\text{CR}}(f_m)\varphi_n(f_m), \tag{15.51}$$

where H_{CR} is the loop transfer function, and after the output frequency divider it will be

$$\varphi_n''(f_m) = H_{CR}(f_m)\varphi_n(f_m). \quad (15.52)$$

Therefore, the signal that is demodulated, by being compared to the recovered carrier, will be accompanied by phase noise,

$$\varphi_{\text{dem}}(f_m) = \varphi_n(f_m) - \varphi_n''(f_m) = [1 - H_{CR}(f_m)]\varphi_n(f_m). \quad (15.53)$$

Since $\varphi_n(f_m)$ can be a narrow band of phase noise,

$$\varphi_n^2(f_m) = S_\varphi \delta f_m, \quad (15.54)$$

Eq. (15.53) also implies that the phase power spectral density for the demodulated signal $S_{\varphi, \text{dem}}$, in response to an input density $S_{\varphi, \text{in}}$, is

$$S_{\varphi, \text{dem}} = S_{\varphi, \text{in}} |1 - H_{CR}|^2. \quad (15.55)$$

Thus this noise is filtered by the error response of the carrier recovery loop, as discussed in Section 15.1.1. One would minimize the deleterious effects of this noise by maximizing the bandwidth of the carrier recovery loop. Equivalent performance can be expected from the Costas loop.

We would like to write the additive noise in terms of equivalent phase noise using Eq. (13.4). Unfortunately the relationship between a particular additive noise component and the signal changes as the data changes, and this de-correlates the phase noise at the carrier recovery output from the noise that accompanies the signal. For example, a particular sideband may be in quadrature with the signal, but, when the signal changes phase by 90° , it is suddenly in phase with the signal. Thus the effect of additive noise is not attenuated by the loop because the noise into the demodulator is not coherent with the noise at the output of the carrier recovery loop. In fact, that noise adds to the other noise in the demodulator.

The phase variance, due to additive noise, of the recovered clock in a Costas loop or multiply-and-divide recovery circuit (Fig. 10.9) is

$$\sigma_\varphi^2 = S_L^{-1} / \rho. \quad (15.56)$$

Here S_L is the “squaring loss,” ρ is the signal-to-noise ratio in the recovery loop bandwidth, and the PLL is preceded by a filter that is wide compared to the PLL [Lindsey and Simon, 1973, pp. 57–75]. Here S_L^{-1} is greater than unity and represents the additional noise due to mixing that is inherent in the multiplier. With (small) phase noise, the multiplier acts linearly on the phase deviation, but this is not the case with additive noise.

For a biphasic Costas loop, (Fig. 10.7) having a one-pole low-pass filter with a noise bandwidth of B_1 ,

$$S_L^{-1} = 1 + N_0 B_1 / (2P). \quad (15.57)$$

If the filter is an ideal rectangular filter,

$$S_L^{-1} = 1 + N_0 B_1 / P. \quad (15.58)$$

For the squaring circuit (Fig. 10.9 with $M = 2$), B_1 is the equivalent low-pass bandwidth (half the actual bandwidth) of a filter that is inserted between the multiplier and the PLL and Eqs. (15.57) and (15.58) hold as for a PLL with the same low pass.

Thus the optimum bandwidth is a tradeoff between wide bandwidth to reduce the effects of phase noise and a narrow bandwidth to minimize the effects of additive noise.

15.C APPENDIX: SPECTRUM OF CLICKS

Here we develop the frequency power spectral density due to clicks [Eq. (15.39)]. The frequency change due to a single click occurring at time T_i can be represented by the function $\Delta f(t - T_i)$. If this frequency deviation is processed linearly by the discriminator and then passed through the video filter F_v , the output voltage due to the click will equal the discriminator gain constant times the convolution of the pulse with the impulse response of the filter,

$$\Delta v_i(t) = K_d \Delta f(t - T_i) * F_v(t). \quad (15.C.1)$$

If the half bandwidth of the filter that precedes the discriminator is wide compared to the video filter bandwidth, $W/2 \gg B_v$, the typical click will have much of its energy at frequencies that are outside of the video bandwidth—that is, the click will be fast compared to $1/B_v$. Then the response of F_v to Δf will be similar to its response to an equivalent impulse. Substituting an impulse with area equal to one cycle for the pulse,

$$\Delta f(t - T_i) \approx 1\text{-cycle } \delta(t - T_i), \quad (15.C.2)$$

(15.C.1) becomes

$$\Delta v_i(t) = K_d \delta(t - T_i) * F_v(t) \text{Hz} = K_d F_v(t - T_i) \text{Hz}. \quad (15.C.3)$$

The Fourier transform of this is

$$\Delta v_i(f) = K_d F(f) \exp(-j2\pi f T_i) (\text{Hz-sec})^{1/2}. \quad (15.C.4)$$

The units here are consistent with equal areas under the squared moduls in the two Fourier domains [see Rayleigh’s Theorem (Bracewell, 1965)]. Because of the randomness of T_i , the spectrum of $\Delta v_i(f)$ has random phase and the transform of the sum of many pulses cannot be written at any frequency except zero, where it represents the sum of frequency changes. (There the sum will be zero if the number of positive and negative clicks are equal). However, the energy spectral density can be written at other frequencies. The variance of the sum of $\Delta v_i(f)$ in a narrow bandwidth δf for n pulses is⁴

$$\sigma_{vi}^2 = nK_d^2 |F(f)|^2 \delta f \text{ Hz-sec.} \tag{15.C.5}$$

This is an energy (for a 1-Ω system). The power is obtained by dividing by the elapsed time, which converts the total number of pulses n to the pulse rate \bar{F} /cycle. Dividing also by δf to get spectral density, we obtain

$$S_{v2} = \bar{F}K_d^2 |F_v(f)|^2. \tag{15.C.6}$$

Note that this is a two-sided density, since we have been dealing with Fourier transforms. This would be the output of the discriminator if it were driven by a one-sided frequency power spectral density of

$$S_\omega = 2\bar{F}. \tag{15.39}$$

15.E APPENDIX: erfc

Figure 15.E.1 is a plot of the complementary error function, $\text{erfc}\sqrt{\rho}$.

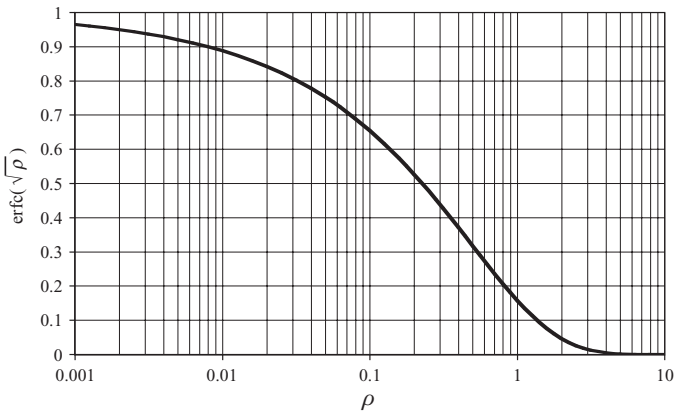


Fig. 15.E.1 Complementary error function of the square root of ρ .

⁴ In this section we are using $\delta(t)$ as the traditional symbol for an impulse and δf for a narrow frequency bandwidth.

CHAPTER 16

PARAMETER VARIATION DUE TO NOISE

The received signal strength can affect the loop parameters through its influence on the phase detector gain constant K_p . To maintain control of the loop parameters, we can employ various techniques to maintain them constant, but input noise can modify the effectiveness of these techniques. Here we will study gain-control techniques and how they are affected by noise.

16.1 PREVIEW

Before delving into the details let us consider gain-control at a higher level.

16.1.1 Automatic Gain Control (AGC)

We have seen [Eq. (13.17)] that K_p tends to be proportional to $\sqrt{P_c}$, where P_c is the signal power. If the phase detector were an ideal multiplier, this would be strictly true; but, even with practical balanced-mixer phase detectors, it is true as long as the signal is weak relative to the VCO. Since the loop parameters depend on K , and therefore on K_p , they will change as the signal strength changes. One way to maintain control of the design parameters is to vary the gain preceding the phase-locked loop (PLL) to compensate for changes in signal strength. Such an automatic gain control (AGC) subsystem is shown in Fig. 16.1.

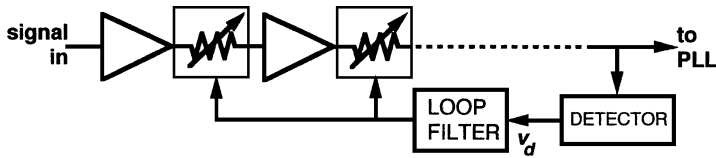


Fig. 16.1 AGC circuit.

Here the amplitude, or the power, of the signal entering the PLL is detected and used to control attenuators that precede the detector. There is a feedback control loop that attempts to maintain the strength of the signal seen by the PLL at a constant level as the input signal level varies. Of course, the degree to which this is accomplished depends on the AGC-loop gain, in a manner analogous to the PLL. Gain may be distributed, as shown, to maintain an optimum tradeoff between noise and saturation level in the chain.

16.1.2 Limiter

Another device to control the input amplitude to the PLL is the limiter. The limiter output is a constant-amplitude signal with zero crossings (the time when the sine wave changes sign) synchronized to those of the signal at its input. Figure 16.2 illustrates such a circuit.

In this simple form of a limiter the input sinusoid v_{in} produces an approximately square wave v_d across the back-to-back diodes. As v_{in} rises through zero, v_d is essentially equal to v_{in} because the diodes do not draw appreciable current at low voltages (assume a high impedance input to the filter). But, as the forward diode drop (≈ 0.7 V for silicon) is approached, the diodes begin to draw current and part of v_{in} is dropped across the resistor. As v_{in} far exceeds the diode forward drop, the diode impedance drops greatly and v_d stays approximately constant until v_{in} drops toward zero again. The same occurs on the negative half cycle and, as a result, an approximately square wave is produced across the diodes. If a sine wave is required, the square wave is passed through a filter that rejects harmonics of the square wave (ideally there are no even harmonics so the filter must primarily cut off the third harmonic). The fundamental component of the square wave is a sine wave with amplitude equal to $4/\pi$ times the amplitude of the square wave. The synchronism between input and output

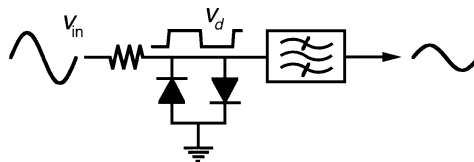


Fig. 16.2 Diode limiter.

zero crossings is obvious and the amplitude varies little as long as the input signal is sufficiently large. Thus the limiter passes phase (frequency) information but, ideally, strips off amplitude variations.

Digital logic gates, with their inputs properly biased, can perform the limiting function also, particularly if several are placed in series. One advantage of the use of gates is that they operate naturally in a “saturated” mode, where the output levels are fixed. Therefore, as the signal level grows, the gates at the high end of the chain begin to operate in normal fashion.

Sometimes the average output of the chain of gates is fed back to provide bias for the input, as shown in Fig. 16.3*a*. Again, this forms a closed loop. A net inversion in the forward path (such as obtained from an odd number of NAND gates) is required for negative feedback. It can be useful in placing the input bias at a level that will ensure operation with small signals, since the steady-state direct current (DC) solution places the input bias in the linear region between the one and zero levels.¹ However, this circuit can have very high gain, and it has the potential to oscillate if not properly designed. In addition, it may not work well with certain signals, such as large pulse trains that do not have 50% duty cycle or even large square waves, since they might have to be biased far from their mid-voltage to establish the duty factor required by the design. Another configuration, one that establishes the bias in an open-loop fashion, is shown at Fig. 16.3*b*.

Flip-flops can also serve as limiters if the signal is input to a properly biased clock line. At microwave frequencies, YIG (yttrium iron garnet) devices have been used for limiting.

16.1.3 Driving the Phase Detector Hard from the Signal

Since the output amplitude to a balanced mixer typically is little affected by small changes in the high level local oscillator (LO) input, why not make the varying signal, rather than the VCO, the high level input? This may be beneficial under some circumstances but it will only work over a limited range, between the power level at which the response to the LO flattens and the level at which the mixer’s diodes are destroyed. (We might also say this for the simple form of the limiter in Fig. 16.2, but the range there is not necessarily as limited.) In addition, if there are multiple

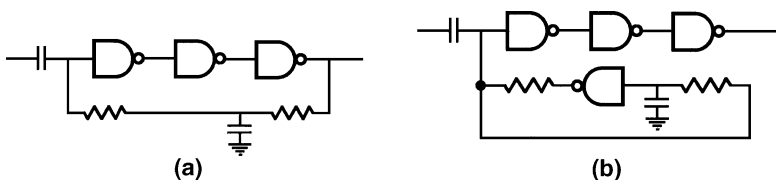


Fig. 16.3 Gate limiter circuit: (a) biased by feedback and (b) separate bias.

¹ Be cautious of possible negative consequences when some types of logic gates are held in this state.

signals at the input, their tendency to produce additional spurious signals in the mixer increases as their levels increase.

If we do use this method, our first representation of additive noise as phase modulation (PM) on the signal will have no problem (assuming the restrictions on this representation are met) and, of course, there will be no problem with true phase noise. However, our second representation of additive noise as multiplicative modulation will not fare well in this analysis because it depends on representation of the phase detector as a multiplier.

16.1.4 Effects of Variations

These control devices have inherent limitations in the presence of noise. The limitations are due to the difficulty in differentiating between the signal and noise. To maintain K'_p constant, the signal must be maintained constant. If the control device holds the sum of signal and noise constant, then K'_p will decrease as the noise increases relative to the signal. We will characterize the control devices by a parameter η , called the suppression factor, that indicates by how much K'_p is reduced due to the presence of noise. Later we will study how these imperfections affect performance.

16.2 AGC

Certain aspects of automatic gain control, in particular the coherent detector and the effect of AGC in aiding acquisition, were previously discussed in Chapter 9.

16.2.1 Types of Detectors

We can choose the detector in Fig. 16.1 from several types. A DC or low-frequency output is generated when current is passed through a nonlinear device such as a diode. All the even-order terms (i.e., those where v is raised to an even power) in Eq. (3.1) generate DC. We can see this in Eq. (3.4), which contains the terms

$$v_d = c \left[\frac{A^2 + B^2}{2} \right]. \quad (16.1)$$

Since we are only considering one signal in an amplitude detector (not two as in the phase detector), let $B = 0$. At low levels (small A), v_d is proportional to A^2 , for which reason the detector is called a square-law detector. Since the DC voltage from the n th-order term will be raised to the n th power, all of these terms will change more rapidly with A than will the second-order term. Therefore we can expect this square-law performance to dominate at low signal levels. As the level increases, however, the other terms come into play and the detector loses its square-law characteristic, commonly taking on a linear characteristic. We can see that, while the detector in Fig. 16.4 might generate a square-law output at low levels, it would become a rectifier at higher levels. When the diode begins operating like a switch, by being in either a

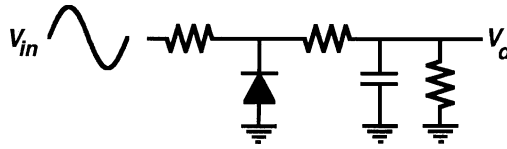


Fig. 16.4 Diode amplitude detector.

relatively high- or relatively low-impedance state during most of the input wave, it allows the capacitor to charge only when v_{in} is positive. Thus the capacitor tends to hold something proportional to the amplitude of v_{in} .

The coherent detector, shown in Fig. 16.5, can be a linear amplitude detector even at low levels. The detected output comes from a balanced mixer whose LO is in phase with the detected signal. By Eq. (3.5), this phase will produce an output proportional to the signal amplitude. The LO for the coherent detector's mixer is produced by the VCO in a PLL, which is locked to the signal being detected. If the PLL gain is high, its input to the phase detector (PD) will be in quadrature with the signal, but the 90° hybrid circuit ensures that quadrature at the PD mixer implies phase alignment at the other, amplitude detecting mixer. Unlike the simple diode detector, noise or other signals to which the PLL is not locked will produce a non-DC output that can be attenuated by a low-pass filter.

When the coherent detector is used to control an AGC, the simultaneously operating PLL and AGC loops can present a complex analysis problem; however once proper lock has been achieved, the coherent detector can provide accurate detection. Conversely, it does not provide such detection before the PLL is locked, since the output from the balanced mixer, due to the signal, will then be roughly a sinusoid.

16.2.2 Square-Law Detection

With square-law detection the output of the detector is proportional to total power so the AGC circuit will attempt to keep it constant. The power gain of the AGC circuit

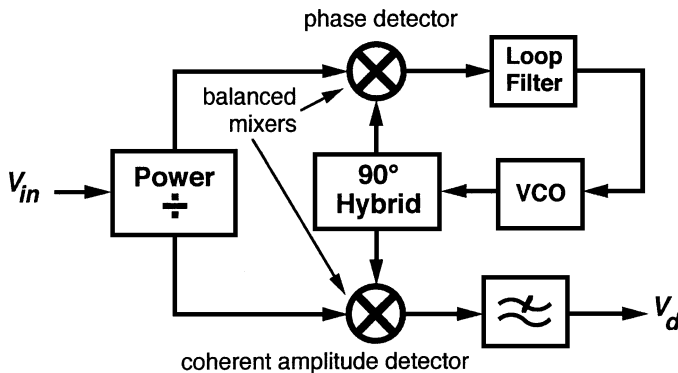


Fig. 16.5 Coherent amplitude detector.

can be written

$$G_p = \frac{P_o}{N_i + S_i} = \frac{P_o/N_i}{1 + S_i/N_i}, \tag{16.2}$$

where P_o is the total output power, S_i is the input signal power, and N_i is the input noise power. The output signal power is

$$S = G_p S_i = \frac{P_o}{N_i + S_i} S_i = \frac{P_o S_i/N_i}{1 + S_i/N_i} \tag{16.3}$$

while the noise power output is

$$N = G_p N_i = \frac{P_o}{N_i + S_i} N_i = \frac{P_o}{1 + S_i/N_i}. \tag{16.4}$$

These relationships are plotted in Fig. 16.6. When the signal power dominates (right-hand side of the figure), the total output power P_o is largely signal power; thus, when the AGC controls the gain to achieve constant P_o , it is also controlling for constant signal level, which is what we want. Since the signal is held constant at the output, any change in input signal-to-noise (S/N) is reflected as a change in output noise. Conversely, when the noise power dominates, it is held constant at P_o and the output signal level changes with S/N .

Consider what happens if the noise level is constant. If the maximum gain is not high enough to amplify the noise to the design output level P_o , the gain will go to its maximum value when a signal is absent or very small. As the input signal power increases, it will be amplified linearly until the output power reaches P_o , at which point the gain will begin to decrease to hold the total output power at P_o .

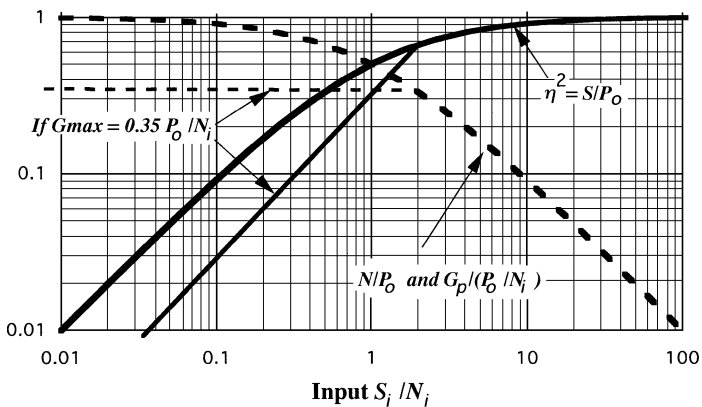


Fig. 16.6 AGC characteristics normalized to P_o , the limiting power level at the output, square-law detector.

The gain-limited condition is represented by the straight lines in Fig. 16.6. It occurs at signal levels where the gain G_p is less than the required value, P_o/N_i , implying $S_i < (P_o/G_p) - N_i$.

From here we will consider operation in the region where sufficient gain is available to maintain constant power output (i.e., not within the straight-line regions). We can find the maximum gain required from the AGC circuit from Eq. (16.2) with S_i and N_i set at minimums.

We define a factor η_A , called the AGC “suppression” factor, that describes the reduction in signal amplitude relative to the level set by the AGC in the absence of noise,

$$\eta_A^2 \triangleq \frac{S}{P_o}. \quad (16.5)$$

The output noise can be written in terms of η_A as

$$N = P_o - S = P_o(1 - \eta_A^2). \quad (16.6)$$

From these two equations the S/N at the output, which must equal the input S_i/N_i , since an AGC provides linear amplification, can be written in terms of η_A as

$$\rho \triangleq \frac{S}{N} = \frac{\eta_A^2}{1 - \eta_A^2} = \rho_i \triangleq \frac{S_i}{N_i}, \quad (16.7)$$

and, in turn, the suppression factor can be written in terms of ρ as

$$\eta_A^2 = \frac{\rho}{1 + \rho} = \frac{\rho_i}{1 + \rho_i}. \quad (16.8)$$

This is plotted in Fig. 16.10. Recall that η_A applies to the use of square-law detectors. Noise does not suppress the signal ($\eta = 1$) when coherent detection is employed.

16.3 LIMITER

A second way to obtain constant input amplitude is through use of a limiter, which does not allow the voltage to rise above a level A' nor to drop below a level $-A'$. This is illustrated in Fig. 16.7. A sine wave is limited to excursions that are $\pm A'$ relative to its average value. If the wave is large compared to the limits (which we will assume), the output is approximately a square wave of amplitude A' . This is sent through a filter that passes only the fundamental component, the amplitude of which is $(4/\pi)A'$.

16.3.1 Limiting in the Presence of Small Noise

If additive noise is small enough so we can approximate it as phase and amplitude noise on the signal, we can represent the phase noise by Eq. (13.3) and (13.4). Hard

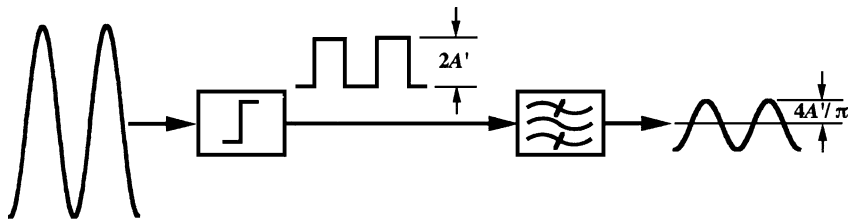


Fig. 16.7 Limiter.

limiting (where a square wave is created as illustrated above) will remove amplitude variations, leaving only the phase modulation. Thus the relative phase power spectral density [e.g., in decibels relative to the carrier per hertz (dBc/Hz)] will be reduced by passage through the limiter, but the noise will be all phase noise. See Fig. 16.8. Since this is the only kind that the PLL follows, the improvement in S/N is only apparent so far as the loop is concerned.

L_φ is unchanged in passing through the limiter and thus so is S_φ . Therefore the equations in Section 13.4 still apply. The main change is in the value of P_c . It becomes set at a level determined by the limiter,

$$P_c = [(4/\pi)A']^2/2 = 0.81(A')^2. \tag{16.9}$$

This permits K'_p to be independent of the signal.

16.3.2 Limiting in the Presence of Large Noise

If S_φ cannot satisfy the small-modulation-index restriction of Eq. (11.29) we must use the alternate representation of Section 13.3, multiplication of quadrature carriers. The signal is represented by v_c in Fig. 16.9, and the in-phase and quadrature noise voltages are represented by v_I and v_Q respectively. The mean square values of these components are

$$\langle v_c^2 \rangle = P'_c \tag{16.10}$$

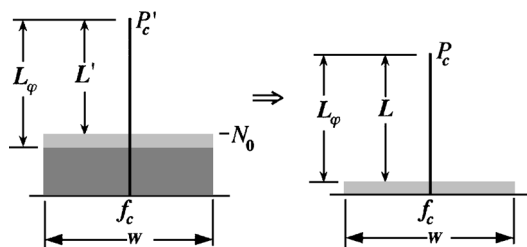


Fig. 16.8 Spectral changes in passing small noise through a limiter. The ordinates are logarithmic (dB).

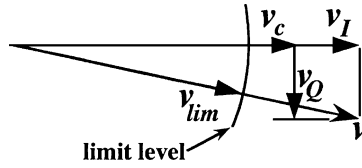


Fig. 16.9 Carrier and noise.

and

$$\langle v_I^2 \rangle = \langle v_Q^2 \rangle = N_0 W. \tag{16.11}$$

The resultant, v , varies in both amplitude and phase due to the noise. The output of the limiter v_{lim} is the same as v except that its amplitude is constant; thus it consists only of signal and phase noise.

The ratio of signal to noise at the output of an ideal limiter is [Blachman, 1966, pp. 86–87]

$$\frac{S}{N} = \frac{a}{b - a}, \tag{16.12}$$

where

$$a = \pi \rho_i \left[I_0 \left(\frac{\rho_i}{2} \right) + I_1 \left(\frac{\rho_i}{2} \right) \right]^2, \quad b = 4e^{\rho_i}, \tag{16.13}$$

where ρ_i is the input S/N and I_0 and I_1 are modified Bessel functions of the first kind. The ratio of signal to total power P_o is

$$\eta_L^2 = \frac{S}{S + N} = \frac{a}{b} = \left\{ \frac{\sqrt{\pi \rho_i}}{2} e^{-\rho_i/2} \left[I_0 \left(\frac{\rho_i}{2} \right) + I_1 \left(\frac{\rho_i}{2} \right) \right] \right\}^2 \tag{16.14}$$

and η_L is the limiter suppression factor. An approximate formula, accurate to 0.1 dB, is [Tausworthe, 1966]

$$\eta_L^2 \approx \frac{0.7854 \rho_i + 0.4768 \rho_i^2}{1 + 1.024 \rho_i + 0.4768 \rho_i^2}. \tag{16.15}$$

For high ρ_i , we can use the first term in the large argument expansion [Abramowitz and Stegun, 1964, p. 377] for I_ν (see Appendix 16.A),

$$\lim_{z \rightarrow \infty} I_\nu(z) = \frac{e^z}{\sqrt{2\pi z}} \left\{ 1 - \frac{4\nu^2 - 1}{8z} + \frac{(4\nu^2 - 1)(4\nu^2 - 9)}{128z^2} - \dots \right\}, \tag{16.16}$$

to give $I_0(\rho_i/2) \approx \exp(\rho_i/2)/\sqrt{\pi\rho_i} \approx I_1(\rho_i/2)$ so Eq. (16.14) approaches $\eta_L^2 = 1$, as it must. Using the first two terms in Eq. (16.16), Eq. (16.12) also approaches $2\rho_i$, under this extreme condition, corresponding to the elimination of amplitude noise. For low ρ_i , $I_\nu(z) \approx (z/2)^\nu / \Gamma(\nu + 1)$ [Abramowitz and Stegun, 1964, p. 375], which approaches 1 for I_0 and approaches 0 for I_1 , so Eq. (16.14) approaches

$$\eta_L^2 \approx (\pi/4)\rho_i. \tag{16.17}$$

That is, the ratio of the output signal to P_o is 1.05 dB lower than ρ_i .

Suppression factors for both the limiter and the AGC are plotted in Fig. 16.10. Since the limiter is not linear, the output S/N can, and does, differ from the input S/N . Because the output is a square wave at the level set by the limiter, the power near the signal frequency, signal plus noise, is constant at P_o . The signal portion is $\eta_L^2 P_o$ the noise power is $(1 - \eta_L^2)P_o$ and the output signal-to-noise ratio is

$$\frac{S}{N} = \frac{\eta_L^2}{1 - \eta_L^2}. \tag{16.18}$$

This is plotted in Fig. 16.11. At high ρ_i , the 3-dB increase in S/N is ignorable, insofar as its effect on a PLL, because the noise that remains is entirely phase noise and thus the improvement is in the type of noise that does not affect the loop anyway. At low ρ_i , the deterioration of S/N as seen by a PLL is not as bad as it would seem from the plot because the additional noise is distributed more heavily toward higher modulation frequencies that the loop is less likely to follow, assuming that the loop is narrow compared to the noise band into the limiter. The effect seen at the output of such a

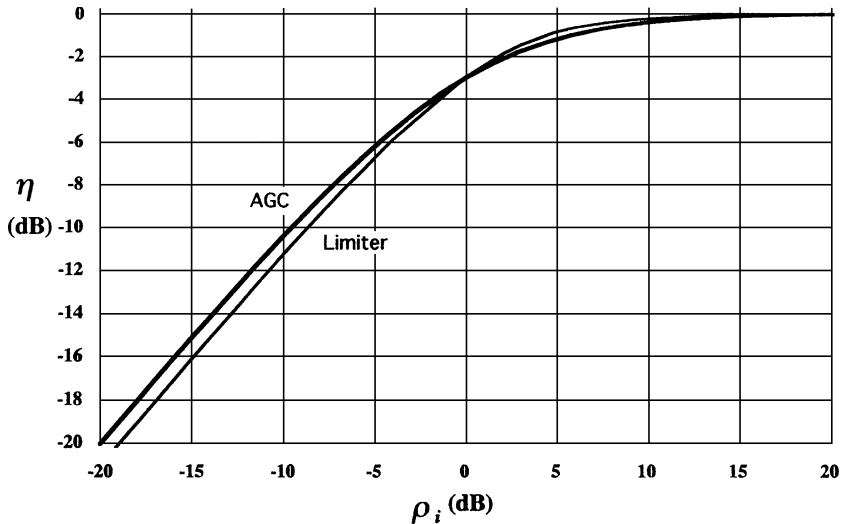


Fig. 16.10 AGC and limiter suppression factors. Note: η in decibels is $20 \log_{10} \eta$.

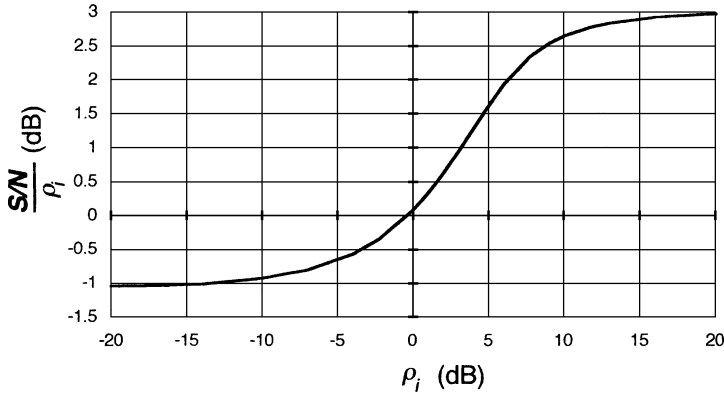


Fig. 16.11 Ratio of output to inputs S/N vs. input $S/N = \rho_i$ with a limiter (equals 0 dB with AGC). The effect at the output of a PLL is much less than this seems to indicate. It depends also on the type of noise and the distribution of the density.

PLL corresponds to a deterioration of ρ_i , compared to what would occur with the same loop parameters in the absence of the limiter, by only 0.25 dB, if the filter preceding the limiter is a single-pole filter, 0.49 dB if it is Gaussian, and approximately 0.64 dB if it is a rectangular filter [Springett and Simon, 1971]. Because this effect is small (and not simply quantified) and because the AGC has no effect, we will neglect any effect on S/N in a PLL due to either the AGC or the limiter.

16.4 EFFECTS OF GAIN VARIATION ON LOOP PARAMETERS

If $K_p = K_{p0}$ with the signal power at the design level P_0 , the value when the signal has been reduced to $\eta^2 P_0$ will be $K_{\rho\eta} = \eta K_{p0}$, where η is η_A or η_L . Therefore, the loop gain will decrease from K_0 to $K_\eta = \eta K_0$ in the presence of noise.

The noise bandwidth for a first-order loop, or a second-order loop with $\alpha = 0$, is given by Eq. (14.7). With the gain multiplied by η , this becomes

$$B_{n\eta} = B_{n0}\eta \tag{16.19}$$

For other second-order loops, the parameters change as follows:

$$\omega_{n\eta}^2 = K_\eta \omega_\rho = \eta K_0 \omega_\rho = \eta \omega_{n0}^2; \tag{16.20}$$

$$\omega_{n\eta} = \sqrt{\eta} \omega_{n0}; \tag{16.21}$$

$$\zeta_\eta = \frac{1}{2} \left(\frac{\omega_p}{\omega_{n\eta}} + \frac{\omega_{n\eta}}{\omega_z} \right) = \frac{1}{2} \left(\frac{\omega_p}{\sqrt{\eta} \omega_{n0}} + \frac{\sqrt{\eta} \omega_{n0}}{\omega_z} \right). \tag{16.22}$$

These effects can be seen in Fig. 4.3 and 4.4 if the various value of K shown there are taken to represent reductions in the effective value of K due to noise. Apparently the

noise bandwidth and other bandwidths generally drop as η drops. While the effect of η on the loop parameters can be computed for any particular case, it is instructive to consider one case. It is not uncommon to have ω_n much higher than the geometric mean of ω_p and ω_z in the manner illustrated by the highest curve in Fig. 4.4 and in Fig. 16.12. (Typically $\omega_p \ll \omega_n$ to keep the bandwidth narrow compared to the acquisition range and $\omega_z \ll \omega_n$ to give phase margin). This restriction,

$$\omega_n \gg \sqrt{\omega_p \omega_z}, \tag{16.23}$$

can be rewritten

$$\frac{\omega_n}{\omega_z} \gg \frac{\omega_p}{\omega_n} \tag{16.24}$$

and therefore

$$\zeta_\eta \approx \frac{1}{2} \left(\frac{\sqrt{\eta} \omega_{n0}}{\omega_z} \right) = \sqrt{\eta} \zeta_0. \tag{16.25}$$

From Eq. (14.10). with Eqs. (6.5c) and (16.23), we obtain

$$B_n = \frac{\omega_n}{4} \left(\frac{1}{2\zeta} + \frac{2\zeta}{\left(1 + \frac{\omega_z \omega_p}{\omega_n^2}\right)^2} \right) \frac{\text{cycle}}{\text{rad}} \approx \frac{\omega_n}{4} \left(\frac{1}{2\zeta} + 2\zeta \right) \frac{\text{cycle}}{\text{rad}}. \tag{16.26}$$

Substituting for ζ from Eq. (16.25), we obtain the noise bandwidth when the loop parameters are modified by $\eta = \eta_A$ or $\eta = \eta_L$,

$$B_{n\eta} |_{\omega_n \gg \sqrt{\omega_p \omega_z}} = \frac{\sqrt{\eta} \omega_{n0}}{4} \left(\frac{1}{2\sqrt{\eta} \zeta_0} + 2\sqrt{\eta} \zeta_0 \right) \frac{\text{cycle}}{\text{rad}} = \frac{\omega_{n0}}{8\zeta_0} (1 + 4\eta \zeta_0^2) \frac{\text{cycle}}{\text{rad}}. \tag{16.27}$$

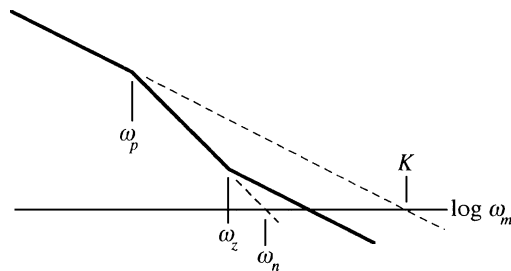


Fig. 16.12 Bode plot of a “high-gain” loop.

Thus the noise bandwidth drops with decreasing suppression factor. Since the latter decreases with increasing noise, the noise at the output of the PLL tends to be more constant than it would in the absence of suppression.

Problem 16.1 illustrates how noise can reduce bandwidth and phase margin in a loop fed by an AGC. It is apparent from Fig. 16.12 why the phase margin would decrease when the gain decreases due to the noise. Under what conditions would noise tend to increase phase margin?

Example 16.1 AGC Suppression Factor A signal has power equal to the noise in a 1-MHz rectangular filter passband. The phase variance at the output of a high-gain loop due to the noise is 0.01 rad^2 and the loop has $\zeta = 1$ under these conditions. Compare the change in variance when the signal drops 10 dB with and without an AGC circuit, one that uses a square-law detector.

From Fig. 16.10, a drop from $\rho_i = 0 \text{ dB}$ to $\rho_i = -10 \text{ dB}$ is accompanied by a 7-dB reduction (i.e., $10^{-7/20}$) in η . According to Eq. (16.27), $B_{n\eta}$ is proportional to $(1 + 4\eta\zeta_0^2) = (1 + 4\zeta_\eta^2)$ and, since ζ_η is initially unity, the factor equals $(1 + 4) = 5$ when $\rho_i = 0 \text{ dB}$ but drops to $[1 + 4 \times 10^{7/20}] = 2.8$ when η drops 7 dB. Thus the noise bandwidth ratio is $B_{n\eta}(\rho_i = -10 \text{ dB})/B_n(\rho_i = 0 \text{ dB}) = 2.8/4 = 0.7$. Without the AGC, the 10-dB decrease in ρ_i will cause a 10-dB increase in the phase variance, to $0.01 \text{ rad}^2 (10) = 0.10 \text{ rad}^2$. With the AGC, the noise bandwidth simultaneously shrinks to 0.7 of its former value so the variance is 0.07 rad^2 .

16.5 EFFECT OF AGC OR LIMITER ON AN OPTIMIZED LOOP

In Section 14.3 we found that we can minimize $\sigma_{\varphi,\text{out}}^2$ under a constraint on the allowed integrated squared transient error E^2 by choosing a particular kind of loop, depending on the type of transient. For example, for a frequency step, that loop is a type 2 second-order loop with $\zeta = 1/\sqrt{2}$ (Section 14.3.4). Once we have chosen ω_n to give the desired loop bandwidth, as long as we maintain that value, E^2 will not change for a fixed input transient, but $\sigma_{\varphi,\text{out}}^2$ will vary inversely with the input signal-to-noise ratio ρ . The loop will remain optimum under constraint, giving minimum $\sigma_{\varphi,\text{out}}$ with the fixed E . One way to ensure that ω_n does not change, regardless of either signal or noise power, is to employ an AGC with a coherent detector.

While it is possible for the value of E to be chosen independently of the resulting value of $\sigma_{\varphi,\text{out}}$, it seems more likely that, in choosing ω_n , we will trade off E against $\sigma_{\varphi,\text{out}}$ and we have obtained expressions to help us do that, Eq. (14.24) in the case of a frequency step. We would then be interested in how E and $\sigma_{\varphi,\text{out}}$ vary when ρ changes from its design value. Jaffe and Rechtin (1955) showed the change in both performance measures with ρ for a particular frequency-step problem where the loop had been optimized at $\rho = 0 \text{ dB}$. They did this for a loop preceded by a coherent AGC, for one preceded by a limiter, and for one that was re-optimized at each new value of ρ . To understand their results, we need to know a little more about the loop optimization process (Appendix i14.B).

The optimization process minimizes a weighted sum.

$$\Sigma^2 = \sigma_{\varphi, \text{out}}^2 + \lambda^2 E^2, \quad (16.28)$$

and is able to determine the kind of loop required without a known value for the constant λ . While we have not shown λ in Section 14.3, it is determined implicitly when we use an equation such as (14.23) to determine ω_n and thus set E and $\sigma_{\varphi, \text{out}}$. For example, when we set the values $E = E_1$ and $\sigma_{\varphi, \text{out}} = \sigma_1$ for a frequency step, this implicitly sets² $\lambda = \sigma_1 / (\sqrt{3}E_1)$. When Jaffe and Rechtin re-optimized the loop to minimize Σ as ρ changed, they did so without changing the value of λ from that set at the design point so Σ was minimized with that weighting factor. Using the fixed λ they showed (Fig. 16.13) the excess of Σ with coherent AGC and limiter relative to its value for a re-optimized loop. We see fairly broad optimums that tolerate considerable variation in ρ . We also see that the limiter keeps Σ closer to its value with a continuously optimized loop than does the coherent AGC. This is due to the beneficial shrinking of the loop bandwidth in the presence of greater noise (as well as to what we are here considering to be optimum). Results in the presence of AGC with a square-law detector should be like those with a limiter since their suppression factors are similar.

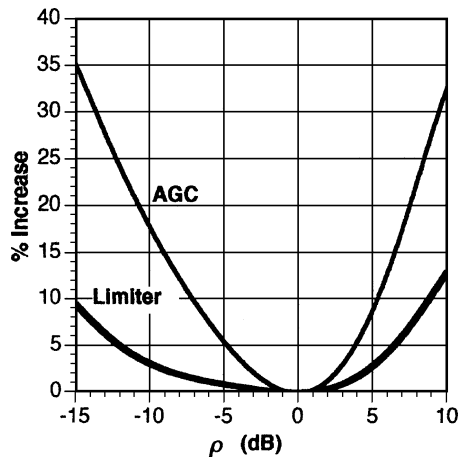


Fig. 16.13 Percent increase in Σ relative to continuously optimized loop as input S/N varies. The loop is optimized for minimum Σ with a frequency step at $\rho = 0$ dB. The percentage by which Σ exceeds the value that would be obtained by reoptimizing for each value of ρ is plotted for loops whose gain is controlled by coherent AGC and by a limiter. [Based on Jaffe and Rechtin (1955, Fig. 10) © 1955 IEEE.]

²By combining the three equations preceding Eq. (1) in Jaffe and Rechtin (1955).

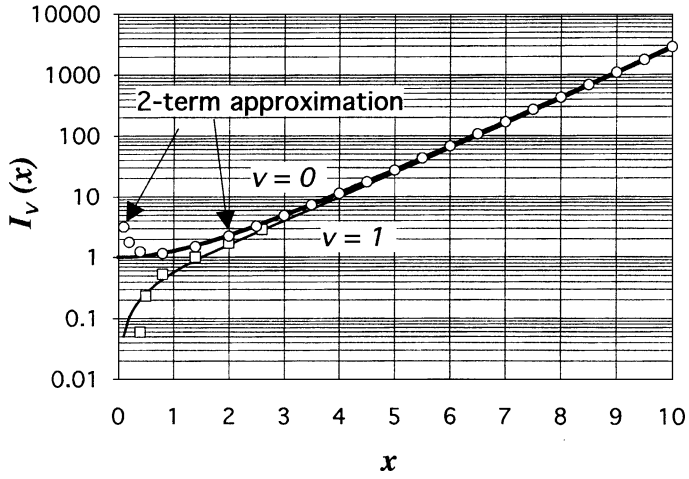


Fig. 16.14 Modified Bessel functions.

16.A APPENDIX: MODIFIED BESSEL FUNCTIONS

Figure 16.14 shows I_0 and I_1 plus points obtained from an approximation using the first two (of the three) terms in Eq. (16.16).

CHAPTER 17

CYCLE SKIPPING DUE TO NOISE

In this chapter we consider performance of the PLL when the signal-to-noise ratio is so poor that it causes cycle skipping. Figure 17.1 shows the phase error of a loop in the presence of significant noise. The noise here has half the power, in the (linear) loop bandwidth, of the signal. For a while the result is phase noise on the VCO, which is reflected in the phase error [see (a)]. Then the deviation becomes large enough that the phase skips a cycle, locking again one cycle from the original phase. Over a longer observation period, multiple cycle skips will be observed, including events during which two or more cycles are skipped without an obvious pause as a potential lock point is passed [see (b)]. Analysis under these conditions is understandably difficult, but some useful results have been obtained and simulation provides a powerful means of ascertaining performance.

Mathematical expressions that describe loop performance are most valuable, where they can be obtained. Accurate simulations offer a less efficient means of determining expected performance but have the advantage of applying where formulas cannot be devised and can be used to find the range of conditions over which the formulas do apply. Experiments are of great value in verifying the accuracy of simulations, and, of course, for the data they provide directly, but they can be difficult (expensive). As a result, relatively little experimental information is available. We will try to combine all of these to the best advantage.

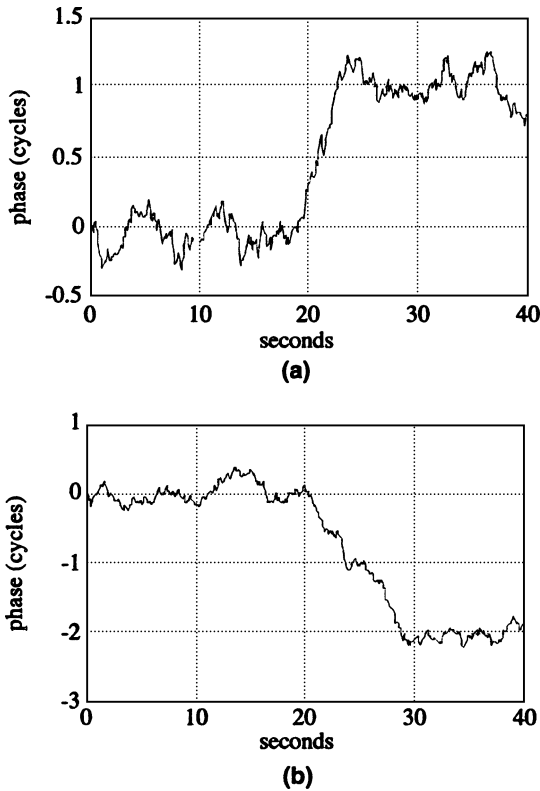


Fig. 17.1 Examples of phase error in the presence of noise, $\rho_{L0} = 2$, $\alpha = 1$, $\zeta = 0.7$, $\omega_n = 1$ rad/sec. (These plots were produced by software described in Appendix i17.M.1.)

17.1 PHASE

17.1.1 Fokker–Plank Method

Viterbi (1963, 1966) employed the Fokker–Plank differential equation to describe the change in the probability density $p(\varphi_n, t)$ of the phase error over time. He attempted to solve that equation for time approaching infinity to obtain the steady-state probability density of phase error. He did obtain the solution for simple loops and suggested an approximation for more complex loops in which we determine the S/N under linear assumptions, ρ_{L0} , and then determine how a simple loop, with that same ρ_{L0} , would react when analyzed exactly.

We assume no input modulation and zero phase error so that u_1 in Fig. 13.4 becomes

$$u_1(t) = -K'_p \sin \varphi_n(t) + u_{ln}(t), \tag{17.1}$$

where φ_n is the output phase in the presence of noise.

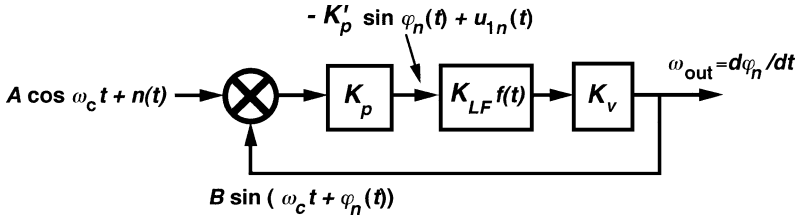


Fig. 17.2 Loop in the presence of noise.

The differential equation describing the loop performance is (see Fig. 17.2)

$$\omega_{out}(t) = \frac{d\varphi_n}{dt} = K' \{-\sin \varphi_n(t) + n'(t)\} * f(t), \tag{17.2}$$

where Eq. (13.20) becomes $n'(t) = u_{in}(t)/K'_p$ under the condition of zero phase error and $*f(t)$ represents convolution with the filter impulse response.

17.1.1.1 Assumption Regarding the Nature of the Noise. Based on the premise that the VCO phase $\varphi_n(t)$ changes slowly compared to the input additive noise, $u_{in}(t)$ is assumed to be Gaussian and white. In other words, the frequencies f_m of the modulation on the noise carriers, due to noise at $f_c \pm f_m$, are predominantly high compared to the loop noise bandwidth B_n . This in turn is based on the premise that the input bandwidth W is much wider than twice the loop bandwidth,

$$W \gg 2B_n. \tag{17.3}$$

In Section 13.5 we concluded that the noise introduced into the loop can be considered random to the degree that the input noise bandwidth greatly exceeds the loop noise bandwidth.

17.1.1.2 First-Order Loop. For the first-order loop, $f(t)$ in Eq. (17.2) equals one, so we have

$$\frac{d\varphi_n}{dt} = K' \{-\sin \varphi_n(t) + n'(t)\}. \tag{17.4}$$

This is a Markov process in that the derivative of φ_n at time t depends only on its value at time t . As a result, its probability density p is described by the Fokker–Plank equation,

$$\frac{\partial p(\varphi_n, t)}{\partial t} = -\frac{\partial}{\partial \varphi_n} \left[-p(\varphi_n, t) K' \sin \varphi_n \right] + \frac{K'^2 N_0}{4 P} \frac{\partial^2 p(\varphi_n, t)}{\partial \varphi_n^2}. \tag{17.5}$$

If we solve this equation, starting with the initial condition that the phase error is zero [i.e., $p(\varphi_n, 0)$ is an impulse at $\varphi = 0$], the solution for a sequence of times, $0, t_1, t_2, t_3,$

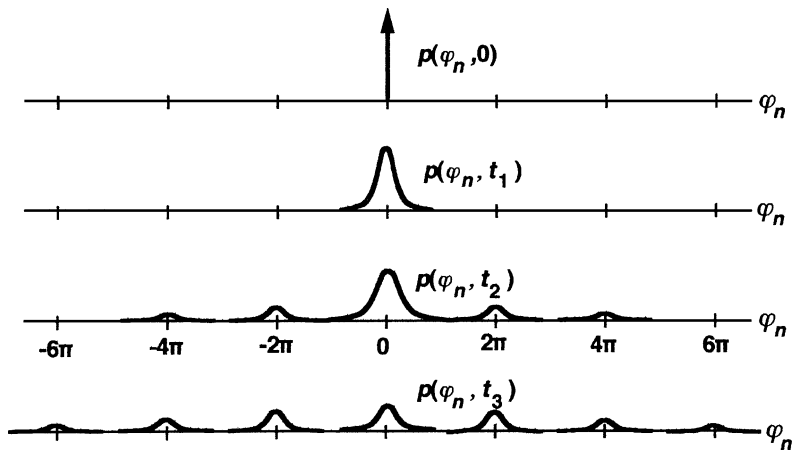


Fig. 17.3 The diffusion of the error probability as time progresses.

is shown in Fig. 17.3. It is easy to see that, as time increases, $p(\varphi_n, t)$ diffuses so the ultimate value approaches zero everywhere. This simply reflects the fact that, if there is any chance that the phase error will skip a cycle, as time approaches infinity the probability that it will be in any particular 2π range approaches zero.

We have obtained the correct answer to the wrong question.

To get a useful answer, we must revise the question. It would be useful to know the density of phase after a given finite time or to know it after a long (infinite) time, mod 2π . We will obtain the density mod 2π and we will later characterize the time dependence of the phase statistics by obtaining the mean time between cycle skips.

We need a function of angle φ that represents the probability that the phase is at $\varphi \pm m2\pi$ for any m . To obtain this function, Viterbi (1966) used the related function

$$P(\varphi_n, t) = \sum_{-\infty}^{\infty} p(\varphi_n + m2\pi, t). \tag{17.6}$$

This is like the original function except that each 2π -wide region contains an identical distribution. P begins with impulses at $m2\pi$ for all m . As P diffuses with time, the probability that the phase will jump out of the primary interval $-\pi < \varphi < \pi$ is the same as the probability that it will jump into that interval. As the edges of the distribution at $\pm\pi$ build up, the probability of the phase moving from $\pi - \varepsilon$ to $\pi + \varepsilon$ is the same as the probability of the phase moving from $-\pi - \varepsilon$ to $-\pi + \varepsilon$. As a result, in the $\pm\pi$ region, Eq. (17.6) can be written

$$P(-\pi < \varphi_n < \pi, t) = p(\varphi_n, t) \text{ mod } 2\pi, \tag{17.7}$$

which is the desired function. Equation (17.6) is substituted for $p(\varphi_n, t)$ in Eq. (17.5) and the initial condition is set as an impulse of probability density $p(0,0)$. Boundary

conditions are set as

$$P(\pi, t) = P(-\pi, t). \quad (17.8)$$

The probability density is normalized such that

$$\int_{-\pi \text{ rad}}^{\pi \text{ rad}} P(\varphi_n, t) d\varphi_n = 1. \quad (17.9)$$

Under the steady-state condition,

$$\lim_{t \rightarrow \infty} \frac{\partial P(\varphi_n, t)}{\partial t} = 0, \quad (17.10)$$

the solution is found to be

$$P(\varphi_n) = \frac{\exp(\rho_{L0} \cos \varphi_n)}{2\pi I_0(\rho_{L0}) \text{ rad}}, \quad -\pi \text{ rad} < \varphi_n < \pi \text{ rad}, \quad (17.11)$$

where I_0 is the modified Bessel function of the first kind, order 0, and ρ_{L0} is the S/N in the loop bandwidth, Eq. (14.12), under a linear assumption. This distribution is not Gaussian but does have a zero mean. The variance from Eq. (17.11) is

$$\sigma_{\varphi_n}^2 = \left[\frac{\pi^2}{3} + 4 \sum_{k=1}^{\infty} \frac{(-1)^k}{k^2} \frac{I_k(\rho_{L0})}{I_0(\rho_{L0})} \right] \text{ rad}^2. \quad (17.12)$$

In Fig. 17.4, $P(\varphi_n)$ is shown as a function of ρ_{L0} .

For high S/N , the probability density of Eq. (17.11) becomes

$$P(\varphi_n) \approx \sqrt{\frac{\rho_{L0}}{2\pi}} \frac{e^{-\varphi_n^2 \rho_{L0}/2}}{\text{rad}}, \quad (17.13)$$

which is Gaussian with a variance $\sigma_{\varphi_n}^2 = 1/\rho_{L0}$, as is the case with a linear analysis. For very low S/N , Eq. (17.11) approaches a uniform distribution

$$P(\varphi_n) \approx 1/(2\pi \text{ rad}), \quad (17.14)$$

indicating that the phase is completely unknown. We can observe these trends in Fig. 17.4.

17.1.1.3 Second-Order Loop. The solution for a second-order loop is more difficult because $d\varphi_n(t)/dt$ is not dependent only upon $\varphi_n(t)$, since the second-order

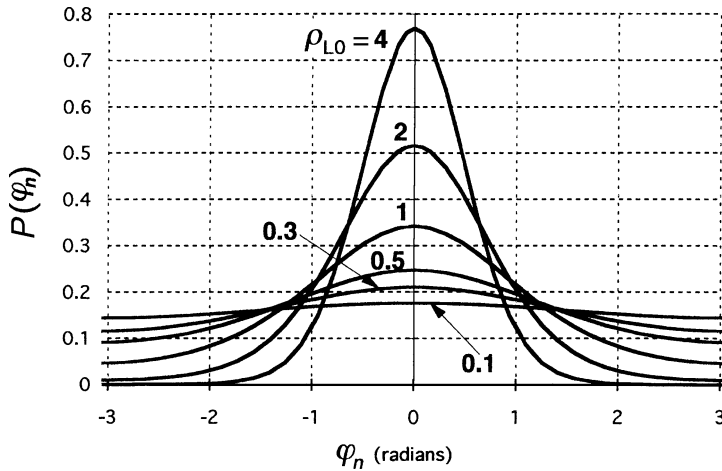


Fig. 17.4 Probability density of phase error versus phase error, first-order loop, or second-order loop with low-pass filter.

loop has memory,¹ and thus $\varphi_n(t)$ is not a Markov process. The solution involves the formation of a two-dimensional vector related to $\varphi_n(t)$. The vector describes the state of the loop but it is Markov because its derivative depends only on its own value. The second-order Fokker–Plank equation is then used to describe the time dependence of the joint probability function of this two-dimensional variable. The solution has been obtained for $\omega_n \gg \omega_z$ [Viterbi, 1963; Viterbi, 1966, Chapter 4], which is similar to a first-order loop, and for a loop with a low-pass filter [Blanchard, 1976, pp. 302–314]. These solutions are again given by Eq. (17.11). In the case with the low-pass filter, the distribution for the frequency error is also obtained. It is

$$P(\omega_n) = \frac{1}{\sqrt{2\pi}\sigma_{\omega n}} \exp\left(-\frac{\omega_n^2}{2\sigma_{\omega n}^2}\right) \tag{17.15}$$

This is Gaussian with a null mean and with variance

$$\sigma_{\omega n}^2 = \frac{K^2\omega_p N_0}{4P} = \omega_n^2\sigma_{\varphi_0}^2. \tag{17.16}$$

Example 17.1 Phase Error with Large Noise, Exact Solution A second-order loop with a low-pass filter has a natural frequency of 1 kHz and a damping factor of 0.25 (in the absence of noise). With an input signal strength of 0 dBm and one-sided noise density of -38 dBm/Hz:

- (a) What is the probability that the phase error will be within $\pm 35^\circ$ of the nominal (90°) value?

¹ Refer to text after Eq. (8.3).

From Eq. (14.10), the (linear) noise bandwidth is

$$B_n = \frac{2\pi \times 10^3 \text{ rad/sec}}{4} \left[\frac{1}{2(0.25)} + 0 \right] \frac{\text{cycle}}{\text{rad}} = \pi \times 10^3 \text{ Hz.}$$

From Eq. (14.11), the noise in the loop's (linear) noise bandwidth is

$$P_n = \pi \times 10^3 \text{ Hz} \times 10^{-3.8} \text{ mW/Hz} = 0.5 \text{ mW.}$$

Since the signal strength is 1 mW, the signal-to-noise ratio in the (linear) loop bandwidth is $\rho_{L0} = 2$. To get the exact answer, we would integrate Eq. (17.11) from -0.61 to $+0.61$ rad (35° is 0.61 rad) but we can obtain an approximate answer using Fig. 17.4. The average probability density over ± 0.61 rad is approximately $0.47/\text{rad}$. (This can be obtained by inspection or by properly averaging a number of values over the range.) Multiplying this density by the full width of the range, 1.22 rad, we obtain a probability of 0.57 that the phase is within $\pm 35^\circ$. [By "exact" we mean that we are applying Eq. (17.11), which is the mathematical solution that applies to the type of loop we are analyzing. Nevertheless, we have done an approximate integration.]

(b) What is the root mean square (rms) frequency deviation of the loop's VCO?

Using Eq. (17.16) with $\rho_{L0} = 1/\sigma_{\varphi 0}^2$ from (a) above, we obtain

$$\sigma_{\omega n}^2 = (2\pi \times 10^3 \text{ rad/sec})^2 \left(\frac{1}{2} \right),$$

$$\sigma_{\omega n} = 4443 \text{ rad/sec.}$$

Viterbi (1966, p. 118) has suggested that the performance of other second-order loops, in the presence of large amounts of noise, be estimated by using equations for a first-order loop [e.g., Eq. (17.11)] with $\sigma_{\varphi 0}$ obtained using B_{n0} for the second-order loop. This procedure produces the correct results for the two cases for which exact solutions have been obtained. We will be able to judge its accuracy in other cases by comparing it to the results of experiments and simulations.

17.1.2 Experimental Results

Charles and Lindsey (1966) have provided experimental results² for a high-gain second-order loop ($K \gg \omega_z$) with $\zeta = 0.707$ and six values of ρ_{L0} from 0.41 dB to 6.41 dB and have compared these results to three different theoretical models. Figure 17.5 shows their data for $\rho_{L0} = 1.41$ dB along with the first-order-loop results

²The data are also in Lindsey and Simon (1973, pp. 33–35) and Viterbi (1966, pp. 112–115) (4 of 6 data sets)

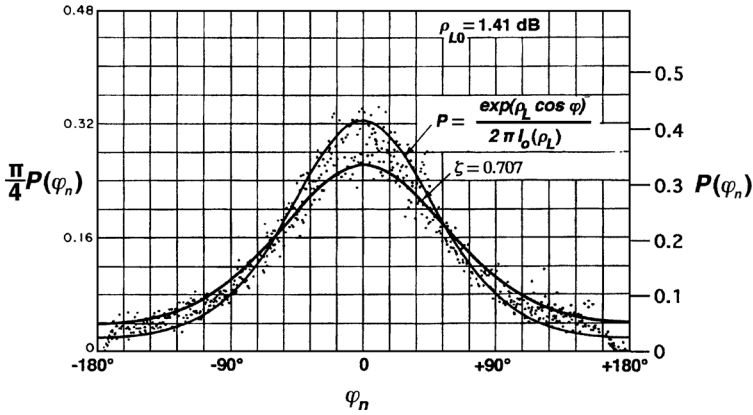


Fig. 17.5 Theoretical and experimental density (per radian) of phase errors. Data points, a first-order loop curve P , and Lindsey's approximation for a high-gain second-order loop are shown. [From Lindsey and Simon, 1973, p. 33. See also Lindsey and Simon, 1978, p. 285; Charles and Lindsey, 1966, p. 1161; Viterbi, 1966, p. 115.]

of Eq. (17.11). Lindsey has also given an approximate solution³ that tends to form a lower bound in the center of Fig. 17.5 and an upper bound further out (lighter curve).

Figure 17.6 shows the experimentally determined probability distribution function of phase error (integral of the density) for several values of ρ_{L0} from the same source. Note how the distribution approaches Gaussian at high S/N , corresponding to the integral of Eq. (17.13), and uniform at low S/N , corresponding to the integral of Eq. (17.14).

17.1.3 Simulation

The simulation method is described in Appendix i17.M.

17.1.3.1 First-Order and $\alpha = 0$. Figure 17.7 shows the true variance of the output phase of a first-order loop according to Viterbi's results ["theory (FP)"]. Points from a simulation are also plotted and correspond well to the theory. We see that the variance approaches an upper limit of $\pi^2/3$, which represents random phase, as a consequence of the phase being defined mod 2π . This is the integrated density under the flat curve in Fig. 17.4 as $p \Rightarrow 0$, and it represents an upper limit of the variance for all loops,

$$[\sigma_\varphi^2]_{\max} = \int_{-\pi}^{\pi} P(\varphi)\varphi^2 d\varphi = \int_{-\pi}^{\pi} \frac{1}{2\pi} \varphi^2 d\varphi = \frac{1}{2\pi} \left[\frac{\varphi^3}{3} \right]_{-\pi}^{\pi} = \frac{\pi^2}{3}. \quad (17.17)$$

³ Lindsey and Simon (1973, pp. 31–36). They reference Lindsey (1972).

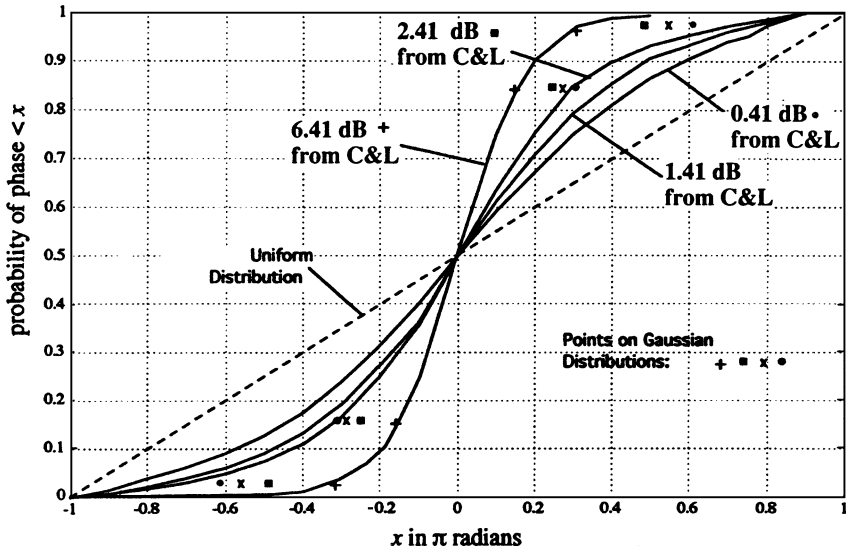


Fig. 17.6 Probability distribution of phase with parameter $\rho_{L0}, \zeta = 0.707, \alpha \approx 1$. [From Charles and Lindsey (C&L), 1966, p. 1161 (©1966 IEEE)].

Figure 17.8 is an expansion of Fig. 17.7 for smaller variances. Results from the simulation of a second-order loop with a low-pass filter are also plotted and also fall on the curve, in accordance with Blanchard’s findings.

17.1.3.2 Second-Order, High-Gain. Figure 17.9 shows one of the cumulative probability curves from Fig. 17.6. Also shown are simulated results with the same parameters ($\alpha \approx 1$) and simulated results with $\alpha = 1$.

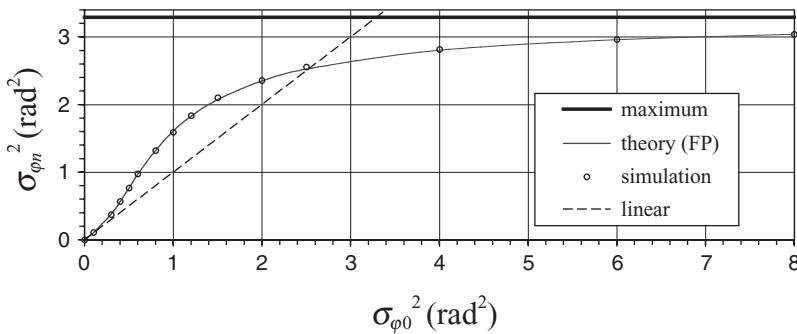


Fig. 17.7 True output variance versus output variance in B_{n0} for first-order loops. Output variance is limited to a value of $\pi^2/3$, representing random phase. [Script NStat1, Appendix i17.M.2.]

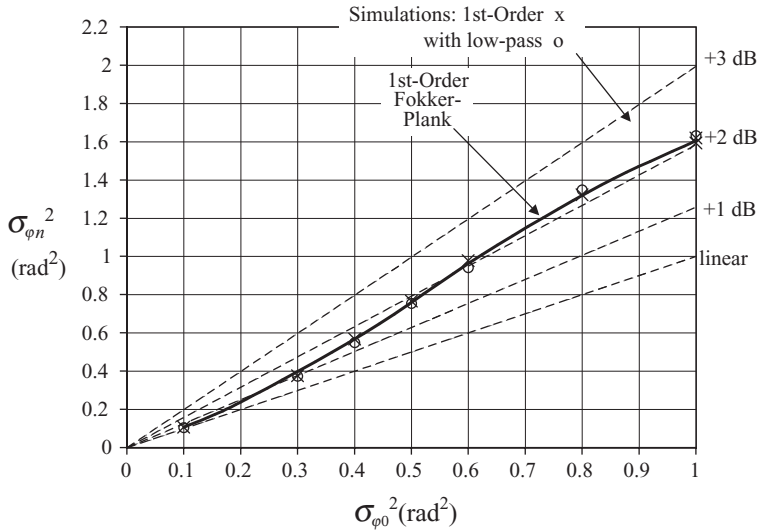


Fig. 17.8 Expansion of Fig. 17.7 at low variances plus simulated results for second-order loop with $\alpha = 0, \zeta = 0.707$. [Script NStat2, Appendix i17.M.2.]

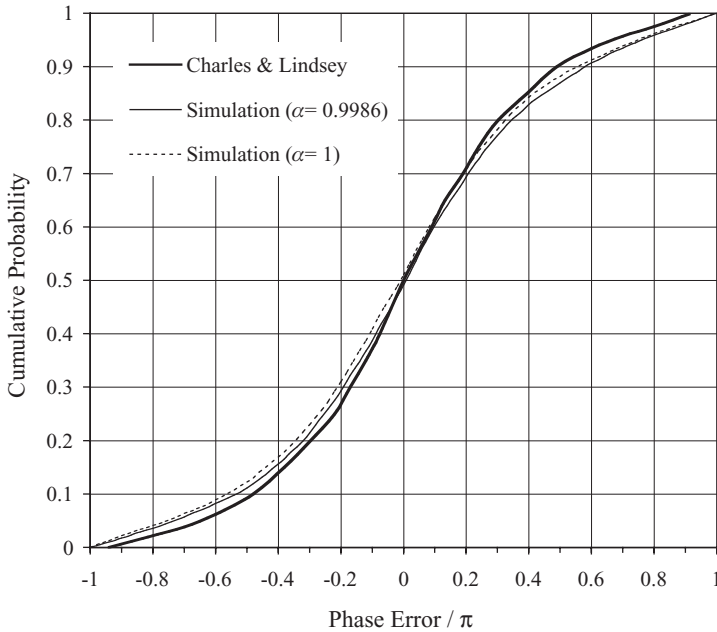


Fig. 17.9 Cumulative probability distributions of phase for $\alpha = 0.9986, \zeta = 0.707, \rho_{L0} = 1.41$ dB from Charles and Lindsey (1966) plus simulation results. (One of the simulation curves does not quite go through the center; it would probably come closer with a longer simulation.) [Software is described in Appendix i17.M.1.]

The simulated results are close to the measured results, the most significant difference being that the measured results reach the extremes before $\pm\pi$ radians. This corresponds to the lack of data points near $\pm\pi$ in Fig. 17.5. Since there does not appear to be a reason why this region should be lacking in measured points—this is, after all, a point of unstable equilibrium—the simulation may be more accurate. { The experiment measured phase using a flip-flop phase detector (Fig. 3.1). The sudden change in output, as $\pm\pi$ radians is crossed (at 0 or 1 in Fig. 3.1c), would be slowed by the low-pass filter that would be required to attenuate the high content of signal at the reference frequency [see $v(C)$ in Fig. 3.1b], thus spreading these points over the rest of the phase range. }

Figures 17.10 and 17.11 show output phase variance versus the variance that would be produced if the loop remained linear (i.e., N_0B_{n0}/P_c). These results (from simulations) are given for various combinations of α and ζ . The curve for $\alpha = 0.9986$ in Fig. 17.10 is about 0.6 dB more pessimistic than the results from the experiment over much of the range (see Appendix i17.D.1). This is in a direction that would correspond to the difference in results indicated by Fig. 17.9.

The straight dashed lines that radiate from (0,0) are loci of constant $\sigma_{\varphi n}/\sigma_{\varphi 0}$, with the ratio given in dB. For example, in Fig. 17.10, for $\alpha = 0.96$ at $x = \sigma_{\varphi 0}^2 = 0.9$, $y = \sigma_{\varphi n}^2 = 1.8$ and the ratio is 3 dB.

While only a finite number of combinations of parameters are available from these graphs, any combination can be simulated using the same software.

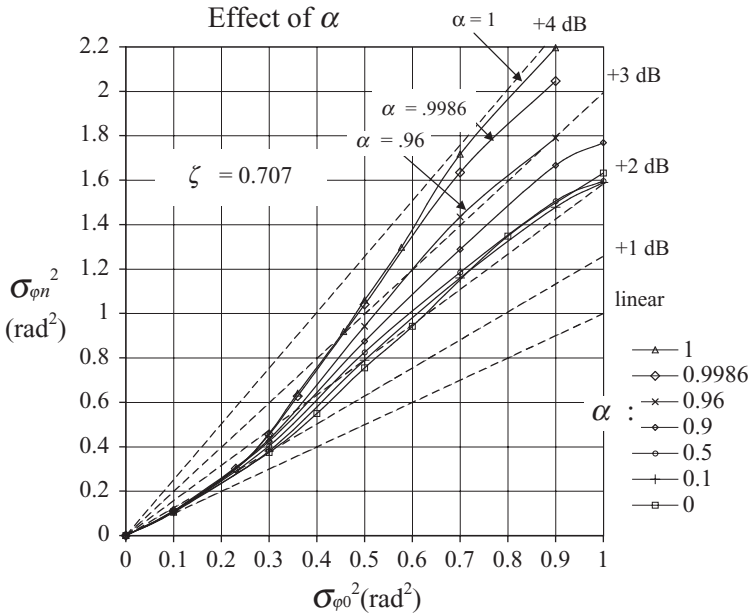


Fig. 17.10 True variance versus variance in B_{n0} for $\zeta = 0.707$ with various values of α simulated results. [Script NStat2, Appendix i17.M.2.]

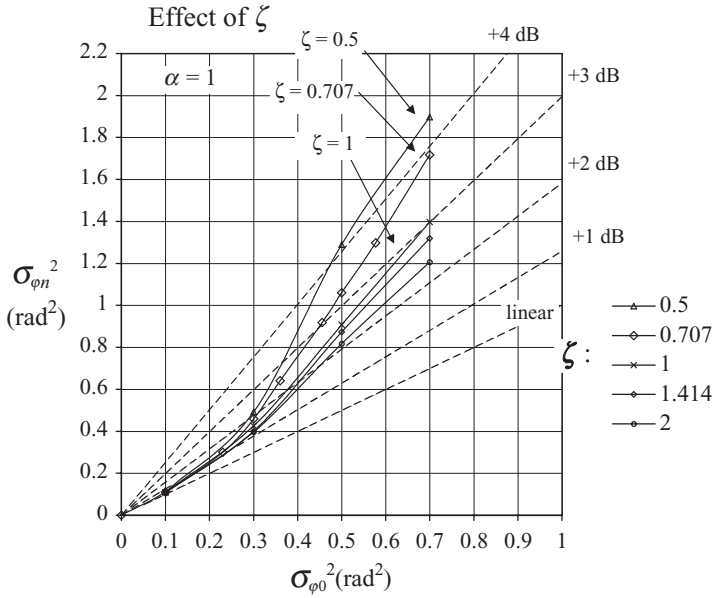


Fig. 17.11 True variance versus variance in B_{n0} for $\alpha = 1$ with various values of ζ , simulated results. Note that the lowest curve is approaching the first-order results in Fig. 17.8. [Script NStat2, Appendix i17.M.2.].

Example 17.2 Phase Error in Second-Order Loop, Large Noise Estimate the probability of phase error within $\pm 35^\circ$ ($= \pm 0.61$ rad) if ρ_{L0} is 1.384 (1.41 dB), comparing the results for a first-order loop and a second-order loop. Evaluate Viterbi’s suggestion of using results from the first-order loop at the same σ_{ϕ_0} ($\sigma_{\phi_0}^2 = 1/1.384 = 0.723 \text{ rad}^2$).

If we repeat the process of Example 17.1 for this value of ρ_{L0} , we obtain a probability of 0.31 for a first-order loop or for a second-order loop with $\alpha = 0$. For $\zeta = 0.707$, we can obtain the probability of ϕ_n from the $\alpha = 1$ curve in Fig. 17.9, subtracting the cumulative probability at $(-0.61 \text{ rad} / \pi \text{ rad}) = -0.194$ from the value at $+0.194$ to obtain a probability of 0.395 for phases in the $\pm 35^\circ$ range, about 27% higher than the value for a first-order loop.

If the shapes of the distributions were the same, we would expect the probabilities to be proportional to the values of σ . In Fig. 17.10, the curve for $\alpha = 1, \zeta = 0.707$ indicates a value of $\sigma_{\phi_n}^2 = 1.8 \text{ rad}^2$ at $\sigma_{\phi_0}^2 = 0.723 \text{ rad}^2$ while Fig. 17.8 shows $\sigma_{\phi_n}^2 = 1.2 \text{ rad}^2$ for a first-order loop. The value of σ is therefore about 22% higher with the second-order loop, not too different from the 27% difference in probabilities.

Figure 17.11 shows a 27% increase in σ relative to the first-order loop for the highest point, for which $\zeta = 0.5$, while the increase is smaller at higher ζ . The first-order loop does appear to provide an optimistic, but reasonable, approximation for the second-order loop, in accordance with Viterbi’s suggestion, at least for $\sigma_{\phi_0}^2 \leq 0.7 \text{ rad}^2$ and $\zeta \geq 0.5$.

17.2 CYCLE SKIPPING, MEAN TIME

Here we are concerned with the average time for the loop to skip a cycle—that is, to reach a phase error of one cycle—after starting at zero phase and zero frequency error.

17.2.1 First-Order Loop

Viterbi solved this problem for a first-order loop. We begin again with Eqs. (17.4) and (17.5), but this time the condition (17.9) is not applied. This time, $p(\varphi_n, t)$ represents the probability density of the phase of something that is “eliminated” if it reaches the phase limits of $\pm\varphi_L$. The value of such a density function is that it can be integrated over $\pm\varphi_L$ to determine the probability that the phase has never reached those limits. It is not a true probability density because its integral over all values is less than one. We (and Viterbi) designate it $q(\varphi_n, t)$ to differentiate it from the function discussed in Section 17.1.1.2. The Fokker–Plank equation still describes the density and its dynamics while the phase stays within the $\pm\varphi_L$ limits. Since the phase q cannot exist at the limits (being eliminated upon reaching there), the boundary conditions are $q(\pm\varphi_L, t) = 0$ for all time. Viterbi indirectly solves for $q(\varphi_n, t)$, using the function to determine T_m , the mean time between cycle skips (with $\varphi_L = 2\pi$). He also shows that the frequency of reaching $\pm\pi$ ($\varphi_L = \pi$) is twice the frequency of skipping cycles. This is consistent with the equal likelihood, in the first-order loop, that the phase, having reached $\pm\pi$, will return to its original region or skip a cycle.

For a first-order loop with zero mean phase and frequency error, he found that the mean time between cycle skips is

$$T_{m1} = \frac{\pi^2 \rho_{L0} I_0^2(\rho_{L0})}{2B_{n0}} \text{cycle.} \quad (17.18)$$

Figure 17.12 shows $\text{cycle}/(T_{m1} B_{n0})$ versus ρ_{L0} from this equation plus points from a simulation.

For high ρ_{L0} , Eq. (16.16) for $v = 0$ approaches⁴

$$I_0(\rho_{L0}) \approx \frac{\exp(\rho_{L0})}{\sqrt{2\pi\rho_{L0}}} \left(1 + \frac{1}{8\rho_{L0}} \right). \quad (17.19)$$

So Eq. (17.18) then becomes

$$T_{m1} \approx \frac{\pi}{4} \frac{\exp(2\rho_{L0})}{B_{n0}} \text{cycle.} \quad (17.20)$$

This approximation is shown as the dashed line in Fig. 17.12.

⁴The second term is shown here, even though it will be dropped, to make it easier to estimate the validity of the approximation for a given value of ρ .

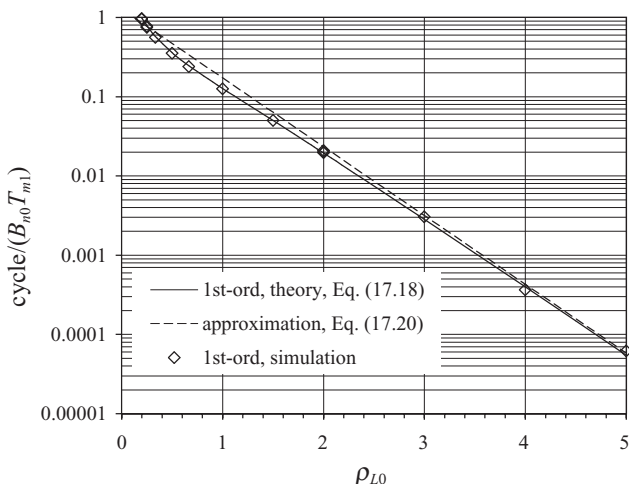


Fig. 17.12 Mean time to first skip, first-order loop, normalized to linear noise bandwidth. The simulation deviates from the curves more at low skip rates where fewer data were taken. [Script Skip1, Appendix i17.M.3].

17.2.2 Second-Order Loop, $\alpha = 0$

For a second-order loop with a low-pass filter, the first-order curve again gives a good estimate of the mean time to first skip for highly damped loops, but mean times are longer (curves are shifted down and left) for lower values of ζ . Gupta, et al. (1968) gave an estimate of T_m for $\alpha = 0$ as

$$T_{m20} \approx T_{m1}/F_2, \tag{17.21}$$

where T_{m1} is the approximate value from Eq. (17.20) and

$$F_2 = 2\zeta[\sqrt{\zeta^2 + 1} - \zeta]. \tag{17.22}$$

F_2 approaches 1 for large ζ . This factor (Fig. 17.13) appears to be better applied to the first-order solution, Eq. (17.18), than to its approximation. Figure 17.D.2 (in Appendix i17.D) compares simulation results to Eq. (17.21), where T_{m1} is the value from Eq. (17.18). The correspondence is best for lower values of ρ_0 , possibly because much longer simulations are required to produce abundant data at high ρ_0 .

17.2.3 Second-Order Loop, $\alpha = 1$

Simulation results for $\alpha = 1$ (Fig. 17.14) again match the first-order solution Eq. (17.18) for high damping factors, but T_m is shorter (in contrast to the case where $\alpha = 0$) for lower damping factors and high values of ρ_{L0} .

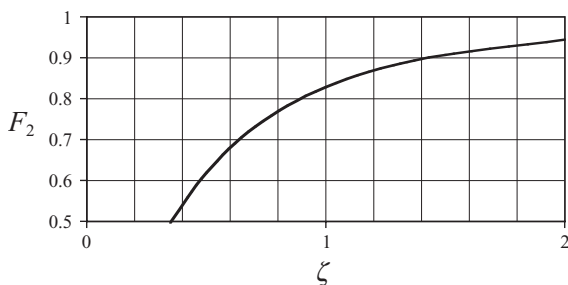


Fig. 17.13 Divisor F_2 to apply to first-order solution, Eq. (17.18), for mean time to first skip to obtain an estimate of mean time for a second-order loop with a low-pass filter ($\alpha = 0$).

The results correspond better to Eq. (17.18) if we use a modified value of ρ_{L0} ,

$$\rho'_{L0} = \rho_{L0} \min(1, \sqrt{B_{n0}/\omega_{n0}}), \tag{17.23}$$

when ρ_{L0} is greater than 1 (or perhaps some number between 1 and 2), as is shown in Fig. 17.15. In other words, the value of $1/(T_m B_{n0})$ for a high-gain loop is found by using ρ'_{L0} in place of ρ_{L0} in Eq. (17.18) for such values (ρ'_{L0} is lower than ρ_{L0} for $0.13 < \zeta < 1.87$).

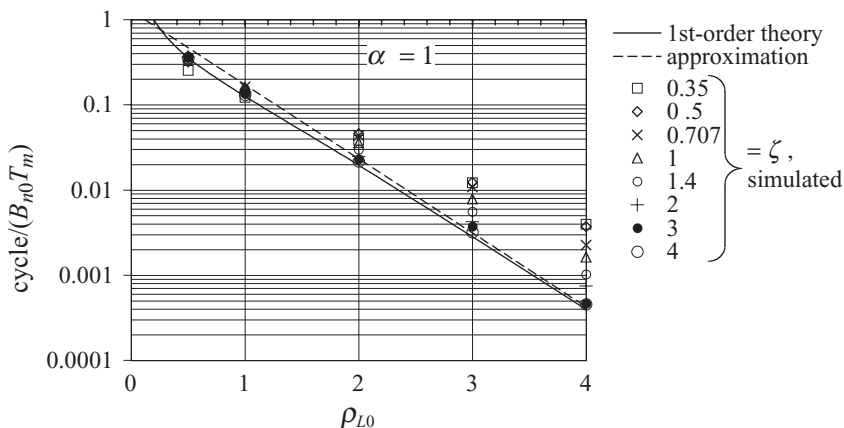


Fig. 17.14 Mean time to first skip (normalized) for a second-order loop with $\alpha = 1$, versus signal-to-noise in linear loop. Points from simulation and theoretical curves for first-order loop are shown. (See Fig. i17.D.3 for an expansion of the region near $\rho_{L0} = 0.5$.) [Script Skip2i, Appendix i17.M.3].

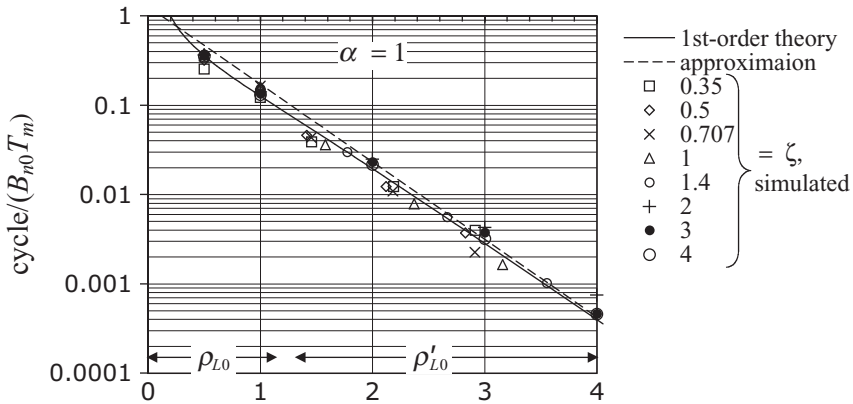


Fig. 17.15 Mean time plotted against modified S/N .

17.2.4 Second-Order Loop, $0 < \alpha < 1$

We might assume that T_m at intermediate value of α would be between the value at $\alpha = 0$ and that at $\alpha = 1$ with some degree of proportionality to α . Simulations (Fig. i17.D.4) indicate that this is true, at least for $\zeta = 0.707$ and $\rho_{L0} = 2$.

17.2.5 Probability of Cycle Skipping in a Given Time

The probability that a cycle will be skipped within a time T (timed either from the previous burst of skipped cycles or from the occurrence of initial zero phase and frequency error) has been given as [Rowbotham and Sanneman, 1967; Smith, 1966]

$$P(T) = 1 - e^{-T/T_m} \tag{17.24}$$

17.3 CYCLE SKIPPING, MEAN FREQUENCY

The mean frequency of cycle skipping is

$$\bar{F} \geq 1/T_m, \tag{17.25}$$

with equality in the absence of multiple skips, or “bursts.” If the phase and frequency should return to zero error, the initial conditions for T_m , after each cycle skip, equality would hold. This is certainly true for a first-order loop since it has only one independent state variable, so once it has change phase by 1 cycle, it is at the initial condition. However, second-order loops have memory; their state depends on their recent history. It is determined by both frequency and phase. When a second-order loop has skipped a cycle, it may have a frequency error that will increase its likelihood of skipping another cycle. Figure 17.1*b*, for example, shows a case where a second skip happens

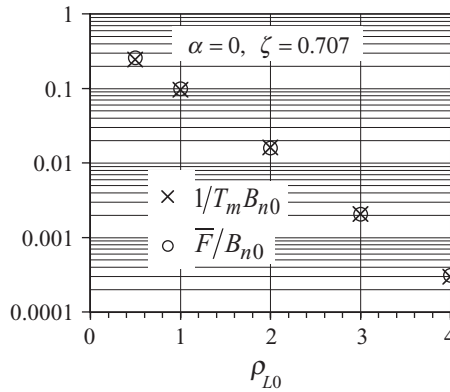


Fig. 17.16 Normalized values of \bar{F} and $1/T_m$ for second-order loop with low-pass filter, $\zeta = 0.707$, from simulation. Equal values indicate no multiple skips. [Scripts NStat2, Appendix i17.M.2, and Skip2, Appendix i17.M.3.]

immediately. Therefore, in addition to knowing T_m , it may be important to also know \bar{F} for a particular loop.

Figure 17.16 shows that \bar{F} and T_m are equal for a second order loop with $\alpha = 0$ and $\zeta = 0.707$, implying that there are no multiple skips for this case.

Figure 17.17 is a plot of experimental and simulated values of \bar{F} for two high-gain second-order loops and shows a high degree of agreement between experiments and simulations. (Additional data from the literature are shown in Appendix i17.D.3.) Values of T_m were also obtained from simulations, and the product of \bar{F} and T_m in Fig. 17.18 indicates that there are multiple skips at higher noise levels.

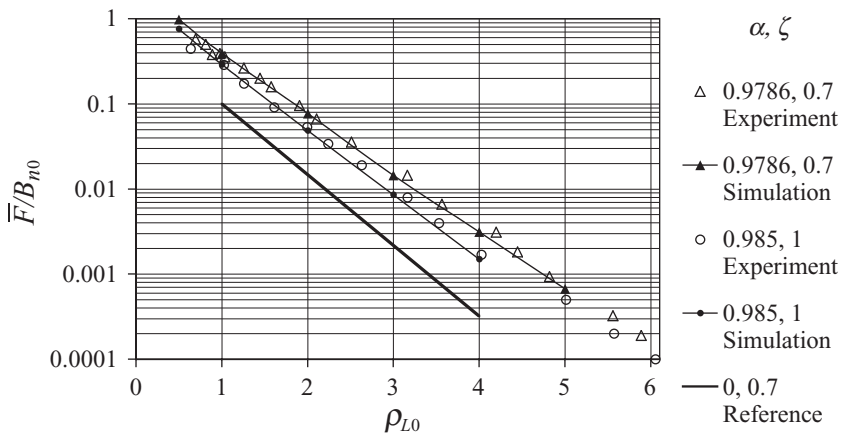


Fig. 17.17 Experimental \bar{F} for two high gain loops [Ascheid and Meyr (1982), Fig. 22] plus simulated values for the same loops and line corresponding to data in Fig. 17.16. [Script SigOff2i, Appendix i17.M.4.]

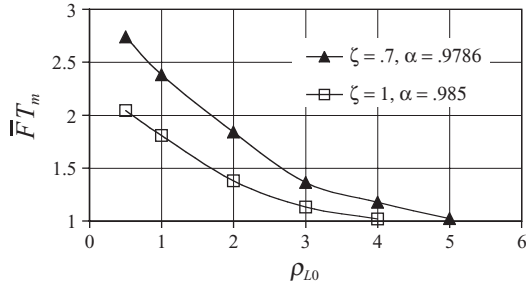


Fig. 17.18 $\bar{F}T_m$, simulation results, for loops in Fig. 17.17. Multiple skips are indicated at higher noise levels. [Scripts SigOff2i in Appendix i17.M.4 and Skip2i in Appendix i17.M3.]

Figure 17.19 indicates the degree of multiple skips as a function of α and ζ that was observed in a series of simulations with $\rho_{L0} = 2$. No curve is shown for zero multiple skips, a condition that is difficult to determine accurately and may not even exist, since it seems that there would always be some probability of a second skip.

The heavy dashed line is the locus where $B_{n0} = 0.5\omega_{n0}$ [Eq. (14.10)]. Gupta et al. (1968), in studying low-gain loops, indicate that multiple skips will not occur if the hold-in range equals the pull-in range, a condition they say occurs above this line. That seems to be consistent with the simulated data in the low- α region.

Figure 17.19 indicates that multiple skips are promoted by low ζ and high α , which is consistent with the experimental results in Fig. 17.18. Smith (see Appendix i17.D.3)

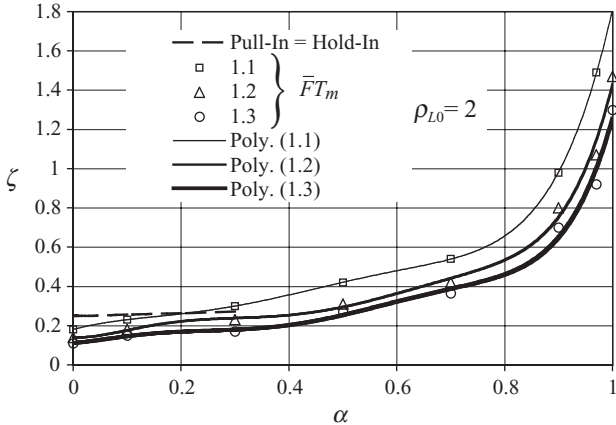


Fig. 17.19 Combinations of α and ζ that produce 10%, 20%, and 30% skips in excess of $1/T_m$, from simulations with $\rho_{L0} = 2$. Points are interpolated from curves of $\bar{F}T_m(\zeta)$ at fixed α (“Poly.” refers to sixth-order polynomial curve fit.) [Scripts Skip2i, Appendix i17.M.3 and NStat2, Appendix i17.M.2.]

also noted a reduction in multiple skipping as α was reduced from one, corresponding to ω_p increasing from zero, but Blanchard [1976, p. 320] indicates that a low value of ω_p is beneficial in preventing multiple skips. Which of these apparently contradictory observations is true depends on the loop mistuning.

17.4 CYCLE SKIPPING WITH MISTUNING

When the loop skips a cycle, its filter capacitor (Fig. 3.23g) holds a voltage that represents both the frequency error, which must exist for a cycle to be skipped, and the required frequency offset, if any. The capacitor discharges that voltage through the shunt resistor R_{2p} . The discharge rate is determined by the time constant, $\tau_1 = 1/\omega_p = R_{2p}C$ [Eq. (3.27)]. Without the locked loop to control it, the voltage on C will decay toward its value at the VCO center frequency. Most of the results given in the literature (and in this chapter) are for loops with zero mistuning. In that case, it is *desirable* for the loop to relax back to its center frequency, where it will rapidly pull into lock again. *However*, if there is a frequency offset, it may be desirable to maintain this voltage in order to maintain the required mistuning.

17.4.1 Effect of α

Here we will investigate this dichotomy and learn more about some limitations as ω_p approaches the idealized, but unattainable, value of zero. We will see that there can be an advantage to making K larger, in order to decrease the phase offset that corresponds to a required tuning range, but not so large that that offset becomes dominated by the uncompensated offsets in real loop components (i.e., deviations from design values).

It will be convenient to characterize the loop by α rather than by ω_p . With a fixed ω_n , as ω_p decreases, K increases [Eq. (4.7)] and, if ζ is also fixed, α then increases [Eq. (6.5a)], so the decrease in ω_p , which slows the decay of the capacitor voltage, corresponds to an increase in α .

Figure 17.20 shows that, at any given phase offset (i.e., the mean phase during lock), T_m is little affected by small deviations of α from unity. However, the occurrence of multiple skips increases with α and, therefore, so does the mean frequency of cycle skipping \bar{F} . This can be seen in Fig. 17.21, where the curves become steeper with increasing α . As α approaches 1, \bar{F} approaches infinity for any nonzero offset.

We typically begin simulations with zero phase and frequency errors, so the loop starts in lock. If noise causes lock to break, the loop tends to relax toward its center frequency—that is, toward the frequency mistuning that corresponds to the phase offset. As long as the mistuning is within the pull-in range, the loop will relock. As α approaches 1, the mistuning increases faster than the pull-in range. [That is why there is an optimum, nonzero value for ω_p , in accordance with Eqs. (8.76) and (8.24).] Experimentally, α cannot be one, but when α is set to one in a simulation where there is an offset, there may be a long period before the loop permanently loses lock, but eventually it will. At that time, the frequency error will begin to increase continuously. Therefore, the true skip rate, over all time, for $\alpha = 1$, is infinite. Simulations with any

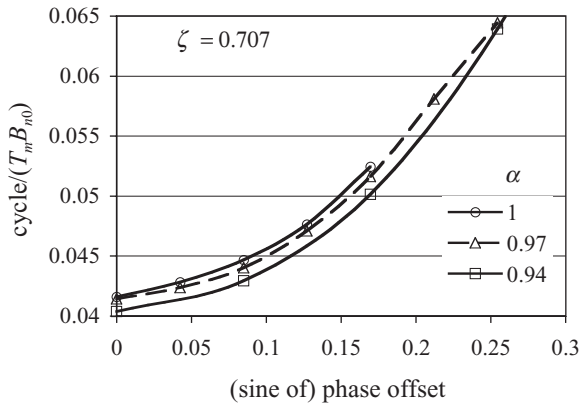


Fig. 17.20 The effect of phase offset on T_m for various high values of α , $\rho_{L0} = 2$. Note that the ordinate is not shown below 0.04 so the total change illustrated is less than 2:1. While an offset can cause a gain reduction that modifies loop parameters (e.g., ζ and α), the parameters given here are nominal design values at zero offset and are given for the linear loop. This is generally true throughout the chapter. [Script Skip2i, Appendix i17.M.3.]

offset and $\alpha = 1$ may appear to settle to a stable skip rate, but that can only last for a limited time. This is illustrated in Fig. 17.22.

Here the skip rate appears to have settled (value not visible at this scale) until more than 150,000 sec (nearly 2 days) have gone by. This is for a loop with a noise bandwidth of 0.53 Hz, but the equivalent time would be only 150 msec for a loop with 0.5-MHz noise bandwidth.

From all this we might conclude that α should be small to minimize \bar{F} , but there are reasons not to make α too small. In Section 8.7 we found that α should have a certain nonzero value for maximum pull-in range when there is an uncompensated

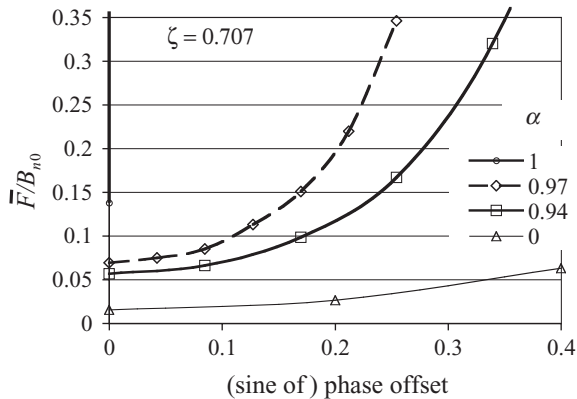


Fig. 17.21 Mean frequency of skipping cycles versus phase offset, $\rho_{L0} = 2$. [Script SigOff2i, Appendix i17.M.4.]

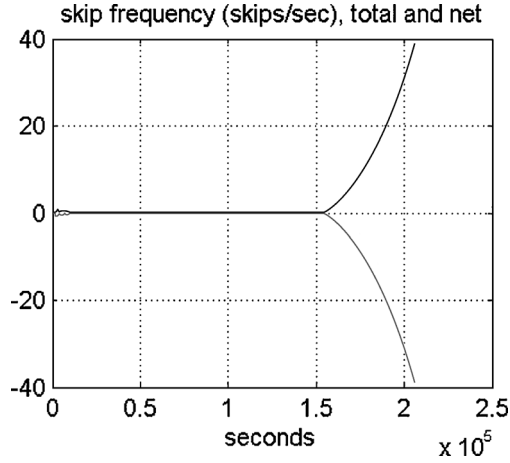


Fig. 17.22 Mean skip frequency with 2° offset, $\rho_{L0} = 2$, $\alpha = 1$. $\zeta = 0.707$, $\omega_n = 1$ rad/sec. 1520 samples/sec. Upper curve is total skip rate, and lower curve is net skip rate.

offset and a fixed bandwidth. This optimum may change in the presence of noise, but it is still unlikely to be zero. Moreover, if there is a required range of mistuning because the reference frequency varies, the corresponding phase offset will decrease as α , along with K , increases. Therefore, at a given mistuning, while \bar{F} increases with both α and phase offset, the phase offset decreases with α . To see the *net* effect, we plot \bar{F} against mistuning in Fig. 17.23.

The values of mistuning in Fig. 17.23 are related to the phase offsets in Fig. 17.21 by K , so, as K increases at higher α , any range of mistunings in Fig. 17.23 correspond to a smaller range of phase offset in Fig. 17.21. This causes the curves to flatten out more at higher value of α in transitioning from Fig. 17.21 to Fig. 17.23. As α approaches

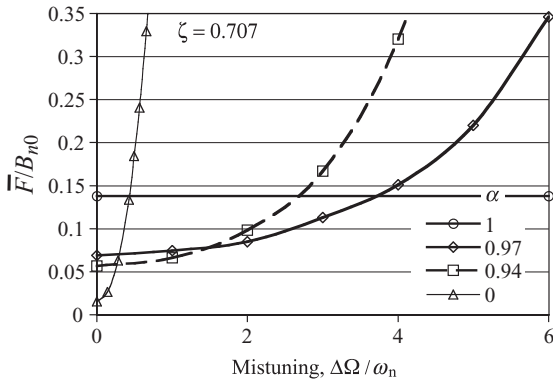


Fig. 17.23 Mean frequency of skipping cycles versus mistuning, $\rho_{L0} = 2$. [Script SigOff2i, Appendix i17.M.4].

1, K approaches infinity and the values of mistuning in Fig. 17.23 all correspond to a phase offset near zero in Fig. 17.21. Therefore, \bar{F} becomes almost constant in Fig. 17.23, and its locus approaches a horizontal line.

As a result, the value of α that produces minimum \bar{F} depends on the amount of mistuning. At small values of mistuning, α should be small to minimize \bar{F} . (Figure 17.17 suggests $\alpha = 0$ when there is no offset.) At greater mistunings, the optimum α is larger.

We began by showing that \bar{F} increased with increasing α at a given offset, but we then saw that increasing α could be beneficial if it resulted in a smaller offset. As we increase α to reduce \bar{F} at a given mistuning, we may reduce the corresponding phase offset to the point where other, uncompensated phase offsets (e.g., due to the phase detector or opamp) become dominant. At that point, increasing α further will have a detrimental effect on \bar{F} .

Note that Figs. 17.21 and 17.23 are two different presentations of the same data. We could also present the values of $1/(T_m B_{n0})$ from Fig. 17.20 in the manner of Fig. 17.23. We would see roughly similar curves that would lead to similar conclusions about $1/T_m$.

17.4.2 Comparison to Other Results

We have seen results from experiments by Ascheid and Meyr (1982) in Fig. 17.17, but they also obtained experimental data for \bar{F} with mistuning (phase offset). Their results are more optimistic than our simulated results, but the match is better if we reduce the mistuning by about 14% in our simulations. In contrast, Holmes (1971) obtained values for T_m from a simulation that appear to be pessimistic compared to ours (about 14% at the point checked). More details on the comparisons are available in Section 17.D.4.

We, and these other sources, define a cycle as being skipped when the error reaches ± 1 cycle. It might seem more logical, in the present of a phase offset, to say that a cycle is skipped when the error deviates by ± 1 cycle from the offset. It should be simple to accommodate that change in the simulation script.

17.5 SUMMARY

17.5.1 Phase Variance

- Viterbi used the Fokker–Plank equation to find the probability density of phase, mod 2π , for a first-order loop with broadband additive noise. The probability density is given by Eq. (17.11) and the variance by Eq. (17.12).
- The solution also applies to a second-order loop with $\alpha = 0$.
- Variances are shown graphically in Figs. 17.7 and 17.8 versus variance in a linear loop.
- The variances increase with α and decrease with ζ for other second-order loops [Figs. 17.10 and 17.11].

17.5.2 Cycle Skipping

- Viterbi again used the Fokker–Plank equation to obtain mean time to first cycle skip T_{m1} for a first-order loop [Eq. (17.18), Fig. 17.12].
- For $\alpha = 0$, T_{m2} is longer than T_{m1} [Eq. (17.21)], especially at low damping factors.
- For $\alpha = 1$, T_{m2} is shorter than T_{m1} [Fig. 17.14], especially at higher values of ρ .

17.5.3 Skip Frequency

- In a first-order loop, there are no multiple skips, so the skip frequency is $\bar{F} = 1/T_m$.
- For $\alpha = 0$, with $\zeta = 0.7$, $\bar{F} = 1/T_m$ again.
- For $\alpha \approx 1$, $\bar{F} > 1/T_m$ [Figs. 17.18, 17.19], approaching infinity as α approaches 1.

17.5.4 Mistuning in High-Gain Loops ($\zeta = 0.7$)

- Phase offset decreases T_m [Fig. 17.20] and increases \bar{F} [Fig. 17.21].
- At higher values of α , frequency mistuning produces less phase offset, leading to a slower increase in \bar{F} with mistuning until there is no change at all with mistuning when $\alpha \Rightarrow 1$ [Fig. 17.23].
- When frequency mistuning is small, \bar{F} is highest for α near 1, but this reverses at greater mistunings [Figs. 17.23].
- \bar{F} increases with phase offset [Figs. 17.21] until, at (the theoretical) $\alpha = 1$, \bar{F} becomes infinite; the loop eventually unlocks and stays unlocked there.
- Higher DC gain (and associated α) can decrease bursting by decreasing the offset caused by mistuning, but too much gain can accentuate the effects of other, fixed phase offsets, thereby increase bursting.

i17.D APPENDIX: ADDITIONAL DATA

This appendix is available from the Wiley Internet site.

i17.M APPENDIX: MATLAB[®] SCRIPTS

This appendix describes MATLAB scripts that are used to generate simulations for this chapter as well as the theory on which they are based. It is available from the Wiley Internet site.

CHAPTER 18

NONLINEAR OPERATION IN A LOCKED LOOP

We have seen that the operation of the loop in the presence of large amounts of noise can be difficult to analyze, forcing us to experiment and simulate. However, we have also seen (Chapter 16) analysis methods that modify parameter values as a function of the amount of noise. Here we will extend the process of parameter modification in an attempt to treat nonlinear effects, caused by broadband additive noise and by modulation, analytically, but we will be able to compare our results to those from experiment and simulation to determine their limits of applicability.

18.1 NOTATION

To differentiate the various phase modulations that we consider, we define the input signal v_{in} and output (VCO) signal v_{out} as shown below.

$$\text{Loop in: } v_{\text{in}} = A \cos[\omega_c t + \Phi_e + \phi_i(t)] + \underbrace{v_i(t) \sin \omega_c t + v_r(t) \cos \omega_c t}_{\text{Input additive noise-related}} \quad (18.1)$$

↑

↑

Input phase modulation-related

Input additive noise-related

↓

↓

$$\text{Loop out (VCO): } v_{\text{out}} = -B \sin[\omega_c t + \phi_o(t) + \varphi_n(t)] \quad (18.2)$$

Here Φ_e is a steady error relative to the desired phase quadrature, $\phi_o(t)$ is the output phase in response to an input phase modulation $\phi_i(t)$, and $\varphi_n(t)$ is the output phase response to the additive input noise. Since the noise is broadband, $\varphi_n(t)$ responds only to v_i , which modulates the carrier that is maintained in phase with the VCO (Section 13.5). In cases where the loop is considered linear, we have

$$\varphi_n(t) \Rightarrow \varphi_0(t). \tag{18.3}$$

Even in cases where the noise level is high enough to cause the loop to be non-linear, we designate φ_0 as the linear response. We may compute it, based on linear assumptions, as a step in computing a more accurate value φ_n that takes the nonlinearity into account.

18.2 PHASE-DETECTOR OUTPUT u_1

The terms of interest that are generated in a balanced-mixer phase detector are obtained by retaining the difference-frequency parts of the product of Eq. (18.1) and (18.2), as was done in Chapter 3.

$$u_1(t) = \eta_p B \left\{ \begin{array}{l} A \sin[\Phi_e + \phi_i(t) - \phi_o(t) - \varphi_n(t)] \\ -v_i(t) \cos[\phi_o(t) + \varphi_n(t)] - v_r(t) \sin[\phi_o(t) + \varphi_n(t)] \end{array} \right\} \tag{18.4}$$

$$= K'_p \left\{ \begin{array}{l} \sin[\Phi_e + \phi_i(t) - \phi_o(t) - \varphi_n(t)] \\ -\frac{v_i(t)}{A} \cos[\phi_o(t) + \varphi_n(t)] - \frac{v_r(t)}{A} \sin[\phi_o(t) + \varphi_n(t)] \end{array} \right\}. \tag{18.5}$$

The noise terms represent a voltage $u_{1n}(t)$, as described by Eq. (13.19). Therefore, if the additive noise is sufficiently wideband, $u_1(t)$ (Fig. 13.4) can be represented by

$$u_1(t) = K'_p \sin[\Phi_e + \phi_i(t) - \phi_o(t) - \varphi_n(t)] + u_{1n}(t). \tag{18.6}$$

If the phase deviation is small, this becomes

$$u_1(t) = K'_p \sin[\Phi_e + \phi_i(t) - \phi_o(t) - \varphi_n(t) + n'(t)], \tag{18.7}$$

where

$$n'(t) \equiv u_{1n}(t)/K'_p \tag{13.20}$$

is the input phase that would produce $u_{1n}(t)$.

Note that the linear relationship between $u_1(t)$ and $u_{1n}(t)$ does not depend on small phase deviation, since it exists in Eq. (18.6). This means that, while $n'(t)$ acts like an equivalent phase modulation, as we can see by Eq. (18.7), the voltage produced at the phase detector output is not diminished by the nonlinear nature of the sine function as it is with true phase modulation.

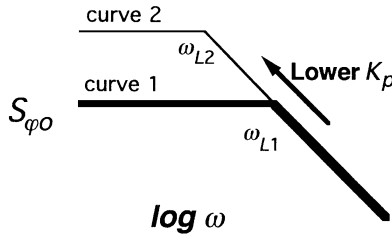


Fig. 18.1 Change in output phase noise with K_p .

Note also that nonlinearity due to modulation will be minimized at low modulation frequencies where the gain is high and ϕ_0 will more closely follow ϕ_i , thus minimizing the argument of the sine. However, since n' is not a true phase and does not appear as part of that argument, ϕ_n will contribute to nonlinearity even at low frequencies.

18.3 CHANGES IN THE OUTPUT SPECTRUM

Consider what happens to the output phase noise $\phi_0(\omega_m)$ in response to input white additive noise as K_p decreases due, for example, to a reduction in signal strength or to a static phase error. Assume a first-order loop for simplicity. The straight-line approximation of the output phase-power spectral density is shown in Fig. 18.1, curve 1. As K decreases, it causes ω_L to decrease. Since the corner between the flat and sloped phase noise regions moves lower, one or the other of these regions must change level. The question is, Which one? Since only K_p , and not the rest of $|G(\omega_m)|$, has decreased, and since $u_{1n}(t)$ is not really affected by K_p , the output noise density at high frequencies will remain unchanged, as illustrated in Fig. 18.1, curve 2. Therefore, the low-frequency output phase noise density increases inversely as K_p . This is also because, within ω_L , the output phase moves to cancel u_1 but more phase deviation is required to produce a voltage to counter u_{1n} when K_p is lower. Another way to look at it is that the equivalent phase noise, $n'(t)$ in Eq. (13.20), increases as K_p decreases.

18.4 GAIN SUPPRESSION, QUASI-LINEAR APPROXIMATION

We will now approximate the nonlinear effects produced by noise as a decrease in K_p . This has been called the Booton quasi-linearization method.¹

Example 18.1 Effective Gain with Noise Determine the effective gain of a phase detector in the presence of uniformly distributed phase noise.

¹ Blanchard (1976, p. 314) credits Develet (1963) with proposing this method for PLL study. Develet, in turn, credits the basis for his method to Booton (1952).

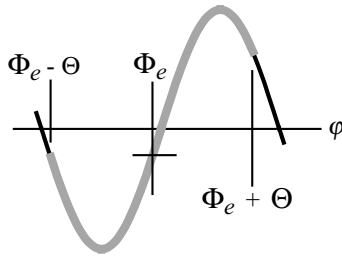


Fig. 18.2 Locus of operating point with noise.

Figure 18.2 represents the operating point on a sinusoidal PD characteristic where the average phase is Φ_e and noise causes equiprobable phase deviations from $-\Theta$ to $+\Theta$. If the noise is sufficiently high in modulation frequency so the loop does not respond to it, the average phase will be the important loop variable. For $\Theta = \pi$, $K \Rightarrow 0$ because the average phase is

$$\frac{1}{2\pi} \int_{\Phi_e - \pi}^{\Phi_e + \pi} \sin \varphi \, d\varphi = \frac{-\cos \varphi}{2\pi} \Big|_{\Phi_e - \pi}^{\Phi_e + \pi} = 0.$$

For any magnitude Θ ,

$$\begin{aligned} \frac{1}{2\Theta} \int_{\Phi_e - \Theta}^{\Phi_e + \Theta} \sin \varphi \, d\varphi &= \frac{-\cos \varphi}{2\Theta} \Big|_{\Phi_e - \Theta}^{\Phi_e + \Theta} = \frac{1}{2\Theta} [\cos(\Phi_e + \Theta) - \cos(\Phi_e - \Theta)] \\ &= \frac{\sin \Theta}{\Theta} \sin \Phi_e. \end{aligned}$$

Therefore the average PD output would be given by the PD characteristic in the absence of noise multiplied by $\sin(\Theta)/\Theta$ (Fig. 18.3), which equals one when $\Theta = 0$ (no noise).

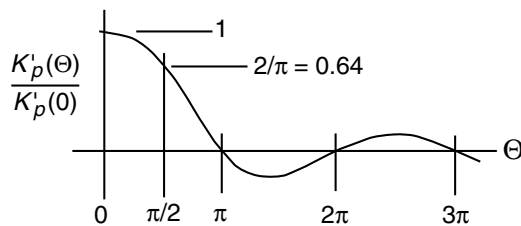


Fig. 18.3 Change in effective PD gain with extent of equiprobable noise excursion.

18.4.1 Basics

We now assume that the frequency of the input modulation $\phi_i(t)$ is too high for the loop to follow so $\phi_o(t) \Rightarrow 0$ in Eq. (18.2) and Eq. (18.6) becomes

$$u_1(t) = K'_p \sin[\Phi_e + \phi_i(t) - \varphi_n(t)] + u_{1n}(t). \quad (18.8)$$

By means of trigonometric identities, this can be expanded as

$$u_1(t) = K'_p \{ \sin[\Phi_e + \phi_i(t)] \cos[\varphi_n(t)] - \cos[\Phi_e + \phi_i(t)] \sin[\varphi_n(t)] \} + u_{1n}(t) \quad (18.9)$$

with a mean value

$$\overline{u_1(t)} = K'_p \{ \sin[\Phi_e + \phi_i(t)] \overline{\cos[\varphi_n(t)]} - \cos[\Phi_e + \phi_i(t)] \overline{\sin[\varphi_n(t)]} \} + \overline{u_{1n}(t)} \quad (18.10)$$

As long as Φ_n and u_{1n} have zero mean, the last two terms are zero, leading to

$$\overline{u_1(t)} = K'_p \{ \sin[\Phi_e + \phi_i(t)] \times \eta_n \}, \quad (18.11)$$

where

$$\eta_n \triangleq \overline{\cos[\varphi_n(t)]}. \quad (18.12)$$

We now expand Eq. (18.11) and average again, giving

$$\overline{u_1(t)} = K'_p \eta_n \{ \cos[\Phi_e] \overline{\sin[\phi_i(t)]} + \sin[\Phi_e] \overline{\cos[\phi_i(t)]} \}. \quad (18.13)$$

Since any average value in the modulation is included in Φ_e , the first average is zero and we can write

$$\overline{u_1(t)} = K'_p \eta_n \sin[\Phi_e] \overline{\cos[\phi_i(t)]} \quad (18.14)$$

$$= K'_p \eta_m \eta_n \sin[\Phi_e], \quad (18.15)$$

where

$$\eta_m \triangleq \overline{\cos[\phi_i(t)]}. \quad (18.16)$$

We can use the relationship of $\overline{u_1(t)}$ to Φ_e to define an effective value of phase-detector gain over the averaging time,

$$K_{p\eta} = \frac{d\overline{u_1(t)}}{d\Phi_e} = K'_{p0} \eta_m \eta_n \cos \Phi_e \quad (18.17)$$

$$= K_{p0} \eta_m \eta_n, \quad (18.18)$$

where we have changed K_p to K_{p0} to emphasize that it is the value before suppression, the linear value, and we have used our previous convention (see Section 1.6),

$$K_p = K'_p \cos \Phi_e. \quad (18.19)$$

Equation (18.18) is a linearized equation giving the slope at the operating point, whereas Eq. (18.15) shows the nonlinear nature of the phase detector.

We can also look at $\cos \Phi_e$ in Eq. (18.17) as a suppression factor,

$$\eta_e \triangleq \cos \Phi_e, \quad (18.20)$$

and rewrite $K_{p\eta}$ as

$$K_{p\eta} = K'_{p0} \eta_m \eta_n \eta_e, \quad (18.21)$$

but this requires that our reference state be at $\Phi_e = 0$ and Eq. (13.20) would then be written relative to K'_p , rather than K_p , and $\sigma_{\varphi 0}^2$ would occur with $\Phi_e = 0$. This is another way of expressing the difference between K_p and K'_p , and we must take care not to use both methods together.

In Chapter 16 the noise was affected by the AGC or limiter in the same way as was the signal, and multiplication of K by the suppression factor η represented the reduction of the effects of both upon the loop. It represented an actual reduction in K relative to the value that pertains if the AGC or limiter operates as desired, unperturbed by noise.

Now, however, the additive noise appearing in the loop is unaffected (as if only the signal strength had changed). Therefore, $n'(t)$ must be multiplied by $1/\eta$, at the same time that K_p is multiplied by η , in order that $u_{1n}(t)$ in Eq. (13.20) not be reduced. Thus, the effective input phase power spectral density, given by Eq. (13.4), must be modified to

$$S_{\varphi}(f_m) = \frac{N_0(f_c \pm f_m)}{P_c \eta_s^2}, \quad (18.22)$$

where η_s includes (is the product of) η_m and η_n (and η_e , if we should choose to employ that convention) but not η_A or η_L . Thus

$$\eta = \eta_A \eta_L \eta_s, \quad (18.23)$$

where

$$\eta_s = \eta_n \eta_m \eta_e \quad (18.24)$$

and, if y represents any of the subscripts above, $\eta_y = 1$ if there is no suppression due to that effect.

Note that, while all of the suppression factors modify loop parameters, such as ζ , nominal design parameters are generally used in labeling curves throughout this book. Loops are not redesigned to maintain effective parameters (e.g., ζ_n) as the independent variable changes (except for the “continuously optimized loop” of Fig. 16.13).

Example 18.2 Effect of Modulation on Phase Offset A loop has a phase offset from quadrature of $\Phi_{e0} = 30^\circ$ due to mistuning ($\omega_{in} \neq \omega_c$). If the input is then phase modulated at a frequency too high for the loop to follow, such that Eq. (18.16) gives $\eta_m = 0.8$, what will the offset become?

Since the mistuning has not changed, the same tuning voltage is required after the modulation begins. Therefore, $u_1(t)$ in Eq. (18.15) must remain unchanged, implying

$$\sin[\Phi_e]_{\text{no mod}} = \eta_m \sin[\Phi_e]_{\text{with mod}} \tag{18.25}$$

$$\sin[\Phi_e]_{\text{with mod}} = \sin(30^\circ)/0.8 \tag{18.26}$$

$$\Phi_e = 38.7^\circ$$

18.4.2 Phase Variance with Additive Noise

The values of both η_m and η_n are dependent upon the input to the loop but, as we have seen, the nature of the dependence of φ_n on the additive noise input is complex. For small $\sigma_{\varphi 0}^2$, the loop is approximately linear and responds to the Gaussian input noise by producing Gaussian φ_n . At this extreme, Eq. (18.12) becomes the mean value of the cosine of a Gaussian distribution,

$$\lim_{\sigma_{\varphi 0}^2 \rightarrow 0} \left[\eta_n = \int_{-\infty}^{\infty} p(\varphi) \cos \varphi \, d\varphi \right] = \int_{-\infty}^{\infty} \frac{1}{\sigma_{\varphi 0} \sqrt{2\pi}} \exp\left(-\frac{\varphi^2}{2\sigma_{\varphi n}^2}\right) \cos \varphi \, d\varphi, \tag{18.27}$$

so the suppression factor is approximately the solution to this integral,

$$\eta_n \approx \exp\left(-\frac{1}{2}\sigma_{\varphi n}^2\right), \tag{18.28}$$

which is unity for $\sigma_{\varphi n}^2 = 0$. However, we have seen (Section 17.1) that the distribution deviates from Gaussian at low ρ_L (high $\sigma_{\varphi n}^2$). For this, and another reason that will become evident, Eq. (18.28) is accurate only for small values of $\sigma_{\varphi 0}^2$. How small we will determine by comparison with the simulated results.

The output phase variance with suppression $\sigma_{\varphi n}^2$ can be written in terms of $\sigma_{\varphi 0}^2$, using Eq. (18.22) and the effective noise bandwidth (Section 16.4), as

$$\frac{\sigma_{\varphi n}^2}{\sigma_{\varphi 0}^2} = \frac{B_{n\eta} N_0 / \eta_s^2}{P_c} \bigg/ \frac{B_{n0} N_0}{P_c} = \frac{B_{n\eta}}{B_{n0}} \frac{1}{\eta_s^2}. \quad (18.29)$$

The factor η within $\beta_{n\eta}$ represents any form of suppression, but η_s represents only forms that modify n' . The subscript s cannot be A or L ; for those form of suppression, $\eta_s \Rightarrow 1$

For a first-order loop and a loop with $\alpha = 0$, substituting Eq. (16.19) in this expression gives

$$\frac{\sigma_{\varphi n}^2}{\sigma_{\varphi 0}^2} = \frac{\eta}{\eta_s^2}, \quad (18.30)$$

which is

$$\left. \frac{\sigma_{\varphi n}^2}{\sigma_{\varphi 0}^2} \right|_{\eta \neq \eta_s} = \eta \quad (18.31)$$

for AGC and limiting, but is

$$\left. \frac{\sigma_{\varphi n}^2}{\sigma_{\varphi 0}^2} \right|_{\eta = \eta_s} = \frac{1}{\eta_s} \quad (18.32)$$

for the suppressions discussed here. For the particular case of $\eta = \eta_n$, we use Eq. (18.28) to give, for $\sigma_{\varphi n}^2$ sufficiently small,

$$\left. \frac{\sigma_{\varphi n}^2}{\sigma_{\varphi 0}^2} \right|_{\alpha=0} = \exp\left(\frac{1}{2}\sigma_{\varphi n}^2\right). \quad (18.33)$$

From this equation, $\sigma_{\varphi 0}^2$ can be computed as a function of $\sigma_{\varphi n}^2$ and plotted as in Fig. 18.4, where it is shown along with the theoretical, Fokker–Plank (FP), solution.

The quasi-linear solution is within about 0.4 dB of the exact solution until the non-linear variance exceeds 1 rad². It then rapidly approaches the vertical and eventually becomes double valued in $\sigma_{\varphi 0}^2$. Several causes for inaccuracy can be ascertained.

We have assumed a Gaussian distribution in using Eq. (18.28), and this becomes inaccurate for large $\sigma_{\varphi 0}^2$. Moreover, we have seen that the true variance for all time, taking cycle skipping into account, is infinite and that we are saved from this infinity by taking $\varphi \bmod 2\pi$. But our quasi-linear development has not accounted for either effect and, while they may tend to compensate each other at small values of $\sigma_{\varphi 0}^2$, this is not likely to occur at higher values. Equation (18.33) provides a solution for $\sigma_{\varphi 0}^2$ for

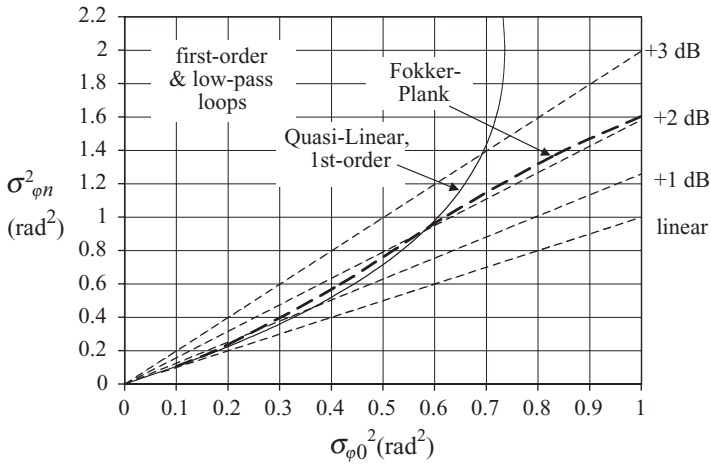


Fig. 18.4 Output phase variance vs. variance under linear assumption, first-order loop, or loop with low-pass filter, quasi-linear, and exact solutions.

any value of $\sigma_{\varphi n}^2$, but $\sigma_{\varphi n}^2$ cannot go above the $\pi^2/3$, the value for equal probability over $\pm\pi$ no matter what is the value of $\sigma_{\varphi 0}^2$.

If we expand Eq. (18.29), using Eq. (16.27) for B_{nn} at $\alpha \approx 1$ instead, we obtain

$$\left. \frac{\sigma_{\varphi n}^2}{\sigma_{\varphi 0}^2} \right|_{\alpha \approx 1} = \frac{(1 + 4\zeta_0^2 \eta)}{1 + 4\zeta_0^2} \frac{1}{\eta_s^2}, \tag{18.34}$$

which, for the particular case $\eta_s = \eta_n$ according to Eq. (18.28), is

$$\left. \frac{\sigma_{\varphi n}^2}{\sigma_{\varphi 0}^2} \right|_{\alpha \approx 1} = \frac{(1 + 4\zeta_0^2 \exp(-\frac{1}{2}\sigma_{\varphi n}^2))}{1 + 4\zeta_0^2} \exp(\sigma_{\varphi}^2 n) \tag{18.35}$$

for this high-gain loop. This equation produces a set of curves, each bearing a similar relationship to the true results (Fig. 17.11) as does the quasi-linear curve in Fig. 18.4. Each quasi-linear curve in the range $0.5 < \zeta < 2$ is low while $\sigma_{\varphi n}^2$ is less than about 1 rad^2 , at which point it rises above the true curve. There is little value in plotting the quasi-linear approximation for noise effects when accurate simulation results are available since we rely on plots in both cases. However, the quasi-linear approximation may supply useful results for small enough variances and parameters for which simulation results are not available. Equations such as (18.33) and (18.35) can be solved iteratively (e.g., using Excel’s “solver”) to find $\sigma_{\varphi n}(\sigma_{\varphi 0})$.

Equations (18.30) and (18.34) do give accurate representations of the effects of phase offset and high-frequency phase modulation, as well as AGC or limiting, as long as the noise level is small.

Example 18.3 Use of Suppression Factors Suppose we choose the parameters, α , ζ , and ω_n , to provide some desirable transient or modulation response, say $\alpha = 0$, $\zeta = 1/\sqrt{2}$, for a maximally flat modulation response [Section 7.6.4]. Then, based on the expected signal strength and noise density, we determine [Eqs. (14.4) and (14.10)] that the phase variance with these parameters will be $\sigma_{\varphi_n}^2 = 0.8 \text{ rad}^2$. The corresponding suppression factor [Eq. (18.28)] is $\eta_n \approx 0.67$. We can get a more accurate value from Fig. 17.8 or 18.4, which show that [Eq. (18.32)] $\sigma_{\varphi_0}^2/\sigma_{\varphi_n}^2 = 0.64 = \eta_n$. Therefore, we increase the voltage gain in an amplifier preceding the loop by $1/0.64 = 1.56$ to compensate for the decrease in $K_\eta = \eta K$ and reestablish the design values (all assuming that the PD can be characterized as a multiplier).

The loop output phase variance $\sigma_{\varphi_n}^2$ will equal 0.8 rad^2 with the assumed input signal, noise levels, and amplifier, implying $\sigma_{\varphi_0}^2 = 0.51 \text{ rad}^2$ in Fig. 17.8, and $\sigma_{\varphi_n}^2$ will follow the locus there as $\sigma_{\varphi_0}^2$ decreases in proportion to N_0/P_c . If α were not zero, we would have to use the values of α and ζ without noise to find the corresponding locus [e.g., from figures like 7.10 or 7.11].

If the input noise should go away, ω_n would become higher than the design value by $\sqrt{1.56} = 1.25$ [Eq. (4.7)] and ζ would decrease by the same factor to 0.57 [Eq. (4.15b)]. The response would now be between the curves for $\zeta = 0.5$ and $\zeta = 0.7$ [Fig. 7.8] and the -3-dB bandwidth would be at $\omega_m/\omega_n = 1.187$ [using Eq. (7.41) for the exact value]. Taking into account the increase in ω_n , this is 48% higher than the original bandwidth [$1.187 \times 1.25 = 1.484$]. The response would also have slight peaking, rather than being maximally flat [Fig. 7.8].

Suppose, however, that the loop is preceded by a square-law AGC, in order to maintain the signal strength at the design value, and the S/N is only -12 dB in the relatively wide filter preceding the limiter. The signal will be attenuated 12 dB ($\eta_a = 1/4$) [Fig. 16.10], requiring another fourfold increase in the voltage gain before the loop, giving then a total voltage gain of 6.24. But, now, if the noise goes away, K_η will increase by 6.24, causing ω_n to increase by $\sqrt{6.24} = 2.5$ and ζ to decrease by $\sqrt{6.24}$ to 0.28. The response will go from maximally flat to having a peak of about 5 dB [Fig. 7.8], and the 3-dB bandwidth will increase by about 3.7 times from the design value [$1.467 \times 2.5 = 3.67$]. We might consider using a coherent AGC to prevent such a severe change.

18.4.3 Tracking the Carrier

18.4.3.1 With Phase Modulation. PLLs are often used to lock to the carrier frequency of a signal that is phase modulated. In such cases, the loop bandwidth is made narrow compared to the modulation frequency so it will not follow the modulation. We have shown that this affects the loop parameters through the suppression factor η_m . Here we will compute the value of η_m , as given by Eq. (18.16), for several modulations.

For biphas modulation, where $\phi_i(t) = \pm\Theta$, Eq. (18.16) becomes

$$\eta_{m,\text{biphase}} = \cos \Theta. \quad (18.36)$$

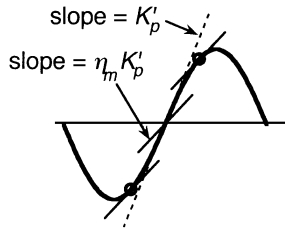


Fig. 18.5 Gain suppression by bi-phase modulation.

This effect is illustrated in Fig. 18.5. The slope of the sinusoidal phase-detector characteristic at the operating points is more shallow than at the center so changes in the mean phase produce less voltage change.

For Gaussian modulation with phase variance σ_{mod}^2 , Eq. (18.16) becomes

$$\eta_{m,\text{Gaussian}} = \exp\left(-\frac{1}{2}\sigma_{\text{mod}}^2\right). \quad (18.37)$$

This must be out-of-band modulation, since we are assuming that the VCO does not track it. Perhaps it represents a communication signal that has the properties of random noise.

For sinusoidal phase modulation, Eq. (18.16) becomes

$$\eta_{m,\text{sine}} = \overline{\cos[m \sin(\omega_m t + \theta_m)]} = \overline{J_0(m) + 2J_2(m) \sin(2\omega_m t + 2\theta_m) + \dots} \quad (18.38)$$

$$= J_0(m). \quad (18.39)$$

Since $J_0(2.4) = 0$, $\eta_m = 0$ at $m = 2.4$. At that modulation index the carrier disappears and there is nothing to lock to. Similarly, in the case of MPSK digital phase modulation, in which there are M equiprobable states at phases $(m \times 360^\circ / M)$, where $m = 0, 1 \dots M$, the mean value of $\cos \Theta$ is zero and

$$\eta_{m,\text{MPSK}} = 0. \quad (18.40)$$

These results just indicate the reduction in the carrier strength relative to the total signal power due to the modulation. We can accomplish the same thing by considering the carrier to be the signal, using its power in equations like (13.4), (13.19), and (14.4).

18.4.3.2 With Phase Modulation and VCO Noise. Although the out-of-band modulation does not directly produce VCO modulation, the reduction in effective gain that it causes produces an increase in $\sigma_{\varphi m}^2$, (properly then $\sigma_{\varphi nm}^2$), exacerbating the inaccuracy of the quasi-linear approximation at higher additive noise levels.

Figures 18.6 and 18.7 show simulated results for a first-order loop and for a loop with $\alpha = 0$ with additive noise plus binary phase modulation of $\pm\Theta$ [Eq. (18.36)].

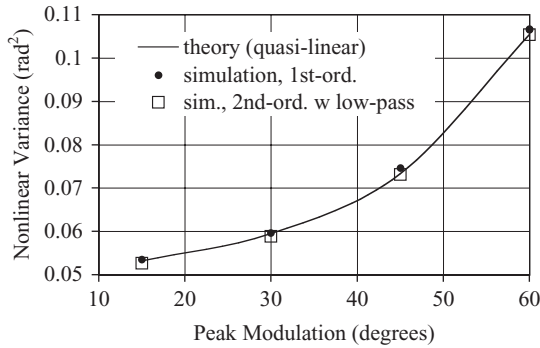


Fig. 18.6 Output variance due to additive noise and modulation versus peak value of binary phase modulation for $\sigma_{\varphi_0}^2 = 0.05 \text{ rad}^2$ in linear, unmodulated first-order loop and second-order loop with $\alpha = 0$, $\zeta = 0.707$ (no phase offset). [Scripts Sig1i and SigOff2i in Appendix i17.M.4.]

Figure 18.6 shows that the quasi-linearization approximation, Eq. (18.32) with $\eta_s = \eta_m \eta_n$, tracks the simulation over a wide range of peak deviations for binary phase modulations if the variance $\sigma_{\varphi_0}^2$ due to noise in the linearized loop without modulation is small.

Figure 18.7 shows, for higher noise levels and two values of Θ , how the quasi-linear approximation breaks away from the true values around 1- rad^2 output variance. However, if we use η_m in the Fokker–Plank solution, dividing ρ_{L0} in Eq. (17.12) by η_m , the results track the simulation accurately, as can be seen in Fig. 18.7. We can start with $\sigma_{\varphi_0}^2$ (the variance, due to noise, without nonlinear effects) and compute $\sigma_{\varphi_m}^2$

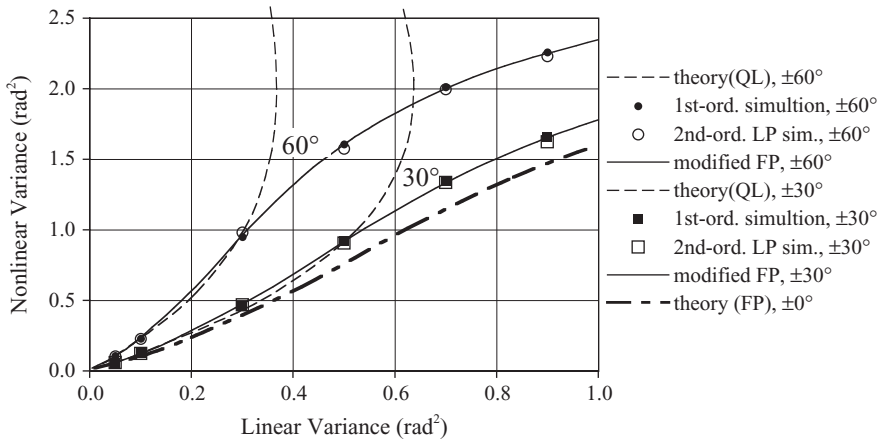


Fig. 18.7 Output variance versus variance due to additive noise in linear, unmodulated first-order loop and second-order loop with $\alpha = 0$, $\zeta = 0.707$, for two peak values of binary phase modulation (no phase offset). [Scripts Sig1i and SigOff2i in Appendix i17.M.4.]

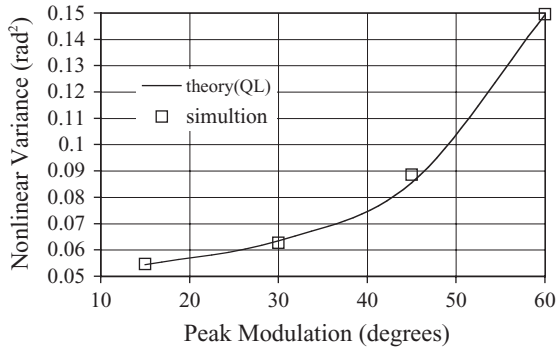


Fig. 18.8 Output variance versus peak value of binary phase modulation for 0.05 rad² due to additive noise in linear, unmodulated, high-gain second-order loop, $\zeta = 0.707$ (no phase offset). [Script SigOff2i in Appendix i17.M.4.]

(the variance as increased just due to modulation) using Eq. (18.32), and, from that, compute $\sigma_{\varphi nm}^2$ (the variance as increased also by the nonlinear effect of noise), using the Fokker–Plank relationship, Eq. (17.12). Alternately, we can start with $\sigma_{\varphi nm}^2$ and proceed in the opposite direction or we can start with $\sigma_{\varphi m}^2$ in the middle. We cannot, however, use $\sigma_{\varphi n}^2$ (the variance as increased only by the nonlinear effect of noise) as the intermediate variable because Eq. (17.12) defines a relationship that depends on the actual output variance.

Simulated results for $\alpha = 1$ are shown in Figs. 18.8 and 18.9. Again, the quasi-linear results match the simulated results only for small $\sigma_{\varphi n}^2$. We might consider using simulation results to give $\sigma_{\varphi nm}^2$ ($\sigma_{\varphi m}^2$), as we used the Fokker–Plank solution above, but ζ , on which results depend, will also be changed by η_m , so it may be simpler just to simulate the combined noise and modulation.

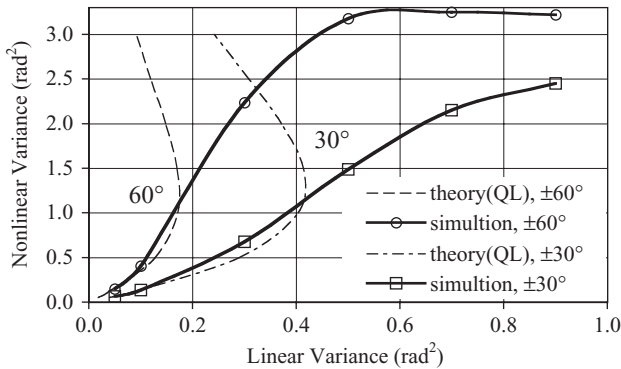


Fig. 18.9 Output variance versus variance due to additive noise in linear, unmodulated, high-gain second-order loop, $\zeta = 0.707$, for two peak values of binary phase modulation (no phase offset). [Script SigOff2i in Appendix i17.M.4.]

Example 18.4 Effect of Modulation on Phase Noise A high-gain second-order loop with $\zeta = 0.707$ has an output phase variance $\sigma_{\varphi 0}^2 = 0.05 \text{ rad}^2$. What is the phase variance after high-frequency (untracked) input phase modulation producing $\eta_m = 0.8$ is added?

From Eq. (18.28), assuming that $\sigma_{\varphi n}^2 \approx \sigma_{\varphi 0}^2$, $\eta_n \approx 0.975$. This is small compared to η_m so let us approximate it as unity. Then $\eta = \eta_s = \eta_n \eta_m \approx \eta_m = 0.8$ and Eq. (18.34) gives

$$\sigma_{\varphi n}^2 = \frac{(1 + 4 \times 0.5 \times 0.8)}{(1 + 4 \times 0.5)} \frac{1}{0.8^2} \sigma_{\varphi 0}^2 = 1.35(0.05 \text{ rad}^2) = 0.068 \text{ rad}^2.$$

By Eq. (18.36), if η_m is produced by biphasic modulation, $\Theta = 37^\circ$ and Fig. 18.8 shows approximately the same variance for that deviation.

18.4.4 Effect on Phase Offset

We can generalize Eq. (18.15) as

$$\overline{u_1(t)} = \eta K'_{p0} \sin \Phi_e, \quad (18.41)$$

where η can be η_A , η_L , η_n , or η_m , or a product thereof.

From this we can write the phase error required to tune the oscillator from center frequency,

$$\Phi_e = \sin^{-1} \left[\frac{\overline{u_1(t)}}{\eta K'_{p0}} \right] = \sin^{-1} \left[\frac{\omega_{\text{out}} - \omega_c}{\eta K'_0} \right]. \quad (18.42)$$

(See also Example 18.2.) Similarly, applying also Eq. (16.20), Eq. (9.5) for the error in a second-order loop during a frequency sweep would become

$$\Phi_e = \sin^{-1} \left\{ \frac{\dot{\Omega}}{\eta} \left[\frac{1}{\omega_{n0}^2} - \frac{t}{K'_0} \right] \right\}. \quad (18.43)$$

18.4.4.1 With High-Frequency Phase Modulation. With binary phase modulation of $\pm\Theta$ at a frequency much higher than the loop bandwidth, Eq. (18.41) can be used to give the phase offset in terms of the offset when the modulation is removed. We use subscript m to indicate a condition during modulation and 0 to indicate it with no modulation. If the loop stays locked, $u_{1m} = u_{10}$, in order that the frequency be unchanged, leading to

$$\sin \Phi_{em} = \frac{\overline{u_{1m}}}{\eta_m K'_{p0}} = \frac{u_{10}}{\eta_m K'_{p0}} = \frac{\sin \Phi_{e0}}{\eta_m} = \frac{\sin \Phi_{e0}}{\cos \theta}, \quad (18.44)$$

where the last equation uses Eq. (18.36).

This is accurately reflected in simulations of first-order loops, where the theory holds, even in the extreme case of $u_1 = 0.5K'_p$ and $\Theta = 60^\circ$ (i.e., $\sin \Phi_{e0} = 0.5$ and $\cos \Theta = 0.5$). In that case, the instantaneous phase eventually alternates between about 30° and 150° while the mean value Φ_m slowly approached 90° , where the gain would be zero. Initial phase and frequency errors were zero for these simulations. Equation (18.44) also holds for second-order loops with $\zeta = 0.707$ and $\alpha = 0$ or 1 , except that, with $\alpha = 1$, the loop does not stay locked at $\Theta = 60^\circ$. It does stay locked at $\Theta = 45^\circ$, however. Thus we should expect Eqs. (18.41) to (18.43) to be accurate for a broad range of high-frequency modulation. This is not surprising since η_m just reflects a reduction in the magnitude of the spectral line to which the loop locks, and thus changes K_p . However, when η includes η_n , we must be concerned about the inaccuracies that we have observed at higher noise levels.

18.4.4.2 With Additive Noise. When there is both offset and additive noise, we must not only determine the value of the mean phase error (i.e., phase offset) in the presence of $\sigma_{\varphi n}^2$, but must also determine the value of $\sigma_{\varphi n}^2$ in the presence of the phase offset.

The theoretical equation for phase offset in the presence of additive noise is analogous to Eq. (18.44),

$$\sin \Phi_{en} \frac{\sin \Phi_{e0}}{\eta_n} = \exp\left(\frac{1}{2}\sigma_{\varphi n}^2\right) \sin \Phi_{e0}, \quad (18.45)$$

where the last equality requires small $\sigma_{\varphi n}^2$.

In order to study the effect of offset on phase noise, it will be convenient to adopt the definition in Eq. (18.20), in which case $\sigma_{\varphi 0}^2$ is the variance with zero offset, as well as zero noise and

$$\eta = \eta_n \eta_e = \exp\left(-\frac{1}{2}\sigma_{\varphi n}^2\right) \cos(\Phi_{en}). \quad (18.46)$$

Combining with Eq. (18.45), this becomes

$$\eta = \sqrt{\exp(-\sigma_{\varphi n}^2) - \sin^2(\Phi_{e0})}. \quad (18.47)$$

We can substitute this for η_s in Eq. (18.32) to obtain the quasi-linear solution for a first-order or $\alpha = 0$ loop,

$$\sigma_{\varphi 0}^2 = \sqrt{\exp(-\sigma_{\varphi n}^2) - \sin^2(\Phi_{e0})} \sigma_{\varphi n}^2. \quad (18.48)$$

This is shown in Fig. 18.10 for two offsets along with the simulation results and the Fokker–Plank solution (which was obtained in the absence of offset). The simulated results are little affected by the smaller offset but an offset equal to half of the

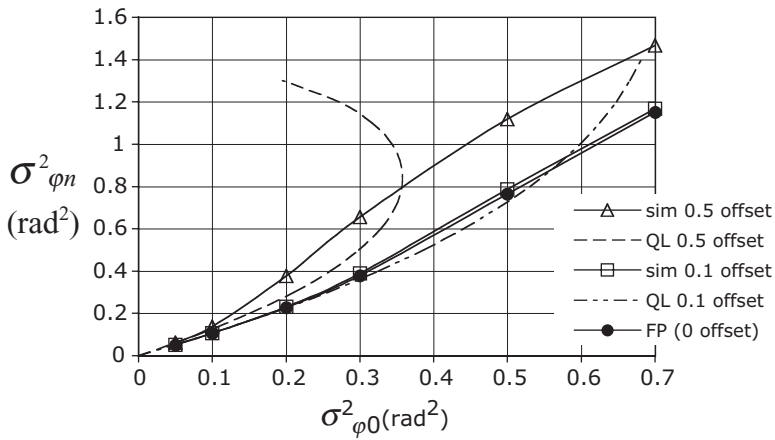


Fig. 18.10 Output variance versus variance for linear loop with additive noise, $\alpha = 0$, $\zeta = 0.707$ (before offsets), at two phase offsets. Offsets equal $\sin \Phi_e$. Simulation results, quasi-linear (QL) and Fokker–Plank (FP) solutions are shown. [Script SigOff2i in Appendix i17.M.4.]

maximum PD output does increase the output variance significantly. The quasi-linear approximation is considerably worse with the larger offset.

For the smaller of the offsets shown here, the mean output phase is shown as a function of the output variance in Fig. 18.11. We see good correlation between the quasi-linear [Eq. (18.45)] and simulated results only for small variances. The reduction in offset in Fig. 18.11 at large variances may reflect the tendency for the average phase (mod 2π) to approach zero when the loop is out of lock. Plots of mean phase for two other cases in Appendix i18.S show similarities.

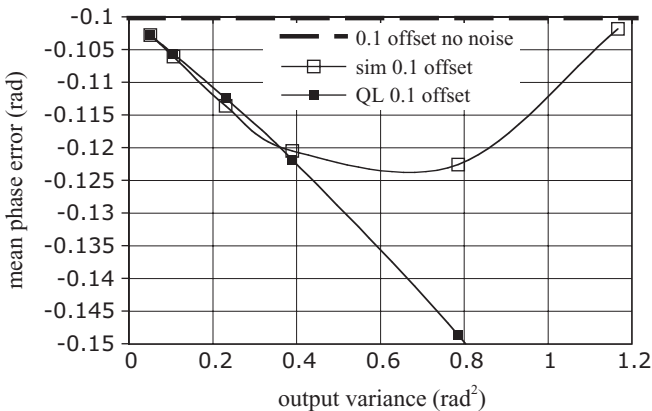


Fig. 18.11 Mean phase error versus output variance when sine of phase offset equals 0.1, $\alpha = 0$, $\zeta = 0.707$ (before noise and offset). [Script SigOff2i in Appendix i17.M.4.]

18.4.4.3 With Limiting. Mean phase is also affected by gain reduction due to limiting, which causes the gain to be multiplied effectively by the suppression factor η_L . We can most easily see the effect of η_L on the offset by observing it with a filter bandwidth that is much wider than the loop bandwidth, so the noise can produce a significant gain reduction without also producing significant nonlinearity in the loop (i.e., $\eta_L < 1$ while $\eta_n \approx 1$). Figure 18.12 shows results under these conditions for $\sin \Phi_{e0} = 0.2$. (The simulation of band-limited noise will be covered in Chapter 20.) If we substitute η_L for η_m in Eq. (18.44), we obtain

$$\Phi_{eL} = \sin^{-1} \left(\frac{\sin \Phi_{e0}}{\eta_L} \right) = \sin^{-1} \left(\frac{0.2}{\eta_L} \right), \tag{18.49}$$

which is labeled “suppression” in the figure. This curve represents only the effect of additive noise in a limiter. It does not include the effect of the noise modulation of the VCO. The simulated results are slightly lower than this curve until they break away due to cycle skipping. Differences from Eq. (18.49) are apparently due to nonlinear effects since they are reduced when the filter bandwidth, in which the variance is converted to density, is widened, thus reducing the variance within the loop bandwidth.

18.4.5 Summary of Suppression

Suppression due to AGC (η_A) simply reflects the change in signal strength, to which K'_p is assumed to be proportional. The additive noise that causes the suppression may, or may not, produce significant nonlinearity. Significant suppression without

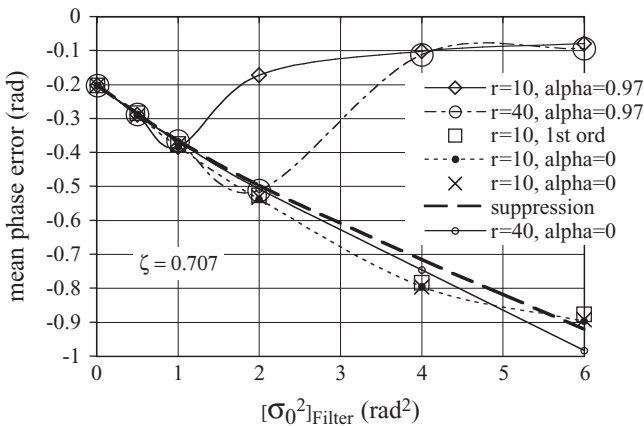


Fig. 18.12 Mean phase error with limiting versus noise variance in the equivalent input video bandwidth. Phase offset is $\sin^{-1}0.2 \approx 0.2$ rad. The ratio of filter video noise bandwidth to loop noise bandwidth is r ; α and ζ are linear parameters before noise and offset. (The dips before the two upper-right curves cross the others may be artifacts of curve fitting.) [Scripts NB1bw in Appendix 20.M.3 and NB2bw or NB2bwi in Appendix 20.M.5]

significant nonlinearity requires that the noise bandwidth preceding the AGC be much greater than the loop noise bandwidth.

Suppression due to limiting (η_L) is basically different because all of the additive noise is converted to phase noise. The frequency distribution of the phase noise is not simply expressed. But, fortunately, the loop tends to respond to that noise as if the suppression were caused by AGC, at least when the loop bandwidth is narrow compared to the prefilter bandwidth. This permits us to estimate the phase variance at the loop output without considering the exact character of the noise at the interface between the limiter and the loop. Otherwise, we can simulate.

Suppression factors η_A and η_L represent a reduction in both signal and additive noise, but other suppression factors, η_m and η_n , do not reduce additive noise. We have given them the general symbol η_s . While η_A and η_L simply multiply K'_0 , η_s must also divide the *equivalent* phase deviation, in order that the additive noise that it represents not change.

Suppression due to high-frequency modulation (η_m) represents a decrease in the effective value of K'_p , due to either (a) an averaging effect produced by modulation to which the loop does not respond or (b) a decrease in the power of the carrier that is being tracked, also due to that modulation. The results are the same but η_m may be easier to calculate with one or the other of these views. To the degree that the loop's assumed unresponsiveness to the modulation is true, the phases of neither the VCO nor the carrier, to which the loop is locked, are affected by the modulation, so it does not produce nonlinearity, although the reduction in K'_p can make the loop more susceptible to nonlinearity from other causes (like additive noise).

Suppression due to phase offset ($\cos \Phi_e$) differs from other suppressions in that it does not represent a decrease in K'_p , but only the reduced gain that we have represented by K_p . Nevertheless, that does reduce the effective noise bandwidth without reducing the additive noise and is therefore treated as a factor in η_s . The concept of suppression represented by η_e and the reduction of K'_p to K_p due to lack of quadrature are two different ways of representing the same effect and are mutually exclusive.

Suppression due to additive noise (η_n) is the only type that we have studied that represents a process wherein the phase error in our linearized loop equation changes dynamically. Suppression due to η_A and η_L involve noise, but they do not represent the effect of that noise in the loop. Neither η_m nor offset represents dynamic changes in the phase of the VCO or of the signal being tracked (the carrier). But, additive noise, as represented by η_n , causes the VCO phase, and thus the phase error, to vary dynamically. Since the additive noise is Gaussian, there is always some probability, in any interval, that it will cause cycle skipping. Thus, whenever we use η_n , the results are subject to inaccuracy, especially at large values of output phase variance.

i18.S APPENDIX: ADDITIONAL OFFSET SIMULATION DATA

This appendix is available from the Wiley Internet site at ftp://ftp.wiley.com/public/sci_tech_med/phase_lock.

CHAPTER 19

ACQUISITION AIDS IN THE PRESENCE OF NOISE

In this section we revisit some of the acquisition aids of Chapter 9 and consider how their performance is affected by noise. In the two cases we will consider, the loop is closed during the sweep. In the first case the sweep voltage continues after lock, whereas in the second case a coherent detector stops the sweep voltage when it detects lock. In all cases $\alpha \approx 1$ so that we will not be concerned with finite hold-in ranges or changes in phase offset with mistiming.

19.1 SWEEPING WITH PLAIN CLOSED-LOOP

19.1.1 Maximum Sweep Rate in Noise

In the presence of Gaussian noise, a loop will not have a 100% chance of maintaining lock over a given time, much less of acquiring lock during a sweep. Therefore, we must settle for some lower probability of acquiring lock in noise. Based on extensive tests, Frazier and Page (1962) gave the sweep rate corresponding to 90% probability of lock as

$$R_{90\%} \triangleq \frac{|\dot{\Omega}_{90\%}|}{\omega_n^2} \approx 2 \frac{(\sqrt{0.5} - \sigma_{\varphi 0})}{1 + \exp(-\pi \zeta / \sqrt{1 - \zeta^2})} \quad (19.1)$$

for $\alpha \approx 1, 1 \geq \zeta \geq 0.5$. (All of the loop parameters in this section will be for the linear loop; thus, for example, $\zeta \equiv \zeta_0$.) Since their data appear to be mainly for $\sigma_{\varphi 0} \leq 0.5$ rad, we might also use that restriction.

Gardner (2005, pp. 195, 197) believes that there is an error of a factor of 1.4 in their work and suggests a preliminary design value of

$$R = 0.5 - \sigma_{\varphi 0} \tag{19.2}$$

with $1 \geq \zeta \geq 0.7$.

Based on new simulations, it appears that

$$R_{90\%} \approx \sqrt{\zeta} \left(1 - \sqrt{2}\sigma_{\varphi 0} \right) \tag{19.3}$$

is a fair (but rough) approximation for $\alpha \approx 1, 1 \geq \zeta \geq 0.5$, and $\sigma_{\varphi 0} \leq 0.5$. In the simulations, the input frequency is swept and the VCO frequency is observed to see if it follows the swept input. This is basically equivalent to sweeping the VCO center frequency and observing whether the VCO frequency stops at the input frequency. The three estimates are compared in Fig. 19.1. We first note that $R > 1$ is not possible (Section 9.5.1), since R equals 1 when the phase detector is producing its maximum possible output, K'_p rad. This supports Gardner's contention that a factor-of-1.4 error is present in Frazier and Page's results. If the value of R from Eq. (19.1) is reduced by $\sqrt{2} \approx 1.4$, it is equivalent to Eq. (19.3) when $\zeta = 1$. Although Eq. (19.3) indicates a 90% probability of lock with no noise and $\zeta = 1$, the probability is actually zero under exactly those conditions. However, it rises to 90% if the sweep rate is reduced just slightly, less than 2%. As ζ increases beyond 1, R may continue to increase, but it still cannot go above 1.

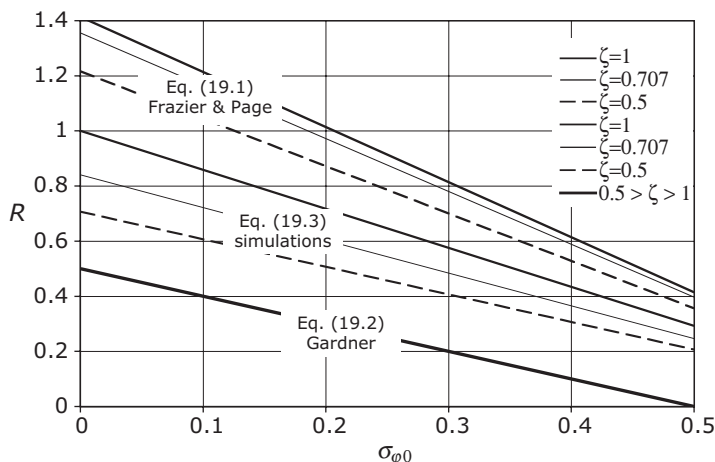


Fig. 19.1 Three approximations for maximum sweep rate in noise.

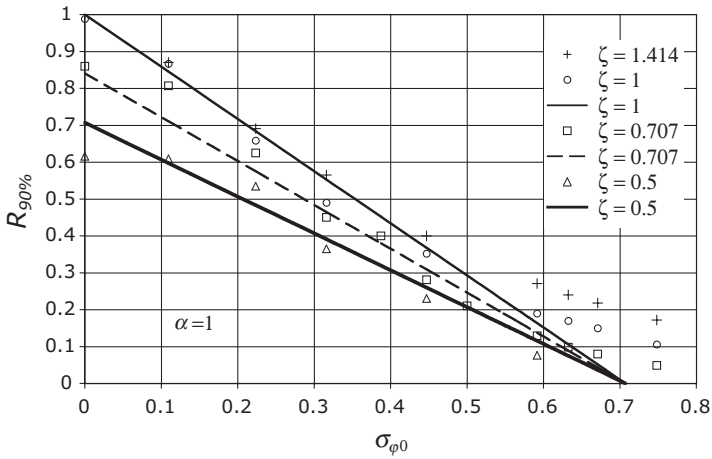


Fig. 19.2 $R_{90\%} = |\dot{\Omega}|/\omega_n^2$ for 90% probability of lock at several values of ζ , simulation results [points] and Eq. (19.3) [lines].

Gardner’s Eq. (19.2) gives a value in the absence of noise that we know produces 100% probability of lock when $\zeta = 0.707$ (Table 9.1). However, this is a recommended preliminary design value rather than an estimate for 90% probability.

Figure 19.2 shows data on which Eq. (19.3) is based while Fig. 19.3 shows the actual probability of lock obtained with sweep rates given by Eq. (19.3). From these we can get an idea of the degree of approximation in Eq. (19.3).

In the region where $\sigma_{\phi 0} \leq 0.5$ rad, the data correspond roughly to Eq. (19.3) and the data for $\zeta = 1.414$ is not much higher than the predicted values for $\zeta = 1$. At

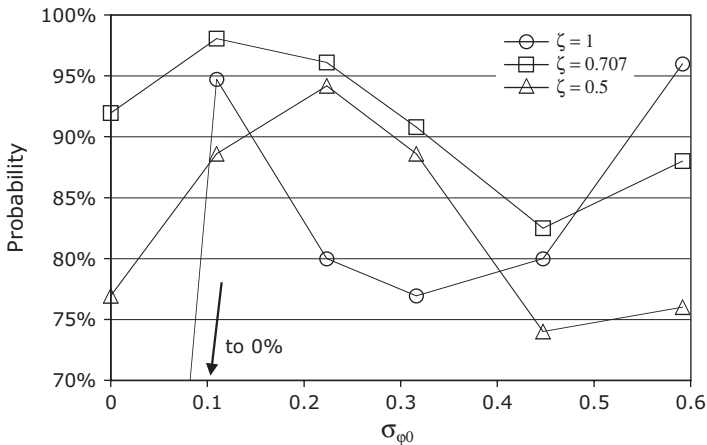


Fig. 19.3 Probabilities of lock from simulations using the theoretical $R_{90\%}$ from Eq. (19.3).

higher values of σ_{φ_0} and ζ , however, the slope of the trend line is shallower and a different approximation would be appropriate. We will see that this is important in determining the optimum sweep strategy.

The question of maximum sweep rate is more complex than it might appear to be; $R_{90\%}$ depends on both the initial mistuning and the criteria for declaring successful acquisition. It is worthy of more detailed consideration, which may improve our understanding as well as reveal possible problems in using the information.

19.1.2 Effect of Initial Mistuning

Viterbi found most of the facts in Table 9.1 using phase-plane plots generated by analog-computer solutions of an equation like Eq. (8.67), except with a linearly increasing mistuning, that is $|\Omega| = kt$. Figure 19.4 represents such a solution. (The phase plane is discussed in Section 8.5.)

Note that this loop locks at $\varphi_e = 60^\circ$ during sweep, so $\sin \varphi_e = \dot{\Omega}/\omega_n^2 = \sqrt{3}/2$ in accordance with Eq. (9.5). We can see that about 2/3 of the values of x at $y = 0$ (values of φ_e at $\omega_e = 0$) lead to lock, whereas the other 1/3 moves to higher values of y where lock is not attained. By drawing these plots for different sweep rates, Viterbi was able to determine that for $\alpha = 1$, $\zeta = 1/\sqrt{2}$, no values of x at $y = 0$ led to lock when $R > 1$ (i.e., $|\dot{\Omega}| > \omega_n^2$) and all of them did if $R < 0.5$ (i.e., $|\dot{\Omega}| < \omega_n^2/2$); Viterbi was also able to discern a pull-in range that exists at lower sweep rates, all of which are shown in Table 9.1. While one can identify the conditions for 0 and 100% probability of lock from the corresponding phase-plane plots, it is more difficult to draw useful conclusions about intermediate probabilities. Figure 19.4 indicates that if we connect the reference to the loop when the frequency error is zero (which practically requires $\alpha < 1$, so we can control the VCO frequency), since the initial phase is presumably unknown, the chances of lock are 2/3. However, if we could do this (i.e., set $\omega_e = 0$), we would not need to sweep. We sweep to accommodate a frequency uncertainty.

If we connect the input when $\omega_e = -2\omega_n$ ($y = -2$), however, the probability of lock appears to be much higher, since all of the loci that miss the lock point at $y = 0$ are concentrated in a very narrow region at $y = -2$. As we move to even greater frequency errors, it becomes more difficult to determine which loci lead to lock. In fact, eventually, at any given large initial frequency error, there are multiple regions of phase that lead to lock separated by regions that do not.

Because of these difficulties and to enable the introduction of noise, we turn to measurements or simulations for determination of intermediate lock probabilities. However, confirming a formula like Eq. (19.3) is particularly difficult because we must obtain multiple lock probabilities at various sweep speeds for any given parameter combination in order to estimate the value of $R_{90\%}$ for that set of parameters and we must do simulations at multiple initial phases to determine each of those probabilities. Moreover, we must use sufficiently fine phase steps and start with a sufficient frequency error that the results are only weakly dependent on either. Higher noise levels will reduce that dependence on initial phase by increasing randomness in a simulation that is actually deterministic in the absence of noise.

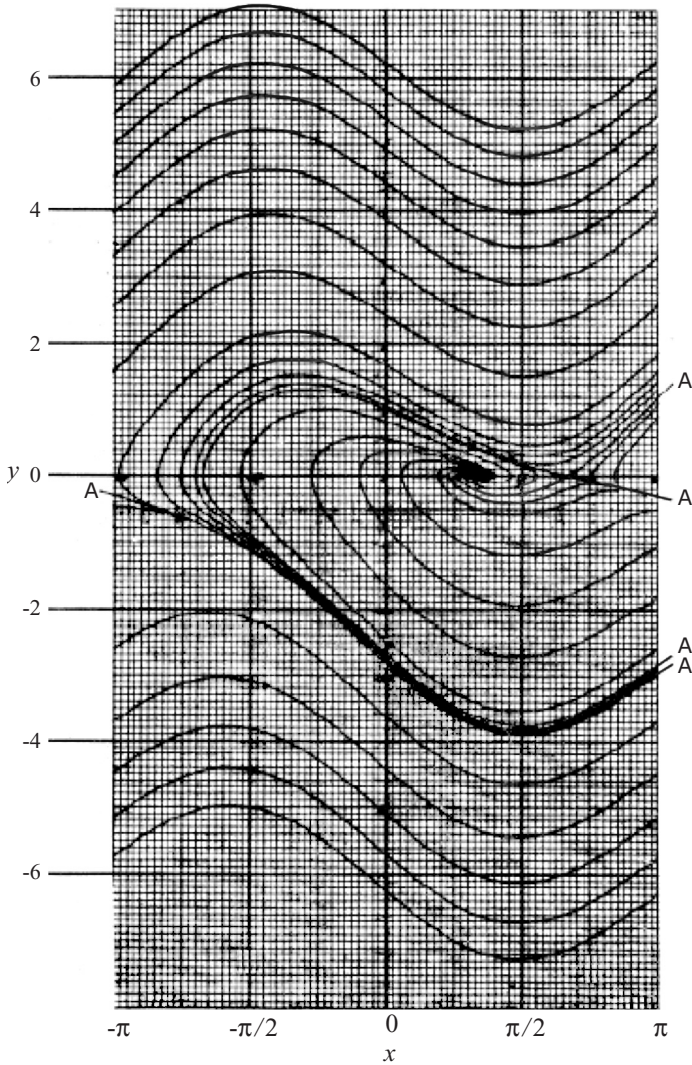


Fig. 19.4 Phase plane plot for $\hat{\Omega}/\omega_n^2 = \sqrt{3}/2, \alpha = 1, \zeta = 1/\sqrt{2}$ where x is φ_e and y is ω_e/ω_n . [From Viterbi (1959, Fig. 11, p. 49), courtesy of NASA/JPL/CALTECH. (Also in Viterbi, 1966, Fig. 3.10, p. 57.)]

The dependency of probability of lock on the initial frequency offset is apparent in Fig. 19.5. A few points here have been plotted for $\alpha < 1$, but no great difference is seen from the plots with $\alpha = 1$. The sampling rate used for most simulations is $f_s = 200\pi f_L$, which is high enough to produce less than 0.3° extra phase shift at unity gain. However, when simulating a loop with a large initial frequency error, higher sampling rates may be necessary to avoid delays that produce significant phase

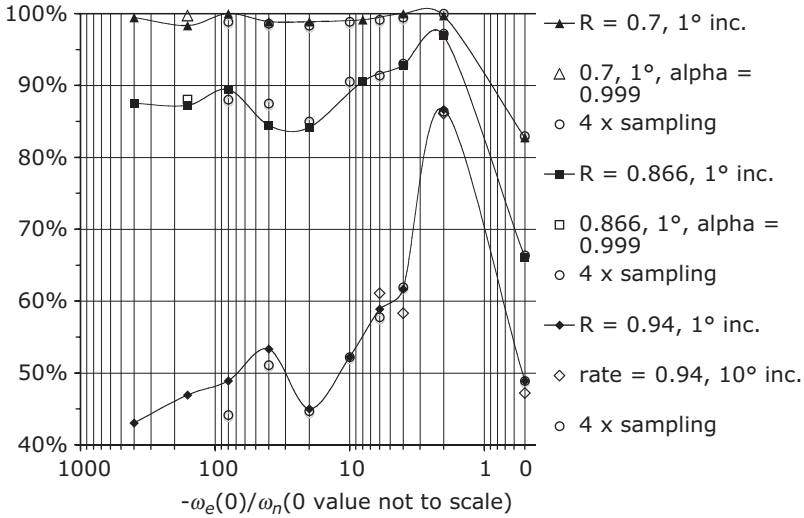


Fig. 19.5 Probability of lock without noise at 3 values of $R \triangleq |\dot{\Omega}|/\omega_n^2$ versus initial frequency error. $\zeta = 0.707, \alpha \approx 1$. The logarithmic scale is distorted to show 0.

shift at the initial offset frequency or even phantom (aliased) error frequencies. The simulations for Figs. 19.2 and 19.3 were performed with an initial frequency offset of $f_e(0) = -80f_n$ and a sampling rate that was up to four times the normal rate of $f_s = 100\omega_L$ in order to maintain $f_s > 30|f_e(0)|$.

Most simulation used 1° phase increments; some used 0.1° increments. Using finer increments gives more statistical samples as well as greater accuracy.

19.1.3 Determining Successful Acquisition

Without noise, it is easy to tell if the VCO is following the input ramp, but at higher noise levels it can be difficult.

Exactly when the observation should be made depends on the application. For example, when the lock is detected, the sweep might be stopped, the phase error might be reduced (e.g., by adjusting the VCO center frequency) to reduce cycle skipping, and the loop bandwidth might then be reduced to improve the output variance. The observations required to determine that lock had been achieved would be similar to those made here. Time is required for the changes resulting from lock (e.g., stabilization of phase error and the change in the trend of the VCO frequency) to become significant compared to the noise. However, longer times can lead to higher probabilities that the lock will be lost.

For the simulation described here, the VCO frequency value was low-pass filtered to reduce the noise in the observation and then compared to an input frequency ramp. In most cases, the simulation began at a mistuning of $-80\omega_n$ and ended at about

+80 ω_n . Thus, zero mistuning occurred near the middle of the display. The mean square phase error over the last 20% of the run was also observed.

There is some tradeoff between accuracy and the time required to analyze the results for a given sweep speed. (1) The fastest method is to base the judgment on histograms of final relative frequency errors and mean square phase errors. (2) Making a plot of those values for all of the initial phases can be helpful. (3) Judging whether lock was achieved for each individual run by inspecting each pair of final values can improve accuracy. (4) The most accurate, and time consuming, method is to observe the data plotted at the end of each individual run; this displays what is happening during the run, not just at the end. Results given here were based on a mixture of the first three methods with insight gained from the fourth. More details are given in Appendix i19.M.

19.1.4 Optimum Sweep Parameters

From basic considerations, we have confidence that, for a given lock probability, ω can be normalized to ω_n^2 and that the result is a function only of α , ζ , and $\sigma_{\varphi 0}$. For a fixed α , which we will set to 1, this is expressed by the functional relationship

$$R \triangleq \frac{|\dot{\Omega}|}{\omega_n^2} = f(\zeta, \sigma_{\varphi 0}). \tag{19.4}$$

This relationship is important because it means that a formula or simulation or experimental data can be obtained at any value of ω_n but can be applied at any other value. If we know all of these parameters, knowledge of $f(\zeta, \sigma_{\varphi 0})$ allows us to determine how fast we can sweep. But, if we want to maximize $|\dot{\Omega}|$ in a given broadband noise environment by adjusting these parameters, we would like $|\dot{\Omega}|$ to be normalized to some function of the only parameter that is fixed, $S = N_0/P_c$. Fortunately, $|\dot{\Omega}|$ can be normalized to S^2 ,

$$R' \triangleq S^2 |\dot{\Omega}| = f(\zeta, \sigma_{\varphi 0}). \tag{19.5}$$

Let us outline how we can show that this is true. We can use Eq. (14.10) to replace ω_n in Eq. (19.4) by $4B_n \text{ rad}/(2\zeta + 1/(2\zeta)) \text{ cycle}$ and then moving everything but B_n to the right side, giving

$$\frac{|\dot{\Omega}|}{B_n^2} = f(\zeta, \sigma_{\varphi 0}). \tag{19.6}$$

Then we can use Eq. (14.4) to replace $1/B_n$ by $S/\sigma_{\varphi 0}^2$ and move $\sigma_{\varphi 0}^4$ to the right-hand side, giving the functional relationship in Eq. (19.5). This applies to both the simulation results and approximate formulas (Fig. 19.2). Applying this transformation

to Eq. (19.3) [using Eqs. (14.4) and (14.10)], we obtain

$$R'_{90\%} = 16F_{\sigma}(\sigma_{\varphi 0})/F_{\zeta}^2(\zeta), \tag{19.7}$$

where

$$F_{\sigma}(\sigma_{\varphi 0}) = \sigma_{\varphi 0}^4(1 - \sqrt{2}\sigma_{\varphi 0}) \tag{19.8}$$

and

$$F_{\zeta}(\zeta) = 2\zeta^{0.75} + 0.5\zeta^{-1.25}. \tag{19.9}$$

By setting the derivatives of Eqs. (19.8) and (19.9) to zero, we find that the fastest sweep, in the region in which the formula applies, occurs when the loop bandwidth is adjusted to produce a 5 dB S/N in the linear loop ($\sigma_{\varphi 0} = 0.57$ rad) and $\zeta = 0.65$. This can be seen in Fig. 19.6. There each curve peaks at $\sigma_{\varphi 0} = 0.57$ rad and each curve corresponds to two values of ζ , both values getting further from 0.65 as the curves become lower. Thus the curves represent a (three-dimensional) peak of $R'_{90\%}$ in both the $\sigma_{\varphi 0}$ and ζ dimensions.

However, the peak in Fig. 19.6 is near the upper limit of $\sigma_{\varphi 0}$ for applicability of Eq. (19.3) and we can see that values of $R'_{90\%}$ obtained from simulations continue to rise at higher $\sigma_{\varphi 0}$. The true optimum is affected by the deviations of the sweep rate from Eq. (19.3), which are apparent in Fig. 19.2. It could be found by following a similar procedure on an equation that approximates the sweep rate over the whole

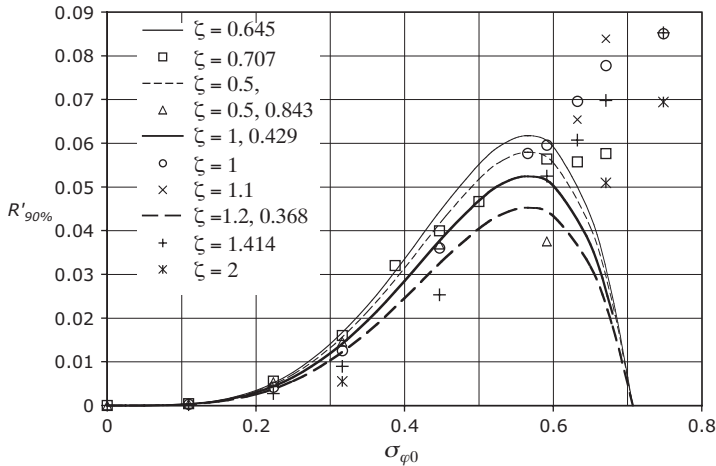


Fig. 19.6 Sweep rate, normalized to S . Curves are theoretical values from Eq. (19.3) and points are from simulations. A curve for $\zeta = 0.707$ would be within 1% of the highest curve shown here ($\zeta = 0.645$), so the corresponding data points can be compared to it.

range of $\sigma_{\varphi 0}$ or from a three-dimensional plot of a sufficient number of data points for $R'_{90\%}$.

It does appear that 90% probability of lock can be attained in a shorter time by opening the loop wider and using a higher ζ than the above theoretical optimums. Determination of how much wider requires further investigation of that region, where, unfortunately, simulations require more time because of the low values of R . Some results are discussed in Appendix i.19.S and these suggest that a normalized rate of $R'_{90\%} \approx 0.12$ can be obtained for $\zeta = 2$ and $\sigma_{\varphi 0} = 1.2$ rad, but it is possible that a higher value could be obtained for greater values of both ζ and $\sigma_{\varphi 0}$. The following should be noted:

- There is no independent confirmation of these results.
- The phase (and frequency) is very noisy in this region, making determination of lock difficult.
- We may be getting close to bandwidths where acquisition without sweeping is a significant factor. We can often see the output locking to the sweeping frequency well before the occurrence of zero frequency error.

19.2 REDUCTION OF COHERENT DETECTOR OUTPUT (CLOSED-LOOP SWEEPING)

We have seen that u_1 is reduced in the presence of noise by the factor η_n . We have also seen that this results in an increased phase error given by Eq. (18.43) as

$$\varphi_e = \sin^{-1} \left\{ \dot{\Omega} \left[\frac{1}{\omega_n^2} - \frac{t}{K} \right] \frac{1}{\eta_n} \right\}.$$

As a result, the output of the coherent detector [Eq. (9.6)] is modified to be

$$v_{CD} = \eta_n K'_{pd} \cos \varphi_e = \eta_n K'_{pd} \sqrt{1 - \sin^2 \varphi_e} \quad (19.10)$$

$$= K'_{pd} \eta_n \sqrt{1 - \left\{ \dot{\Omega} \left[\frac{1}{\omega_n^2} - \frac{t}{K} \right] \frac{1}{\eta_n} \right\}^2} = K'_{pd} \sqrt{\eta_n^2 - \left\{ \dot{\Omega} \left[\frac{1}{\omega_n^2} - \frac{t}{K} \right] \right\}^2}. \quad (19.11)$$

The approximation for η_n in Eq. (18.28) would then lead to

$$v_{CD} \approx K'_{pd} \sqrt{e^{-\sigma_{\varphi n}^2} - \left\{ \dot{\Omega} \left[\frac{1}{\omega_n^2} - \frac{t}{K} \right] \right\}^2}. \quad (19.12)$$

19.3 CLOSED-LOOP SWEEPING IN NOISE WITH COHERENT DETECTOR

19.3.1 Successful Acquisition

Blanchard (1976, pp. 293–296) has obtained experimental data showing the probability of successful acquisition in a circuit such as that shown in Fig. 9.1. For his test circuit, $\zeta = 1$ and $\alpha \approx 1 (\omega_p \ll K)$. The “Lock Indication” is used to stop the sweep. The threshold is set at half of the value given by Eq. (19.12) [with $K \Rightarrow \infty$],

$$v_T = v_{CD}/2, \tag{19.13}$$

to obtain the least dispersion in the frequency offset when the sweep stops. The bandwidth ω_d of the low-pass decision filter at the output of the coherent detector was varied to change the probability of a false stop. The wider is that filter the more likely are false stops but the more readily will the sweep be stopped on the signal. The probability p of a false stop is a parameter. Figure 19.7 illustrates his results for

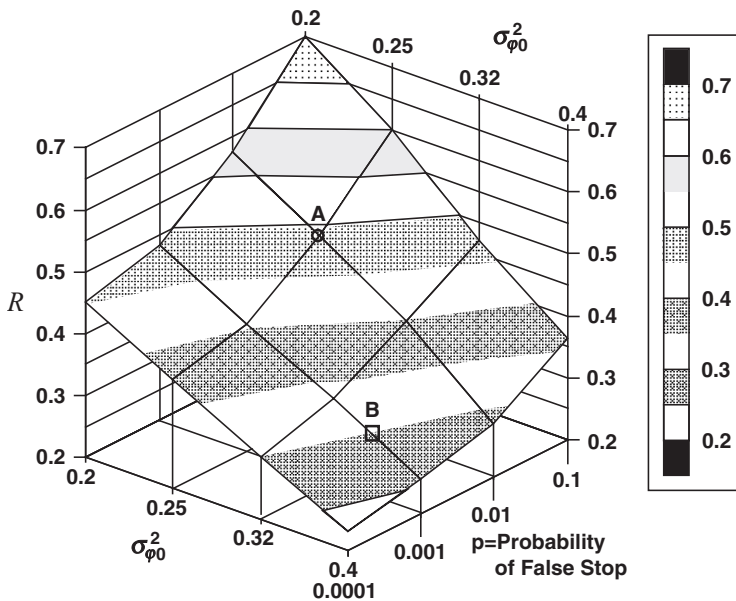


Fig. 19.7 Sweep speed for 99% probability of acquisition vs. phase variance and probability of false stop. The loop has $\zeta = 1$ and $\alpha \approx 0.9975$. The x and y axes (horizontal plane) are logarithmic. In this three-dimensional plot, read the intersection of planes representing the three variables with the data surface. The intersections of the constant $x (= \sigma_{\phi_0}^2)$ planes and the constant $y (= p)$ planes with the data surface are marked by lines. Shading on the data surface indicates vertical ranges of $z = R \triangleq |\dot{\Omega}|/\omega_n^2$, corresponding to the legend on the right. (Points A and B are used in the Example 19.1.) [Based on data taken from Blanchard (1976).]

a 99% probability of acquisition. (Blanchard plots probability of acquisition against sweep speed for four noise levels and four probabilities of a false stop.)

One interpretation of Fig. 19.7 for some particular loop (i.e., once its parameters have been chosen) is as follows. The noise in the loop bandwidth identifies a particular vertical plane—for example, $\sigma_{\varphi_0}^2 = 0.32$. As we increase the bandwidth of the decision filter, more noise passes through the filter and false stops become more likely. Some filter bandwidth will produce a false stop probability of $p = 0.0001$ for example. The intersection of the $\sigma_{\varphi_0}^2 = 0.32$ plane and the $p = 0.0001$ plane defines a vertical line, which is easily visible in the figure. (Sweep speed does not affect p because false stop occurs when the signal is not passing through the decision filter.) As sweep speed increases, probability of acquisition decreases, moving the locus along that vertical line. Probability of acquisition drops to 99% when the locus on the vertical line reaches the surface.

Example 19.1 Sweep Acquisition in the Presence of Noise A high-gain second-order loop has $\zeta \approx 1$, $\omega_n = 1600$ rad/sec, and a ratio of additive noise power density to signal of -36 dbc/Hz. The closed loop is swept using a circuit of the type described above.

- a. How fast can it be swept if the chances of acquisition are to be 99% and the probability of false stop is to be 1%?

From Eq. (14.10), the noise bandwidth is

$$B_n = \frac{1600 \text{ rad/sec}}{4} \left[\frac{1}{2} + 2 \right] \frac{\text{cycle}}{\text{rad}} = 1000 \text{ Hz.}$$

Multiplying this by the relative power density [Eq. (14.4)], we obtain

$$\sigma_{\varphi_0}^2 = 1000 \text{ Hz} \times 10^{-3.6} = 0.25 \text{ rad}^2.$$

We locate this value of $\sigma_{\varphi_0}^2$ on the x -axis of Fig. 19.7 and note that the plane representing 0.25 rad^2 , which is parallel to the left wall, intersects the data surface in a line that contains point **A**. We also note that the plane representing $p = 0.01$ probability, which is parallel to the right wall, intersects the data surface in a line that also contains point **A**. Point **A** occurs where the height is $z \approx 0.48$, as can be seen by noting its position near the top of the shaded stripe and comparing that to the legend on the right. Therefore, the corresponding sweep speed is

$$|\dot{\Omega}| = z\omega_n^2 = 0.48 (1600 \text{ rad/sec})^2 = 1.23 \times 10^6 \text{ rad/sec}^2.$$

- b. By how much will the signal power be decreased if we are to have 0.1% probability of false stop when the sweep speed is set at $7.7 \times 10^5 \text{ rad/sec}^2$?

Dividing $|\dot{\Omega}|$ by ω_n^2 , we obtain $z = 0.3$. This value lies along the top of the lowest stripe on the data surface in Fig. 19.7. The intersection of this value of z

with the 0.001 probability plane occurs at a point **B**. This point lies in the plane where $\sigma_{\varphi 0}^2 \approx 0.35 \text{ rad}^2$. Thus the signal power has been multiplied by a factor of $0.25 \text{ rad}^2 / 0.35 \text{ rad}^2 = 0.7$, or -1.5 dB , to produce the corresponding increase in phase variance.

19.3.2 False Stops

We can compute how the probability of false stop, p , depends on the detector threshold and the decision-filter bandwidth. While the loop is attempting to lock, the noise in a differential bandwidth at the output of the coherent detector is given by Eq. (13.19), where K'_p is here the gain K'_{pd} of the coherent detector.¹ Integrating this noise density over the band of the low-pass decision filter gives the mean square voltage at its output,

$$\sigma_d^2 = (K'_{pd})^2 B_{nd} N_0 / P, \tag{19.14}$$

where B_{nd} is the noise bandwidth of the decision filter and P is the signal power. The noise voltage is characterized by a Gaussian probability density with this variance and the probability that the threshold will be exceeded is obtained by integrating the density that exceeds the threshold V_T . (See Fig. 19.8.) We can do this by integrating the standard normal density function from $x_T = V_T / \sigma_d$ to infinity.

$$\begin{aligned} p &= \frac{1}{\sqrt{2\pi}} \int_{x_T}^{\infty} \exp\left(\frac{-x^2}{2}\right) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-x_T} \exp\left(\frac{-x^2}{2}\right) dx \\ &\equiv \text{SNCD}(-x_T) = 1 - \text{SNCD}(x_T), \end{aligned} \tag{19.15}$$

where

$$x_T = \frac{V_T}{\sigma_d} = \frac{V_T}{K'_{pd} \sqrt{(B_{nd} N_0) / P}}, \tag{19.16}$$

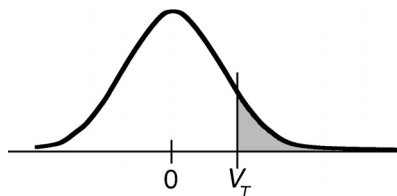


Fig. 19.8 Probability of noise exceeding threshold before lock.

¹ We have used K_p for the gain of both the coherent detector and the phase detector. The meaning is identical in both cases, but it is not unlikely that because of differing signal levels, for example, the values will be different for the two detectors in a given application. Therefore, at this point, we use K'_{pd} for that particular value of K'_p that applies to the coherent detector.

and B_{nd} is the noise bandwidth of the single-pole low-pass decision filter $[(\omega_d/4)c$; see Eqs. (14.7) and (14.8)], and SNCD is the standard normal cumulative distribution function, for which tables are available.

The consequence of false stops is to slow up the sweep by a factor $(1 - p)$. The length of each stop would tend to be about $1/\omega_d$. If, to reduce false restarts during lock, the threshold is lowered or the decision filter is narrowed after the sweep stops, then the sweep will be slowed more. If this is not done, then there is the possibility that the loop will begin sweeping again because noise causes the output from the coherent detector to drop below threshold after lock is achieved.

Example 19.2 False Stop What was the decision-filter bandwidth in Example 19.1(a)?

The probability of false stop in that example was $p = 0.01$. From a table of normal cumulative distribution functions we find that Eq. (19.15) gives 0.01 when $x_T = 2.33$ [SCND(2.33) = 0.99]. We will use this in Eq. (19.16) to solve for B_{nd} .

In Example 19.1, we determined that $\sigma_{\varphi_0}^2 = 0.25 \text{ rad}^2$. From Fig. 17.11, this implies an actual variance of $\sigma_{\varphi_n}^2 \approx 0.33 \text{ rad}^2$. Using, in Eq. (19.12), this phase variance with the sweep speed and ω_n from Example 19.1, we find the steady-state output voltage, when the sweeping loop is locked, to be $v_{CD} = 0.7K_{pd}$. Setting the threshold at half this voltage, as in Eq. (19.13), we obtain $v_T = 0.35K_{pd}$ as the value used in the experiment. Using, in Eq. (19.16), these values plus the relative noise power density of -36 dBc/Hz from Example 19.1, we obtain

$$2.33 = \frac{0.35}{\sqrt{B_{nd} \times 10^{-3.6}/\text{Hz}}},$$

or $B_{nd} = 90 \text{ Hz}$.

19.3.3 False Restart

The probability p' , that the coherently detected output will drop below threshold after lock, is given by the shaded area under the probability density curve in Fig. 19.9.

$$p' = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-x'_T} \exp\left(\frac{-x^2}{2}\right) dx \equiv \text{SNCD}(-x'_T) = 1 - \text{SNCD}(x'_T) \quad (19.17)$$

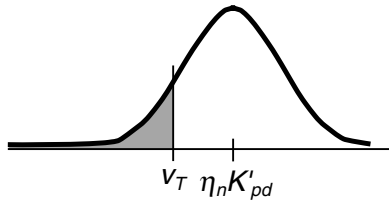


Fig. 19.9 Probability of noise dropping below threshold after lock.

The mean output with the loop locked to a signal in the presence of noise is K_{pd}' multiplied by η_n . From this, the normalized integration limit is

$$x_T' = \frac{K_{pd}'\eta_n - V_T}{\sigma_d} \approx \frac{K_{pd}' \exp\left(-\frac{1}{2}\sigma_{\varphi n}^2\right) - V_T}{K_{pd}'\sqrt{(B_{nd}N_0)/P}}. \quad (19.18)$$

Lock will not necessarily be lost when the detector output drops—this depends on how long it stays below threshold—but some transient will be caused by restart of the sweep, even if it is momentary.

19.3.4 False Stop versus False Restart

Since both p and p' depend on the noise power and on the sweep speed and the loop's natural frequency (because v_T depends on them), both will be varied by these parameters. However, the dependence is not the same in the two cases, so we may want to select these parameters based on the required values of both p and p' . Blanchard (1976, p. 297) plots p against p' for fixed values of

$$Q_1 \equiv Re^{0.5\sigma_{\varphi n}^2}. \quad (19.19)$$

Such a plot helps us trade off p and p' for a given noise level and sweep rate. We can generate any of its curves as follows. Combine Eq. (19.12), (19.13), and (19.19), assuming a high gain loop (where $K \gg \omega_n^2 t$), to give

$$V_T = 0.5K_{pd}'\sqrt{1 - Q_1^2}e^{-0.5\sigma_{\varphi n}^2}. \quad (19.20)$$

Then write Eq. (19.16) as

$$x_T = Q_2\sqrt{1 - Q_1^2}, \quad (19.21)$$

where

$$Q_2 \equiv \frac{\exp(-0.5\sigma_{\varphi n}^2)}{2\sqrt{(B_{nd}N_0)/P}}, \quad (19.22)$$

and Eq. (19.18) as

$$x_T' = Q_2\left(2 - \sqrt{1 - Q_1^2}\right). \quad (19.23)$$

If we begin with a probability of false stop p , we can obtain x_T from Eq. (19.15). Then, for a given value of Q_1 , if we can obtain x_T' by combining Eq. (19.21)

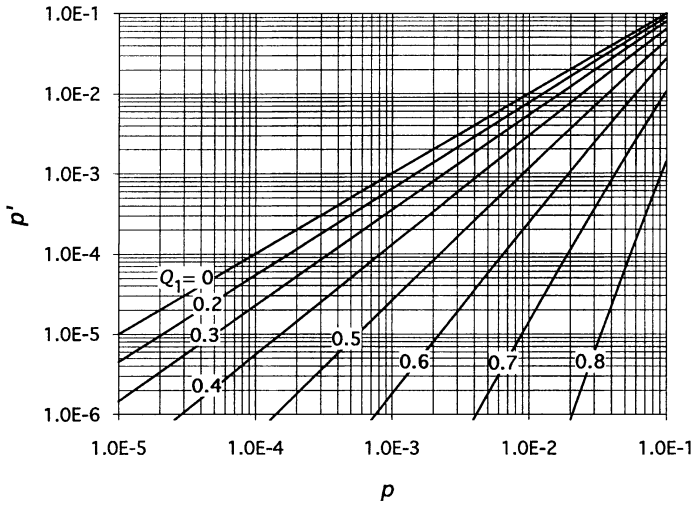


Fig. 19.10 Probability of false restart (p') vs. probability of false stop (p), fixed Q_1 . Thresholds set at 50% of maximum output during sweep.

and (19.23),

$$x'_T = x_T \left(\frac{2}{\sqrt{1 - Q_1^2}} - 1 \right). \tag{19.24}$$

From this we can obtain p' by Eq. (19.17). By these means, p' has been plotted against p in Fig. 19.10.

Equation (19.19) shows that noise can be traded for sweep speed along any of the curves in Fig. 19.10 but, for a given combination of noise and speed (fixed Q_1), the probability of false stop and the probability of false restart depend on the decision bandwidth in a similar fashion. As the bandwidth is widened, both probabilities increase. However, the probability of acquisition also increases, a fact that cannot be seen from Fig. 19.10.

If the operating point in Fig. 19.7 has been determined—for example, by choice of loop parameters and sweep rate in the presence of a given noise environment—then Q_1 and p are no longer independent. For given values of $\sigma_{\varphi_0}^2$ and p , we can read R from Fig. 19.7. If we then convert $\sigma_{\varphi_0}^2$ to $\sigma_{\varphi_n}^2$ —using Fig. 17.11, for example—we can compute Q_1 from Eq. (19.19). This enables us to plot the constant- $\sigma_{\varphi_0}^2$ curves in Fig. 19.11,² which give the relationship between p and p' for 99% probability of

²Theoretically, we can locate the points by knowing p and Q_1 . Practically, we compute p' from p and Q_1 and plot p' versus p . Straight-line approximations are plotted along with the computed points.

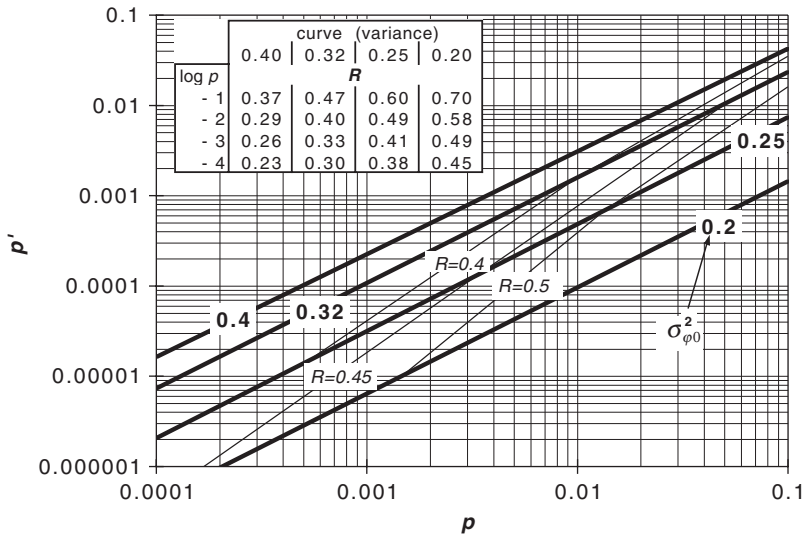


Fig. 19.11 Probability of false restart (p') vs. probability of false stop (p), $\sigma_{\varphi_0}^2$ fixed for 99% acquisition probability. Normalized sweep speed $R = \Omega/\omega_n^2$ is also shown. Thresholds set at 50% of maximum output during sweep. $\zeta = 1, \alpha = 0.9975$.

acquisition. As the decision bandwidth is widened, the probabilities of false stop and restart increase as does R (see Fig. 19.7).

Consideration of Fig. 19.7 and the two sets of curves in Fig. 19.11 leads us to the conclusion that, under the conditions that have been set,³ we can choose independently only three of the five parameters, $p, p', \sigma_{\varphi_0}^2, R$, and probability of acquisition.

Example 19.3 False Stop and Restart What are the conditions that correspond to a 1% chance of false stop and a 0.1% chance of false restart after stopping? The loop is a high-gain second-order loop, $\zeta = 1$, and is sweeping with the loop closed.

We see from Fig. 19.11 that $R \approx 0.43$ and $\sigma_{\varphi_0}^2 \approx 0.29$ at the intersection of $p = 0.01$ and $p' = 0.001$. This is approximately the same value of R obtained from Fig. 19.7 for the same values of $\sigma_{\varphi_0}^2$ and p . Also, from Fig. 19.10, $Q_1 \approx 0.51$ while from Fig. 17.11 we estimate $\sigma_{\varphi_n}^2 \approx 0.4$. Using these values, Eq. (19.19) gives $R \approx 0.42$. The two results are closer than we might expect, considering all of the graphs involved. Note that while we have used simulated results to relate $\sigma_{\varphi_0}^2$ and $\sigma_{\varphi_n}^2$, we have also depended upon Eq. (18.28), which is most accurate for small $\sigma_{\varphi_n}^2$. As always, the

³ Coherent detector, 50% threshold setting.

proof is in test results on completed designs, but we need the best estimates we can get to produce them efficiently.

i19.M APPENDIX: MATLAB[®] SCRIPT, `swp1`

This appendix, which describes the MATLAB script that was used to generate simulations of swept acquisitions for this chapter, is available from the Wiley Internet site.

i19.S APPENDIX: OPTIMUM SWEEP FOR A FIXED NOISE DENSITY

This appendix is available from the Wiley Internet site at ftp://ftp.wiley.com/public/sci_tech_med/phase_lock.

CHAPTER 20

BANDLIMITED NOISE

Previously we have assumed broadband noise, noise of effectively unlimited bandwidth, at the input to the loop. While those studies have the advantage of applying to any situation where the input noise bandwidth is “wide enough,” and therefore not introducing an additional parameter to complicate the results, there will, in fact, always be a finite noise bandwidth preceding the loop. One benefit of studying the effects of this limited bandwidth is an improved understanding of the degree of approximation that accompanies the broadband assumption. We will again have the benefit of theory and measured results to which to compare our simulated results.

The theory is built on the effects of pulse noise (“clicks”) in the filter preceding the loop, as described in Section 15.7. We begin by considering signals that are centered in a band of noise.

20.1 SIGNALS CENTERED IN A NOISE BAND

20.1.1 Clicks with a First-Order PLL

Hess [1968] studied the reaction of a first-order loop to these clicks to determine the skip rate \bar{F} at the loop output. He observed that the magnitude of the clicks was

$$\omega = \dot{\varphi} \approx 2\gamma, \quad (20.1)$$

where γ is defined by Eq. (15.41). Based on this, he was able to determine (through analysis of the nonlinear loop equation) that the loop would follow the click if

$$\frac{\omega_L r_n(t_i)}{\omega(t_i)} = \frac{\omega_L r_n(t_i)}{2\gamma} \geq 0.52, \quad (20.2)$$

where $r_n(t_i)$ is the ratio of the signal-plus-noise amplitude to the signal amplitude at the time of the click. He further determined, through probability considerations, that this results in a skip rate at the loop output of

$$\bar{F} = \frac{\gamma}{2\pi} \operatorname{erfc} \left[\sqrt{\rho_{\text{in}}} \left(1 + \frac{1.04\gamma}{\omega_L} \right) \right] \text{ cycle}, \quad (20.3)$$

where ρ_{in} is the S/N in the input (RF) filter. Note that this approaches the value for a standard discriminator, Eq. (15.40), for very wide loops, where $\omega_L \gg \gamma$.

For a rectangular input filter, where Eq. (15.43) gives

$$\gamma_{\text{rec}} = \frac{\pi W}{\sqrt{3} \text{ cycle}}, \quad (20.4)$$

this becomes

$$\frac{\bar{F}}{W} = \frac{1}{2\sqrt{3}} \operatorname{erfc} \left[\sqrt{\rho_{\text{in}}} \left(1 + \frac{0.3W}{f_L} \right) \right]. \quad (20.5)$$

Hess obtained experimental values with $W = 13$ kHz and the results are shown in Fig. 20.1 along with plots of Eq. (20.5).

20.1.2 Simulation Results Compared to Measured Data

An important use of the data in Fig. 20.1 is to verify our simulations. When that is done, we can obtain results for different conditions. Figure 20.2 shows the theoretical curves of Eq. (20.5), which we have seen to be a good match to data in Fig. 20.1, along with simulation results. From this we see good agreement between the theory (and, thus, the data) and simulations using the two-source model of Fig. 13.5. We also see the necessity of using that model from the errors that occur when the simple one-source model (where u_{1n} equals u_{ns} or u_{nc}) is used.

Having established the validity of our simulations by comparison with data, we will now employ simulations to expand our description of the performance of the loop in response to bandlimited noise. We will employ input noise filters that are centered on the signal and have a half (video-equivalent) noise bandwidth that is related to the loop noise bandwidth by

$$[B_n]_{\text{Filter}} = r[B_{n0}]_{\text{Loop}}. \quad (20.6)$$

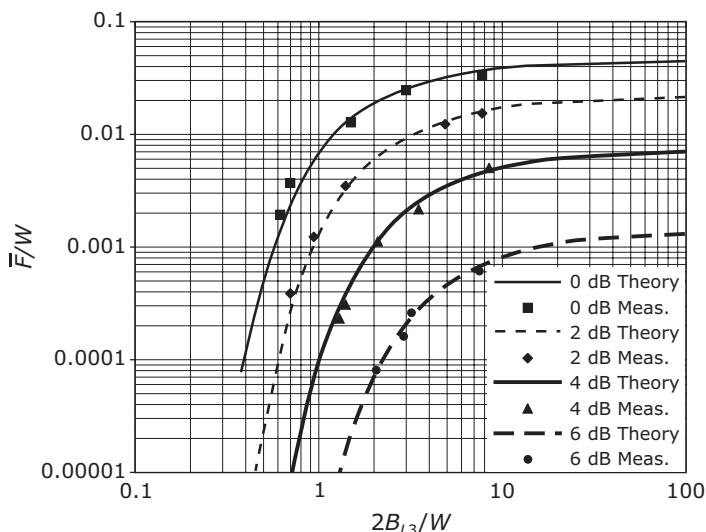


Fig. 20.1 Skip rate versus ratio of loop (linear) -3 -dB bandwidth to RF video-equivalent bandwidth, theory, and measured results for first-order loop. Curve parameter is S/N in the filter’s RF noise bandwidth.

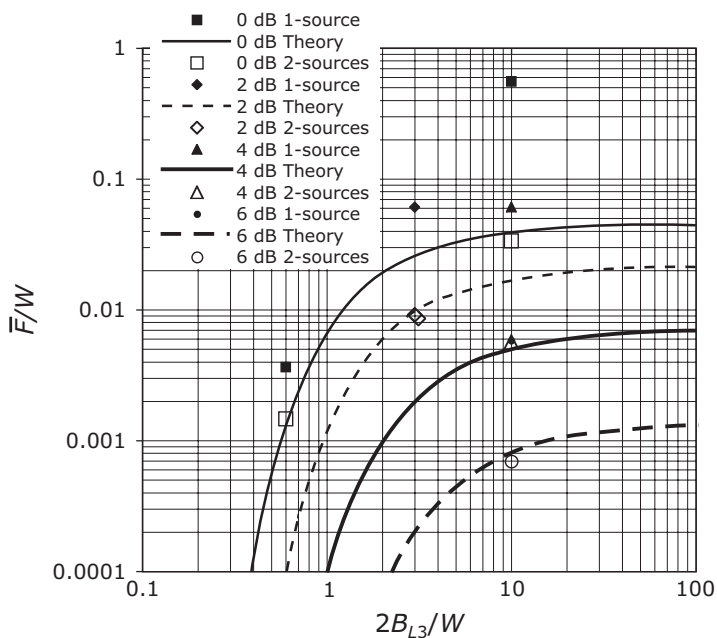


Fig. 20.2 Comparison of simulation results to theory, first-order-loop. Theory curves match measured data. “1-source” and “2-sources” refer to simulations with a simple single noise source or with two sources, respectively. [Scripts like NB1bw, NB1bwi, and NB1ch1 in Appendix i20.M.3.]

TABLE 20.1 Methods of Computing the Net Noise Bandwidth, $[B_n]_{\text{net}}$.

	First-Order RF Filter	High (Fourth)-Order RF Filter
First-order loop	The product of their responses, $F_1(f_m) H_1(f_m)$, is second-order and the net noise bandwidth can be obtained from Eq. (14.10).	Approximating the filter as being very sharp compared to the loop response, we use $[B_n]_{\text{net}} = \int_0^{[B_n]_{\text{filter}}} H_1(f_m) df_m, \quad (20.7)$ where H_1 is the loop transfer function.
Second-order loop	Contour integration of the product of the filter and loop transfer functions, $\oint F_1(s) H_2(s) ds.$	Approximating both responses as rectangular, which is most accurate when the two noise bandwidths differ significantly (see Fig. 7.8). $[B_n]_{\text{net}} = \min [[B_n]_{\text{loop}}, [B_n]_{\text{filter}}]$

Therefore, for near-rectangular input filters with bandwidth B_{RFIn} ,

$$[B_n]_{\text{Filter}} \approx B_{\text{RFIn}}/2 \quad (20.8)$$

(= $W/2$ for rectangular filters). We will plot results against the noise in the net noise bandwidth, the noise bandwidth of the total response of the linear loop preceded by the input filter. The net noise bandwidth, $[B_{n0}]_{\text{net}}$, was computed as shown in Table 20.1. [While we use the subscript, zero, here to indicate that we are referring to the linear bandwidth of the loop, all of the loop parameters (e.g., ω_L , B_{L3}) in this chapter are for the linear loop.]

We will continue with consideration of normalized skip rate, now normalized to the total or net noise bandwidth of the filter and loop, $\bar{F}/[B_{n0}]_{\text{net}}$.

20.1.3 Skip Rate

Figure 20.3 shows skip rate for several filter bandwidth ratios with a rectangular input filter and a first-order loop, conditions similar to those of Fig. 20.1 but with somewhat different normalization and plot variables. Note that the noise in B_{RFIn} is twice the noise in the equivalent video noise bandwidth, so

$$\rho_{\text{in}} = \rho_0/2 \quad (20.9)$$

(i.e., ρ_0 represents the ratio of signal power to the equivalent input phase noise power).

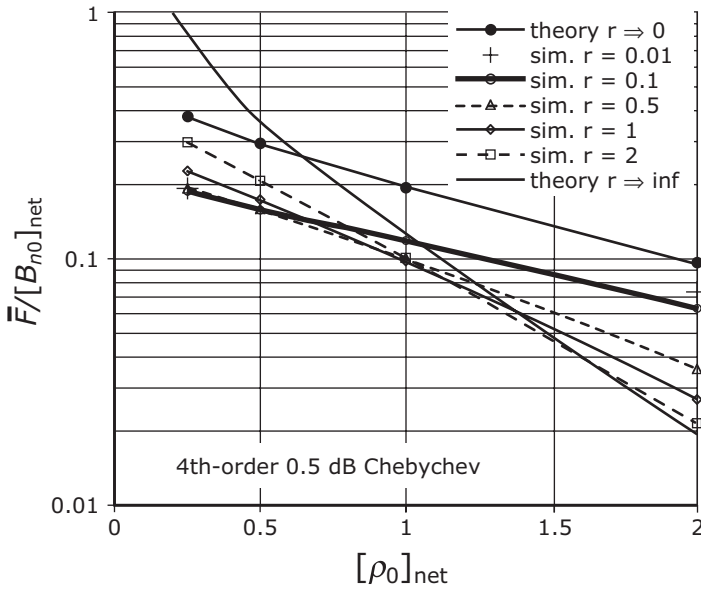


Fig. 20.3 Skip rate (normalized) versus S/N in the net video bandwidth for several values of r , rectangular input filter and first-order loop. [Scripts like NB1bw, NB1bwi, and NB1ch1 in Appendix i20.M.3.]

We cannot show all possible conditions, but these first-order-loop curves appear to be representative; they do not differ much from those obtained for second-order loops having $\zeta = 0.707$ and $\alpha = 0$ or 0.97 . For example, the $r = 0.1$ curve has a total spread of about 4% on the left end and 30% on the right end for the three conditions (i.e., first-order loop and second-order loop with two values of α). Even if the curve for a first-order filter with a first-order loop is included, the spread goes to only about 40% on the left without increasing on the right. The latter is particularly interesting, because γ in Eq. (15.41) is infinite for a first-order filter.

As r increases, we expect the skip rate to approach that of a first-order loop in broadband noise,

$$\lim_{r \rightarrow \infty} \frac{\bar{F}}{[B_{n0}]_{\text{net}}} = \lim_{r \rightarrow \infty} \frac{\bar{F}}{[B_{n0}]_{\text{Loop}}} = \frac{1}{[T_{m1} B_{n0}]_{\text{Fokker-Plank}}}, \tag{20.10}$$

which can be obtained, as a function of $\rho_{L0} = [\rho_0]_{\text{Loop}}$, from Eq. (17.18) and is also plotted in Fig. 20.3 (“theory $r \Rightarrow \text{inf}$ ”). All of the plots tend to pivot about a point near $[\rho_0]_{\text{net}} = 1$, with the broader-band curves changing more with $[\rho_0]_{\text{net}}$ than do the narrower-band curves. At both plotted extremes of $[\rho_0]_{\text{net}}$, they progress toward the broadband value as r increases, as we would expect.

As r decreases toward zero, we expect the skip rate to approach the value given by Eq. (20.5), with $B_{\text{RFin}} \ll f_L$. As $r \Rightarrow 0$, $\bar{F}/[B_{n0}]_{\text{net}} \Rightarrow \bar{F}/[B_{n0}]_{\text{filter}}$, which for a

rectangular filter becomes $\bar{F}/[W/2]$. Using Eq. (20.5), with $W \ll f_L$, this gives

$$\lim_{r \rightarrow 0} \frac{\bar{F}}{[B_{n0}]_{\text{net}}} = 2 \lim_{r \rightarrow 0} \frac{\bar{F}}{W} = \frac{\text{erfc}(\sqrt{\rho_{\text{in}}})}{\sqrt{3}} = \frac{\text{erfc}(\sqrt{[\rho_0]_{\text{net}}/2})}{\sqrt{3}}. \tag{20.11}$$

(The last equality is due to the 2:1 ratio between W , in which ρ_{in} is measured, and $[B_n]_{\text{filter}}$, in which ρ_0 is measured.) This is also plotted [but with a slight increase (6.4%) in value because γ is slightly higher here than with a true rectangular filter]. However, the low- r curves do not appear to approach this theoretical ($r \Rightarrow 0$) curve on the left (at low $[\rho_0]_{\text{net}}$). To do so, the theoretical curve would have to cross the other curves and drop below them on the left, much as the $r \Rightarrow \infty$ curve crossed them in the opposite direction. The apparent improvement in performance with the PLL, even when its bandwidth is large compared to the filter bandwidth, may be related to a difference in the way cycle skips are defined. The simulations (and the experiment) require a full one-cycle phase change, whereas the theoretical equation requires that the phase difference between signal and signal-plus-noise merely cross ± 0.5 cycle [Schwartz, 1966, pp. 148–150]. These may be equivalent at high S/N , but probably not at low S/N .

Perhaps we can observe these effects more clearly by plotting $\bar{F}/[B_{n0}]_{\text{net}}$ against r , which is done in Figs. 20.4 and 20.5 for the two extremes values of $[\rho_0]_{\text{net}}$ that were used in Fig. 20.3. (Data for some other prefilterers are also included.) We again see that $\bar{F}/[B_{n0}]$ remains below the value predicted by Eq. (20.11), even for very low r -values. Note that the lowest S/N in Fig. 20.1 is $\rho_{\text{in}} = 1 \Rightarrow [\rho_0]_{\text{filter}} = 2$, the

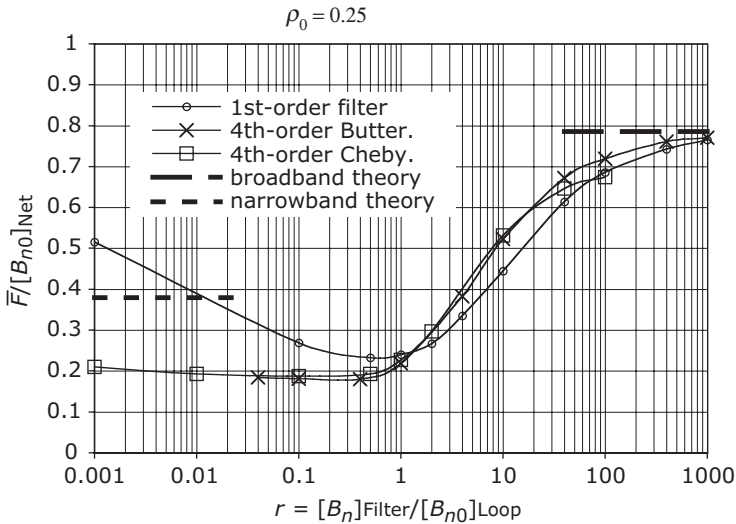


Fig. 20.4 $\bar{F}/[B_{n0}]_{\text{net}}$ versus r for a rectangular filters and for a first-order (single-pole) filter at $[\rho_0]_{\text{net}} = 0.25$, first-order loop. [Scripts like NB1bw, NB1bwi, and NB1ch1 in Appendix i20.M.3.

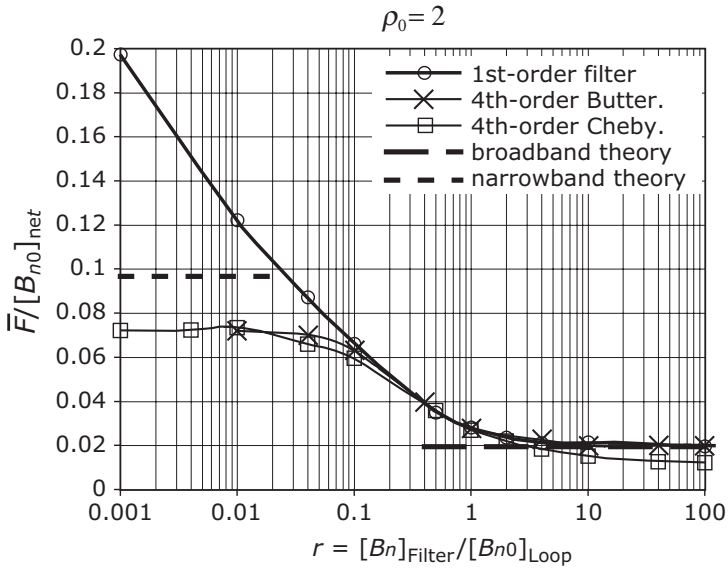


Fig. 20.5 $\bar{F}/[B_{n0}]_{\text{net}}$ versus r for rectangular filters and for a first-order (single-pole) filter at $[\rho_0]_{\text{net}} = 2$, first-order loop. [Scripts like NB1bw, NB1bwi, and NB1ch1 in Appendix i20.M.3.]

value used in Fig. 20.5, whereas Fig. 20.4 is for a S/N lower than any included in Fig. 20.1. Taking into consideration the ratio between the 3-dB and noise bandwidths of a first-order loop, Eq. (14.9), the data points on the 0-dB curve in Fig. 20.1 lie in the region $0.08 < r < 0.9$ in Fig. 20.5. For these reasons and because Fig. 20.1 is a log-log plot, the discrepancy at low r is not very apparent there.

With a first-order filter, $\bar{F}/[B_{n0}]_{\text{net}}$ continues to climb right through the value predicted for the rectangular filter at low r (although \bar{F} is still falling), but this is consistent with the infinite value of γ for the first-order filter.

The normalized skip rate $\bar{F}/[B_{n0}]_{\text{net}}$ does approach the theoretical wideband value, Eq. (20.10), as $r \Rightarrow \infty$, except that it goes below theory with the fourth-order Chebyshev filter. This discrepancy appears to be due to ripple, which causes a drop in noise near the center of the filter, the region influencing the PLL at high value of r . To show this, the filter was changed to a Butterworth (“maximally flat”) filter, whose response does not drop in the middle, and the theoretical value was then approached, as it is for the first-order filter. (Changing to an odd-order Chebyshev filter should also increase $\bar{F}/[B_{n0}]_{\text{net}}$ at high r).

Note that r does not have to be much greater than unity for the broadband result to be a close approximation when $\rho_0 = 2$ (Fig. 20.5), but this is not so for low S/N (e.g., 0.25 in Fig. 20.4).

Additional results and comparisons for various loops and filters at larger values of r , as well as details on the various bandwidths used in the bandlimited noise simulations, are available in Appendix i20.S.

Example 20.1 Frequency Discriminator Output S/N An FM transmission has a maximum frequency deviation of 80 kHz and a maximum modulation frequency of 20 kHz. The noise power density relative to the signal is 2×10^{-6} /Hz. If the minimum usable output S/N is 0 dB, what is the dynamic range (i.e., the ratio between 80-kHz deviation and the deviation that produces 0 dB S/N)? We will set the input bandwidth to just meet Eq. (15.33),

$$W = 2[\Delta f_{\text{MAX}} + F_m] = 2[8 \times 10^4 + 2 \times 10^4] \text{ Hz} = 2 \times 10^5 \text{ Hz}.$$

The S/N in the input bandwidth is then

$$\rho_{\text{in}} = \frac{P}{WN_0} = \frac{1}{(2 \times 10^5 \text{ Hz})(2 \times 10^{-6}/\text{Hz})} = 2.5 \Rightarrow 4.0 \text{ dB}.$$

Without clicks, the output S/N would be given by the numerator of Eq. (15.47),

$$\left[\frac{S}{N} \right]_{\text{linear}} = \frac{3}{2} \frac{P}{N_0} \frac{\Delta f^2}{F_m^3} = 1.5 \frac{1}{2 \times 10^{-6}/\text{Hz}} \frac{\Delta f^2}{(2 \times 10^4 \text{ Hz})^3},$$

which is unity at $\Delta f = 3266$ Hz. The corresponding dynamic range is

$$\Delta f_{\text{MAX}}/\Delta f = 80 \text{ kHz}/3.266 \text{ kHz} = 24.5 \Rightarrow 27.8 \text{ dB}.$$

With a standard discriminator, the denominator in Eq. (15.47) changes from 1 to

$$1 + \sqrt{3} \frac{W^2}{F_m^2} \rho_{\text{in}} \operatorname{erfc}(\rho_{\text{in}}) = 1 + \sqrt{3} \frac{(2 \times 10^5 \text{ Hz})^2}{(2 \times 10^4 \text{ Hz})^2} 2.5 \operatorname{erfc}(2.5) = 11.98,$$

requiring Δf to increase by the square root of that value in order to maintain 0 dB S/N . Thus the dynamic range reduces, by 20 dB $\log_{10}(\sqrt{11.98}) = 10.78$ dB, to 17.0 dB.

We will now employ a first-order loop to try to get the performance closer to the ideal.

We will set the loop bandwidth equal to that given by Eq. (15.34), giving unity gain at 10^5 Hz (so $W/f_L = 2$). Then we can use Eq. (20.5) to give the skip rate,

$$\frac{\bar{F}}{W} = \frac{1}{2\sqrt{3}} \operatorname{erfc} \left[\sqrt{\rho_{\text{in}}} \left(1 + \frac{0.3W}{f_L} \right) \right] = 10^{-4}.$$

This agrees with Fig. 20.1 and, while there are no data points at or below $2B_{L3}/B_{\text{RFIn}} = 1$, there is a point at about 1.3, which gives us some confidence that the theory applies, at least approximately, at this low level.

Substituting $\bar{F} = 10^{-4}W \approx 20$ Hz into Eq. (15.49), we obtain

$$\left[\frac{S}{N} \right]_1 = 8.15 \times 10^{-8} \Delta f^2$$

with this first-order loop, leading to $\Delta f = 3502$ Hz for 0-dB S/N , a dynamic range of 27.17 dB. This is only about 0.6 dB worse than without clicks and an improvement of more than 10 dB on the standard discriminator.

Based on Fig. i20.S.2b, we might expect some improvement using a second-order loop with $\alpha = 0, \zeta = 0.707$. We find, by simulating such a loop, one that has the same noise bandwidth as the first-order loop, that \bar{F} does drop, from 20 to about 13.7 Hz. However, the effect on dynamic range is small, an improvement of about 0.2 dB, because we are already close to the linear value; that is, the skipped cycles are having little effect.

20.1.4 Output Phase Variance

The variance of the output phase is plotted against the variance in the net, linear noise bandwidth for several bandwidth ratios r in Fig. 20.6. This is for a first-order loop but the bandlimited results are close to these for low-gain ($\alpha = 0$) and high-gain ($\alpha = 0.97$) second-order loops, $\zeta = 0.707$. (The spread is about 2% at $[\sigma_{\varphi 0}^2]_{\text{net}} = 0.25$ and 10% at $[\sigma_{\varphi 0}^2]_{\text{net}} = 2$. See Appendix i20.S.) The responses to band-limited noise appear to depend far less on the type of loop than do the wideband results that we have discussed. The lower output variance with bandlimited noise might result from the lower modulation frequencies that can be better tracked by the wider loop, thus reducing nonlinearity. Note that the narrowband results diverge from the broadband values at lower S/N (higher $[\sigma_{\varphi 0}^2]_{\text{net}}$), as was true for skip rate.

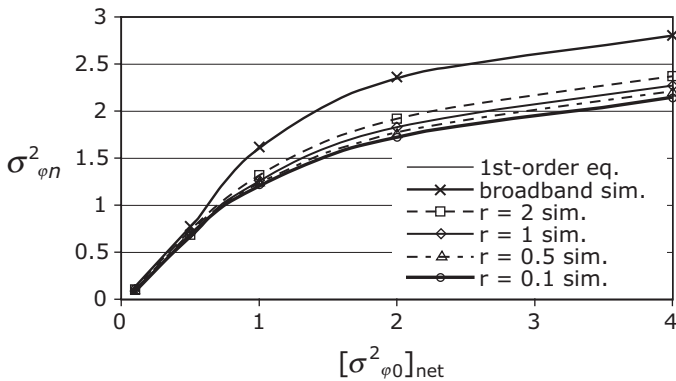


Fig. 20.6 Output phase variance versus variance in the net noise bandwidth for several ratios of input noise bandwidth (video equivalent) to loop noise bandwidth, first-order loop with first-order prefilter. [Script NB1, Appendix i20.M.2.]

20.1.5 Effect of a Limiter

When a limiter precedes the loop, only a phase variation is seen by the loop, so r_n becomes 1 in Eq. (20.2). As a result, there should be skipped cycles only if

$$\frac{\omega_L}{2\gamma} \geq 0.52. \tag{20.12}$$

The procedure that produced Eq. (20.3) now produces a skip rate of zero when Eq. (20.12) is not met or

$$\bar{F} = \frac{\gamma}{2\pi} \operatorname{erfc} \sqrt{\rho_{in}} \tag{20.13}$$

when it is. This is the same rate as with a standard discriminator [Eq. (15.40)]. Figure 20.7 shows this theoretical step function with measured data and simulation results.

The data are closer to the theoretical step function than it is without the limiter but the same value of \bar{F}/W is approached when the filter bandwidth becomes relatively narrow. Simulation results in Fig. 20.7 closely follow the measured data, although they would be even closer if the values of the RF input bandwidth for the simulation were decreased slightly or the value for the measured results were increased a bit. In both cases the noise bandwidth was estimated since the filters were both sharp.

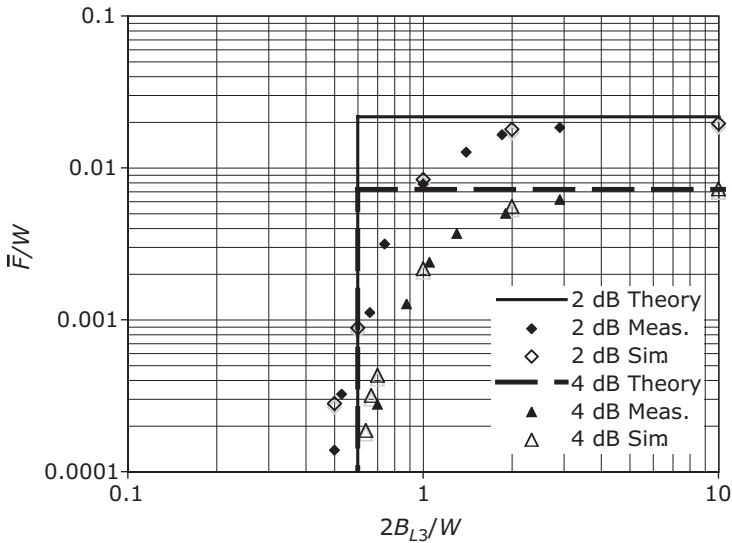


Fig. 20.7 Skip rate with a hard limiter following the input filter, theory, measured data, and simulation results for first-order loop. Curve parameter is S/N in the filter's RF noise bandwidth, W . [Script NB1bwi, Appendix i20.M.3.]

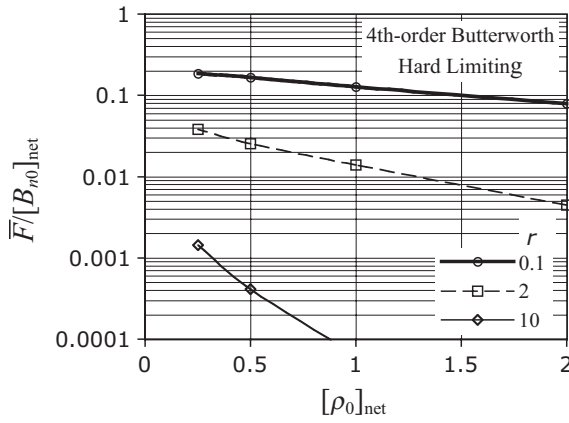


Fig. 20.8 Normalized skip rate with a limiter versus S/N in combined noise bandwidth for three ratios of filter to loop noise bandwidths (r), first-order loop. [Script NB1bwi, Appendix i20.M.3.]

Skip rates and phase variances from simulations with a limiter are shown in Figs. 20.8 and 20.9. These can be compared to similar plots without limiting in Figs. 20.3 and 20.6. Performance is similar for narrow bandwidth ratios but much improved with a limiter at high bandwidth ratios. This is because the noise in the limiter is much greater than it is in the loop bandwidth under that condition, so signal strength is significantly reduced, leading to narrower loop bandwidths.

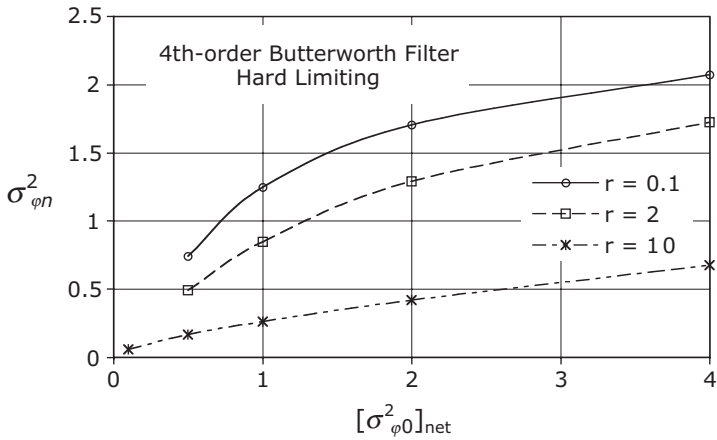


Fig. 20.9 Output phase variance with a limiter versus linear variance in combined noise bandwidth for three ratios of filter to loop noise bandwidths (r), first-order loop. [Script NB1bwi, Appendix i20.M.3.]

20.1.6 Comparison to AGC

Figure 20.10 shows loop output variance versus variance in the net linear bandwidth for a filter video bandwidth that is 10 times the loop noise bandwidth. Simulated results are with a limiter and with square-law AGC following the filter and without a gain-control device. These can be compared to (“theory”) curves based on linear-loop analysis that takes into account the decrease in gain (η) caused by the AGC or limiter. The ratio $r = 10$ is great enough to show these effects without masking them

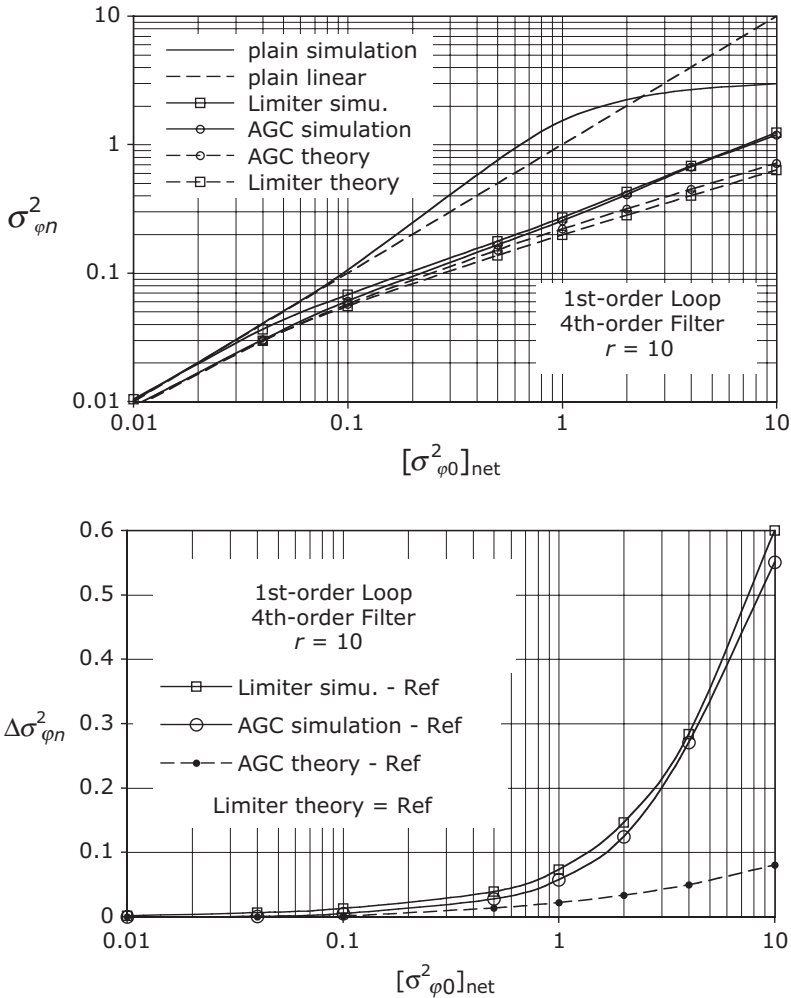


Fig. 20.10 Output variance versus variance in the net linear bandwidth with AGC and limiter and with no preconditioning, first-order loop. The results of simulation and of linear analysis that accounts for the effect of the preconditioning on the gain are shown at *a*. Differences between some close curves are plotted on a linear scale at *b*. [Script NB1bw, Appendix i20.M.3.]

by driving the loop too far into nonlinearity. Nevertheless, the simulations do show increases in $\sigma_{\varphi_n}^2$ relative to linear-loop theory at the higher variances. Also apparent is the slightly greater variance with limiting compared to the theory, which neglects any change in S/N through the limiter (Section 16.3.2). While the limiter theoretically produces a greater reduction in variance over most of the range than does the AGC, consistent with Fig. 16.10, the relative size of the reductions is actually reversed in the simulations.

20.1.7 Higher-Order Loops

Some phase variances and skip rates for higher-order loops are compared to those for a first-order loop in Appendix i20.S, and scripts described in Appendix i20.M are capable of producing simulations for higher-order loops with or without limiting. Because of the greater number of parameters that must be specified when a finite bandwidth is introduced, we must depend on simulations to a greater extent to reveal performance under particular conditions.

20.1.8 Summary, Symmetrical Narrowband Noise

- The click rate \bar{F} in a standard discriminator with a rectangular prefilter is given approximately by Eq. (15.40) for $\rho > 1$.
- The skip frequency \bar{F} of a first-order PLL, preceded by a rectangular filter, is given by Eq. (20.5) and Fig. 20.1.
- Figure 20.2 illustrates the necessity of using the two-source model in simulating narrowband additive noise.
- \bar{F} is shown over a wider range of parameters, based on simulations of a first-order loop preceded by a filter, in Figs. 20.3–20.5.
- The degradation of phase variance at high noise levels is reduced when the loop is preceded by an RF filter [Fig. 20.6].
- With a limiter between the rectangular filter and the first-order loop, \bar{F} is similar to \bar{F} without limiting at wider loop bandwidths but stays higher and then falls precipitously as the loop bandwidth narrows [Fig. 20.7].
- Output phase variance can be improved by either AGC or limiter [Fig. 20.10].
- The tendency, with broadband noise, for high-gain second-order loops to have higher output variances and skip rates than first-order loops and for loops with low-pass filters to be somewhat better is also true with narrowband noise [Appendix i20.S].

20.2 ECCENTRIC SIGNALS

In all our work in this chapter, the signal has been centered in the bandpass filter. While the filter may be tuned to center it on the signal, how are we to treat the more general case, where the signal is not centered?

20.2.1 Expected Performance

Once the bandpass filter is wide enough that the response is essentially the same as the response to broadband noise of the same density, then noise beyond that band is ineffective and further widening the band on one side of the signal or the other should have no effect.

If this is not the case, however, the loop will perform as if the signal were embedded in additive noise of bandwidth $\pm w/2$, where $w/2$ is the separation between the signal and the nearest noise-band edge (Fig. 20.11), with an additional perturbation due to the remaining noise. The remaining noise would be in an offset sideband, extending from an offset of $w/2$ to one of $W - w/2$ on one side of the signal. If the power in this sideband is small enough compared to the signal power, it can be decomposed into a pair of AM sidebands and a pair of FM sidebands, as indicated in Appendix i13.A [see also Egan, 2000, pp. 75–78; or Egan, 2003, pp. 220–222]. Each of the FM sidebands would contain one-quarter of the power of the original band and, since the loop responds primarily to FM (assuming phase quadrature), it could be treated as if driven by this additional FM noise. This would produce FM sidebands on the VCO at offsets equal to the offsets of the offset sidebands, $w/2$ to $W - w/2$. If the effects of both the bandlimited ($\pm w/2$) additive and FM ($f_m = W/2 \pm (W - w)/2$) noise are small enough for the loop to respond linearly to them, their effects can be superimposed.

For greater levels of power in the single sideband, we can judge the effect from studies of interference by a CW (single-frequency) signal at some frequency separation from the reference [Blanchard, 1974; Kliger and Olenberger, 1976; Olenberger, 1976]. As the sideband power level is raised, the FM sidebands continue to appear at the VCO output but at power levels that are not linearly proportional to the interfering power. A DC, or average, voltage will be introduced into the PD output and it must be canceled by a phase shift in the VCO signal. This should not be too surprising if we recall how such a voltage was developed during acquisition (Section 8.4) while the VCO was not at the same frequency as the reference. Eventually, as the power of

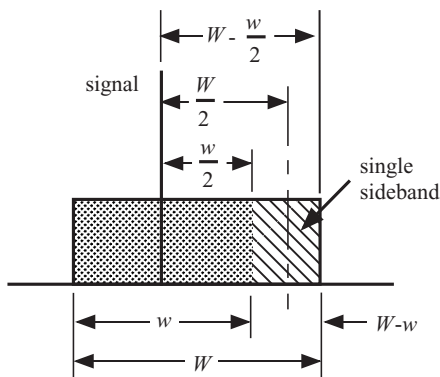


Fig. 20.11 Eccentric signal in a noise band.

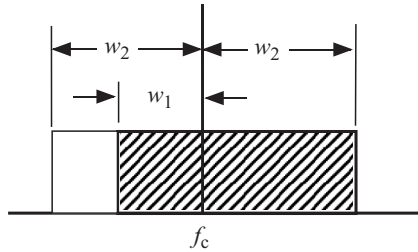


Fig. 20.12 Is noise in $\pm w_2$ worse?

the sidebands increases, the loop will tend to break lock so it can attempt to lock on the eccentric interference. Performance depends on the power and frequency separation of the interference and, in our case, also on the level and width of the symmetrically located (within w) noise. We might hope that no parameter, such as output phase variance or mean time to skip, would be made worse by *removal* of some noise on one side of the signal. That is, we would hope that a band of noise extending from $f_c - w_1$ to $f_c + w_2$ would be no worse than a band extending from $f_c - w_M$ to $f_c + w_M$, where w_M is the larger of w_1 and w_2 (Fig. 20.12). However, this may not be the case for large noise levels. Simulation or experimentation is required. We will look at a few results of simulation to get the flavor of the effects.

20.2.2 Simulating Eccentric Noise

If the noise is not centered on the signal, then the phase φ_{out} in Fig. 13.5 changes with time. The rate of change is the frequency difference between the signal, to which the VCO is locked, and the center of the noise band. This changing phase can be added in simulations (see Appendix i20.M.8). While it is difficult to find experimental data on offset noise, there are reported experiments on the effects of discrete interference offset in frequency from the signal. When we simulate a narrow band of offset noise, we obtain results that correlate well with those results (see Appendix i20.O).

Figure 20.13 defines some of the terms that describe the offset. It shows the variable o , which is the offset of the center of the noise band relative to the signal frequency in units of loop noise bandwidth. It shows the RF equivalent of the loop noise bandwidth B_{n0} , located relative to the input additive noise band. The area of overlap is shaded. Since $r = 2$ in the figure, when $o = -1$, the loop band of width $2B_{n0}$ is at one edge of the input noise band. It becomes centered when $o = 0$ and progresses to the other edge when $o = 1$. At $o = 2$, only half of the input noise band overlaps the RF equivalent of the loop bandwidth and, when o reaches 3, there is no overlap.

20.2.3 Some Simulation Results

We will concentrate here on the second-order loop and fourth-order Butterworth prefilters with equivalent video noise bandwidth twice that of the loop ($r = 2$, as in

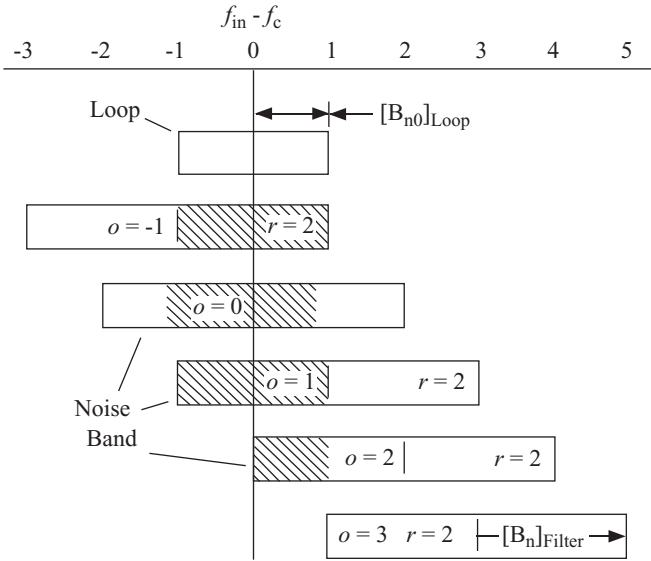


Fig. 20.13 Variables used to define offset.

Fig. 20.13) and unity S/N in the loop bandwidth ($\rho_{L0} = 1$). The effects we observe will worsen with poorer S/N .

20.2.3.1 Cycle Skipping. Figure 20.14 shows the effect of eccentricity on skip rate. The VCO frequency is pulled toward the noise frequencies. With negative noise offsets, the VCO frequency drops, producing a positive average frequency error at the PD output. Not surprisingly, the net (difference between phase increases and decreases) skip rate is affected even at the smallest noise offsets, whereas the effect of the eccentricity on the total skip rate is not obvious until the net rate becomes a significant part of the total.

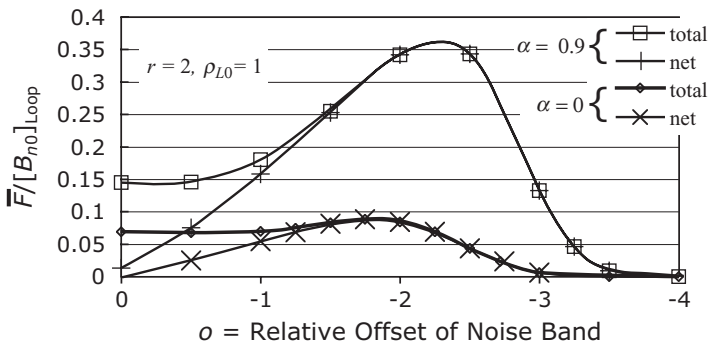


Fig. 20.14 Normalized skip rate versus eccentricity of the noise band.

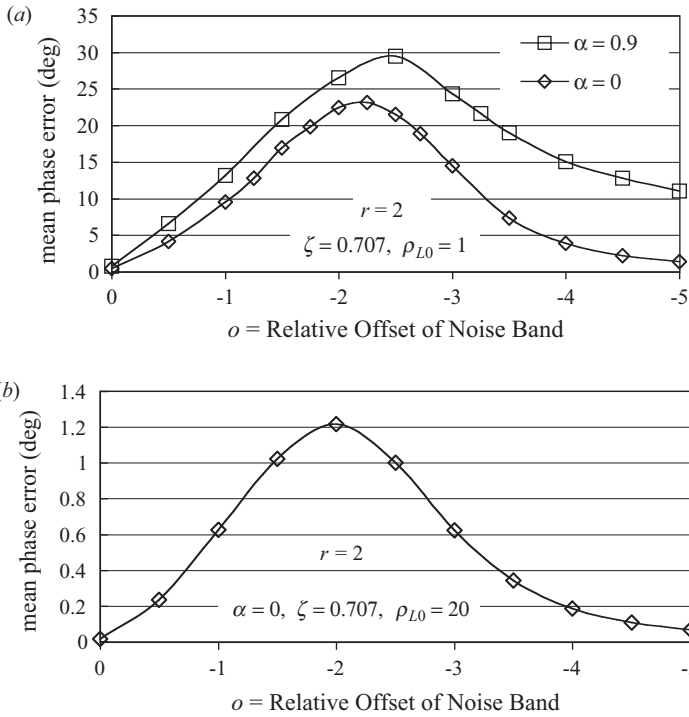


Fig. 20.15 Phase error in the loop caused by eccentricity of the noise band.

20.2.3.2 Phase Offset. Figure 20.15 shows the effect of noise eccentricity on average loop phase error. The phase error increases with eccentricity until $-o$ equals or exceeds 2, peaking at a point where the noise band only partially overlaps the loop's noise bandwidth. The phase error does not completely counter the voltage produced by the offset noise in Fig. 20.15a, as evidenced by the net cycle skipping shown in Fig. 20.14. Based on the values of K that correspond to the parameters of these curves, we can compute that the peak skip rates correspond to 8° and 8.5° of phase error, respectively for the lower and upper curves; that much additional phase error would apparently be required to reduce the net \bar{F} to zero. Figure 20.15b shows the mean phase error (for $\alpha = 0$) at 20 times higher S/N , where there is no skipping. The maximum phase error is smaller by about 20 times.

20.2.3.3 Output Variance. For $\alpha = 0.9$, the output variance (Fig. 20.16) peaks about where the mean phase error peaks. For $\alpha = 0$, the output variance falls as the noise offset increases (even for $\rho = 0.3$, the variance is highest at $o = 0$).

20.3 EXTENSION TO OTHER TYPES OF INTERFERENCE

In addition to the more obvious extensions of these narrowband simulations by combining with features already employed for wideband simulation (e.g., limiting, phase

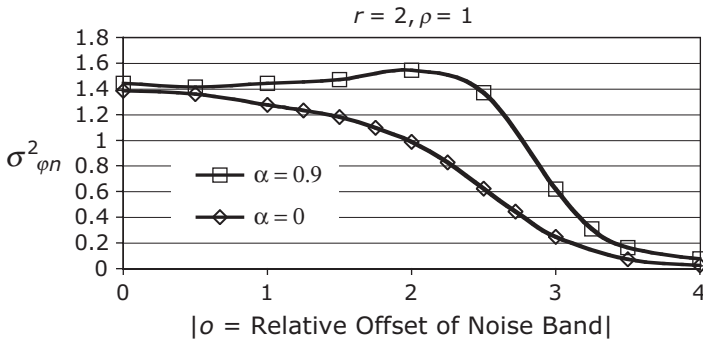


Fig. 20.16 Output phase variance versus eccentricity of the noise band.

offsets), it seems likely that discrete interferers and multiple noise bands could be incorporated.

The bandlimited noise (Fig. 20.17a) simulation might be extended by using multiple noise bands (Fig. 20.17b) of different widths and power levels. The eccentric noise band (Fig. 20.17c) might be extended to multiple eccentric noise bands (Fig. 20.17d) having different widths, power levels, and offsets, effectively removing restrictions on the shape of the simulated noise. (The detailed shapes of these noise bands would depend on the filters used.)

The offset noise band (Fig. 20.17e), a special case of the eccentric noise band, might be replaced with a discrete interferer (Fig. 20.17f). (See Appendix i20.O.)

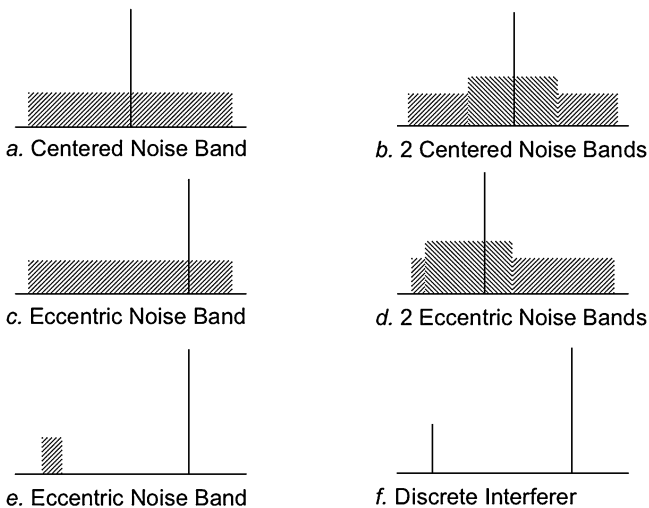


Fig. 20.17 Possible configurations to simulate. Rectangular noise bands are used for illustration.

i20.M APPENDIX: MATLAB® SCRIPTS

MATLAB scripts used to generate simulations for this chapter are described in this appendix, which is available from the Wiley Internet site.

i20.O APPENDIX: OFFSET INTERFERENCE DATA CORRELATION

This appendix shows correlation between simulations of offset noise and data from the literature relating to discrete offset interference. It is available from the Wiley Internet site.

i20.S APPENDIX BANDLIMITED SIMULATION DATA

This appendix contains additional data from simulations of bandlimited noise. It is also available from the Wiley Internet site at ftp://ftp.wiley.com/public/sci_tech_med/phase_lock.

CHAPTER 21

FURTHER INFORMATION

21.1 SOURCES FOR ADDITIONAL STUDIES IN PHASE LOCK

Blanchard's (1976) book ends with "Appendix B, Example of a Rough Draft of a Coherent Receiver," a 30-page description of the initial design of a receiver that includes several phase-locked loops. It illustrates the use of many of the concepts in his book and presents an excellent opportunity for review and for gaining a better feel for how these concepts might be used.

An established method for gaining more knowledge on any subject is to find related articles or papers and to follow the trail of references until one understands the bases. This can be daunting and can be made much more so by the tendency of different authors to use different terminology and different symbols for the same variable. It is much easier to start with a book that covers the area of interest, using common symbols and terminology throughout. Eventually, though, one is usually forced into the more difficult process as one delves into more specific or advanced topics, but fortified by what has already been learned.

Directions in which to pursue further study in particular areas are suggested by the sources referenced throughout this book. Entering both "phase" and "lock" as keywords for a title search in an online bookstore (e.g., www.amazon.com or www.search.barnesandnoble.com) will bring up many pertinent titles, often with descriptions and readers' comments. The IEEE Xplore[®] web site gives access to many papers for further study. However, the same two words, used in a search at www.ieeexplore.ieee.org, will match more than 5000 documents, so a more targeted search is advisable.

Here we will mention two useful books.

Gardner's (2005) book is the third edition of a classic, one of the first on the subject, and is a model of succinctness. The expanded coverage in the third edition includes extensive material on digital implementations and a short annotated bibliography.

Stensby (1997) goes into considerable depth in analyzing the nonlinear PLL, devoting half of the book to nonlinear analysis in the presence of noise, an area we have treated largely by simulation.

21.2 SOURCES COVERING PHASE-LOCKED FREQUENCY SYNTHESIS

Books that treat phase-locked loops as they apply to frequency synthesizers include those by Crawford (1994), Egan (1981, 2000), Kroupa (1973, 2003), Manassewitsch (1987), Rhode (1983, 1997), and Stirling (1987). While these have been listed alphabetically, they nevertheless tend to move, left to right, from the more to the less theoretically oriented; that is from those that explain more of the mathematical basis to those that concentrate more on circuit details. Both of Kroupa's books cover frequency synthesis and phase lock, but the earlier book emphasized synthesis more (it was the first book on frequency synthesis). Manassewitsch is the most encyclopedic, and Stirling specializes on microwave synthesizers.

Egan (2000) and Kroupa (2003) discuss sigma-delta techniques in frequency synthesis. Books by De Muer and Steyaert (2003) and Shu and Sánchez-Sinencio (2005) describe the implementation of these techniques in CMOS ICs.

21.A APPENDIX: MODULATIONS AND SPECTRUMS

This appendix illustrates the relationships between various representations of signals: continuous, amplitude-modulated, and phase-modulated. Specifically, formulas, waveform pictures, phasors, modulation spectrums, signal spectrums, power spectrums for modulation and signals, and single-sideband spectrums are shown. Table 21.A.1 covers discrete modulations, and Table 21.A.2 covers noise modulation.

TABLE 21.A.1 Discrete Modulation

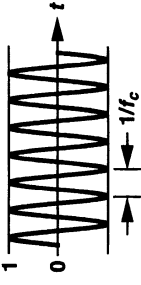
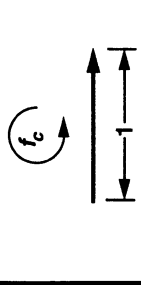
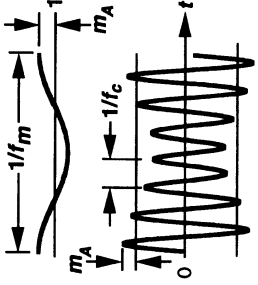
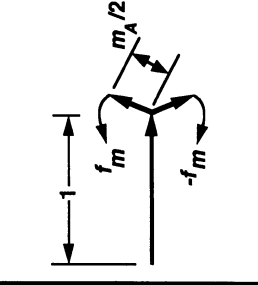
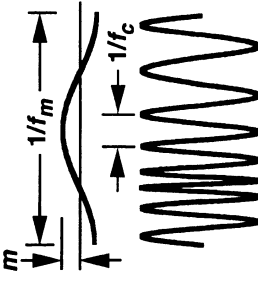
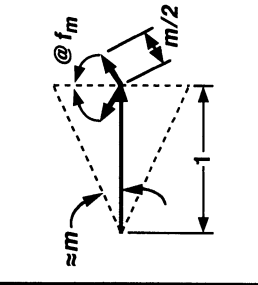
	1. Formula	2. Waveform	3. Phasor
CW	$g_c(t) = \cos \omega_c t$		
AM	$g_{m_A}(t) = 1 + m_A \cos \omega_m t$ $g_A(t) = g_{m_A}(t)g_c(t) = (1 + m_A \cos \omega_m t)\cos \omega_c t$ $= \cos \omega_c t + \frac{m_A}{2} \cos(\omega_c + \omega_m)t + \frac{m_A}{2} \cos(\omega_c - \omega_m)t$		
PM small m	$g_{m_P}(t) = m \sin \omega_m t$ $g_P(t) = \cos(\omega_c t + g_{m_P}(t)) = \cos(\omega_c t + m \sin \omega_m t)$ $= J_0(m)\cos \omega_c t$ $+ J_1(m)[\cos(\omega_c + \omega_m)t - \cos(\omega_c - \omega_m)t]$ $+ J_2(m)[\cos(\omega_c + 2\omega_m)t + \cos(\omega_c - 2\omega_m)t]$ $\rightarrow \cos \omega_c t + \frac{m}{2} [\cos(\omega_c + \omega_m)t - \cos(\omega_c - \omega_m)t]$ <p style="text-align: right;">as $m \rightarrow 0$</p>		

TABLE 21.A.1 Discrete Modulation (Continued)

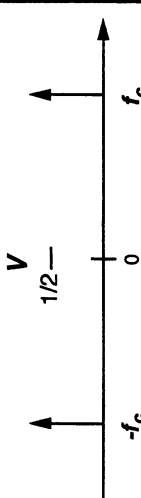
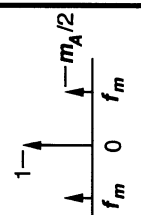
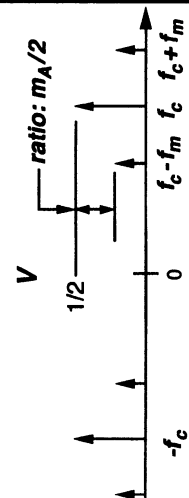
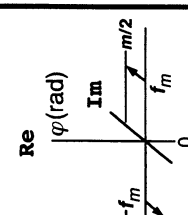
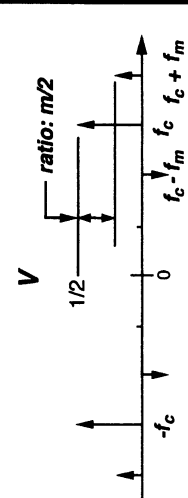
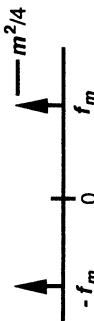
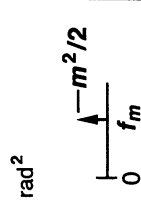
	4. Mod Spectrum	5. Signal Spectrum	6. Power Spectrum of Modulation (1Ω System)	
CW		 $\cos \omega_c t = \frac{e^{j\omega_c t} + e^{-j\omega_c t}}{2}$		
AM			2-sided	1-sided
PM small m				

TABLE 21.A.1 Discrete Modulation (Continued)

	7. Power Spectrum of Signal		8. SSB Spectrum
	Two-Sided	One-Sided	
CW			
AM			
PM small <i>m</i>			

TABLE 21.A.2 Noise Modulation

1. Formula		2. Waveform		6. Power Spectral Density of Modulation	
				Two-Sided	
				One-Sided	
AM	$g_a(t) = g_{ma}(t)g_c(t) \quad ((V, A))$ $\sigma_{ma}^2, \sigma_{ma} = 0$				
PM small <i>m</i>	$g_p(t) = \cos(\omega_c t + g_{mp}(t)) \quad ((V, A))$ $\sigma_{mp}^2, \sigma_{mp} = 0$				

7. Power Spectral Density of Signal		8. SSB Spectrum	
		One-Sided	
AM			
PM small <i>m</i> higher-order s.b.s not shown			

21.B APPENDIX: GETTING FILES FROM THE WILEY INTERNET SITE

To download the files listed in this book and other material associated with it, use an ftp program or an appropriate Web browser.

Ftp Access

If you are using an ftp program, type the following at your ftp prompt:

```
ftp://ftp.wiley.com
```

Some programs may provide the first part for you, in which case you might enter

```
ftp.wiley.com
```

Log in as anonymous (e.g., User ID: anonymous). Leave password blank. After you have connected to the Wiley ftp site, navigate through the directory path of

```
/public/sci_tech_med/phase_lock
```

Web Access

If you are using a standard Web browser, type URL address of:

```
ftp://ftp.wiley.com/public/sci_tech_med/phase_lock
```

If you are using Macintosh® OS X you can also enter this information in the “Connect to Server” dialog.

If you need further information about downloading the files, you can go to www.wiley.com/techsupport or call 317-572-3994.

REFERENCES

- Abromowitz, M., and Stegun, I. (1964). *Handbook of Mathematical Functions*. Washington, D.C.: U.S. Government Printing Office.
- Allan, D. W. (1966). Statistics of atomic frequency standards, *Proc. IEEE*, vol. 54, no. 2 (Feb.), pp. 221–230.
- Allan, D., Hellwig, H., Kartaschoff, P., Vanier, J., Vig, J., Winkler, G., and Yannoni, N. (1988). Standard terminology for fundamental frequency and time metrology, *Proceedings, 42nd Annual Frequency Control Symposium, 1988*, pp. 419–425.
- Ascheid, G., and Meyr, H. (1982). Cycle slips in phase-locked loops: A tutorial survey, *IEEE Trans. Commun.*, vol. COM-30, Oct., pp. 2228–2241. Also in Lindsey and Chie (1986), pp. 29–42.
- Blachman, N. M. (1966). *Noise and Its Effect on Communications*. New York: McGraw-Hill.
- Blanchard, A. (1974). Interference in phase-locked loops, *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-10, no. 5, Sept., pp. 686–697. Corrections are in vol. AES-11, March 1975, pp. 285–287.
- Blanchard, A. (1976). *Phase-Locked Loops*. New York: Wiley; also Malabar, FL: Krieger, 1992.
- Bois, G. P. (1961). *Table of Indefinite Integrals*. New York: Dover.
- Booton, R. C., Jr. (1952). Nonlinear Control Systems with Statistical Inputs, Massachusetts Institute of Technology, Cambridge, MA, Report No. 61; March 1 [referenced in Develet (1963)].
- Bracewell, R. (1965). *The Fourier Transform and Its Applications*. New York: McGraw-Hill.
- Buchanan, J. (1983). Chip capacitors and dielectric absorption, *Electron. Prod.*, May 12, pp. 107–109.
- Byrne, C. J. (1962). Properties and design of the phase controlled oscillator with a sawtooth comparator, *The Bell Syst. Tech. J.* (March), pp. 559–602.

- Cahn, C. R. (1962). Piecewise linear analysis of phase-locked loops. *IRE Trans. Space Electron. Telem.*, vol. SET-8, no. 1, March, pp. 8–13. Also in Lindsey and Simon (1978), pp. 8–13.
- Charles, F. J., and Lindsey, W. C. (1966). Some analytical and experimental phase-locked loop results for low signal-to-noise ratios, *Proc. IEEE*, vol. 54, Sept., pp. 1152–1166. Also in Lindsey and Simon (1978), pp. 276–290.
- Chen, Chi-Tsong (1970). *Introduction to Linear System Theory*. New York: Holt, Rinehart and Winston.
- Crawford, J. (1994). *Frequency Synthesizer Design Handbook*. Boston: Artech House.
- Davenport, W. B., Jr., and Root, W. L. (1965). *An Introduction to the Theory of Random Signals and Noise*. New York: McGraw-Hill.
- De Muer, B., and Steyaert, M. (2003). *CMOS Fractional-N Synthesizers*. Boston: Kluwer.
- Develet, J. A., Jr. (1963). An analytic approximation of phase-lock receiver threshold, *IEEE Trans. Space Electron, Telem.*, vol. SET-9, March, pp. 9–12.
- Dorf, R. C. (1965). *Time-Domain Analysis and Design of Control Systems*. Reading, MA: Addison-Wesley.
- Egan, W. F. (1981). *Frequency Synthesis by Phase Lock*. New York: Wiley; also Malabar, FL: Krieger, 1990.
- Egan, W. F. (1991). Sampling delay—is it real?, *RF Design*, Feb., pp. 114–116.
- Egan, W. F. (2000). *Frequency Synthesis by Phase Lock, 2nd Ed.* New York: Wiley.
- Egan, W. F. (2003). *Practical RF System Design*. New York: Wiley.
- Endres, T., and Kirkpatrick, J. (1992). Sensitivity of fast settling PLLs to differential loop filter component variations, *Proceedings of the 1992 IEEE Frequency Control Symposium*. New York: IEEE Press, pp. 213–223.
- Franco, S. (1989). Current-feedback amplifiers benefit high-speed designs, *EDN*, Jan. 5, pp. 161–172.
- Frazier, J. P., and Page, J. (1962). Phase-lock loop frequency acquisition study, *IRE Trans. Space Electron. Telem.*, vol. SET-8, Sept., pp. 210–227. Also in Lindsey and Simon (1978), pp. 132–149.
- Gardner, F. M. (1977). Hangup in phase-lock loops, *IEEE Trans. Commun.*, vol. COM-25, Oct., pp. 1210–1214. Also in Lindsey and Chie (1986), pp. 43–47.
- Gardner, F. M. (1979). *Phaselock Techniques*, 2nd ed. New York: Wiley.
- Gardner, F. M. (1980). Charge-pump phase-lock loops, *IEEE Trans. Commun.*, vol. COM-28, Nov., pp. 1849–1858. Also in Lindsey and Chie (1986), pp. 321–332 and in Razavi (1996), pp. 77–86.
- Gardner, F. M. (2005). *Phaselock Techniques*, 3rd ed. Hoboken, NJ: Wiley.
- Gill, W. (1981). Easy-to-use BASIC program, *EDN*, Feb. 18, pp. 133–136.
- Goldstein, A. J. (1962). Analysis of the phase-controlled loop with a sawtooth comparator, *Bell Syst. Tech. J.*, March, pp. 603–633.
- Gray, P., and Meyer, R. (1977). *Analysis and Design of Analog Integrated Circuits*. New York: Wiley.
- Greenstein, L. J. (1974). Phase-locked loop pull-in frequency, *IEEE Trans. Commun.*, Aug., pp. 1005–1013. Also in Lindsey and Simon (1978), pp. 150–158.
- Gupta, S. C., Bayless, J. W., and Hammels, D. R. (1968). Threshold investigation of phase-locked discriminators, *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-4, no. 6, Nov., pp. 855–863.

- Gupta, S. C., and Solem, R. J. (1965). Optimum filters for second- and third-order phase-locked loops by an error function criterion, *IEEE Trans. Space Electron. Teleme.*, vol. SET-11, June, pp. 54–62.
- Hess, D. T. (1968). Cycle slipping in a first-order phase-locked loop, *IEEE Trans. Commun. Tech.*, vol. COM-16, no. 2, April, pp. 255–260.
- Hewlett-Packard (1989). *Spectrum Analysis, Application Note 150*. Rohnert Park, CA: Hewlett-Packard Co., pp. 31–33.
- Holmes, J. K. (1971). First slip times versus static phase error offset for the first- and passive second-order phase-locked loop, *IEEE Trans. Commun. Technol.*, April, pp. 234–235. [Loops are 1st order and 2nd order with $\alpha \approx 1$ implied by his ignoring ω_p in the equation for ζ . His symbol \Rightarrow ours: $\omega_L \Rightarrow 2B_{n0}$; $r \Rightarrow (2\zeta)^2$; $C \Rightarrow$ sine of phase offset.]
- Jaffe, R., and Rechtin, E. (1955). Design and performance of phase-lock circuits capable of near-optimum performance over a wide range of input signal and noise levels, *IRE Trans. Inform. Theory*, vol. IT-1, March, pp. 66–76. Also in Lindsey and Simon (1978), pp. 20–30.
- Jeong, D., Borrello, G., Hodges, D., and Katz, R. (1987). Design of PLL-based clock generation circuits, *IEEE J. Solid-State Circuits*, vol. SC-22, April, pp. 255–261. Also in Razavi, B., Ed. (1996). *Monolithic Phase-Locked Loops and Clock Recovery Circuits*. New York: IEEE Press, pp. 326–332.
- Johnson, M., and Hudson, E. (1988). A variable delay line PLL for CPU-coprocessor synchronization, *IEEE J. Solid-State Circuits*, vol. SC-23, October, pp. 1218–1223. Also in Razavi, B., Ed. (1996). *Monolithic Phase-Locked Loops and Clock Recovery Circuits*. New York: IEEE Press, pp. 333–338.
- Klapper, J., and Frankle, J. (1972). *Phase-Locked and Frequency Feedback Systems*. New York: Wiley.
- Kliger, I. E., and Olenberger, C. F. (1976). Phase-lock loop jump phenomenon in the presence of two signals, *IEEE Trans. Aero. Electron. Syst.*, vol. AES-12, no. 1, Jan., pp. 55–63.
- Kroupa, V. (1973). *Frequency Synthesis*, New York: Wiley.
- Kroupa, V. (2003). *Phase Lock Loops and Frequency Synthesis*. Chichester, England: Wiley.
- Kuo, B. C. (1987). *Automatic Control Systems*, 5th ed. Englewood Cliffs, NJ: Prentice-Hall.
- Lee, T. H. (1992). A 155-MHz clock recovery delay- and phase-locked loop, *IEEE J. Solid-State Circuits*, vol. SC-27, Dec., pp. 1736–1746. Also in Razavi, B., Ed. (1996). *Monolithic Phase-Locked Loops and Clock Recovery Circuits*. New York: IEEE Press, pp. 421–430.
- Lindsey, W. (1972). *Synchronization Systems in Communication and Control*. Englewood Cliffs, N.J.: Prentice-Hall.
- Lindsey, W., and Chie, C. (1981). A survey of digital phase-locked loops, *Proc. IEEE*, vol. 69, April, pp. 410–431. Also in Lindsey and Chie (1986), pp. 296–317.
- Lindsey, W., and Chie, C., Eds. (1986). *Phase-Locked Loops*. New York: IEEE Press.
- Lindsey, W. C., and Simon, M. K. (1973). *Telecommunication Systems Engineering*. Englewood Cliffs, NJ: Prentice-Hall.
- Lindsey, W. C., and Simon, M. K., Eds. (1978). *Phase-Locked Loops and Their Applications*. New York: IEEE Press.
- Little, A. (1990). Using current feedback amplifiers, *RF Design*, Dec. 1990, pp. 47–52.
- Manassewitsch, V. (1987). *Frequency Synthesizers Theory and Design*, 3rd ed. New York: Wiley.

- Olenberger, C. F. (1976). Effects of automatic gain control on phase-locked loop behavior in the presence of interference, *IEEE Trans. Aero. Electron. Systems*, vol. AES-12, no. 1, Jan., pp. 803–805.
- Pease, R. A. (1982). Understand capacitor seakage to optimize analog system, *EDN*, Oct. 13, pp. 125–129.
- Pouzet, A. (1972). Characteristics of phase detectors in presence of noise, *Proceedings of the 1972 International Telemetry Conference*, Los Angeles, CA, pp. 818–828.
- Protonotarios, E. N. (1969). Pull-in performance of a piecewise linear phase-locked loop. *IEEE Trans. Aerosp. and Electron. Syst.*, vol. AES-5, no. 3, May, pp. 376–386.
- Razavi, B., Ed. (1996). *Monolithic Phase-Locked Loops and Clock Recovery Circuits*. New York: IEEE Press.
- Rey, T. J. (1960). Automatic phase control, theory and design, *Proceedings of the IRE*, Oct., pp. 1760–1771. Corrections are in *Proc. IRE*, March 1961, p. 590. Also in Lindsey, W. C., and Simon, M. K., *Phase-Locked Loops & Their Application*. New York: IEEE Press, 1977, pp. 309–320.
- Rhode, U. (1983). *Digital PLL Frequency Synthesizers*. Englewood Cliffs, NJ: Prentice Hall.
- Rhode, U. (1997). *Microwave and Wireless Synthesizers, Theory and Design*. New York: Wiley.
- Rice, S. O. (1963). Noise in FM receivers, in *Proceedings Symposium of Time Series Analysis*. New York: Wiley, Chapter 25, pp. 395–424.
- Richman, D. (1954a). Color-carrier reference phase synch. . . . *Proc. IRE*, Jan., p. 125. Also in Lindsey and Simon (1978), p. 401.
- Richman, D. (1954b). The DC quadricorrelator: A two-mode synchronization system, *Proc. IRE*, January, pp. 288–299.
- Rowbotham, J. R., and Sanneman, R. W. (1967). Random characteristics of the type II phase-locked loop, *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-3, no. 4, July, pp. 604–612.
- Sanneman, R. W., and Rowbotham, J. R. (1964). Unlock characteristics of phase-locked loops, *IEEE Trans. Aerosp. Navig. Electron.*, vol. ANE-11, March, pp. 15–24. Also in Lindsey and Simon (1978), pp. 250–259.
- Schwartz, M., Bennett, W., and Stein, S. (1966). *Communication Systems and Techniques*. New York: McGraw-Hill.
- Shoaf, J. H., Halford, D., and Risley, A. S. (1973). *Frequency Stability Specifications and Measurement: High Frequency and Microwave Signals*, NBS Technical Note 632, Jan.
- Shu, K., and Sánchez-Sinencio, E. (2005). *CMOS PLL Synthesizers: Analysis and Design*. New York: Springer.
- Smith, B. M. (1966). The phase-lock Loop with filter: Frequency of skipping cycles, *Proc. IEEE*, Feb., p. 296.
- Spilker, J. J., Jr. (1963). Delay-lock tracking of binary signals, *IEEE Trans. Space Electron. Telemetry*, vol. SET-9, March, pp. 1–8. Also in Lindsey and Simon (1978), pp. 226–233.
- Spilker, James J., Jr. (1977). *Digital Communications by Satellite*. Englewood Cliffs, NJ: Prentice-Hall, pp. 443–445, 537–596.
- Spilker, J. J., Jr., and Magill, D. T. (1961). The delay-lock discriminator—An optimum tracking device, *Proc. IRE*, vol. 49, Sept. pp. 1403–1416. Also in Lindsey and Simon (1978), pp. 234–247.

- Springett, J. C., and Simon, M. K. (1971). An analysis of the phase coherent–incoherent output of the bandpass limiter, *IEEE Trans. Commun. Technol.*, vol. COM-19. no. 1, February, pp. 42–49.
- Stensby, J. L. (1997). *Phase-Locked Loops, Theory and Applications*. Boca Raton, FL: CRC Press.
- Stirling, R. (1987). *Microwave Frequency Synthesizers*. Englewood Cliffs, NJ: Prentice-Hall.
- Tausworthe, R. C. (1966). *Theory and Practical Design of Phase-Locked Receivers*, J.P.L. Technical Report No. 32–819, Feb., as quoted in Blanchard (1976).
- Thomas, G., Jr., and Finney, R. (1984). *Calculus and Analytic Geometry*, 6th ed. Reading, MA: Addison-Wesley, pp. 483–489.
- Troha, D. G. (1994). Texas Instruments Application Note sdla005a, *Digital Phase-Locked Loop Design Using SN54/74LS297*.
- Truxal, J. G. (1955). *Control System Synthesis*. New York: McGraw-Hill.
- Viterbi, A. (1959). *Acquisition and Tracking Behavior of Phase-Locked Loops*. JPL External Publication 673, July 14, Pasadena, CA: Jet Propulsion Laboratory.
- Viterbi, A. (1963). Phase-locked loop dynamics in the presence of noise by Fokker–Planck technique, *Proc. IEEE*, vol. 51, Dec., pp. 1737–1753. Also in Lindsey and Simon (1978), pp. 90–106.
- Viterbi, A. (1966). *Principles of Coherent Communication*. New York: McGraw-Hill.
- Wolaver, D. H. (1991). *Phase-Locked Loop Circuit Design*. Englewood Cliffs, NJ: Prentice-Hall, p. 75.
- Young, I. A., Greason, J. K., and Wong, K. L. (1992). A PLL clock generator with 5 to 110 MHz of lock range for microprocessors, *IEEE Journal of Solid-State Circuits*, SC-27, pp. 1599–1607 (November). Also in Razavi, B., Ed., *Monolithic Phase-Locked Loops and Clock Recovery Circuits*, New York: IEEE Press, pp. 339–346.
- Ziemer, R. E., and Peterson, R. L. (1985). *Digital Communications and Spread Spectrum Systems*. New York: Macmillan, pp. 278–280.

INDEX

Italic page numbers lead to topics that are available only online.

- 1minHgen, 121
- Acq2, 167
- Acqpz, 167
- Acquisition of lock, 137
 - aids, 171
 - frequency discriminator, 177
 - logic, 179
 - parameter modification, 176
 - phase-frequency detector, 179
 - sweeping. *See* Sweep, sweeping
 - analysis, 149, 153
 - bias, example, 152
 - changing loop parameters during, 174
 - experiments using MATLAB, 167
 - false lock, 156
 - first-order loop, 141, 144
 - noise, in the presence of, 377
 - range, 155, 157
 - defined, 138
 - sine PD, 146
 - second-order loop
 - example, 148
 - formulas, 146
 - summary table, 167
 - sine PD, 146
 - spreadsheets, 167
 - time, 147, 158, 160
 - by phase plane, 164
 - using MATLAB, 167
 - various PDs, formulas, 168
 - waveforms during, 140
- Added pole, effect
 - on modulation response, 134
 - on step response, 108
- Additional studies, 415
- ADPLL, 204, 229
 - by pulse addition and removal, 212, 228
 - ripple control, 228
 - stability, 228
 - choice of values, example, 209
 - example of first-order, 209
 - higher order loops, 210
 - hold-in range, 214
 - sampling, effects on stability, 206
 - second-order, 228
 - simulation, 229
 - SN74LS297, 212
- AGC, 319
 - coherent, 174
 - effect on optimized loop, 331
 - loop output variance with, 406
 - suppression factor, 325
 - example, 331
 - with square law detection, 323

- All digital PLL. *See* ADPLL
- alpha (loop parameter), xxiii, 88
 - initial conditions specified
 - linear response unaffected, 107, 109
 - nonlinear response affected, 167
- AM sidebands, 273
- Amplitude detector, 322
- Analog to the PLL, electrical, 23
 - equivalent circuit
 - first-order loop, 18
 - second-order loop, 102
- Analysis
 - classical, 16
 - Laplace, 21
 - state space, 103
 - Z-transform, 208
- Answers to problems, xx
- Applications, xix, xxiii, 189
- Assumption re nature of the noise, 337
- Autocorrelation, 279
- Automatic gain control, 319. *See also* AGC
- Automatic sweep circuit, 186
 - ... *See also* Sweep, sweeping
- Band-limited noise, 395
- Bandwidth
 - noise, 287
 - unity gain with high ζ , 67
- Bd2ordT1, 86
- Bessel functions, modified, 333
- Bit synchronizer, 197
- Blanchard, Alain, xxi
- Block diagram, mathematical, 21, 24
- B_n , 288
- Bode plot(s), 24, 72
 - determining noise response from, 252
 - drawing, 24, 81
 - example, 83
 - in MATLAB, 86
 - tangential approximation, errors in, 85
- Broadband noise, 280
- CAE, xxii
- Capacitor in loop filter, 56
 - reducing the size of, 54
- Carrier recovery, 195
 - BPSK, 195
 - loop, 314
 - noise in, 314
 - M -phase signals, 196
 - QPSK, 196
- Carrier, tracking. *See* Tracking carrier
- Carson's rule, 304, 307
- Changing loop parameters to lock, 174
- ChsPlot, 86
- Click noise, 310, 395
 - with a PLL, 313, 395
- Clock recovery, symbol, 227
- Clock synchronization, 202
- Clock timing control, 200
- Closed-loop
 - equations, 22, 60
 - first-order loop, 22
 - second-order loop, 60
 - third-order loop, 217
 - sweeping. *See* Sweep, sweeping
- Coherent
 - AGC, 174
 - detector, 171, 323
 - output in noise, 385
- Complementary error function, 317
- Component noise, example, 260
- Correlation between noise samples, 279
- Costas loop, 195, 315
 - bi-phase, 195
 - N phases, 196, 227
- Cycle skipping, 335, 347
 - due to noise, 335
 - eccentric signal in noise, 410
 - effect of α , 353
 - first-order loop, 347
 - frequency, band-limited noise, 395
 - limiter, effect on, 404
 - mean frequency, 350
 - mean time, 347
 - narrow-band noise, 407
 - observing using NLPn, 167
 - probability for a given time, 350
 - second-order loop
 - $0 < \alpha < 1$, 350
 - $\alpha = 0$, 348
 - $\alpha = 1$, 348
 - with mistuning, 353
 - experimental results, 356

- Data synchronization, 197, 203
- Data timing control, 200
- dB, 25, 251
- dBc/Hz, 243
- DC gain in loop filter, 49, 52
- DDS, 211
- Decibels, 25, 251
- Decision filter, 181
- Decomposition of single sideband, 286
- Delay, representing in stability analyses, 80
- Delay-and-multiply synchronizer, 200
- Delay-locked loop, 202
- Demodulator, loop as, 113, 128, 297
 - ... *See also* Frequency demodulator
 - ... *See also* Phase, demodulation
- Densities, integrating, 249
- Desensitization, 256
 - example, 256
- Detector(s)
 - amplitude, 322
 - coherent, 171, 323, 385
 - linear, 322
 - phase-frequency, 179
 - square law, 322
 - types of, 322
 - with sweep
 - full wave, 182
 - half wave, 182
 - quadrature, 183
- Dielectric absorption in filter capacitors, 57
- Diffusion of error probability, 338
- Digital PLL (ADPLL), 204
- Direct digital synthesizer, 211
- Distortion of demodulated FM signal, 299
- Division, frequency, 247
- DLL, 200
- Early-late gate bit synchronizer, 197
- Eccentric signal in noise, 407
 - effect on phase error, 411
 - simulating, 409
- Electrical analog to the PLL, 23, 102
- Equivalent circuit
 - first-order loop, 23
 - second-order loop, 102
- erfc, 317
- Error probability, diffusion of, 338
- Error response(s), 121
 - higher-order loops, approximate, 118
 - modulation, 118
- Error, 2π , xxiii
- ErTrn..., 109
- ErTrn1a, 109
- Ev2ordT1, 86
- Evans (root locus) plot, 72, 76
 - in MATLAB, 86
- Examples
 - 1.1 Preservation of Units, 7
 - 1.2 Combined Gain, 10
 - 1.3 Hold-In Range, 13
 - 2.1 Transient Response, 19
 - 2.2 Modulation Response, 20
 - 2.3 Modulation Response, 24
 - 3.1 Effect of Current Source Z, 50
 - 5.1 Bode plot(s), 83
 - 6.1 Step Response, 91
 - 6.2 Envelope of Step Response, 94
 - 6.3 Ramp Response, 96
 - 6.4 Response to a Parabolic Input, 100
 - 7.1 Modulation in a Simple Loop, 116
 - 7.2 Use of Table 7.1, 119
 - 7.3 Modulation Response Curves, 126
 - 7.4 Use of Related Modulation Response Curves, 127
 - 7.5 Frequency Demodulator, 129
 - 8.1 Acquisition formulas, 148
 - 8.2 Pull-In Bias, 152
 - 9.1 Sweeping to Acquire, 184
 - 10.1 First-order ADPLL . . . , 209
 - 10.A.1 Synthesizer Frequency Error, Third-Order Loop, 222
 - 10.A.2 Synthesizer Phase Error, Third-Order Loop, 224
 - 11.1 Integrating Noise Density, 239
 - 11.2 Spectral Display of Noise SBs, 244
 - 11.3 Spectral Display of Phase Noise, 245
 - 12.1 Loop Response to Reference Phase Noise, Standard Plots, 252
 - 12.2 Loop Response to Reference Phase Noise, Bode Plot, 253
 - 12.3 Loop Response to VCO Phase Noise, 255
 - 12.4 Desensitization, 256
 - 12.5 Optimum Loop Bandwidth, 258
 - 12.6 Component Noise, 260
 - 12.7 Calibration ... Sine Character, 264

Examples (*Continued*)

- 12.8 Calibration Using Modulation, 265
 - 12.9 Measurement With a PD, 268
 - 13.1 PPSD from ... Noise PSD, 274
 - 13.2 u_1 Due to Additive Noise, 280
 - 13.3 PD Characteristic Δ by Noise, 285
 - 14.1 Loop Response... Add. Noise, 289
 - 15.1 S/N in a Freq. Discriminator, 308
 - 16.1 AGC Suppression Factor, 331
 - 17.1 Phase Error with Large Noise, Exact Solution, 340
 - 17.2 Phase Error in Second-Order Loop, Large Noise, 346
 - 18.1 Effective gain with noise, 361
 - 18.2 Effect of Modulation on Phase Offset, 365
 - 18.3 Use of Suppression Factors, 368
 - 18.4 Effect of Modulation on Phase Noise, 372
 - 19.1 Sweep Acquisition in Noise, 387
 - 19.2 False Stop, 389
 - 19.3 False stop and restart, 392
 - 20.1 Frequency Discriminator S/N , 402
- Excel, 5, 167
- spreadsheets, 167
 - 1minHgen, 121
 - Acq2, 167
 - Acqppz, 167
 - IntFNs, 250
 - IntPDens, 249
 - IntPhNs, 249
- Extending modulation frequency range, 131
- Extensions of theory, 189
- ExtrPole, 109
- False alarm, probability, 388
- False lock, 156, 229
- False restart, probability, 389
- False stop and restart, example, 392
- False stop versus false restart, 390
- False stop, probability, example, 389
- Files, getting from internet site, 421
- Filter, loop
 - active, 51
 - capacitors in, 54, 56
 - DC gain in, 49, 52
 - driven by current source, xix, 45
 - equations, form of, 56
 - higher order, 57
 - impractical, 55
 - integrator, 47
 - integrator-and-lead, 48, 66
 - lag, 48, 65
 - lag-lead, 49, 62
 - low-pass, 48, 65
 - modification during acquisition, 175
 - op-amps in, 57
 - passive, 45
 - reference voltage, 55
 - stability, 52
 - transimpedance, DC, 49
 - unintended poles, 53
- Filtering effect of PLL, 3
- Final value, 87
 - acquisition by sweeping, 383
 - approaching, 139
 - special 3rd-order loop, 219
- floor, oscillator noise, 274
- FM
 - improvement factor, 304, 306
 - power spectrum of signal, 242
 - sidebands, 273
 - threshold, 314
- Fokker-Plank
 - equation
 - first-order loop, 337
 - second-order loop, 339
 - method, 336
- Fractional- N , xix, 228
- FreqRsp, 135
- Frequency
 - demodulator, discriminator, 4, 128
 - acquisition aid, 177
 - bandwidth set by a filter, 301
 - example, 129
 - first-order loop, 305
 - loop linearity, 307
 - phase error, 307
 - S/N
 - example, 308
 - narrow-band noise, example, 402
 - summary table, 308
 - second-order loop, 306
 - output at u_1 , 306
 - standard, 299
 - with noise, 310
 - traditional, 299
 - division, 190, 247

- mixing, 192
- modulation, 4, 111, 130
- multiplication, 247
- power spectral density, 235
- ramp input
 - optimized loop, 294
- response. *See* Modulation response
- step input
 - optimized loop, 294
- synthesis, xix, 4, 190
 - fractional- N , 228
 - references, 416
 - sigma-delta, 228
- variance, 237
- Frequency Synthesis by Phase Lock*, xxi
- FRSPDorI, 136
- gain
 - combined, 10
 - example, 10
 - effect of noise on, example, 361
 - generalized VCO, 190
 - loop filter, 10
 - margin, 73
 - open-loop, computing, 81
 - phase detector, 8
 - suppression, 361
 - variation
 - effect on output spectrum, 361
 - effects on second-order loop, 329
- VCO, 10
- Generalized VCO, 189
- Glossary, xxv
- Hangup, 166
- Heterodyning, 192. *See also* Mixer, mixing
- Hgeneric, 121
- Higher-order loops, 189
 - approximate response
 - error, 118
 - output, 117
 - estimating response, 60
 - in ADPLL, 228
 - MATLAB, responses using, 228
 - modulation response, 252
 - stability analysis, 77
- High-frequency poles, 152
- Hold-in range, 11
 - ADPLL, 214
 - defined, 138
 - example, 13
 - first-order loop, 146
 - second-order loop, 168
 - sine PD, 146
- IF filter, 156
- Improvement factor, FM, 304
- Infinite variances, 240
- Initial value of \mathbf{X} from \mathbf{C}^{-1} , 107
- Injection points for signals, various, 117
- Integrating noise density
 - example, 239
 - formulas, 238
- Integrator, 47
- Integrator-and-lead filter, 48, 66
- Internet site, getting files from, 421
- IntFNs, 250
- IntPDens, 249
- IntPhNs, 249
- L , 243
- \mathcal{L} , 247
- Limiter, limiting, 320, 325
 - by high signal level, 321
 - effect on cycle skipping, 404
 - effect on optimized loop, 331
 - loop output variance with, 406
- Limits on the noise spectrum, 240
- Linear detector, 322
- Lock indicator, 171
- Lock-in range. *See also* seize range
 - defined, 137
- Locking in noise, 280
- Log of error, step response
 - in 2nd-order loop, 92
 - in special 3rd-order loop, 222
- Logic
 - acquisition aiding, 179
 - gates as limiters, 321
- Long loop, 193
- Long-term (steady-state) response, 102
- Loop stability, 71
- Loop filter, 44. *See* Filter, loop
- Loop(s)
 - combined PLL and DLL, 204
 - delay-locked, 200, 202
 - higher order, xxiii, 4, 189
 - ... *See also* Higher order loops
 - long, 193

- Loop(s) (*Continued*)
 - nonlinear operation in locked, 358
 - order, 59
 - parameters
 - changing to lock, 174
 - measuring, 134
 - response, 59, 217
 - equations, summary, 67
 - first-order loop, 113
 - second-order loop, 116
 - special third-order, 215, 217
 - to additive noise, 287
 - example, 289
 - to reference phase noise
 - example, 252, 253
 - to VCO phase noise, example, 255
 - stability, 71
 - third-order, special case, 189
 - type, 59
- Magnitude of transfer functions, 120, 226
- Margin, gain and phase, 73
- Mathematical block diagram, 21, 24
- MATLAB[®], xxii, 4, 86, 109, 134, 228, 357, 393, 413
 - Bode plots, 86
 - Evans (root locus) plots, 86
 - function: *residue*, 103
 - higher-order responses using, 228
 - modulation response using, 134
 - nonlinear simulation, 167
 - using scripts, 109, 167
 - with noise, 357
 - scripts (programs)
 - Bd2ordT1, 86
 - ChsPlot, 86
 - erdis, 357
 - ErTrn..., 109
 - ErTrn1a, 109
 - Ev2ordT1, 86, 109
 - ExtrPole, 109
 - FreqRsp, 135
 - FRspDorI, 136
 - LMatPhPn, 109
 - LMPHPhHz, 109
 - NB1, 403
 - NB1bw, 375, 397, 399, 400, 401, 406
 - NB1bwi, 397, 399, 400, 401, 404, 405
 - NB1chi, 397, 399, 400, 401
 - NB2bw, 375
 - NB2bwi, 375
 - NLPhN, 357
 - NLPhP, 167
 - NLPhx, 167
 - NLSaw, 167
 - NqLagLd, 86
 - NStat1, 343
 - NStat2, 344, 345, 346, 351, 352
 - Nyquist, 86
 - Open2c4, 228
 - Open2c4x, 228
 - Open2cls, 228
 - Ord3spcl, 228
 - OutTrn..., 109
 - root locus. *See* Evans *above*
 - Sig1i, 370
 - SigOff2i, 351, 352, 354, 355, 370, 371, 374
 - Skip1, 347, 348
 - Skip2 (i), 349, 351, 352, 354
 - SSaddedP, 109
 - SSstep, 109
 - step, 109
 - swpi, 393
 - Transnts, 109
 - simulation, xix
 - of ADPLL, 228
 - of nonlinear PLL, 167
 - in noise, 357
 - Maximum sweep rate
 - ... *See* Sweep, sweeping, speed
 - Mean frequency of cycle skip, 350
 - Mean square value, 233
 - Mean time to skip cycle, 347
 - Mean value of the cosine, 365
 - Measurement
 - of loop parameters, 134
 - of phase noise, 261
 - ... *See* also Phase, noise, measuring
 - Minimization under constraint, 296
 - Mistuning, 138
 - effect on cycle skipping, 353
 - Mixer, mixing, 33
 - analog multiplier, 39
 - doubly balanced, 38
 - frequency in generalized VCO, 193
 - IC, 40

- reciprocal, 256
- simple, 34
- singly balanced, 35
- Modified Bessel functions, 333
- Modulation
 - discrete, table of, 417
 - in a first-order loop, 115
 - index, 112
 - noise. *See* Noise modulation
 - phase. *See* Phase, modulation(s)
 - processing by PLL, 236
 - response, 111
 - higher-order loops, 252
 - in a first-order loop, 19, 113
 - example, 20, 24, 116
 - in a second-order loop, 116
 - in a special 3rd-order loop, 224
 - plots, $\alpha = 0, 1$
 - $\angle H(\Omega), \zeta \leq 1, 124$
 - $\angle H(\Omega), \zeta \geq 1, 126$
 - $|1 - H(\Omega)|, \zeta \leq 1, 123$
 - $|1 - H(\Omega)|, \zeta \geq 1, 125$
 - $|1 - H(\Omega)|/\Omega, \zeta \leq 1, 128$
 - $|H(\Omega)|, \zeta \leq 1, 123$
 - $|H(\Omega)|, \zeta \geq 1, 125$
 - $|H(\Omega)|\Omega, \zeta \leq 1, 127$
 - plots, various α
 - $\angle H(\Omega)$ & $\angle 1 - H(\Omega), \zeta = 1, 122$
 - $|H(\Omega)|$ & $|1 - H(\Omega)|, \zeta = 1, 122$
 - using curves, example, 126, 127
 - using MATLAB, 134
 - small, 245
- Modulations and spectrums, 416
- Modulator, loop as, 113
- Modulator, phase or frequency
 - extended modulation frequency range, 131
 - loop as, 128
- MPSK, 369
- Multiple loops, 260
- Multiple noise bands, 412
- Multiplication, frequency, 247
- Multiplicative modulation, 275
- Multiplier, analog, in phase detection, 39
- Multiply and divide, 196

- Natural frequency, 61
 - on Bode plot, 63
- NCO, 211

- NLPhN, 357
- NLPhP, 167
- NLPhx, 167
- NLSaw, 167
- Noise, 280
 - additive, 271
 - decomposition of, 271
 - experimental results, 335, 341, 345, 351, 356
 - large, 335
 - summary of effects, 356
 - loop response to, example, 346
 - no lock to, 280
 - PPSD from, 274
 - example, 274
 - representation
 - as AM, 275
 - as multiplicative modulation, 275
 - as PM, 273
 - on quadrature carriers, 275
 - restrictions on, 277
 - simulation, 342
 - tracking signals in, 282
 - assumptions regarding, 337
 - at the phase detector output, 276
 - band(s), 407
 - general shape, 412
 - signal centered, 395
 - multiple, 412
 - signal not centered, 407
 - band-limited, 395
 - bandwidth, 287
 - carrier recovery loop, 314
 - first-order loop, 288
 - net, 398
 - RF input, 398
 - second-order loop, 290
 - video equivalent, 396
 - broadband, 280
 - component, example, 260
 - correlation between samples, 279
 - eccentric
 - simulating, 409
 - formulas for integrating curves, 238
 - frequency synthesizer response, 192
 - in standard discriminator, 310
 - injected at various points, 260
 - large amounts, xix
 - models, restrictions on, 277

Noise (*Continued*)

- modulation
 - processing by the PLL, 236
 - table of, 420
 - narrowband, xx
 - narrowband symmetrical, summary, 407
 - nonlinear loop operation due to, 358
 - parameter variation due to, 319
 - phase modulation by, 233
 - power from
 - frequency demodulator, 303
 - first-order loop, 305
 - second-order loop, 307
 - phase demodulator, 299
 - power spectral density, 235, 272
 - resistor, 260
 - simulation, 342
 - source, single or dual, 280
 - u_1 due to, example, 280
- Nonlinear operation in a locked loop, 358
- Nonlinear simulation, 166
- with noise, 342
- `NqLagLd`, 86
- Numerically controlled oscillator, 211
- Nyquist
- diagram, 72, 74
 - plots in MATLAB, 86
- OA, 211
- Observing the open-loop response, 71
- Offset
- component, effect on pull-in, 164
 - phase, effect of modulation, example, 365
- Op-amp, 51
- considerations, 52
 - current feedback, 57
 - differential response, 57
 - in loop filters, 57
 - non-inverting input, 57
 - voltage feedback, 57
- `Open2c4`, 228
- `Open2c4x`, 228
- `Open2c1s`, 228
- Open-loop
- equations, 62
 - in terms of closed-loop param., 103
 - gain, computing, 81
 - phase, computing, 81
 - response, observing, 71

- transfer function
 - integrator-and-lead filter, 66
 - lag filter, 65
 - lag-lead filter, 62
- Operating range, 11
- Optimization in the Presence of Noise, 291
- simplified formula, 296
- Optimized loop
- effects of AGC or limiter on, 331
 - frequency ramp, 294
 - frequency step, 294
 - phase step, 293
- Optimum loop bandwidth, 258
- example, 258
 - with input and VCO noise, 257
- Order of loop, 59
- Oscillator, 41
- ring, 44
 - spectrums, typical, 238
- Other responses, 100
- Out of lock, 3
- Output
- accumulator, 205, 211
 - responses, 117, 121
 - spectrum, effect of gain change due to noise, 361
- `OutTrn...`, 109
- Parabolic input, response to
- first-order loop, 27
 - plots. *See* Transient, response, plots
 - second-order loop, 98
 - example, 100
- Parameter(s)
- in graphs labels, 365
 - modifications, 176
 - nominal design, 365
 - variation due to noise, 319
- Phase
- definitions, 6
 - demodulation, 4, 130, 297
 - distortion, 299
 - with linear PD characteristic, 300
 - with sinusoidal PD characteristic, 299
 - detector, 8, 29
 - characteristic change by noise, 283
 - example, 285
 - charge pump, 31
 - exclusive-OR, 30

- flip-flop, 29
- noise response of various types, 281
- output u_1 , 360
- phase-frequency, 179
- sinusoidal, 33
 - mixer. *See* Mixer
- detector, 29
- error, peak, in frequency demod., 307
- frequency detector, 179
- locked loop
 - basic, 15
 - higher order, 189
 - second-order, 59
 - special third-order, 189
 - with noise, 231, 413
 - without noise, 3, 229
- margin, 73
 - versus damping factor, 86
- modulation, 4, 111, 131
 - by noise, 233
 - effect on carrier tracking, 368
 - nonlinearity due to, 363, 368
 - offset, effect on, example, 365
 - sidebands, 273
 - symbols for, 358
 - with VCO noise, 369
- noise
 - effect of modulation, example, 372
 - gain, effect on, example, 361
 - harmful in radio receivers, 255
 - in multiple loops, 260
 - measuring, 261
 - calibration, 263
 - modulation, 265
 - example, 265
 - with a sinusoidal characteristic, 264
 - example, 264
 - using freq. discriminator, 269
 - using phase detector, 263, 268
 - example, 268
 - using receiver, 270
 - using spectrum analyzer, 270
 - optimum loop bandwidth for, 257
 - example, 258
 - processing of
 - reference, 251
 - VCO, 254
 - example, 255
 - superposition of, 256
 - offset
 - due to eccentric signal in noise, 411
 - effect of suppression on, 372
 - open-loop, computing, 81
 - plane, 161
 - using MATLAB, 109, 167
 - power spectral density, 235
 - effective input, 274, 364
 - from additive noise, 273
 - example, 274
 - other representations, 248
 - superposition of, 256
 - two-sided, 248
 - probability with noise
 - ... *See also* Fokker-Plank
 - diffusion of, 350
 - from MATLAB script, 357
 - second-order loop
 - example, 340, 346
 - experimental results, 341
 - simulations, 357
 - theory, Fokker-Plank method, 336
 - reversals, 26
 - statistics, gathering, 357
 - step, optimized loop for, 293
 - variance, 237
 - loop output, band-limited noise, 403, 407
 - loop output, eccentric noise, 411
 - loop with integrator & lead filter, 365
 - spreadsheet to compute, 249
 - with additive noise, 365
 - with suppression, 366
 - $\alpha = 0$, 366
 - $\alpha \approx 1$, 367
 - first-order loop, 366
- Phase of transfer functions, 120, 226
- PM. *See* Phase, modulation
- power spectral density, 234, 235
- Power spectrum
 - of FM signal, 242
 - shape, 246
 - "Powers", 233
- PPSD, 235
 - ... *See* Phase power spectral density

- Preface
 - to the first edition, xxi
 - to the second edition, xix
- Probability
 - density, 233
 - of false restart, 389
 - of false stop, 388
 - example, 389
 - of false stop and restart, example, 392
 - of phase versus phase error, 339
 - that a cycle will be skipped, 350
- Problems, xx
 - answers to, xx
- Processing of reference phase noise, 251
- Pseudorandom bit sequence, synch to, 198
- Pull-in. *See also* Acquisition
 - range defined, 138
- Pull-out range, 164
- Pulse addition and removal, 212

- QPSK, 301
- quadratorrelator, 177
- Quasi-linearization method, 361
 - basics, 363
- Ramp, response to
 - first-order loop, 26
 - plots. *See* Transient, response, plots
 - second-order loop, 96
 - example, 96
 - special third-order loop, 221
- Range
 - acquisition
 - ... *See* Acquisition of lock, range
 - hold-in. *See* Hold-in range
 - operating, 11
 - pull-out, 164
 - synchronization. *See* Hold-in range
- Reciprocal mixing, 256
- Recovery, carrier, 195
 - ... *See also* Carrier recovery
- References, 423
- residue, 103
- Resistor noise, 260
- Response
 - approximate, 60
 - modulation. *See* Modulation, response
 - open-loop, observing, 71
 - parabolic. *See* Parabolic input, response
 - ramp. *See* Ramp, response to
 - second-order loop, summary, 68
 - transient. *See* Transient response
- Restrictions on the noise models, 277
- Ripple control in ADPLL, 228
- Root locus, 76

- S/N. *See* Signal-to-noise ratio
 - ... *See also* Phase, variance
- Sampling, effect on stability
 - and nonlinear simulation, 166
 - in ADPLL, 206
 - in frequency synthesizer, 191
- Saturation, component, 165
- Script L, 247
- Scripts
 - MATLAB. *See* MATLAB, scripts
- Seize range
 - defined, 137
 - second-order loop, 168
 - formulas, 146, 148
 - summary table, 168
 - sine PD, 146
- separatrix, 162, 166
- S_f , 239
- Sigma-delta, xix, 228
- Signal-to-noise ratio, 298
 - “in the loop”, 290
 - ... *See also* Phase, variance
 - frequency discriminator
 - example, 308
 - table, 308
 - in the loop bandwidth, 290
 - limiter, 325, 327
- Simulation, nonlinear, 166, 228
 - approximate method, 167
 - sampling and, 166
 - with noise, 357
- Single sideband
 - decomposition of, 286
- Single-sideband
 - density, 243
 - power spectral density, 243
- SN74LS297, 212, 228
- Special-case third-order loop, 215
- Spectral density
 - frequency power, 235
 - noise power, 272

- phase power, 235
- power, 235
- Spectral display
 - of noise sidebands, example, 244
 - of phase noise, example, 245
- Spectral shape of power spectrum, 246
- Spectrum, 240
- Spectrum(s)
 - analyzer, 262
 - measuring noise with, 244
 - and modulations, 416
 - FM signal, 242
 - for small m , 242
 - loop output, change due to noise, 361
 - noise, limits on, 240
 - oscillator, typical, 238
- Spreadsheet. *See* Excel, spreadsheets
- Square-law detection, 323
- Square-law detector, 322
- Squaring loss, 315
- SSaddedP, 109
- SSstep, 109
- Stability
 - analysis methods, 72
 - Bode plot(s). *See* Bode plot
 - delay, representing, 80
 - effect of sampling on, 191, 206
 - first-order loop, 77
 - in ADPLL, 206, 228
 - MATLAB, using to analyze, 86
 - measures of, 73
 - Nyquist plot. *See* Nyquist
 - of frequency synthesizer, 191
 - of the loop, 71
 - over wide gain ranges, 79
 - root-locus. *See* Evans (root locus) plot
 - second-order loop, 77
 - third-order loop, 77
 - conditionally stable, 86
 - various PLL configurations, 76
- Standard discriminator with noise, 310
- State
 - of the system, 104
 - space
 - analysis, 103
 - approximate solution, 108
 - initial conditions, 107
 - MATLAB, using, 109
 - method, xxiii
 - output, 106
 - vector equation, 105
- step, 109
- Step response, 87
 - effect of added pole, 108
 - envelope, 92
 - example, 91
 - first-order loop, 18, 26
 - example, 19
 - log of error, 2nd-order loop, 94
 - second-order loop, 89
 - example, 91
 - special 3rd order loop, 220
- Superposition of phase noise, 256
- Suppression factor, 322
 - AGC, 325
 - plot, 328
 - due to
 - additive noise, 364, 365
 - modulation, 363
 - biphase, 368
 - Gaussian, 369
 - sinusoidal, 369
 - phase noise, 363
 - PM plus additive noise, 369
 - effect on phase variance, 366
 - limiter, 327
 - plot, 328
 - phase offset, 364
 - summary, 375
 - total, 364
 - types that affect effective noise, 364
 - use of, example, 368
- Sweep, sweeping, 180
 - circuit, 185
 - automatic, 186
 - switched current, 185
 - closed-loop, 180
 - combined techniques, 185
 - comparison, closed-, open-loop, 183, 187
 - in noise
 - acquisition, example, 387
 - with coherent detector
 - acquisition probability, 386
 - false restart, 389
 - false stop, 388
 - false stop and restart, example, 392

- Sweep, sweeping (*Continued*)
 - false stop versus false restart, 390
 - output reduction, 385
- open-loop, 181
- speed, xx
 - closed loop, 180
 - with noise, 377
 - determining success, 382
 - maximum, 377
 - mistuning, effect of, 380
 - optimum parameters, 383
 - open vs. closed loop, 187
 - open-loop, 183
- stop probability
 - closed loop, 181
 - in noise, 377
 - with coherent detector, 386
 - open loop, 183
 - to acquire lock, example, 184
- $S_y(f)$, 235
- Symbol clock recovery, 227
- Symbols, list of, xx
- Synchronization range, 11
 - ... *See also* Hold-in range
 - defined, 138
- Synchronization, synchronizer
 - data, 197
 - delay-and-multiply, 200
 - early-late gate, 197
 - to pseudorandom sequence, 198
- Synthesizer, frequency, 190
- S_φ , 236

- Table 2.1 Summary of First-Order Loop Transient Responses, 27
- Table 4.1 Summary of Second-Order Loop Equations, 68
- Table 7.1 Transfer Functions Between Loop Variables, 119
- Table 8.A.1 Acquisition Formulas, 168
- Table 9.1 Lock Probability vs. Sweep Speed, 181
- Table 15.1 Frequency Discriminator Characteristics, 309
- Table 21.A.1 Discrete Modulations, 417
- Table 21.A.2 Noise Modulations, 420
- telecommunications, 4
- Third-order loop, special case, 189, 215–27
- Time
 - pull-in, second-order loop, 147, 158–61
 - by phase plane, 164
 - transient, first-order loop, 143
- Timing
 - adjustment with PLL, ix
 - control, clock and data, 200
- Tracking
 - carrier
 - with modulation and noise, 369
 - with phase modulation, 368
 - signals in additive noise, 282
- Transfer function
 - ADPLL, 213, 228
 - first-order loop, 23
 - second-order loop, 59, 116
 - special third-order loop, 217
 - magnitude and phase, 226
- Transient response
 - first-order loop, 17
 - example, 19
 - frequency synthesizer, 191
 - MATLAB, using, 109
 - parabola. *See* Parabolic input, response
 - plots for second-order loops
 - error, parabola input
 - $\alpha = 1$ vs. $\omega_n t$, 99
 - $\zeta = 0.5, 2$ vs. $\omega_n t$, 100
 - error, ramp input
 - $\zeta > 1, \alpha = 0$, vs. $\omega_n t$, 97
 - $\zeta > 1, \alpha = 1$, vs. $\omega_n t$, 98
 - $\zeta \leq 1, \alpha = 0, 1$, vs. $\omega_n t$, 97
 - error, step input
 - log, $\zeta > 1$ vs. $\omega_n t$, 95
 - log, $\zeta \leq 1, \alpha = 0, 1$, vs. $\omega_n t$, 95
 - $\zeta < 1, \alpha = 0, 1$, vs. $\omega_n t$, 93
 - $\zeta \geq 1, \alpha = 0, 1$, vs. $\omega'_L t$, 94
 - $\zeta \geq 1, \alpha = 0, 1$, vs. $\omega_n t$, 92
 - output, step input
 - $\zeta = 1$, vs. $\omega_n t$, 90
 - units, note on, 101
 - ramp. *See* Ramp, response to
 - second-order loop, 87
 - special third-order loop
 - final values, 219
 - ramp, 221
 - step, 220
 - synthesizer, 215
 - example, 222, 224
 - step. *See* Step response

- Transient time, first-order loop, 143
- transimpedance, 45
- Transients, 109
- Transportation lag, 80
- Triple roots, special 3rd order loop, 220
- Two-sided phase PSD, 248
- Type (loop), 59
- Typical oscillator spectrums, 238
- Units
 - and the Laplace variable s , 13
 - preservation of, example, 7
 - radian vs cycles, 303
 - rule for handling, 13
- Unity-Gain Bandwidth with high ζ , 67
- Values in terms of α , ζ , and ω_n , 167
- Variance, 233
 - frequency, 237
 - phase. *See* Phase, variance
 - spreadsheet to integrate power, 249
- VCO, 3, 41
 - generalized, 189
 - phase noise, processing of, 254
 - example, 255
- Video filter, 299, 304
- Weiner filter, 296
- wireless, 4
- Z-transform, 208