3103

2342458

TR 3104

High-Performance Frequency-Demodulation Systems

High-Performance Frequency-Demodulation Systems

BIBLIOTHEEK TECHNISCHE UNIVERSITEIT
Prom [illegible] 1
DELFT

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof.ir. K.F. Wakker,
in het openbaar te verdedigen ten overstaan van een commissie,
door het College voor Promoties aangewezen,
op maandag 23 maart 1998 te 10:30 uur

door

Michael Hendrikus Laurentius KOUWENHOVEN

elektrotechnisch ingenieur
geboren te Delft.

Dit proefschrift is goedgekeurd door de promotor:
Prof.dr.ir. A.H.M. van Roermund

Samenstelling promotiecommissie:

Rector Magnificus, voorzitter
Prof.dr.ir. A.H.M. van Roermund, Technische Universiteit Delft, promotor
Dr.ir. C.J.M. Verhoeven, Technische Universiteit Delft, toegevoegd promotor
Prof.dr.ir. J. Davidse, Technische Universiteit Delft
Prof.dr.ir. P.M. Dewilde, Technische Universiteit Delft
Prof.dr. R. Prasad, Technische Universiteit Delft
Prof.ir. K.H.J. Robers, Technische Universiteit Delft
Prof.ir. A.J.M. van Tuijl, Universiteit Twente

Published and distributed by:

Delft University Press
Mekelweg 4
2628 CD Delft
The Netherlands
Telephone: +31 15 2783254
Fax: +31 15 2781661
E-mail: DUP@DUP.TUDelft.NL

ISBN 90-407-1641-2 / CIP
NUGI 832

Copyright © 1998 by M.H.L. Kouwenhoven

All rights reserved.
No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without permission from the publisher:
Delft University Press.

Printed in the Netherlands

Contents

List of Symbols

Symbol	Meaning
t	time
$s(t)$	noise-free demodulator input FM carrier wave
$\vec{s}$	phasor corresponding to $s(t)$
$A(t)$	FM carrier amplitude
ω_o	carrier frequency
$\Phi(t)$	instantaneous phase of $s(t)$
$\varphi(t)$	message component of the instantaneous FM carrier phase
$s_q(t)$	imaginary component of $\vec{s}$: wave in quadrature with $s(t)$
$s_i(t)$	real component of $\vec{s}$: equal to $s(t)$
$\Delta\omega$	RMS frequency deviation in (rad/s)
$\Delta\omega_{\max}$	maximum frequency deviation in (rad/s)
$m(t)$	message signal at the FM modulator input
W	bandwidth of the FM message signal in (rad/s)
ω	spectral frequency in (rad/s)
W_{FM}	FM transmission bandwidth in (rad/s)
$S_s(\omega)$	power spectral density of FM carrier $s(t)$
$S_{\dot{\varphi}}(\omega)$	power spectral density of the FM message
$n(t)$	demodulator input noise
$\vec{n}$	phasor corresponding to $n(t)$
$n_i(t)$	real component of the low-pass equivalent demodulator input noise
$n_q(t)$	imaginary component of the low-pass equivalent demodulator input noise
$n_{s,i}(t)$	low-pass equivalent demodulator input noise component in-phase with $s(t)$
$n_{s,q}(t)$	low-pass equivalent demodulator input noise component in quadrature with $s(t)$
$R_n(\tau)$	autocorrelation function of $n_i(t)$ and $n_q()t$
$S_n(\omega)$	power spectral density of $n_i(t)$ and $n_q(t)$

$S_{n,s}(\omega)$	power spectral density of $n_{s,i}(t)$ and $n_{s,q}(t)$
W_n	bandwidth of $n(t)$, $n_i(t)$ and $n_q(t)$
N_o	value of $S_n(\omega)$ at $\omega = 0$
σ_n	standard deviation of $n(t)$, $n_i(t)$, and $n_q(t)$
$r(t)$	noisy demodulator input FM carrier wave
$\vec{r}$	phasor corresponding to $r(t)$
$R(t)$	noisy amplitude of $r(t)$
$\Theta(t)$	instantaneous phase of $r(t)$
$\theta(t)$	phase noise contained in $r(t)$
$S_{\dot{\theta}}(\omega)$	frequency noise power spectral density
p	demodulator input carrier-to-noise ratio
$K_{\text{FM-AM}}$	FM-AM conversion gain
$K_{\text{FM-PM}}$	FM-PM conversion gain
$K_{\text{PM-AM}}$	PM-AM conversion gain
K_{AM}	AM demodulator conversion gain
K_{PM}	PM demodulator conversion gain
K_{dem}	FM demodulator conversion gain
$v(t)$	velocity corresponding to $s(t)$
$\vec{v}$	phasor corresponding to $v(t)$
$v_{\text{rad}}(t)$	radial velocity component
$v_{\text{tan}}(t)$	angular velocity component
$s_o(t)$	FM-AM, FM-PM, PM-AM converter, or oscillator output wave
$\vec{s}_o$	phasor corresponding to $s_o(t)$
$A_o(t)$	amplitude of $s_o(t)$
$\varphi_o(t)$	message component of the instantaneous phase of $s_o(t)$
$s_{o,i}(t)$	real component of $\vec{s}_o$
$s_{o,q}(t)$	imaginary component of $\vec{s}_o$
$s_r(t)$	reference wave used during demodulation
$\vec{s}_r$	phasor corresponding to $s_r(t)$
$s_{r,i}(t)$	real component of $\vec{s}_r$
$s_{r,q}(t)$	imaginary component of $\vec{s}_r$
τ_d	fixed time-delay
τ	variable time-delay
ω_{offs}	frequency offset component
z	complex frequency of a the FM-AM converter zero
Q	quality factor of a filter
$\Gamma_{\text{IF}}(\mathrm{j}\omega)$	low-pass equivalent RF/IF filter transfer
$u(\cdot)$	Heaviside's step-function
$G(\cdot)$	amplitude compressor transfer
$C_{n,1}(A)$	inverse first-order compression factor
$C_{n,2}(A)$	inverse second-order compression factor
$y_{\text{dem}}(t)$	FM demodulator output signal

$y_{\mathrm{dem},1}(t)$	first-order FM demodulator output noise
$y_{\mathrm{dem},2}(t)$	second-order FM demodulator output noise
N_+	rate of positive (counter-clockwise) clicks/cycle-slips
N_-	rate of negative (clockwise) clicks/cycle-slips
ξ	average click pulse area
$S_{\dot{\theta}_{\mathrm{click}}}(\omega)$	click noise power spectral density
$\mathrm{SNR}_{\mathrm{out}}$	FM receiver output signal-to-noise ratio
K	inverse soft-limiter gain (width of linear region)
x	inverse limiter overdrive factor
$G_{\mathrm{sl}}(\cdot)$	soft-limiter amplitude compression transfer
λ	$\frac{1}{2}C_{n,1}(A)$
μ	$\frac{1}{2}C_{n,2}(A)$
r	radius of gyration of the input noise $n(t)$
ρ_0	radius of gyration of the first-order demodulator output frequency noise in (Hz)
ρ_2	radius of gyration of the second-order demodulator output frequency noise in (Hz)
$B_{N,0}$	noise bandwidth of the first-order demodulator output amplitude noise in (Hz)
$B_{N,1}$	noise bandwidth of the baseband filter
$B_{N,2}$	noise bandwidth of the second-order demodulator output amplitude noise in (Hz)
$B_{N,3}$	noise bandwidth of the second-order noise×noise component
$B_{N,\mathrm{IF}}$	noise bandwidth of the IF filter in (Hz)
$C_{n,1,\mathrm{sl}}(A)$	inverse first-order compression factor of a soft-limiter
$C_{\mathrm{cr},1}$	critical compression level
$S_{n,n}(\omega)$	power spectral density of the second-order noise×noise component
$S_{n^2}(\omega)$	power spectral density of the second-order amplitude noise
$H_b(\mathrm{j}\omega)$	baseband filter transfer
$\dot{\theta}_{\mathrm{click}}(t)$	click noise component of the demodulator output frequency noise
R_n	stochastic noise amplitude
φ_n	stochastic noise phase
$\dot{R}_n$	stochastic noise amplitude rate
$\dot{\varphi}_n$	stochastic angular noise frequency
$H_{\mathrm{lf}}(s)$	loop-filter
$H_{\mathrm{pl}}(s)$	post-loop filter
φ_e	phase error
$g\left(\varphi_e\right)$	phase detector nonlinearity

K_d	phase detector gain
K_o	oscilator tuning constant
$n'(t)$	low-pass equivalent noise in phase feedback demodulator loop
$\mathcal{J}_0$	probability current density of φ_e
$\mathcal{J}_{0,\text{drift}}$	drift-component of the probability current density of φ_e
$\mathcal{J}_{0,\text{diff}}$	diffusion component of the probability current density of φ_e
$U_0\left(\varphi_e\right)$	potential function of φ_e
$h_0\left(\varphi_e\right)$	restoring force acting on φ_e
$\varphi_{e,\text{ss}}$	steady-state phase error
B_L	double-sided closed-loop noise bandwidth in (Hz)
W_L	double-sided closed-loop noise bandwidth in (rad/s)
α	SNR inside the closed-loop bandwidth
β	$\alpha\times$ steady state phase error
$\varphi_l(t)$	message component of the instantaneous oscillator output phase in frequency feedback demodulators
$\theta_l(t)$	phase noise contained in the oscillator output in frequency feedback demodulators
Γ_c	carrier suppression factor
Γ_{id}	ideal threshold extension factor

Chapter 1

Introduction

Modulation and demodulation are inseparably related to the notion of communication. Communication is the unifying term that refers to the collection of actions required for the transport of information. Communication systems are the mechanizations of these actions. The current epoch shows an unprecedented infiltration of all kinds of such communication systems into daily life. From postal mail, wired and wireless telephones, radio and television, to relatively new means as Internet, electronic mail, video conferencing, and innumerable others, they all provide us the means to communicate, i.e. to *transmit* and/or *receive* information.

Despite its virtually unbounded diversity, communication of any kind essentially consists of two actions; transmission and reception of information. The corresponding mechanizations of these actions are generally referred to by the terms "transmitters" and "receivers".

The appearance of information, as produced by an information source, is generally rather unsuited for direct transmission, but instead requires some appropriate preprocessing. This is were *modulation* comes into play. For example, direct radio transmission of a voice signal, without any preprocessing other than recording by a microphone and subsequent amplification, would require an antenna of several hundreds of kilometers long [1]! Another inconvenience of this way of transmission is, that it does not allow simultaneous transmission of several signals, inside the entire range of the transmitting antenna, without severe interference between these signals. Since currently, daily life experience shows the possibility of simultaneous reception of various radio broadcasting stations by people that are (almost) the same place, using very handsome radio receivers, there must be some means to overcome the aforementioned inconveniences. This is exactly what modulation provides; it adapts the appearance of the information, such that its transport allows efficient use of the communication channel, i.e. the 'medium' that connects the transmitter and the receiver.

Modulation is the action that maps information on a so called *carrier wave*, which transports the information through the channel from the transmitter to the receiver, by systematic alteration of one or more of the wave characteristics, according to the original information signal. *Demodulation* is the inverse of the modulation action, that is performed inside the receiver in order to recover the original information signal. The corresponding systems are called "modulators" and "demodulators". For example, when we enjoy a film in the cinema the retina in our eyes '*demodulates*' the light received from the image on projection screen, which is the 'modulated carrier', and transforms it into stimuli for the brain. The image is established through light transmitted by film projector, after *modulation* of its wave-length spectrum (color) and intensity (brightness) in accordance to the information contained in the celluloid film strip.

The observation that information is inevitably corrupted by noise and disturbances during transmission through the communication channel, has been the motive for the development of a wide variety of modulation schemes and communication systems. Besides efficient usage of the channel, modulation, and frequently also additional coding, should generally also provide *protection of the message information* against corruption by disturbances inside the channel. A measure for the success of a scheme in fulfillment of both tasks is given by the efficiency of its use of the so called *channel capacity*, defined by Shannon [2], that relates the maximum possible flow of information through the channel to its bandwidth and the relative strength of the information carrier and the disturbances. Communication system design is therefore concerned with the search for a maximally efficient modulation scheme, and, moreover, with the search for transmitter and receiver architectures that maximally exploit the abilities of such a scheme.

This thesis is concerned with the design of electronic demodulators, and receivers, for reception and demodulation of frequency modulated (FM) waves. These waves are characterized by a repetition speed that is altered in accordance to the information. Special attention is devoted to *high-sensitivity* demodulators, capable to reconstruct information received in the presence of violent noise and disturbances. In comparison to other modulation schemes, the efficiency of analog and digital FM schemes, in their usage of the channel capacity, is quite good. This is reflected by their since long recognized ability to establish considerable improvement of the transmission performance, in comparison to other schemes. However, full benefit of this ability is attained only by application of a thereto properly designed FM demodulator.

Since the design of FM demodulators is an issue that has received scientific attention already for almost a century, it seems appropriate first to place the subject in its historic perspective, as described in Section 1.1. Subsequently, the objective, scope and organization of the thesis are discussed in Section 1.2 and Section 1.3.

1.1 History of Frequency Modulation

Frequency modulation finds its origin somewhere in the last decade of the 19-th century. To get a grasp of the 'state of the art' at that time; just before this epoch, the experiments performed by the German physicist Heinrich Hertz showed the existence of electromagnetic waves, that were predicted already by Maxwell's theory in 1864. Marconi was among the first ones, that put this new discovered physical phenomenon into practical use. Based on the work of Hertz, he developed the first completely wireless telegraph system, that was patented in 1897. Meanwhile, Sir Oliver Lodge developed the theory of tuning circuits, that, later on, appeared to be of substantial use in receiver design.

Until 1912, radio communication was merely confined to the transmission of Morse codes, generated by so called 'spark systems'. Some of these systems transmitted the Morse codes as a 'wave' of alternating frequency, a primitive kind of frequency modulation that resembles Frequency Shift Keying (FSK) [3]. However, this modulation was actually not yet recognized and intended as such, but merely followed from the solution of a problem with the telegraph key. The human ear served as 'demodulator'. Since electric amplification was an unknown phenomenon, the power observed at the receiver output was entirely due to the energy captured from the electromagnetic wave by the antenna. Armstrong [4] summarized the burden of of this way of reception as: "*In order to hear weak signals it was necessary to use painfully tight headphones, frequently with the equally uncomfortable necessity of holding one's breath for prolonged intervals*".

The same period showed the first experiments with transmission of audible intelligence [5]. First, transmission was attempted by a kind of amplitude modulation (AM). However, due to the absence of amplification, the microphone connected to the wave generator had to deliver the complete AM modulation power, which soon appeared too small to be detectable by the insensitive receivers, that lacked any kind of amplification too. It was proposed to solve this problem by means of *frequency modulation*. In that scheme, the microphone only altered the frequency of the generated wave, and no longer needed to generate the modulation power. The demodulation should be performed by detuned selective circuits. Unfortunately, one failed to obtain satisfactory operating circuits. Therefore, this type of system was never brought into practical use.

The invention of the regenerative circuit in 1912 solved the problems of AM reception, and caused FM reception to fall into oblivion for almost a decade, by providing amplification and heterodyne reception, using internally generated stable oscillations. This circuit used a revolutionary new device called the "audion", currently know as the triode vacuum tube, invented by Lee de Forest around 1906. However, at first, he didn't recognize the amplifying capabilities of the device, and simply considered it as a slightly more sensitive 'detector' than the "Fleming valve" (diode). An explanation of its operation as amplifier, detector and oscillator lasted until the correspondence of Edwin Howard Armstrong with Lee de Forest [6], and his publication in 1914 [7].

Frequency modulation reappeared at the scene only, when the radio spectrum became highly crowded. It was believed, that this type of modulation could relief the high crowding, by application of a very small frequency deviation. This was thought to yield an FM wave with a considerably smaller bandwidth, than the corresponding AM wave. However, the illusion of bandwidth reduction was effectively dispelled by the mathematician John R. Carson in 1922 [8]. He showed, that the bandwidth required for FM transmission at least equals that required for AM. Further, the observation that, in comparison to AM, the 'amplitude' of a narrow-band FM wave contains an integrated copy of the message lead him to the erroneous conclusion that "*Consequently, this type of modulation inherently distorts without any compensating advantages whatsoever*", which nearly caused FM to be doomed forever.

The break-through of FM came in the second half of the 1930's. The problems of "static", i.e. (man made) noise, in the crowded radio spectrum kept bothering AM reception, even with inventions as the regenerative, and the super-regenerative circuit [9]. Again, Carson [10] placed the matter in a theoretical framework, and showed that the output noise level cannot be reduced below that of the noise contained inside the bandwidth of the AM wave. This article inspired Armstrong to attempt heavy clipping of the receiver input wave in wideband FM transmission, instead of the previously proposed narrow-band FM, as a means to reduce disturbances. He soon reached impressive results, a performance that by far exceeded that of any known receiver at that time, and filed a patent proposal in 1927, which was finally granted in 1933 [11]. The first results were received in the scientific society with disbelief, but the effectiveness of the method was finally accredited after extensive field tests in the period 1927-1935, reported in [12]. Even in 1937, apparently unaware or incredulous, Terman wrote in his textbook [13], that "*Frequency modulation is not particularly satisfactory as a means of transmitting intelligence*".

The period 1937-1950 shows a steady increase in scientific activity on the field of FM. In 1937, Crosby [14], and Carson and Fry [15], developed the currently well-know theory that explains the performance improvement achieved with wideband FM. Chaffee [16, 17] invented the FM frequency feedback receiver, based on the new concept of "feedback", invented by Black [18]. Subsequently, the drawback of FM transmission, the *threshold effect*, was subjected to a closer investigation. Based on the thorough theoretical framework developed by Stephen O. Rice at AT&T Bell Labs. [19–21], Stumpers [22], Middleton [23, 24], and Blachman [25, 26] obtained a mathematical model for the threshold effect in FM, and also for the output noise spectrum. The results, however, where not directly suited for application in practice; some results consisted of formulas of over a page in length, full of awkward functions.

The famous ratio detector, an improved version of the Foster-Seeley detector [27] that remained in use until the 1980's, was developed in 1947 [28]. This circuit gave the occasion to a severe conflict between Armstrong and the Radio Company of America (RCA). RCA claimed the invention of a new method of disturbance reduction, that was completely disagreed upon by Armstrong [4], and circumvented Armstrong's patent.

Research in the period 1950-1970 was concerned with improvement of FM reception in strong noise. Cohn [29] and Rice [30] finally succeeded independently in formulating a suitable engineering description for the FM threshold effect. Meanwhile, in the early 1960's, it was discovered that application of clipping/limiting to Chaffee's frequency feedback receiver resulted in significant *threshold extension*, that improved the performance in strong noise. Although the modeling of this improvement has never been resolved completely, frequency feedback receivers were soon encountered in satellite communication, and even in the lunar orbiter [31]. Numerous attempts were made to model the improvement, of which, among many others, the rather erroneous one of Enloe [32] is the best known. The quite unknown work of Bax [33], seems to be the most complete of all. In the same period, research on the threshold behavior of Phase-Locked Loops (PLL) and Delay Locked Loops (DLL), with major contributions of Tikhonov [34, 35], Viterbi [36] and Lindsey [37], did succeed, and resulted in, although complicated, models for the threshold.

In the 1970's and 1980's, the attention was focused on the design of fully integrated FM receivers [38–40]. At the same time, the interest gradually shifted from analog FM to digital equivalents as FSK, (G)MSK, and, digital phase modulation, PSK, where special attention was paid to the spectral efficiency of these schemes [41–44].

In the first half of the 1990's, efforts concentrated on all kinds of digital FM transmission for mobile radio [45, 46]. Simultaneously, adaptive demodulator architectures for interference suppression in car radio were developed [47, 48]. One of the current fields of research is the development of receivers for the Digital Audio Broadcasting (DAB) standard, intended to replace analog FM broadcasting in the future.

1.2 Objective and Scope of the Thesis

In the light of its long and rich history, determined by excellent contributions of numerous honorable scientists and engineers, it might seem that hardly any work in the field of FM reception is left. This is however not the case.

In the first place, much of the material, in particular the work on FM demodulator circuit design, has been developed in rather 'scattered' way. Many cross-links between the various types of demodulators, and their characteristics, are missing. No unifying framework seems to have been found yet, that surveys all principles available for the construction of FM demodulators, and relates them to the characteristics and performance of the various demodulator circuits. Such a frame-work would be of significant value in FM demodulator design, since it allows a deliberate selection of the most suitable type at the very beginning of a design trajectory. Moreover, the separation between demodulation principles and their implementations, opens the way to improvements by use of different implementations, possibly in different and/or new technologies.

Secondly, progress in FM demodulator design seems to have been impeded by rather little communication between theoretical scientists, and electronic designers. At one side, mathematicians and communication theorists treated and solved numerous problems with astonishing accuracy. However, such treatments often omitted a translation of the results to intuitively more appealing descriptions, heuristic explanations for these results, and simplified estimations of the various mechanisms, in such a way that these are accessible to electronic designers. On the other hand, electronic designers simply seem to have ignored much of this material. As a result, the implications of all these theories on circuit design are often missing, while at the same time, the theoretical background of problems encountered in circuit design are lacking.

These problems become especially apparent in the design of high-sensitivity demodulators, intended for applications were operation in the presence of strong noise and interference is required. Examples of such applications are car radio, and various types of wireless communication that generally employ some type of digital FM. For example, at Philips Research Laboratories, it appeared that replacement of the old ratio detector, mounted in car radios until deep in the 1980's, by fully integratable PLLs did not in all respects yield completely satisfactory results. Especially at high interference levels, the performance seemed to have been degraded. Even after the development of an improved demodulator, capable of adaptive interference suppression, there remained some concern on its performance around the threshold. This furnished the idea for a project on the threshold behavior of FM demodulators, and its implications on demodulator design.

This thesis describes a structured approach towards the design of high-performance FM demodulators and FM demodulation systems, that provides insight into the principles available for the construction of such demodulators, and into the various architectural measures that can be taken to improve their performance, especially during reception in strong noise. In cases where it is appropriate, implications on the design of the various electronic circuits are discussed.

Although some implications on circuit design are considered, the thesis concentrates on the design at an architectural level, since is was felt, that insight in this aspect of demodulator design is required first, before a deliberate investigation, and substantial contributions to the design of the various electronic sub-circuits can be realized.

Further, the thesis is mainly concerned with demodulation of sinusoidal, analog FM waves. To a large extend, the characteristics of such waves, and the corresponding principles available for their demodulation, also determine the design of demodulators for digital FM schemes, that currently experience a rapidly increasing interest. However, in addition, such schemes generally allow some typically digital provisions in demodulator design, such as "Viterbi decoding", that are not considered here. Although sinusoidal carrier waves are not essential in the discussion, they have been adopted in all expressions and numerical examples in this thesis, unless stated otherwise. Such waves are by far the most frequently encountered ones of all, due to their spectral efficiency.

1.3 Organization of the Thesis

The organization of the thesis is as follows. Chapter 2 reviews the main characteristics of FM transmission and FM waves, that constitute the investigations in the sequel of the thesis. This chapter also contains the definitions of the symbols and signals that are used throughout the thesis.

Chapter 3 discusses a classification of the principles available for the construction of FM demodulators. As an introduction to this classification, a brief discussion is devoted to the principles of a structured design approach, applied in this thesis, and the importance of a classification in FM demodulator design. Subsequently, the various principles are considered, starting from the intrinsic characteristics of FM waves. Algorithms for the implementation of these principles, and the various sub-functions contained in them, are derived. Finally, numerous types of demodulators encountered in literature are classified, and their performance, as far as determined by the demodulation principles, is briefly considered.

Chapter 4 considers the design of the sub-functions encountered in the various demodulation principles and algorithms. Maximization of the demodulator dynamic range is the key item in this chapter, which results in various rules for the design of the demodulator architecture, the noise behavior and frequency characteristic of important sub-functions.

Chapter 5 outlines qualitatively the possibilities to improve the demodulator performance, by proper design of the FM receiver architecture that embeds it. The principles, capabilities and limitations of the various types of pre-demodulation, post-demodulation and (adaptive) feedback, are outlined. Three types of processing are investigated in detail in Chapter 6 through Chapter 8.

Chapter 6 analyzes the performance improvement achieved by (partial) elimination of noise from the demodulator input carrier amplitude by limiting, or, more general, compression of the amplitude. Literature on this subject is merely confined to hard-limiting, i.e. infinite compression, but it is shown that finite compression may improve the performance in the presence of high noise. The various design rules for the selection of the type and level of compression, and the 'optimal' level of compression, as function of the input carrier-to-noise ratio (CNR) are derived.

Chapter 7 investigates the threshold behavior of phase feedback demodulators, like PLLs, and compares it to the performance of previously discussed 'conventional' demodulators. The essence of sophisticated nonlinear analyses developed in the past for these demodulators is explained, and its implications on demodulator design are discussed.

Chapter 8 investigates the threshold behavior of frequency feedback demodulators. A completely unifying theory for these demodulators has never been developed, but it is shown that combination of several important theories yields a model that clearly describes the threshold behavior, and its implications on demodulator design.

Chapter 9 closes the thesis with conclusions.

References

[1] A. Bruce Carlson, *Communication Systems, An introduction to Signals and noise in Electrical Communication,* McGraw-Hill International Editions, Singapore, 1986.

[2] C.E. Shannon, "A mathematical theory of communication", *The Bell System Technical Journal,* vol. 27, pp. 379–423 and 623–656, 1948.

[3] H.G.J. Aitken, *The Continuous Wave: Technology and American Radio, 1900-1932,* Princeton University Press, Princeton, 1985.

[4] Edwin H. Armstrong, "A study of the operating characteristics of the ratio detector and its place in radio history", *Proceedings of the Radio Club of America,* vol. 64, no. 3, pp. 217–232, Nov. 1990, Reprint of vol. 25, no. 3, 1948.

[5] Edwin H. Armstrong, "Evolution of frequency modulation", *Proceedings of the Radio Club of America,* vol. 64, no. 3, pp. 179–188, Nov. 1990, Reprint of dec. 1940, pp.485-494, embracing the substance of lectures presented before the *American Institute of Electrical Engineers.*

[6] Lee de Forest and Edwin H. Armstrong, "Discussion", *Proceedings of the IRE,* vol. 1, pp. 239–240, 1913.

[7] Edwin H. Armstrong, "Operating features of the audion", *Electrical World,* vol. 64, no. 24, pp. 1149–1152, December 12, 1914.

[8] John R. Carson, "Notes on the theory of modulation", *Proceedings of the IRE,* vol. 10, no. 2, pp. 57–64, Feb. 1922.

[9] Edwin H. Armstrong, "Some recent developments of regenerative circuits", *Proceedings of the IRE,* vol. 10, no. 8, pp. 244–260, 1922.

[10] John R. Carson, "Selective circuits and static interference", *The Bell System Technical Journal,* vol. 4, pp. 265 e.v., 1925.

[11] Edwin H. Armstrong, "Radio telephone signalling", U.S. Patent 1,941,447, December 26, 1933.

[12] Edwin H. Armstrong, "A method of reducing disturbances in radio signaling by a system of frequency modulation", *Proceedings of the IRE*, vol. 24, no. 5, pp. 689–741, May 1936.

[13] Frederick E. Terman, *Radio Engineering*, McGraw-Hill Book Company, New York, 1937.

[14] M.G. Crosby, "Frequency-modulated noise characteristics", *Proceedings of the IRE*, vol. 25, no. 4, pp. 472–514, Apr. 1937.

[15] John R. Carson and Thornton C. Fry, "Variable frequency electric circuit theory with applications to the theory of frequency-modulawtion", *The Bell System Technical Journal*, vol. 16, no. 10, pp. 513–540, Oct. 1937.

[16] J.G. Chaffee, U.S. Patent 2,075,503, March 30, 1937.

[17] J.G. Chaffee, "The application of negative feedback to frequency modulation systems", *Proceedings of the IRE*, vol. 27, no. 5, pp. 317–331, May 1939.

[18] H.S. Black, U.S. Patent 2,102,761, 1937.

[19] S. O. Rice, "Mathematical analysis of random noise-I", *The Bell System Technical Journal*, vol. 23, pp. 282–332, 1944.

[20] S. O. Rice, "Mathematical analysis of random noise-II", *The Bell System Technical Journal*, vol. 24, pp. 46–156, 1945.

[21] S. O. Rice, "Statistical properties of a sine wave plus random noise", *The Bell System Technical Journal*, vol. 27, pp. 109–157, 1948.

[22] F. L. H. M. Stumpers, "Theory of frequency-modulation noise", *Proceedings of the IRE*, vol. 36, no. 9, pp. 1081–1092, Sept. 1948.

[23] David Middleton, "The spectrum of frequency-modulated waves after reception in random noise-I", *Quarterly of Applied Mathematics*, vol. vol. VII, no. 2, pp. 129–174, July 1949.

[24] David Middleton, "The spectrum of frequency-modulated waves after reception in random noise-II", *Quarterly of Applied Mathematics*, vol. vol. VIII, no. 1, pp. 59–80, Apr. 1950.

[25] Nelson M. Blachman, "The demodulation of an F-M carrier and random noise by a limiter and discriminator", *Journal of Applied Physics*, vol. 20, pp. 38–47, Jan. 1949.

[26] Nelson M. Blachman, "The demodulation of a frequency-modulated carrier and random noise by a discriminator", *Journal of Applied Physics*, vol. 20, pp. 976–983, Oct. 1949.

[27] D.E. Foster and S. Seeley, "Automatic tuning, simplified circuits, and design practice", *Proceedings of the IRE*, vol. 25, no. 3, pp. 289–313, Mar. 1937.

[28] Stuart W. Seeley and Jack Avins, "The ratio detector", *RCA Review*, vol. 8, pp. 201–236, 1947.

[29] John Cohn, "A new approach to the analysis of FM threshold reception", in *Proceedings of the National Electronics Conference*, 1956, pp. 221–236.

[30] S. O. Rice, "Noise in FM receivers", in *Proceedings of the Symposium on Time Series Analysis, Brown University, 1962*, M. Rosenblatt, Ed. pp. 395–422, John Wiley and Sons, New York, 1963.

[31] F. Lefrak, H. Moore, A. Newton, and L. Ozolins, "The frequency-modulation feedback system for the lunar-orbiter demodulator", *RCA Review*, vol. 27, no. 12, pp. 563–576, Dec. 1966.

[32] L.H. Enloe, "Decreasing the threshold in FM by frequency feedback", *Proceedings of the IRE*, vol. 50, no. 1, pp. 18–30, Jan. 1962.

[33] F.G.M. Bax, *Analysis of the FM Receiver with Frequency Feedback*, PhD thesis, Catholic University of Nijmegen, Nijmegen, The Netherlands, Oct. 1970.

[34] V.I. Tikhonov, "The effects of noise on phase-lock oscillation operation", *Automatika i Telemakhanika*, vol. 22, no. 9, 1959.

[35] V.I. Tikhonov, "Phase-lock automatic frequency control application in the presence of noise", *Automatika i Telemakhanika*, vol. 23, no. 3, 1960.

[36] Andrew J. Viterbi, "Phase-locked loop dynamics in the presence of noise by Fokker-Planck techniques", *Proceedings of the IEEE*, vol. 51, no. 12, pp. 1737–1753, Dec. 1963.

[37] W. C. Lindsey, "Nonlinear analysis of generalized tracking systems", *Proceedings of the IEEE*, vol. 57, pp. 1705–1722, Oct. 1969.

[38] W.G. Kasperkovitz, "FM receivers for mono and stereo on a single chip", *Philips Technical Review*, vol. 41, no. 6, pp. 169–182, 1983–1984.

[39] Bang-Sup Song and Jeffrey R. Barner, "A CMOS double-heterodyne FM receiver", *IEEE Journal of Solid State Circuits*, vol. 21, no. 6, pp. 916–923, Dec. 1986.

[40] A. Sempel and H. van Nieuwenburg, "A fully-integrated HIFI PLL FM-demodulator", in *Dig. IEEE International Solid State Circuits Conference*, Feb. 1990, pp. 102–103.

[41] Frank de Jager and Cornelis B. Dekker, "Tamed frequency modulation, a novel method to achieve spectrum economy in digital transmission", *IEEE Transactions on Communications*, vol. 26, no. 5, pp. 534–542, May 1978.

[42] Mitsuru Ishizuka and Kenkichi Hirade, "Optimum Gaussian filter and deviated-frequency-locking scheme for coherent detection of MSK", *IEEE Transactions on Communications*, vol. 28, no. 6, pp. 850–857, June 1980.

[43] Tor Aulin and Carl-Erik W. Sundberg, "Continuous phase modulation-part I: Full response signalling", *IEEE Transactions on Communications*, vol. 29, no. 3, pp. 196–209, Mar. 1980.

[44] Tor Aulin and Carl-Erik W. Sundberg, "Continuous phase modulation-part II: Partial response signalling", *IEEE Transactions on Communications*, vol. 29, no. 3, pp. 210–225, Mar. 1980.

[45] Bang-Sup Song and In Seop Lee, "A digital FM demodulator for FM, TV and wireless", *IEEE Transactions on Circuits and Systems-II*, vol. vol. 42, no. 12, pp. 821-825, Dec. 1995.

[46] Israel Korn, "Error probability of digital modulation in satellite mobile, land mobile, and gaussian channels with narrow-band receiver filter", *IEEE Transactions on Communications*, vol. 40, no. 4, pp. 697–707, Apr. 1992.

[47] W. Bijker and W.G. Kasperkovitz, "A top-down design methodology applied to a fully integrated adaptive FM IF system with improved selectivity", in *Proceedings of the European Solid-State Circuits Conference*, Sevilla, Sept. 1993, pp. 53–56.

[48] W. Bijker and W.G. Kasperkovitz, "FM receiver with dynamic intermediate frequency (IF) filter tuning", U.S. Patent 5,404,589, Apr. 1995.

Chapter 2

Characteristics of Frequency Modulation

The characteristics of frequency modulated waves and the frequency modulation scheme inevitably play a dominant role in FM demodulator design. Besides the many differences among them, the correspondence between such demodulators is, that they all account for the FM wave, and FM transmission characteristics. The sequel of this thesis makes extensive use of these characteristics, and assumes them to be known to the reader.

This chapter summarizes the main characteristics of FM waves, and describes the conventions used throughout the sequel of this thesis. Most of this material is contained in standard text books on communication and modulation theory.

Section 2.1 outlines the characteristics of modulated and unmodulated carrier waves in general, and, in particular, defines the FM carrier wave used throughout the thesis. Section 2.2 considers the transmission bandwidth of FM waves, and discusses a useful approximative description of their spectrum. Section 2.3 discusses the main characteristics of FM transmission in the presence of small noise, and compares them to other modulation schemes. Section 2.4 presents the conclusions.

2.1 Modulation Scheme

A modulation scheme is the shortest possible, usually mathematical, description of the essential characteristics of a transmission system. It completely specifies in which way the message information is included into the transmitted, modulated carrier wave. Knowledge of the characteristics of these waves is essential in transmitter and receiver design, since they determine the architecture of these systems.

This section briefly considers the main characteristics of carrier waves in general, and subsequently outlines the FM modulation scheme and description of FM waves, as used in the sequel of this thesis.

Section 2.1.1 outlines the main characteristics of, and conditions to be satisfied by carrier waves in general. Section 2.1.2 discusses the basic formulation of the FM scheme, and the corresponding description of FM waves.

2.1.1 Modulated and Unmodulated Carrier Waves

This section outlines the main characteristics of modulated and unmodulated carrier waves, that follow from the various conditions posed on these waves by the modulation scheme.

Deterministic Nature of Carrier Waves

As expressed by their speaking name, carrier waves 'carry' the message information from the transmitter side to the receiver side of a communication link. The main condition posed on carrier waves by any modulation scheme is, of course, that they guarantee the reversibility of the modulation action, i.e. the possibility of demodulation.

More precisely, it is required that the carrier wave allows error-free recovery of the message information when external disturbances are absent. This implies, that an unmodulated carrier wave is not allowed to contain any information, at least as far as the carrier parameter used by the modulation scheme is concerned. Since, according to Shannon's definition [1], information is uniquely related to uncertainty, this means that unmodulated carrier waves have to be deterministic, in the parameter of interest, in order to satisfy the conditions.

Periodic Nature of Carrier Waves

In order to demodulate the received, modulated carrier wave, receivers somehow have to distinguish between fluctuations of the received carrier intensity due to the carrier wave, and fluctuations due to the message information.

For this purpose, it very helpful if the unmodulated carrier wave is of periodic nature, i.e. when their fluctuations are repeated in time. In that case, it is far easier to recognize these fluctuations during transmission of the message information. This is of interest e.g. when the reception is interrupted by some kind of disturbance, or when the receiver is switched on half-way.

Sinusoidal Nature of Carrier Waves in FDM Transmission

In transmission systems that subdivide the available channel capacity in the frequency-domain by *Frequency Division Multiplexing* (FDM), such as FM, it is generally profitable to use carrier waves that possess an as small as possible bandwidth for a given message signal.

A sinusoidal carrier wave is therefore usually the most appropriate choice. For this reason, the sequel of this thesis is essentially concerned with the reception of sinusoidal FM waves. In case of non-sinusoidal carriers, the appearance of the various signals inside the FM receiver, and the corresponding expressions that describe them, are slightly different, but the essential characteristics of the modulation, and also of the transmitter and receiver, are the same as those in case of sinusoidal carriers.

Modulation of Carrier Waves

Since a carrier wave is characterized by two parameters, the carrier amplitude and the carrier argument, as discussed in detail in Chapter 3, two main classes of modulation can be distinguished:

- amplitude or 'linear' modulation;
- argument, i.e. 'exponential' [2] or 'nonlinear' modulation.

The class of amplitude modulation contains, besides the well-known 'analog' AM scheme, schemes as DSB, SSB and PAM [2]. The characteristic property of these schemes is, that the required transmission bandwidth is *only* determined by the bandwidth of the unmodulated carrier wave, and the message, and *not* by the modulation index.

The class of argument modulation contains, besides the FM scheme, schemes as PM, FSK, (G)MSK [3], PSK, PPM, PFM [2]. These schemes are characterized by the property that the required transmission bandwidth is *not only* determined by the bandwidth of the unmodulated carrier wave and the message, but *also* by the modulation index.

2.1.2 FM Carrier Waves

This section describes the representation of the FM carrier wave, and the meaning of its parameters, that is used throughout the sequel of this thesis.

Sinusoidal Carrier Wave

The input of the various FM demodulator and FM receiver configurations described in the sequel of this thesis is, besides with noise and/or interference, supplied with a noise-free *sinusoidal* FM wave, denoted by $s(t)$. This wave is expressed as

$$s(t) \overset{\text{def}}{=} A(t)\cos\left[\omega_o t + \varphi(t)\right], \tag{2.1}$$

where $A(t)$ denotes the carrier amplitude, ω_o the carrier frequency in (rad/s), and $\varphi(t)$ the phase modulation contained in the wave. Notice the nonlinear nature of this scheme; the message phase $\varphi(t)$ is related to the intensity of the carrier wave $s(t)$ by means of a (co)sine, i.e. by complex exponential functions.

FM Message Signal

In most of the subsequent discussions, a constant carrier amplitude $A(t) = A$ is adopted. In a few cases, however, it is important to demonstrate the effect of a time-dependent amplitude, e.g. caused by fading, on the FM demodulator output signal.

According to the FM scheme, see e.g. [2, 4], the phase modulation $\varphi(t)$ contains an integrated copy of the message signal $m(t)$, i.e.

$$\varphi(t) = \Delta\omega \int_0^t m(\tau)\mathrm{d}\tau. \tag{2.2}$$

the *frequency modulation* contained in $s(t)$ therefore equals $\dot{\varphi}(t) = \Delta\omega m(t)$. For convenience, it is assumed that the message signal $m(t)$ possesses a power contents of unity, such that $\Delta\omega$ represents the *RMS frequency deviation*, i.e. the RMS value of the frequency modulation $\dot{\varphi}(t)$.

FM Modulation Index

The bandwidth of the message signal $m(t)$, and the frequency modulation $\dot{\varphi}(t)$, is denoted by W, in (rad/s). The double-sided spectrum of these signals is therefore located in the frequency interval $\omega \in [-W, W]$, which is usually called the *baseband*. Finally, the FM *modulation index* or *frequency deviation ratio*, defined as

$$m_{\text{FM}} \overset{\text{def}}{=} \frac{\Delta\omega}{W} \tag{2.3}$$

is an important parameter in FM transmission, that determines the upper bound on the output SNR of the FM receiver, as discussed in Section 2.3.

2.2 FM Spectrum and Transmission Bandwidth

Knowledge of the spectrum and bandwidth of modulated carrier waves is particularly important in design of receivers for FDM transmission systems, such as FM receivers. This spectrum, and the corresponding bandwidth is required for the design of the various filters inside FM receivers, and is also of importance in the determination of the FM demodulator response to noisy FM waves.

This section discusses a model for the spectrum of FM waves, and some well-known measures for the bandwidth of these waves.

As an introduction, Section 2.2.1 discusses the characteristics of the FM spectrum for the case of sinusoidal modulation. Section 2.2.2 discusses the *quasi-stationary* approximation of the FM spectrum for arbitrary types of message signals. Finally, Section 2.2.3 discusses the well-known estimates for the bandwidth of FM waves.

2.2.1 FM Spectrum for Sinusoidal Modulation

The spectrum of FM waves that are modulated by a sinusoidal message signal $m(t)$ is encountered in almost every introductory text on FM modulation. It is one of the few FM spectra that can be determined exactly; due to the nonlinear nature of the FM scheme, a general closed-form expression of the FM spectrum for arbitrary message signals cannot be given. In the latter cases, an approximation should be used, as discussed in Section 2.2.2. The discussion in this section serves as an introduction to this approximation. The exact results described here may be used as an estimate for the accuracy of this approximation.

Assume that the signal $\dot{\varphi}(t)$, that, for convenience, is called the FM message signal, is given by

$$\dot{\varphi}(t) = \sqrt{2}\Delta\omega \cos Wt, \tag{2.4}$$

where $\Delta\omega$ and W are chosen in accordance with Section 2.1.2. Thus, $\Delta\omega$ equals the RMS frequency deviation, while the factor $\sqrt{2}$, the *crest-factor* of a sine-wave, relates this RMS value to the maximum value of the sine-wave. The FM wave $s(t)$ may then be expressed as

$$s(t) = A \cos\left[\omega_o t + \frac{\sqrt{2}\Delta\omega}{W} \sin Wt\right], \tag{2.5}$$

where the carrier amplitude is assumed to be constant.

By definition, the FM spectrum is obtained by Fourier transformation of (2.5). As derived originally by Carson [5], and outlined e.g. in [2], this spectrum corresponds to the Fourier series expansion

$$s(t) = A \sum_{n=-\infty}^{\infty} \mathrm{J}_n\left(\frac{\sqrt{2}\Delta\omega}{W}\right) \cos\left(\omega_o + nW\right) t, \tag{2.6}$$

where $\mathrm{J}_n(.)$ denote the Bessel functions of the first kind and order n.

For small FM modulation indices, the impulse-spectrum corresponding to (2.6) is similar to the spectrum of AM modulated waves. The FM spectrum for this case is sketched in figure 2.1. It is observed that the spectrum basically

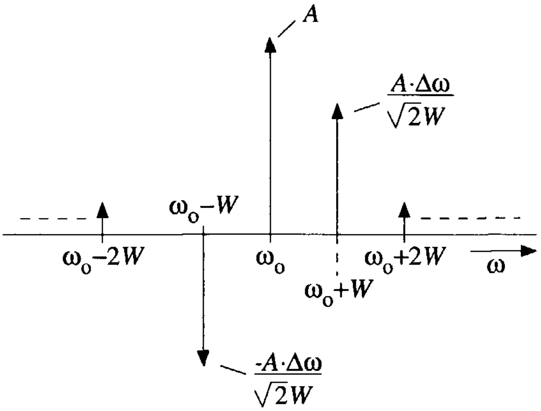


Figure 2.1: FM Spectrum for sinusoidal (tone) modulation and small modulation indices.

consists of a component at the fundamental frequency ω_o, equal to $A\mathrm{J}_0(.) \approx A$, and the first harmonic components at $\omega_o \pm W$, equal to $A\mathrm{J}_{\pm 1}(.) \approx A\Delta\omega/(\sqrt{2}W)$. The spectrum of an AM wave with the same modulation index as the FM wave contains the same spectral components, of the same magnitude. However, in the FM wave, both 'side-bands' at $\omega_o \pm W$, are in anti-phase with each other, and in quadrature with the carrier component at ω_o, while all components in the AM wave are in-phase [2]. This is also reflected by (2.6): for odd n, the Bessel functions satisfy the relation $J_{-n}(.) = -J_n(.)$.

In FM waves, an increase of the modulation index, established by an increase of the frequency deviation $\Delta\omega$, does not only change the magnitude of the spectral components, but also increases the *number* of harmonic components that possesses a significant power contents. Consequently, *an increase of the modulation index increases the bandwidth* of the FM wave, which demonstrates the nonlinear nature of FM. On the contrary, in AM waves, i.e. a linear modulation scheme, an increase of the modulation index results only in an increased power contents of the spectral components, but not in additional components.

2.2.2 FM Spectrum for Arbitrary Modulation

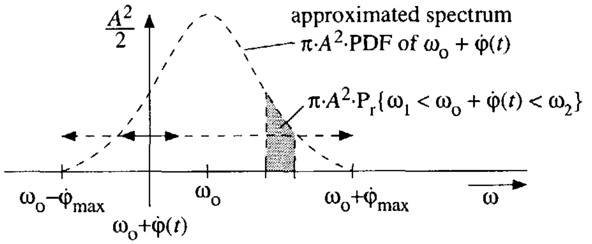
As mentioned in Section 2.2.1, a general and exact expression for the FM spectrum in case of arbitrary types of modulation does not exist, due to the rather complicated nonlinear nature of the FM modulation scheme. However, for so called *wideband* FM waves, characterized by a modulation index that is (much) larger than unity, i.e. a frequency deviation $\Delta\omega$ that exceeds the message bandwidth, a so called 'quasi-stationary' approximation of the spectrum can be obtained.

This section outlines the basics of the quasi-stationary approximation, that is used in the sequel for the calculation of the FM demodulator output noise power density spectrum.

Time-Dependent Spectrum

The central idea of the quasi-stationary approximation is to assume that the instantaneous frequency of the FM wave varies slow in comparison to the bandwidth of the wave. This is equivalent to neglecting the time-dependency of the instantaneous frequency of the FM wave in the calculation of its *power density spectrum*; the time-dependent instantaneous frequency is treated as a constant.

In this way, a time-dependent representation of the FM power density spectrum is attained, that represents the FM wave as an impulse (a "finger") of area πA^2, equal to 2π times the power contents of the FM-wave, that moves along the frequency axis in the rhythm of the instantaneous frequency. The impulsive shape corresponds to the spectrum of an unmodulated carrier wave, with a time-independent instantaneous frequency. This representation is depicted in figure 2.2. At the instant t, the impulse is positioned at $\omega = \omega_o + \dot{\varphi}(t)$, i.e. the instantaneous frequency of the FM wave, and moves along a frequency interval, centered around the carrier frequency ω_o, that is bounded by the maximum (and minimum) value of $\dot{\varphi}(t)$, denoted by $\dot{\varphi}_{\max}$.



$\frac{A^2}{2}$
approximated spectrum
$\pi \cdot A^2 \cdot$PDF of $\omega_0 + \dot{\varphi}(t)$
$\pi \cdot A^2 \cdot \mathrm{P_r}\{\omega_1 < \omega_0 + \dot{\varphi}(t) < \omega_2\}$
$\omega_0 - \dot{\varphi}_{\max}$
ω_0
$\omega_0 + \dot{\varphi}_{\max}$
ω
$\omega_0 + \dot{\varphi}(t)$

Figure 2.2: Representation of the FM spectrum by a moving Dirac-impulse.

This representation clearly demonstrates the previously noticed relation between the frequency deviation $\Delta\omega$, or the modulation index $\frac{\Delta\omega}{W}$, and the FM transmission bandwidth: when $\Delta\omega$, the RMS value of $\dot{\varphi}(t)$, increases, the excursions of the impulse in figure 2.2 extend over a larger frequency range, corresponding to an increased FM transmission bandwidth.

Approximate FM Power Density Spectrum

A time-independent approximation of the FM power density spectrum is obtained by time-averaging of the time-dependent representation of figure 2.2.

By means of this time-average, the area underneath the spectrum in the frequency interval $\omega \in [\omega_1, \omega_2]$ becomes proportional to the fraction of time that $\omega_1 < \omega_o + \dot{\varphi}(t) < \omega_2$, as indicated by the shaded area in figure 2.2.

When the FM message signal $\dot{\varphi}(t)$ is an ergodic, stochastic signal, the time-average may be replaced by the ensemble average. For such signals, the approximated spectrum becomes equal to the probability density (PDF) of $\omega_o + \dot{\varphi}(t)$, times the power contents of the FM wave. When $p_{\dot{\varphi}}(.)$ denotes the PDF of the stochastic signal $\dot{\varphi}(t)$, and $S_s(\omega)$ denotes the *double-sided* power density spectrum of the FM wave $s(t)$, then [2]

$$S_s(\omega) \approx \frac{\pi A^2}{2\Delta\omega}\left[p_{\dot{\varphi}}\left(\frac{\omega-\omega_o}{\Delta\omega}\right)+p_{\dot{\varphi}}\left(\frac{\omega+\omega_o}{\Delta\omega}\right)\right], \tag{2.7}$$

where the frequency deviation $\Delta\omega_o$ is proportional to the RMS deviation $\Delta\omega$, and should be chosen such that the power contents of the spectrum equals $A^2/2$, the power contained in $s(t)$.

The same approximation can be used in case of deterministic FM message signals, when the deterministic signal is replaced with a stationary, ergodic random signal that possesses the same amplitude distribution [2].

Validity of the Approximation

It is evident that the previously described approximation holds only when the FM message signal behaves approximately as a 'constant' frequency offset, i.e. when the impulse in figure 2.2 moves slowly through the spectrum.

A detailed analysis shows that this condition is satisfied when the bandwidth of $\dot{\varphi}(t)$, represented by W, is considerably smaller than the FM transmission bandwidth. This means that the approximation holds for FM modulation indices that are considerably larger than unity, which is the case only for wideband FM waves.

2.2.3 FM Transmission Bandwidth

As a result of the nonlinear nature of FM, the bandwidth of FM waves is, in theory, *infinite.* This was observed already from the spectrum of a sinusoidally modulated FM wave, discussed in Section 2.2.1; the components in the Fourier expansion (2.6) extend over the entire spectral frequency axis. However, it was also observed in that section, that the largest fraction of of the carrier power is concentrated in a *finite* bandwidth, centered around the carrier frequency; the higher harmonics were negligible.

Consequently, any expression for the bandwidth of an FM wave should actually be accompanied by the fraction of the carrier power that is neglected, i.e. located outside the bandwidth. As discussed in Chapter 4 and Chapter 5, this neglection results in distortion of the FM message signal. Therefore, an estimate of the FM bandwidth is associated to a certain level of distortion, introduced into the FM message when filtering of this bandwidth is applied.

A frequently encountered definition of the FM bandwidth, known as *Carson's rule*, defines the FM transmission bandwidth W_{FM} in terms of the *maximum* frequency deviation of a sinusoid, denoted by $\Delta\omega_{\max}$, and the message bandwidth W as

$$W_{\mathrm{FM}} = 2\left(\Delta\omega_{\max} + W\right). \tag{2.8}$$

For non-sinusoidal message signals, however, $\Delta\omega_{\max}$ is usually replaced by the RMS deviation $\Delta\omega$ [4]. This approach is followed in the sequel of this thesis.

The distortion corresponding to the bandwidth given by (2.8) generally lies somewhere between 1-10 % [2], which is somewhat too large for most practical purposes. A more appropriate bandwidth estimate, corresponding to a distortion of around 1%, is given by [2]

$$W_{\mathrm{FM}} = 2\left(\Delta\omega_{\max} + 2W\right). \tag{2.9}$$

For convenience, however, the various examples considered in the sequel of this thesis use Carson's rule, with $\Delta\omega_{\max}$ replaced by $\Delta\omega$, despite the fact that a slightly larger bandwidth is required in practice.

2.3 Performance at High Input CNRs

The wide-spread application of frequency modulation to a large variety of communication systems, is mainly due to the ability of FM to improve the receiver output SNR, i.e. the transmission performance, in comparison to AM and PM transmission. This section briefly outlines the performance of FM transmission systems in the presence of small noise, expressed in terms of the receiver output SNR, and compares it to AM and PM transmission. As explained in Chapter 3, the latter types of modulation are of considerable interest in FM demodulator design.

The investigation of the output SNR requires an appropriate description of the FM demodulator input signal and noise. This description, used throughout the thesis, is discussed in Section 2.3.1. Subsequently, Section 2.3.2 considers the demodulator output noise power spectral density, and the maximum FM receiver output SNR. Finally, Section 2.3.3 compares the performance of AM, PM and FM transmission systems.

2.3.1 Demodulator Input Signal and Noise

This section describes the model for the demodulator input signal and noise, that is used throughout the thesis. Unless stated otherwise, it is assumed that the input signal consists of the FM wave $s(t)$ from (2.1), and band-limited, zero-mean Gaussian noise, denoted by $n(t)$. This noise generally originates from the communication channel, but may also represent the noise produced by various electronic circuits inside the FM receiver.

A phasor representation of the input signal is depicted in figure 2.3. In this figure, the phasor $\vec{s}$ represents the FM wave $s(t)$, $\vec{n}$ represents the input noise $n(t)$, and $\vec{r}$ represents the composite, noisy FM input wave $r(t) = s(t) + n(t)$. Further, all phasors are translated in frequency by the complex factor $\exp(-\mathrm{j}\omega_o t)$.

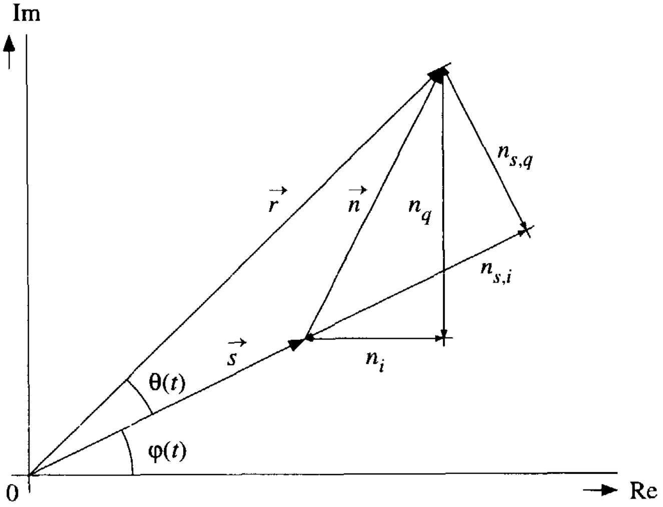


Figure 2.3: Phasor representation of the demodulator input signal.

The meaning and properties of the other variables in this figure are discussed below.

Characteristics of Band-Limited Gaussian Noise

The Gaussian input noise $n(t)$ is assumed to possess an (approximately) rectangular power density spectrum of bandwidth W_n (rad/s), centered around the FM carrier frequency ω_o. This spectrum, denoted by $S_{n,\mathrm{bp}}(\omega)$, is depicted in figure 2.4a. The power contents of this noise, denoted by $P_n = \sigma_n^2$, expressed in terms of the spectral intensity N_o, equals

$$P_n = \sigma_n^2 = \frac{N_o W_n}{2\pi}. \tag{2.10}$$

This noise process can, according to figure 2.3, be expressed in terms of two

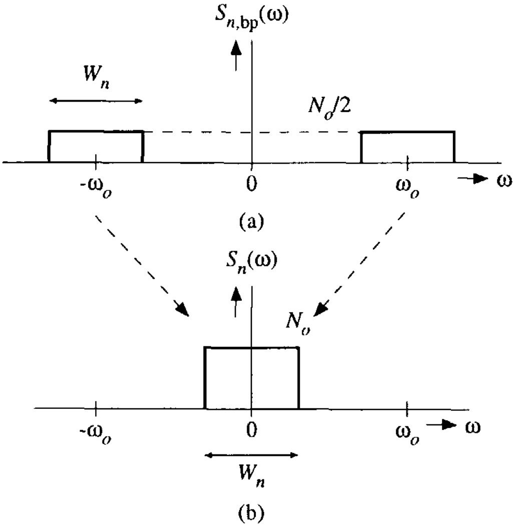


Figure 2.4: Input noise power density spectrum. a) spectrum of $n(t)$, b) spectrum of $n_i(t)$ and $n_q(t)$.

low-pass noise processes $n_i(t)$ and $n_q(t)$ as [2, 4], as

$$\begin{aligned} n(t) &= \mathrm{Re}\left\{[n_i(t) + \mathrm{j}n_q(t)]\exp(\mathrm{j}\omega_o t)\right\} \\ &= n_i(t)\cos\omega_o t - n_q(t)\sin\omega_o t. \end{aligned} \tag{2.11}$$

Thus, $n(t)$ equals the projection on the real axis of $\vec{n}\cdot\exp(\mathrm{j}\omega_o t)$.

It can be shown [2], that when the RF/IF input noise $n(t)$ is zero-mean Gaussian noise with variance σ_n^2, both low pass equivalent processes $n_i(t)$ and $n_q(t)$ are also zero-mean Gaussian noise processes with variance σ_n^2. Furthermore, $n_i(t)$ and $n_q(t)$ are statistically independent. The intensity of their power density spectrum, obtained by frequency translation of the spectrum of $n(t)$, as depicted in figure 2.4b, equals twice the intensity of the spectrum of $n(t)$.

Characteristics of the In-Phase and Quadrature Noise

In order to obtain an expression for the demodulator output frequency noise, i.e. the time-derivative of the phase noise $\theta(t)$ depicted in figure 2.3, the decomposition of $n(t)$ into components in-phase and in quadrature with $s(t)$, denoted by $n_{s,i}(t)$ and $n_{s,q}(t)$ respectively, is is more convenient than the decomposition into $n_i(t)$ and $n_q(t)$.

In terms of the former decomposition, $n(t)$ can be expressed as

$$n(t) = n_{s,i}(t)\cos[\omega_o t + \varphi(t)] - n_{s,q}(t)\sin[\omega_o t + \varphi(t)]\,. \tag{2.12}$$

The first term in this decomposition is in-phase with $s(t)$, while the second term is in quadrature with $s(t)$.

The processes $n_{s,i}(t)$ and $n_{s,q}(t)$ are generally *not* Gaussian distributed, since they depend on the modulation $\varphi(t)$. As considered in detail in Chapter 6 and the corresponding appendices, both processes are *not* mutually independent, but only uncorrelated, since the modulation (instantaneously) disturbs the symmetry of their power density spectrum. The latter is a necessary condition for their independence [2].

An expression for these components follows from the observation that they equal $n_i(t)$ and $n_q(t)$, rotated over the angle $\varphi(t)$, as observed from figure 2.3. Therefore,

$$n_{s,i}(t) = n_i(t)\cos\varphi(t) + n_q(t)\sin\varphi(t), \tag{2.13}$$

$$n_{s,q}(t) = -n_i(t)\sin\varphi(t) + n_q(t)\cos\varphi(t). \tag{2.14}$$

Thus, $n_{s,i}(t)$ and $n_{s,q}(t)$ equal $n_i(t)$ and $n_q(t)$, modulated in phase by the message phase $-\varphi(t)$; in frame of reference corresponding to $n_{s,i}(t)$ and $n_{s,q}(t)$, $n_i(t)$ and $n_q(t)$ seem to be modulated in phase by $-\varphi(t)$, while the FM wave $s(t)$ is considered as an 'unmodulated' carrier. The power contents of $n_{s,i}(t)$ and $n_{s,q}(t)$ also equals $P_n = \sigma_n^2$, which easily follows from (2.13) and (2.14).

Since $\cos\varphi(t)$ and $\sin\varphi(t)$ are FM waves with a zero-valued carrier frequency, it follows from the quasi-stationary approximation (see also Chapter 6), that the power density spectrum of $n_{s,i}(t)$ and $n_{s,q}(t)$ equals the convolution of the spectrum $S_n(\omega)$ of $n_i(t)$ and $n_q(t)$, and the power density spectrum of an FM wave with a zero-valued carrier frequency; it seems that the center-frequency of $S_n(\omega)$ moves around $\omega = 0$ in the rhythm of $\dot{\varphi}(t)$. This is illustrated by figure 2.5. Due to the modulation, the bandwidth of the time-averaged (or ensemble averaged) spectrum, denoted by $S_{n,s}(\omega)$ is slightly larger than the bandwidth of $S_n(\omega)$, but its area is the same. Further, inside the baseband, $S_{n,s}(\omega)$ essentially equals $S_n(\omega)$.

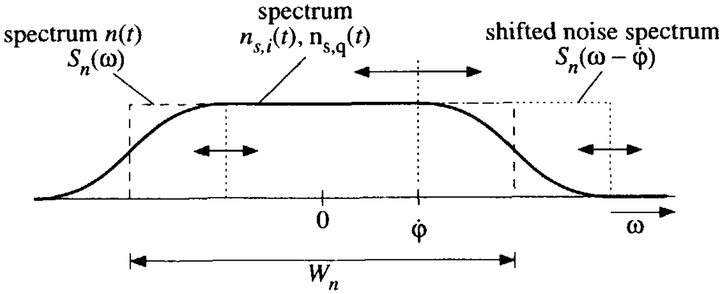


Figure 2.5: Quasi-stationary approximation of the power density spectrum of $n_{s,i}(t)$ and $n_{s,q}(t)$.

Figure 2.6 depicts the quasi-stationary approximation of $S_{n,s}(\omega)$ and the exact spectrum for a sinusoidal message signal, equal to the convolution of $S_n(\omega)$ and the spectrum of the Bessel function expansion given by (2.6). The agreement between the approximation and the exact spectrum is quite good. Further, in the baseband, the spectrum is identical to the rectangular spectrum $S_n(\omega)$.

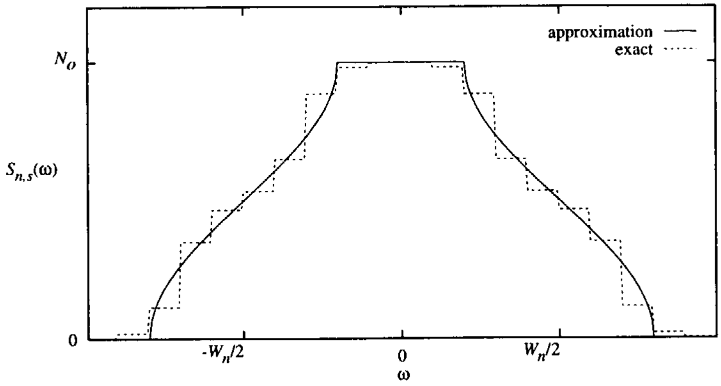


Figure 2.6: Exact and quasi-stationary approximated noise spectrum $S_{n,s}(\omega)$.

Composite Demodulator Input Signal

The composite demodulator input wave can easily be expressed in polar format, i.e. as an amplitude and phase/frequency modulated wave, with the aid of the previously discussed decomposition of the noise.

When $R(t)$ denotes the amplitude of this wave, the length of $\vec{r}$, which is modulated by the noise $n(t)$, and $\theta(t)$ denotes the phase noise (see figure 2.3), then

$$\begin{aligned} r(t) &= s(t) + n(t) \\ &= R(t)\cos\left[\omega_o t + \varphi(t) + \theta(t)\right]. \end{aligned} \tag{2.15}$$

With the aid of figure 2.3, $R(t)$ and $\theta(t)$ can be expressed as

$$R(t) = \sqrt{\left[A + n_{s,i}(t)\right]^2 + n_{s,q}(t)^2}, \tag{2.16}$$

$$\theta(t) = \arctan\left[\frac{n_{s,q}(t)}{A + n_{s,i}(t)}\right]. \tag{2.17}$$

2.3.2 Output Noise Spectrum and Maximum Output SNR

With the aid of the description of the demodulator input signal described in the previous section, the demodulator output noise spectrum and output SNR can be determined quite easily. In this section, we discuss the output noise spectrum and output SNR of an ideal FM demodulator, for small input noise, i.e. high input carrier-to-noise ratios (CNR). This SNR equals the maximum possible SNR, that cannot be exceeded by any FM demodulator.

Response of the Ideal FM demodulator

The response of the ideal FM demodulator to the noisy FM wave $r(t)$ from (2.15) is proportional to the instantaneous frequency of $r(t)$, without the carrier component, i.e.

$$y_{\text{dem,id}}(t) = \dot{\varphi}(t) + \dot{\theta}(t), \tag{2.18}$$

where the frequency noise $\dot{\theta}(t)$, obtained from (2.17), equals

$$\dot{\theta}(t) = \frac{[A + n_{s,i}(t)]\, \dot{n}_{s,q}(t) - n_{s,q}(t)\dot{n}_{s,i}(t)}{[A + n_{s,i}(t)]^2 + n_{s,q}^2(t)}. \tag{2.19}$$

At high input CNRs, i.e. when $A \gg n_{s,i}(t)$, this expression reduces to $\dot{\theta}(t) \approx \dot{n}_{s,q}(t)/A$. Thus, the frequency noise basically consists of the time-derivative of the quadrature noise $n_{s,q}(t)$.

Output Noise Spectrum at High CNRs

At high input CNRs, the power density spectrum of $\dot{\theta}(t)$, denoted by $S_{\dot{\theta}}(\omega)$, can be expressed as

$$S_{\dot{\theta}}(\omega) \approx \left(\frac{\omega}{A}\right)^2 S_n(\omega), \tag{2.20}$$

where A denotes the amplitude of $s(t)$. The details of the calculation of this spectrum are considered in Chapter 6. The spectrum $S_{\dot{\theta}}(\omega)$ is sketched in figure 2.7.

Comparison with the input noise spectrum of figure 2.4 shows that the *differentiation* of the carrier phase performed by the demodulator applies *quadratic shaping* to the input noise spectrum. By virtue of this shaping, the largest part of the output noise power is shifted towards high frequencies, leaving only a very low noise level at low frequencies, i.e. inside the base band region where the message signal resides. This *shaping mechanism is the essence of the SNR improvement achieved with FM*, in comparison to AM.

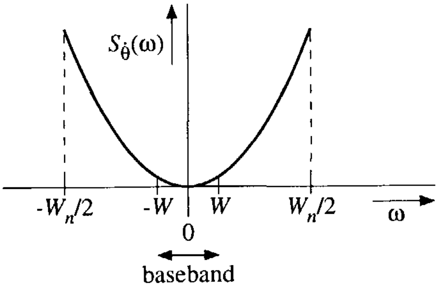


Figure 2.7: Demodulator output frequency noise spectrum.

Output SNR at High CNRs

The maximum possible output SNR, assuming a rectangular low-pass baseband filter of bandwidth W, is easily obtained by integration of $S_{\dot\theta}(\omega)$ over the frequency interval $\omega \in [-W, W]$. A detailed analysis in Chapter 6 shows in which way arbitrary baseband filter characteristics can be included into the output SNR.

The resulting expression for a rectangular filter equals

$$\mathrm{SNR_{max}} = 3p\frac{W_n}{W}\left(\frac{\Delta\omega}{W}\right)^2, \tag{2.21}$$

where p denotes the input CNR, given by

$$p \overset{\text{def}}{=} \frac{A^2}{2\sigma_n^2}. \tag{2.22}$$

The factor pW_n/W represents the demodulator input CNR that includes only the input noise located inside twice the message bandwidth, i.e. a bandwidth of $2W$. This CNR, denoted by γ, is of interest in a comparison of FM with other modulation schemes.

The most apparent property of FM reflected by (2.21) is, that the output SNR can be increased by increment of the frequency deviation $\Delta\omega$. Thus, the frequency deviation $\Delta\omega$ allows an *exchange between the output SNR and the required transmission bandwidth.*

Of course, (2.21) holds only for high input CNRs. For low CNRs, typically below 10 dB, a *threshold* in the output SNR versus input CNR curve is observed. Above this threshold, the output SNR is properly described by (2.21). However, below the threshold, the output SNR decreases much faster than predicted by (2.21) (see Chapter 5).

2.3.3 Transmission Performance of AM, PM and FM

As a final note on the characteristics of FM, it is interesting to compare the transmission performance of FM systems with the performance of AM and PM systems. As discussed in Chapter 3, both amplitude modulation (AM) and phase modulation (PM) play an important role in FM demodulator design.

Transmission Efficiency

As implied by Shannon's information theory [1], a suitable measure for the performance of a modulation scheme is the maximum possible 'distortion-free' *information rate*, denoted by R in (bit/s), at the demodulator output for a given channel capacity, denoted by C in (bit/s). The relation between R and C essentially describes the efficiency of the modulation scheme; a scheme is maximally efficient when R equals C, i.e. when it realizes an as large as possible demodulator output SNR for a given channel bandwidth, and CNR inside the channel.

The channel capacity of a channel with a rectangular spectrum of bandwidth $B = W_n/(2\pi)$ Hz, that contains additive white Gaussian noise (AWGN) of a *single-sided* spectral intensity N_o is given by the well-known expression

$$C = B\log_2\left(1+\frac{S}{N}\right) = B\log_2\left(1+\frac{A^2}{2N_oB}\right), \tag{2.23}$$

where S/N denotes the CNR inside the channel, previously denoted by p.

Performance of the Theoretical Optimum Modulation Scheme

As derived in [2], the maximum possible output SNR that can ever be achieved by communication through this channel with the aid of a theoretical, 'optimum' modulation scheme equals

$$\mathrm{SNR}_{\mathrm{opt}} = \left(1+\frac{\gamma}{b}\right)^b - 1 = (1+p)^b - 1, \tag{2.24}$$

where $\gamma = pW_n/W$ denotes the input CNR inside twice the message bandwidth, and $b = W_n/W$ equals the ratio of the transmission bandwidth and the message bandwidth.

Performance of AM-DSB, PM and FM

For DSB, i.e. AM with suppressed carrier, PM and FM, the output SNR can be expressed [2] as

$$\mathrm{SNR}_{\mathrm{DSB}} = \gamma, \tag{2.25}$$

$$\mathrm{SNR}_{\mathrm{PM}} = \frac{1}{4}(b-2)^2\gamma \leq \pi^2\gamma, \tag{2.26}$$

$$\mathrm{SNR}_{\mathrm{FM}} = \frac{3}{4}(b-2)^2\gamma, \tag{2.27}$$

where Carson's rule, roughly valid for $b > 6$, is used as estimate for the FM and PM transmission bandwidth. The upper bound on PM is due to the fact that the phase deviation in PM is not allowed to exceed π for unambiguous demodulation [2].

Comparison

Expression (2.25) demonstrates the known property of linear modulation schemes, that the output SNR of an AM-DSB system cannot be improved by widening of the transmission bandwidth, since it is independent of b.

Expression (2.26) and (2.27) show, that the output SNR of an FM system is always at least 4.8 dB higher than the output SNR of a comparable PM system, due to the noise shaping. Further, notice that FM is the only scheme of the three that, theoretically, allows an *unlimited exchange between the required transmission bandwidth and the output SNR.*

Figure 2.8 depicts the output SNR versus input CNR curves of the optimal scheme, AM-DSB, PM and FM for the maximum possible phase deviation for PM, corresponding to $b = 2(\pi + 1)$. This figure clearly shows that FM is the most 'efficient' of the three practical modulation schemes. Further, it shows that, since the FM output SNR described by (2.21) and (2.27) crosses through the curve of the optimal scheme, which is impossible, a threshold must occur somewhere. The position of this threshold, and the demodulator response observed in that region of the SNR curve, is determined by the internal structure of the demodulator, and is extensively studied in the sequel of this thesis.

2.4 Conclusions

The modulation scheme, that describes in which way the message information is included into the carrier wave, determines to a large extent the structure of the transmitter and the receiver in a communication system. Further, reliable transmission generally requires a deterministic, periodic carrier wave.

In modulation schemes that divide the available channel capacity by means of Frequency Division Multiplexing (FDM), such as FM, sinusoidal carriers are

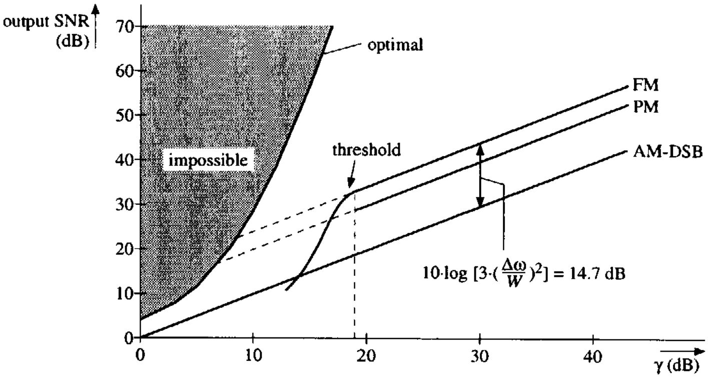


Figure 2.8: Output SNRs versus input CNR of the theoretical optimal modulation scheme, AM-DSB, PM and FM modulation.

generally favorable, since they require the smallest possible transmission bandwidth.

One of the most important properties of the FM scheme is its ability to exchange the required transmission bandwidth for an increase of the demodulator output SNR. This exchange, due to the nonlinear nature of the FM scheme, is controlled by the frequency deviation, or equivalently the FM modulation index.

The bandwidth of FM waves is theoretically infinite. However, in practice, by far the largest portion of the carrier power is concentrated in finite bandwidth. An estimate for this bandwidth, as used in FM receiver design, is always associated to a certain level of distortion introduced into the message by filtering.

For small modulation indices, the spectrum of FM waves closely resembles the spectrum of an AM wave, and possesses roughly the same bandwidth. For large indices, the spectrum resembles the probability density function of the FM message signal.

The improvement of the signal-to-noise ratio (SNR) established by FM transmission in comparison to AM and PM transmission is, besides the theoretical unlimited possibility to exchange bandwidth for SNR improvement, due to the quadratic shaping applied by the (ideal) FM demodulator to the input noise spectrum. This shaping moves the largest part of the noise power to frequencies located outside the message bandwidth.

Finally, on the basis of information theoretical considerations, it follows that a threshold must occur in the FM demodulator output SNR at low input CNRs, that results in a steep decay of the output SNR.

References

[1] C.E. Shannon, "A mathematical theory of communication", *The Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, 1948.

[2] A. Bruce Carlson, *Communication Systems, An introduction to Signals and noise in Electrical Communication*, McGraw-Hill International Editions, Singapore, 1986.

[3] Mitsuru Ishizuka and Kenkichi Hirade, "Optimum Gaussian filter and deviated-frequency-locking scheme for coherent detection of MSK", *IEEE Transactions on Communications*, vol. 28, no. 6, pp. 850–857, June 1980.

[4] David Middleton, *An Introduction to Statistical Communication Theory*, McGraw-Hill Book Company, New York, 1960.

[5] John R. Carson, "Notes on the theory of modulation", *Proceedings of the IRE*, vol. 10, no. 2, pp. 57–64, Feb. 1922.

Chapter 3

FM Demodulation Principles

FM demodulation principles describe the essential operation of FM demodulator circuits and systems. They allow understanding of the operation of the demodulator at a high hierarchical level. Such high level models are extremely valuable in FM demodulator design, since they show the potential capabilities of the demodulator, long before an actual demodulator circuit has been developed. This gives the opportunity to test the demodulator capabilities against the requirements and to include possibly required improvements in a very early design stage.

This chapter develops a classification of FM demodulation principles, that groups FM demodulators operating according to the same or similar principles together, resulting in an overview of the full range of possible FM demodulator operating principles. Such a classification is a powerful instrument in demodulator design, since it allows a deliberate selection of a suitable demodulator for each application, on the basis of high level requirements and design aspects.

An overview of this chapter is as follows. Section 3.1 discusses the general principles of the structured design strategy, applied to FM demodulator design in this thesis. Based on the principles of this strategy, Section 3.2 outlines the main hierarchical levels in the FM demodulator design procedure, and discusses the important function of a classification. Section 3.3 starts the development of the classification and identifies the basic FM demodulator functions. Section 3.4, 3.5 and 3.6 study the characteristics of the classes of FM demodulators that evolve from Section 3.3. The resulting FM demodulator classification is summarized in Section 3.7.

3.1 Structured Design

Circuit and system design may generally be represented as a very complex search process, heading for the circuit/system that best meets the specifications. From a theoretical point of view, such a search is situated in a large 'design space', consisting of all possible system implementations.

Due to their complexity, satisfactorily solution of such design problems requires the application of some design strategy. Without such a strategy, it is very unlikely that the best possible system, or even a suitable system, is found within a finite time.

In this thesis, a so called "structured design" strategy is applied to the design of frequency demodulators. This section outlines the general principles of this strategy. Section 3.2 considers its implications on demodulator design.

We first consider the general objective of the structured design strategy in Section 3.1.1. Subsequently, a high-level view on the design problem and its solution is discussed in Section 3.1.2 and Section 3.1.3 respectively.

3.1.1 Objective of Structured Design

Although many design strategies are able to find suitable system implementations for various applications, it is often uncertain whether the final solution is indeed the best possible, or at least close to it.

In order to reduce this uncertainty, knowledge of, and insight into the fundamental limitations on the system performance is required. *Explicit* relations between the system performance and the performance bounds are often absent in design strategies. However, despite their absence, designers are usually able to apply ***implicit*** knowledge of them, contained in their experience. Although this implicit knowledge often results in improved circuits/systems, it still cannot guarantee that the best possible solution is attained.

It is the objective of the structured design strategy, which has been applied to various types of electronic systems [1–5], to acquire the explicit relations between the system performance and its fundamental limitations. In general, knowledge of these bounds significantly increases the speed and efficiency of the design procedure, and allows quick estimation of the feasibility and performance of the design solution in advance, before it has ever been constructed.

3.1.2 Definition of Design Problem

Formulation of the relation between the system performance and its limitations, requires a proper definition of the various notions involved in a design problem. These definitions are outlined below.

Ideal System Function The *ideal system function* describes the primary function of the system under design. This function, represents the main design objective, and describes the operation of the system at the highest possible hierarchical level, free from errors of any possible cause. The final (physical) design solution is an implementation of this function, that maps it on physical relations. In this view, an electronic FM demodulator is an implementation of the ideal FM demodulation function, that maps this function on relations between electric currents and voltages.

Physical Limitations The mapping of the ideal function on physical relations is inevitably subjected to errors, introduced by the fundamental physical mechanisms that are chosen to constitute the system's operation. These "*physical limitations*" [5–7], such as noise, distortion and bandwidth limitations, define a fundamental upper bound on the system performance, that cannot be overruled by any means, except by selection of other physical mechanisms, materials, or technologies.

Resource Limitations *Resource Limitations* are another cause of errors in the physical system. They limit the types and amount of resources, as e.g. chip area, power consumption and production costs, that may be used to implement the ideal system function. In fact, resource limitations correspond to those specifications that describe the maximum 'costs', of the system. As opposed to physical limitations, performance bounds set forward by these limitations may be overruled by increasing the available amount of resources.

Functional Requirements Due to the inevitability of errors in the physical system, a part of the specifications has to specify the types and magnitude of errors that can be tolerated, often by means of some cost-function. These specifications will be called the "*functional requirements*".

3.1.3 Solution of the Design Problem

The design strategy has to assure that a suitable design solution is obtained, that implements the ideal system function, minimizes the cost-function, comprising the system requirements, and simultaneously complies with the physical limitations and resource limitations. The design approach attempts to satisfy these requirements in a structured way by introduction of *hierarchy*, *simplification* and *orthogonality*.

Hierarchy *Hierarchy* reduces the complexity of the design problem, by subdivision of the original problem into smaller, more readily solved sub-problems.

This procedure encloses the best/optimal design solution, and step by step reduces the 'radius' of the enclosure by gradual inclusion of details, system requirements and imperfections, until finally only the optimal solution is left. A classification of the possible operating principles of the system, as developed for FM demodulators in this chapter, is a valuable instrument to establish this subdivision.

Simplification *Simplification* is a powerful instrument to maintain comprehensibility of the procedure, and allows controlled introduction of detail, limited to the minimum required, in each design step. In this way, the first and most important design steps can be covered by simple models that reveal only the essential system characteristics.

Orthogonality Finally, the approach aims at an arrangement of the system in such a way that *orthogonality* between the various design parameters is established. When this is achieved, iterations in the design procedure are eliminated and the various system characteristics can be optimized independently.

3.2 Design Hierarchy for FM demodulators

The previous section discussed that hierarchy is an efficient means to reduce the complexity of the design problem.

In this section, we focus on the hierarchy in the design of FM demodulators, as schematically depicted in figure 3.1, and show the importance of the classification of FM demodulation principles developed in this chapter.

3.2.1 Ideal FM Demodulation Function

As discussed in Section 3.1.2, the main design objective is represented by the ideal system function. This function describes the system operation at the highest hierarchical level.

The ideal FM demodulation function depicted in figure 3.1 describes the primary function of FM demodulators; retrieval of the instantaneous frequency from FM waves. Suppose that the FM wave to be demodulated is represented by $s(t)$ from (2.1). In that case, the ideal FM demodulation function of the demodulator that operates on $s(t)$, $f_{\text{FM,dem}}(\dots)$, satisfies the equation

$$f_{\text{FM,dem}}\left[s(t)\right] = f_{\text{FM,dem}}\left\{A(t)\cos\left[\omega_o t + \varphi(t)\right]\right\} = \dot{\varphi}(t). \tag{3.1}$$

Unfortunately, this expression is an implicit description of the ideal function whereas an explicit one, that relates the demodulator input signal to its output signal by means of known, (basic) operators and functions, e.q. a differential

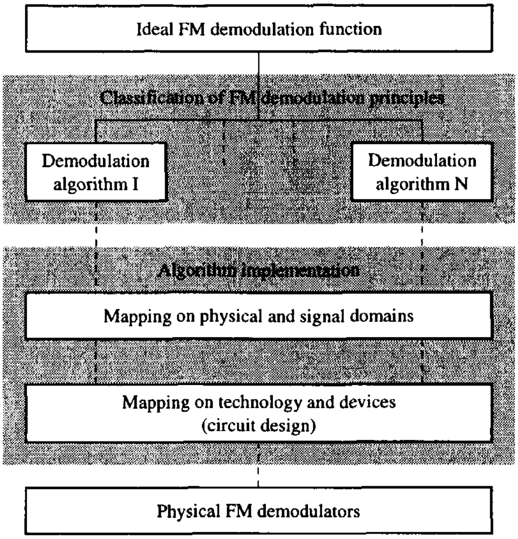

Ideal FM demodulation function
Classification of FM demodulation principles
Demodulation algorithm I
Demodulation algorithm N
Algorithm implementation
Mapping on physical and signal domains
Mapping on technology and devices (circuit design)
Physical FM demodulators

Figure 3.1: Hierarchy in FM demodulator design.

equation, is required for its implementation. Expression (3.1) seems however not to be satisfied by any known basic mathematical function. FM demodulation should therefore be the result of a sequence of such basic operations and functions.

3.2.2 FM Demodulation Principles

The second level in the FM demodulator design hierarchy of figure 3.1 consists of the FM demodulation principles. These principles explicitly describe the algorithms that can be used to implement the ideal FM demodulation function into a physical system. In fact, they 'implement' the ideal FM demodulation function in a mathematical sense as a sequence of basic functions and operations. Due to the complexity of the demodulation function, several valid sequences exist, that describe different FM demodulation principles. The purpose of this chapter is to classify all possible sequences.

The operating principles may thus be represented by e.g. a block schematic or a graph, that connects basic mathematical operators and functions in a predefined order. In many cases, these basic functions correspond to known subsystems that can be readily implemented into a physical system. For example, in electronics, the following set of basic functions is commonly encountered:

- multiplication by a constant, implemented by amplifiers;

- addition, e.g. two current sources that float into the same node;
- multiplication of time-variant signals, implemented e.g. by mixers;
- amplitude and frequency references, implemented by e.g. band-gap references and oscillators respectively;
- differentiation and integration, implemented by capacitors or inductors.

Such basic electronic building blocks can be used to construct an electronic FM demodulator.

Each FM demodulation principle is a high-level description of a particular type of FM demodulator architecture. Demodulators that are based on the same demodulation principle, behave similar for all characteristics, as far as these are intrinsically described by this principle. Consequently, the behavior corresponding to such characteristics is similar for all demodulators based on the same architecture, irrespective of their implementation in e.g. analog electronics, digital electronics or, if appropriate, in pneumatics. An example of such a characteristic is the response to external noise. This response can be found directly from the demodulation algorithm and is therefore basically equal for all demodulators based on the same algorithm.

3.2.3 Implementation of FM Demodulation Principles

The remaining design steps are concerned with the implementation of the FM demodulation principle into a physical system. This starts with the mapping of the various information carrying signals in the mathematical demodulation algorithm on physical quantities. Subsequently, appropriate circuits are designed to process these physical signals.

Mapping of Information on Physical signals

Two selections are required in order to map mathematical signals on physical signals:

- selection of the signal domain used to represent the information;
- selection of the physical domain used to represent the information.

Signal Domain The selection in the signal domain determines the distribution of signal energy, and thus of information, over time/frequency and amplitude. As illustrated in figure 3.2, four different domains can be distinguished:

- the continuous domain, consisting of signals that are continuously distributed in time and amplitude;

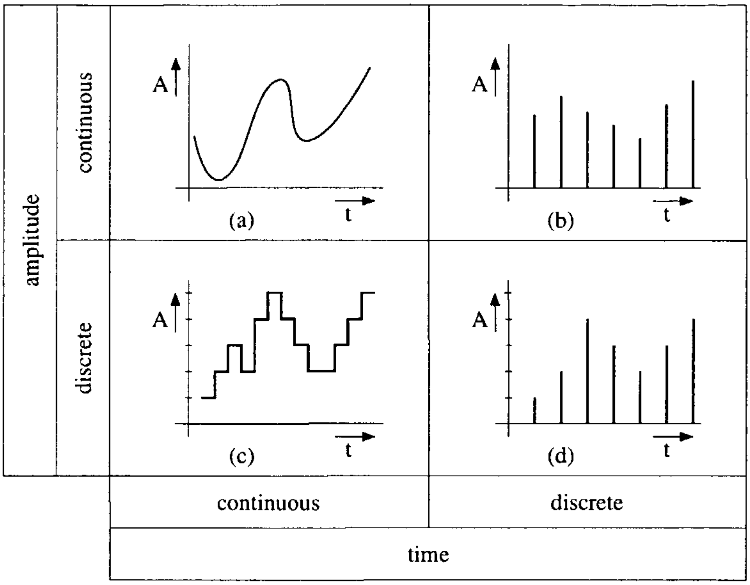


Figure 3.2: The four different signal domains: a) continuous domain, b) sampled domain, c) quantized domain, d) digital domain.

- the sampled domain, consisting of signals distributed discrete in time and continuously in amplitude;
- the quantized domain, consisting of signals that are distributed continuously in time, but discrete in amplitude;
- the digital domain, consisting of signals that are distributed discrete in both time and amplitude.

It should be noted that although the selection of one or several of these domains for representation of the mathematical information signals does strongly influence the appearance of the final FM demodulator system, it does however not affect the demodulator characteristics intrinsically determined by the demodulation principle. Differences observed in the behavior of demodulators based on the same operating principle, but realized in different signal domains are therefore exclusively caused by differences in the character of these domains.

In Section 3.4.6 is shown, for example, that a digital FM demodulator, of which a patent proposal was recently filed, is basically equal to its much older analog counterpart, invented already before 1920. The operating principles are identical, only the signal domains, and the corresponding circuitry, are different.

An illustration of the fact that signals from several domains may be present within one and the same demodulator is described in [8, 9]. These FM demodulators operate as a kind of delta-sigma modulator on the instantaneous frequency of the input FM wave. The input wave is part of the continuous domain, while the output signal is a digital bit stream, and thus belongs to the digital domain.

Physical Domain The selection in the Physical Domain determines the physical quantities used by the demodulator for the representation of information. These quantities may be selected from six different physical domains:

- the radiant domain
- the mechanical domain
- the chemical domain
- the thermal domain
- the magnetic domain
- the electrical domain

Again, the selection of the physical domain(s) used to represent the information does influence the appearance of the demodulator, but it does not affect the demodulator characteristics described by the demodulation principle; the representation of the information does not affect the character of the information processing prescribed by the demodulation principle.

Although the sequel of this thesis is mainly concerned with electronic demodulators, situated in the electrical domain, the various FM demodulation principles can theoretically be implemented in other domains as well.

Illustrations of this fact can be found in literature. For example, up to 1912, when the "regenerative circuit" was invented, magnetic detectors were a popular type of receiver for radio communication [10]. Another example is found in [11], where the so called "Leitungsdemodulator" is described. This FM demodulator is partly realized in the mechanical domain – it uses the geometric wave-length of the received FM wave to determine the frequency – and partly in the electrical domain; the wave length information is converted to a differential voltage.

Circuit Design

Once the physical signals have been selected, the appropriate circuits have to be designed in order to process these signals. In this design step, restrictions are put on the information handling capacity of the demodulator system by physical limitations and resource limitations. The physical limitations originate from the physical mechanisms that constitute the operation of the circuit building blocks, such as transistors, electron tubes and resistors.

3.3 Direct and Indirect Demodulation

According to the previous section, FM demodulation principles describe the algorithms that implement the demodulation function as combinations of basic operators and functions. This section start the development of a classification of FM demodulation principles, by an investigation of the two main classes that can be distinguished: direct and indirect demodulation principles.

Direct demodulation principles straight-out read the message information from the instantaneous frequency of FM waves, as suggested by the ideal FM demodulation function described in Section 3.2.1, and produce an output signal proportional to this frequency. No use is made of other carrier wave parameters, such as the instantaneous phase or amplitude. Consequently, an FM demodulator that implements such a direct FM demodulation principle is able to retrieve the message information from FM waves that are, besides modulated in frequency, also modulated in amplitude and/or phase, e.g. due to noise, interference and fading in the communication channel. Unfortunately, as shown in Section 3.3.1, it appears that physical systems can impossibly read the instantaneous frequency of a carrier wave directly. For this reason, physical FM demodulators that are based on direct demodulation principles do not exist.

Indirect demodulation principles copy the FM message information to the carrier amplitude or phase, and subsequently apply AM or PM demodulation. In this way, they avoid straight-out reading of the instantaneous frequency. These demodulation principles are available for the construction of physical FM demodulators. Section 3.3.2 subdivides the class of indirect demodulation principles into two subclasses, that are separately discussed in Section 3.3.3 and Section 3.3.4.

From the discussion of both classes of FM demodulation principles, a number of basic functions can be identified that necessarily need to be performed by any FM demodulator. These basic functions are discussed in Section 3.3.5.

3.3.1 Direct Demodulation Principles

As stated in the introduction, physical FM demodulators based on direct demodulation principles do not exists, due to the fact that physical systems are unable to read the instantaneous frequency of a carrier straight-out.

This section discusses the underlying mechanisms that hamper a straight-out read of the instantaneous frequency. First, the boundary conditions on the detection of information in general are considered. Subsequently, it is shown that the instantaneous frequency does not comply with these conditions.

Detection of Information

Measurement and detection systems, as FM demodulators, essentially detect the energy supplied to their input. Therefore, detection of information by such systems requires encoding of this information in the instantaneous energy of the wave supplied to their input. Similarly, a human being is only able to read the newspaper, i.e to gather information, when light (optical energy) that is 'modulated' by the newspaper reaches the eye.

Thus, information is detectable only when it 'modulates' the energy of some signal. This statement follows from a fundamental property of information transport in physical systems, described both by information theoretical and physical laws. For example, Shannon's theory [12] states that a nonzero channel capacity, i.e. the capability to transport a nonzero amount of information, exists only when the information signal possesses a nonzero power/energy contents. In quantum mechanics, a similar statement is formulated by Heisenberg's Uncertainty Principle [13, 14]. In essence, this principle states the impossibility to measure the impulse, position, energy, etc. of a particle without interaction, i.e. without exchanging energy with the particle.

Detectability of the Instantaneous Frequency

In the context of demodulators, the previous discussion shows that directly detectable/readable message information necessarily modulates parameters associated with the carrier wave's instantaneous energy. That the instantaneous frequency is not such a parameter, and therefore cannot be detected directly, may be observed with the aid of the phasor representation of the FM wave $s(t)$, depicted in figure 3.3. The instantaneous energy of the FM wave is directly associated with its instantaneous value $s(t)$, i.e. the projection of the phasor $\vec{s}$ on the real axis. Together with the projection on the imaginary axis, $s_q(t)$,

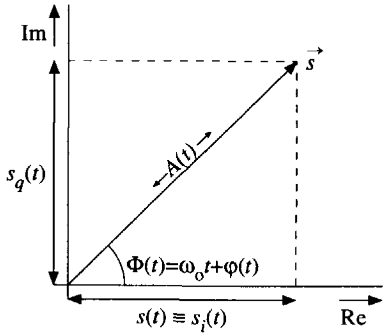


Figure 3.3: Phasor diagram of an FM wave.

$s(t) \equiv s_i(t)$ describes the *position* of the phasor tip in the phasor plane as function of time.

Thus, the instantaneous energy of the FM wave is associated to the Cartesian position coordinates of the phasor tip, which correspond to $s_i(t)$ and $s_q(t)$. Actually, $s_q(t)$ is associated to a copy of $s(t)$ that is shifted instantaneously over 90^o, since it refers to the imaginary axis. However, since such a shifted wave is readily constructed from $s(t)$, both $s_i(t)$ and $s_q(t)$ are, for convenience, considered to be associated with the energy of $s(t)$. These coordinates are therefore directly detectable by physical systems.

An equivalent description of the phasor tip position is given by the polar coordinate system consisting of the carrier amplitude $A(t)$ and phase $\Phi(t) = \omega_o t + \varphi(t)$. This coordinate system is related to the Cartesian system by means of an invertible coordinate transform, and thus contains the same information; no information is lost by changing from Cartesian to polar coordinates or vice versa. Therefore, the mutually independent coordinates $A(t)$ and $\Phi(t)$ and can be expressed directly in terms of $s_i(t)$ and $s_q(t)$. For this reason, the amplitude and phase of a carrier wave can be demodulated directly, without copying the message information to other carrier parameters.

The instantaneous frequency of the FM wave $s(t)$ does not correspond to the position of the phasor tip, but to the angular velocity of the phasor $\vec{s}$ in figure 3.3, i.e. the time derivative of $\Phi(t)$. This velocity is however not directly associated with the energy of $s(t)$, and cannot be detected directly. Therefore, a physical FM demodulator is incapable to read the instantaneous frequency directly. Instead, the message information should be derived from detected position coordinates, i.e. from $A(t)$ and $\Phi(t)$, or copied to a position coordinate in advance of its detection. Both these possibilities result in indirect demodulation principles.

3.3.2 Indirect Demodulation Principles

From the previous section can be concluded that all physical FM demodulators operate according to indirect demodulation principles. As opposed to direct demodulation principles, they do not straight-out read the message information from the FM wave's instantaneous frequency, but copy the information to the amplitude or phase, in advance of AM or PM demodulation.

Therefore, it is possible to distinguish two subclasses of indirect FM demodulation principles, that differ in the conversion and demodulation operations applied to the FM wave's instantaneous frequency:

- conversion to amplitude, followed by AM demodulation;

- conversion to phase, followed by PM demodulation.

Such conversions can be applied successfully in FM demodulators only when respectively the amplitude and phase of the FM wave are unmodulated prior to the conversion. Otherwise, when modulation is present in these parameters, e.g. due to noise or interference, it will mix up with, and irrecoverably corrupt the converted FM message information.

Since the instantaneous frequency represents a velocity in the phasor plane, whereas the amplitude and phase represent the polar position coordinates, FM demodulation may be considered as some special kind of velocity measurement. Two different methods of velocity measurements exist. The two classes of indirect FM demodulation principles each implement one of these methods. Figure 3.4 schematically depicts these classes, and their subclasses. Both classes and the corresponding subclasses are discussed in Section 3.3.3 and Section 3.3.4 respectively.

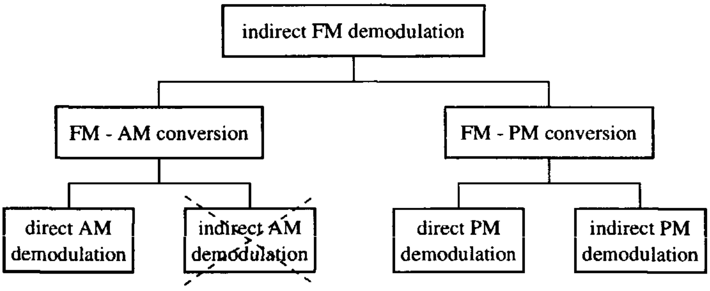


Figure 3.4: Classes of indirect FM demodulation principles.

3.3.3 Indirect Demodulation by FM-AM Conversion

Demodulation through FM to AM conversion and subsequent AM demodulation is equivalent to a velocity measurement, that converts the velocity information to position information (FM-AM conversion), which is subsequently determined by a position measurement (AM demodulation). Both the FM-AM conversion and AM demodulation are briefly discussed below. A detailed discussion is postponed to Section 3.4.

FM to AM Conversion

The conversion of frequency information to amplitude information transforms the original FM wave into a mixed FM-AM wave. Although the frequency modulation remains present, it is disregarded in the sequel of the demodulation process. It is not possible to remove the FM modulation from the FM-AM wave, since this would require an FM demodulator on its own. In fact, any strategy that attempts to remove the FM demodulation without using a separate FM demodulator for this task bounces at the classical "chicken-egg" problem.

AM Demodulation

In succession to the FM-AM conversion, the information has to be detected by means of an AM demodulator in order to obtain the baseband demodulator output signal. As opposed to FM demodulators, necessarily implemented according to indirect demodulation principles, direct and indirect demodulation principles are available for the implementation of AM demodulators. As depicted in figure 3.4, both types of AM demodulation principles each result in a separate subclass of FM demodulation principles.

Direct AM demodulation principles simply read the information contained in the instantaneous amplitude of the wave at their input. According to Section 3.3.1, it is possible to construct physical AM demodulators based on this principle, since the instantaneous amplitude is directly associated with the instantaneous energy of the FM wave.

Indirect AM demodulation principles convert the AM information to PM information, followed by PM demodulation. Although it is possible to implement such a demodulation principle, it cannot be used to construct FM demodulators. This is due to the fact that the phase of FM waves is modulated already, by the integrated FM message, prior to the AM to PM conversion performed by these AM demodulators. Since this conversion operations is principally nonlinear, it is likely to destroy both the converted amplitude information, and the integrated message information in the carrier phase. The AM to PM conversion is therefore not allowed in FM demodulators.

Consequently, FM demodulators based on FM to AM conversion necessarily apply direct AM demodulation.

3.3.4 Indirect Demodulation by FM-PM Conversion

Demodulation through FM-PM conversion and subsequent PM demodulation is equivalent to a velocity measurement, that derives the velocity information (FM-PM conversion) from two consecutive position measurements (PM demodulation) and the elapsed time between them. A detailed discussion of these principles is postponed to Section 3.5.

FM to PM Conversion

The FM-PM conversion of the message information implements the definition formula for the relation between "instantaneous frequency" and "instantaneous phase", i.e. a differentiation to time. The same definition applies to the relation between speed to position. As discussed in Section 3.5, the implementation of this differentiation differs among the various FM to PM conversion principles.

PM Demodulation

Similar to AM demodulation, PM demodulation may be performed according to both direct and indirect demodulation principles, as sketched in figure 3.4.

Direct PM demodulation principles read the information contained in the instantaneous phase of the wave at their input. According to Section 3.3.1, such PM demodulation principle may be used to construct physical PM demodulators, since the phase information can be attained directly from the Cartesian coordinates $s_i(t)$ and $s_q(t)$, without use of other carrier parameters.

Indirect PM demodulation principles convert the PM information to AM information, followed by AM demodulation. As opposed to indirect AM demodulation, indirect PM demodulation may be used to construct FM demodulators. This is due to the fact that, in principle, the FM carrier amplitude is unmodulated prior to the PM-AM conversion performed by these demodulators. This class of demodulation principles is discussed in Section 3.6.

3.3.5 Basic FM Demodulator Functions

At this point in the development of the FM demodulator classification, the three basic FM demodulator operations, to be performed by any physical FM demodulator, can be identified.

The analysis in Section 3.3.1 and Section 3.3.2 showed that FM demodulators are necessarily implemented according to indirect demodulation principles. Therefore, every FM demodulator at least contains a conversion operation, that in some way performs a differentiation to time, and an AM or PM demodulation operation. The third operation, characteristic for any Frequency Division Multiplexing (FDM) modulation scheme, such as FM, is frequency translation. The FM modulator translates the message information from baseband, i.e. zero center frequency, to a center frequency ω_o. The FM demodulator has to perform the reverse translation.

Each FM demodulator therefore performs the following list of basic operations:

- conversion of frequency information to amplitude/phase information;
- amplitude/phase demodulation;
- frequency translation.

3.4 Demodulation by Conversion to AM

This section investigates the principles of operation, and algorithms available for the implementation of FM demodulators based on FM-AM conversion. As discussed in Section 3.3, these demodulators constitute one of the two classes of indirect demodulation principles.

Inspired by the structured design strategy, orthogonalization is applied wherever possible, in order to simplify the design procedure, and to allow separate optimization of the various demodulator sub-functions. The two sub-functions of the FM demodulators considered in this section, FM-AM conversion and AM demodulation, are therefore separately investigated. First, the ideal (sub-) functions are identified. Subsequently, the algorithms available for their implementation are discussed. Some major issues in the implementation of these algorithms into demodulator circuits are addressed in Chapter 4.

An outline is as follows. The ideal FM-AM conversion function and FM-AM conversion algorithm are discussed in Section 3.4.1 and Section 3.4.2 respectively. The ideal AM demodulation function and the two AM-demodulation algorithms obtained from it are considered in Section 3.4.3, Section 3.4.6 and Section 3.4.7. The two subclasses of FM demodulators that evolve from these sections are discussed in Section 3.4.6 and 3.4.7 respectively, and compared with demodulators encountered in literature.

3.4.1 Ideal FM-AM Conversion Function

The ideal FM to AM conversion function establishes a perfectly linear relation, without noise of distortion, between the FM message contained in the instantaneous frequency of the input FM wave and the amplitude of the AM wave at the output. The required transfer is depicted in figure 3.5. For example, the

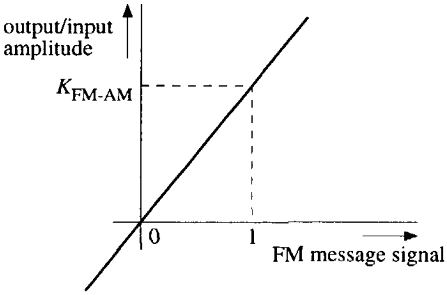


Figure 3.5: Transfer of the ideal FM-AM conversion function.

response to the FM wave $s(t)$ from (2.1) should equal

$$\begin{aligned} s_o(t) &= f_{\text{FM-AM}}\left[s(t)\right] \\ &= K_{\text{FM-AM}}\dot{\varphi}(t)A(t)\cos\left[\omega_o t + \varphi(t) + \varphi_o\right], \end{aligned} \tag{3.2}$$

where φ_o is some fixed phase shift and $K_{\text{FM-AM}}$ denotes the conversion gain.

In practice however, the amplitude of $s_o(t)$ will be proportional to $\omega_o + \dot{\varphi}(t)$, and thus contains an non-informative offset component. The impact of this component on the demodulator output signal is considered in subsequent sections, and in Chapter 4.

3.4.2 FM-AM Conversion Algorithm

This section determines the FM-AM conversion algorithm, required to construct a physical FM-AM converter from basic functions and operators.

First, the algorithm is investigated with the aid of a quasi-stationary approach. Subsequently, FM-AM conversion is related to the phasor representation of the FM wave. Finally, the components of the FM-AM converter response to the FM wave $s(t)$ are analyzed.

Quasi-Stationary Approach

It is illustrative to investigate the FM-AM conversion operation with the aid of the quasi-stationary approximation. As discussed in Section 2.2.2, such an approach describes the FM wave $s(t)$ as a single sinusoid with a frequency that (very) slowly fluctuates in the rhythm of the message information ("moving finger"). In essence, it considers the instantaneous frequency and the spectral frequency to be equivalent.

If this approach is applied to the ideal FM-AM transfer, the characteristic in figure 3.5 may be considered to represent the spectral amplitude characteristic of the system. Obviously, the linearly increasing transfer in this figure suggests that differentiation to time is the required algorithm to implement the FM-AM conversion.

Phasor Representation

By definition, the instantaneous frequency of the FM wave, denoted by $\omega(t)$, which contains the message information, equals the time derivative of the carrier phase $\Phi(t)$, i.e.

$$\omega(t) \overset{\text{def}}{=} \frac{\mathrm{d}\Phi(t)}{\mathrm{d}t}. \tag{3.3}$$

As noted in Section 3.3.1, the phase $\Phi(t)$ equals the angular coordinate of the tip of the FM phasor $\vec{s}$.

Consequently, according to (3.3), $\omega(t)$ corresponds to the angular velocity of $\vec{s}$. The vector that represents the velocity of $\vec{s}$, denoted by $\vec{v}$, may be written as

$$\begin{aligned}\vec{v} &\overset{\text{def}}{=} \frac{\mathrm{d}\vec{s}}{\mathrm{d}t} \\ &= v_{\tan}\vec{u}_{\tan} + v_{\text{rad}}\vec{u}_{\text{rad}},\end{aligned} \tag{3.4}$$

where $v_{\tan}$ denotes the *tangential component* of the velocity vector, directed perpendicular to $\vec{s}$, and v_{rad} denotes the *radial component*, directed parallel to $\vec{s}$. Further, $\vec{u}_{\tan}$ and $\vec{u}_{\text{rad}}$ denote the unit vectors in the tangential and radial direction. This representation of $\vec{v}$ is depicted in the phasor diagram of figure 3.6.

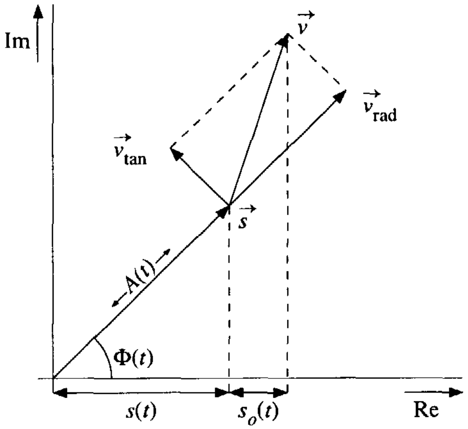


Figure 3.6: Phasor representation of the FM-AM conversion.

The tangential component $v_{\tan}$ represents the angular motion of $\vec{s}$, and is thus proportional to the instantaneous frequency $\omega(t)$, that includes the message. Therefore, the amplitude of the FM-AM converter output signal $s_o(t)$ should be proportional to $v_{\tan}$.

However, differentiation of $\vec{s}$ yields the vector $\vec{v}$, that besides $v_{\tan}$ also includes the radial component v_{rad}. As shown subsequently, the latter component does not contain the message information and should therefore be suppressed. As observed from figure 3.6, the corresponding FM-AM converter output signal $s_o(t)$, the projection of $\vec{v}$ on the real axis, generally equals

$$\begin{aligned}s_o(t) &\overset{\text{def}}{=} \text{Re}\,\{\vec{v}\} \\ &= v_{\tan}(t)\text{Re}\,\{\vec{u}_{\tan}(t)\} + v_{\text{rad}}(t)\text{Re}\,\{\vec{u}_{\text{rad}}(t)\} \\ &= -v_{\tan}(t)\sin\Phi(t) + v_{\text{rad}}(t)\cos\Phi(t).\end{aligned} \tag{3.5}$$

Besides the possibility to eliminate v_{rad} from $s_o(t)$ by exploitation of the phase quadrature between both components in (3.5), alternative possibilities should follow through evaluation of both velocity components.

Tangential Component

By elaboration of (3.4) and (3.5), the tangential component of the velocity vector, $v_{\tan}(t)$, can be expressed as

$$v_{\tan}(t) = A(t)\left[\omega_o + \dot{\varphi}(t)\right]. \tag{3.6}$$

Thus, as stated before, this component is proportional to the instantaneous frequency and contains the message signal $\dot{\varphi}(t)$.

Further, expression (3.6) demonstrates the observation of Section 3.3.2 that modulation contained in the carrier amplitude $A(t)$ prior to FM-AM conversion, e.g. due to noise or fading, corrupts the message information by multiplicative errors.

Finally, since ω_o is usually considerably larger than $\dot{\varphi}(t)$, a significant part of the converter output signal power is generally spoiled to a non-informative carrier component ω_o, that, as shown in Section 4.2, may considerably reduce the FM demodulator Dynamic Range (DR). Its elimination, addressed in Section 4.2, is therefore usually desirable.

Radial Component

The radial velocity component v_{rad} may be expressed as

$$v_{\text{rad}}(t) = \dot{A}(t). \tag{3.7}$$

This component does obviously not contain the FM message information, but represents the rate of change of the carrier envelope. Therefore, it should be suppressed in the converter output signal $s_o(t)$.

As observed from (3.7) this is accomplished by suppression of all modulation and noise in the FM carrier amplitude $A(t)$, e.g. by means of an hard-limiter or AGC with infinite compression, prior to FM-AM conversion. In that way, $A(t)$ becomes constant and its derivative vanishes.

3.4.3 Ideal AM Demodulation Function

The purpose of the AM demodulator is to retrieve the FM message information, i.e. convert it to baseband, contained in the FM-AM converter output signal.

Thus, in general, the basic function of an AM demodulator is to generate a baseband signal that is linearly dependent, i.e distortion-free, on the amplitude of the AM wave supplied to its input. When the input AM wave equals

$$s_o(t) = A_o(t)\sin\Phi(t), \tag{3.8}$$

the AM demodulator output wave $y_{\mathrm{AM}}(t)$ should be proportional to $A_o(t)$:

$$y_{\mathrm{AM}}(t) = K_{\mathrm{AM}} A_o(t), \tag{3.9}$$

where K_{AM} denotes the AM demodulator conversion gain.

In most applications, the phase of the carrier wave $\Phi(t)$ contains only a carrier component $\omega_o t$. The AM wave at the FM-AM converter output however also contains FM modulation. In the remainder of the demodulation process, however, this modulation is neglected, except when synchronization to $s_o(t)$ is required.

The requested linear transfer between the baseband demodulator output signal $y(t)$ and the input AM wave $A_o(t) \sin \Phi(t)$ is depicted in figure 3.7.

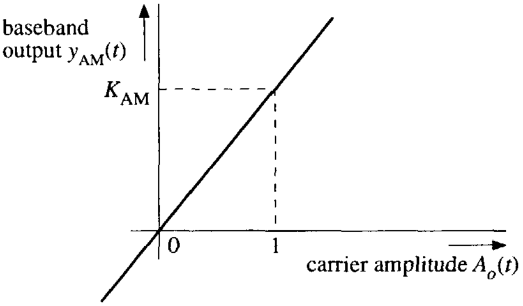


Figure 3.7: Ideal AM Demodulator Transfer Characteristic.

3.4.4 AM-Modulus Demodulation Algorithm

Amplitude demodulation of the FM-AM converter output wave $s_o(t)$ corresponds to determination of the length of the phasor $\vec{s}_o$ in the phasor plane. Two fundamentally different methods to accomplish this, resulting in two different AM demodulation principles, are distinguished.

This section discusses the algorithms that correspond to the first of the two AM demodulation principles. The resulting demodulators will be called "AM-Modulus Demodulators" (AMMD).

Demodulators based on this principle, usually called "full wave" or "half wave rectifiers" [15–17], determine the modulus of $\vec{s}_o$. This principle is illustrated by figure 3.8. According to figure 3.8a, this length can be expressed in terms of the projections of $\vec{s}_o$ on the real and imaginary axis as

$$\begin{aligned} |A_o(t)| &= |\vec{s}_o| \\ &= \sqrt{s_{o,i}(t)^2 + s_{o,q}^2(t)}. \end{aligned} \tag{3.10}$$

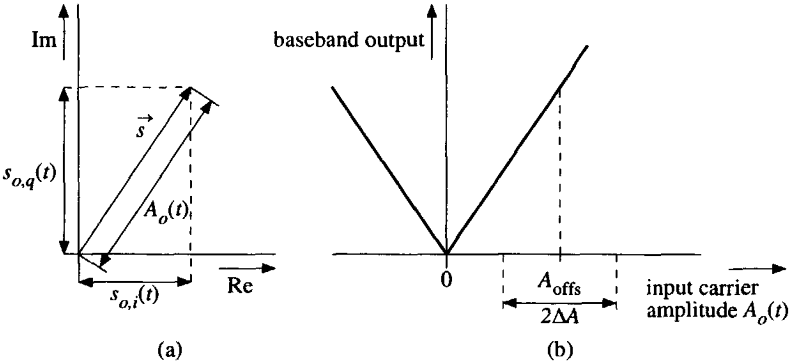


Figure 3.8: AM-modulus detection a) phasor representation, b) demodulator transfer characteristic.

The corresponding transfer characteristic from input amplitude $A_o(t)$ to the demodulator output signal is depicted in figure 3.8b.

This figure shows that the transfer of the class of AM-modulus demodulators contains a nonlinearity at $A_o(t) = 0$. A linear transfer is established only when $A_o(t)$ is unipolar, i.e. when the modulation index of the AM wave, in this respect defined as the ratio of the maximum message signal amplitude ΔA and the amplitude offset A_{ofs} (see figure 3.8 and e.g. [18]), is smaller than unity.

Consequently, for proper operation, such demodulators require the carrier-induced offset component in the FM-AM converter output signal to be at least as large as the maximum value of the FM message $\varphi(t)$, which is rather unfavorable for the DR. Although this offset is also present in the AM demodulator output, it may be eliminated from the FM demodulator output signal, by application of a balanced structure (see Section 3.4.6).

Expression (3.10) shows that generally the following four basic functions are required to construct an AM demodulator according to this demodulation principle:

- generation of a quadrature wave $s_{o,q}(t)$;
- squaring;
- addition;
- a square-root operation.

However, often considerable simplification of this demodulation algorithm is possible by application of *a priori knowledge* of the AM wave $s_o(t)$. The algorithm in (3.10) is general in the sense that it yields the correct result in all possible circumstances. In practice, it is often known in advance that some situations will never occur, which allows parts of the algorithm to be omitted.

The most important AM demodulator architectures obtained by simplification of (3.10), are depicted in figure 3.9, together with the general AM-modulus demodulator architecture. These architectures are discussed below.

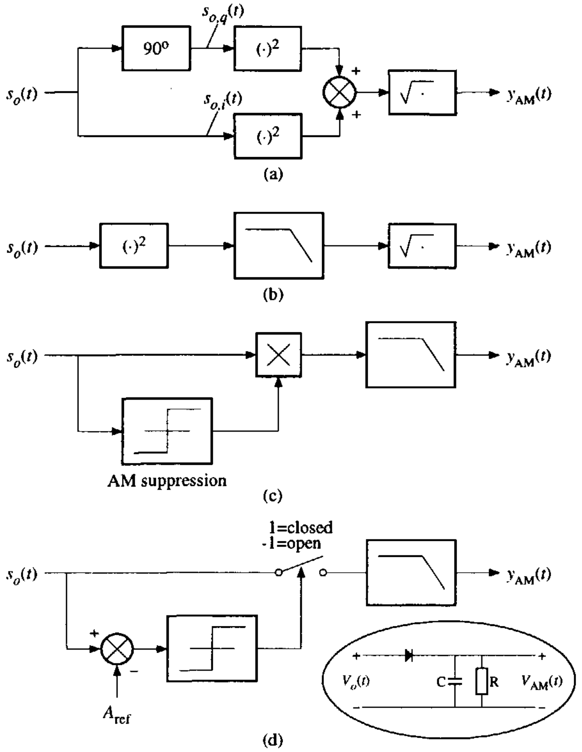


Figure 3.9: AM modulus demodulator architectures obtained by algorithm simplification. a) general FM-modulus demodulator, b) filtering rectifier, c) limiting rectifier, d) sampling rectifier and implementation (peak detector).

Elimination of the Quadrature Generation and Addition

Generation of the quadrature wave $s_{o,q}(t)$ and the addition, can be replaced by low-pass filtering when $s_o(t)$ is a narrow-band wave, i.e. when its bandwidth is substantially smaller than its carrier frequency. This follows by examination of the spectrum of the squared wave $s_o^2(t) = s_{o,i}^2(t)$. For $s_o^2(t)$, we may write

$$\begin{aligned} s_o^2(t) &= \{A_o(t)\cos[\omega_o t + \varphi(t)]\}^2 \\ &= \frac{A_o^2(t)}{2} + \frac{A_o^2(t)}{2}\cos[2\omega_o t + 2\varphi(t)]\,. \end{aligned} \tag{3.11}$$

The first term in this expression is required to construct the baseband demodulator output signal, while the second term represents an AM wave at the double carrier frequency, that should be suppressed.

Figure 3.10 illustrates the filtering operation, and shows the necessity of the narrow-band wave requirement.

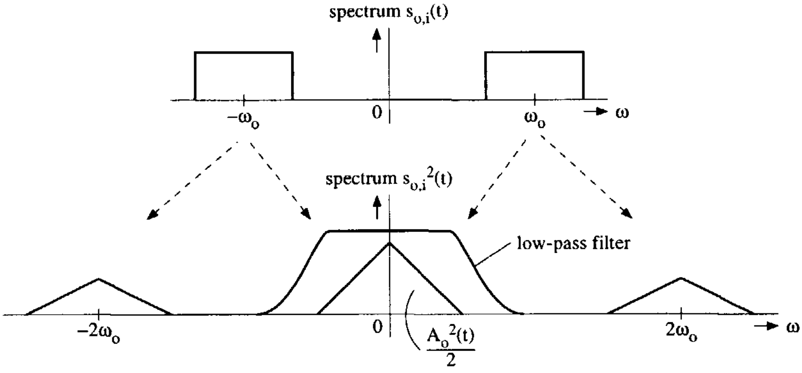


Figure 3.10: Suppression of the double frequency component by means of low pass filtering.

Elimination of the Square-Root

The square-root operation is required due to the presence of the squaring operation, that is usually performed by suppling both inputs of a multiplier with the same wave. When the AM modulation is removed from one of the mixer inputs, e.g. by means of a limiter [19], or a variable-gain amplifier [20], or by application of a switching mixer, the square-root operation may be omitted. Figure 3.9c depicts the version of this architecture that employs low pass filtering to eliminate the double carrier frequency component.

Simplification by Sampling

Further simplification of the algorithm is possible by application of sampling. When the AM wave is narrow-band, the AM message information hardly changes during one cycle of the carrier. In that case, continuous monitoring of the carrier amplitude may be replaced by detection of the carrier top, when $\vec{s}_o$ crosses through the real axis. During these crossings, $s_o(t)$ equals the amplitude $A_o(t)$. The instants of these crossings may be determined synchronously, resulting in an AM projection detector, to be discussed hereafter, or asynchronously, as depicted in figure 3.9d.

Asynchronous detection of the crossing instants, illustrated in figure 3.11, is possible when the AM modulation index is considerably smaller than unity. This method replaces the detection of the crossing instant by a level detection on $s_o(t)$ itself. When $s_o(t)$ exceeds some reference level A_{ref}, as depicted in figure 3.11, the phasor $\vec{s}_o$ must be close to a crossing through the real axis. In that region, $s_o(t) \approx A_o(t)$.

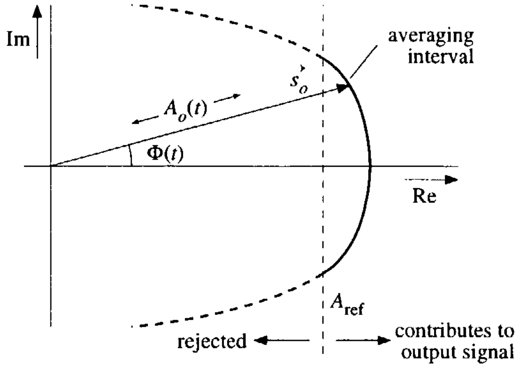


Figure 3.11: Asynchronous Sampled AM detection.

The output signal of the detector, usually implemented by the well-known peak detector (inside ellipse in figure 3.9d) equals the average value of $A_o(t)$, over the time interval where $s_o(t)$ exceeds A_{ref}. The major disadvantage of this approach is the lack of orthogonality between the design parameters; apparently, the architecture has become 'over-simplified'. For example, the choice of the reference level is based on a critical balance between the maximum allowed AM modulation index, and the detector accuracy, determined by the length of the averaging interval. A zero length of the averaging interval, obtained when $A_{\text{ref}} = A_o(t)$, yields the exact crossing instant, but at the same time requires a zero-valued modulation index. Oppositely, a large averaging interval allows large modulation indices, but yields an inaccurate output signal.

3.4.5 AM-Projection Demodulation Algorithm

This section discusses the algorithms of the second AM demodulation principle. The resulting AM demodulators will be called "AM-Projection Demodulators" (AMPD).

This class of AM demodulators determines the length of $\vec{s}_o$ by construction of its projection on a reference phasor $\vec{s}_r$, as illustrated in figure 3.12. The output signal equals the in-product of $\vec{s}_o$ and the reference $\vec{s}_r$. Expressed in

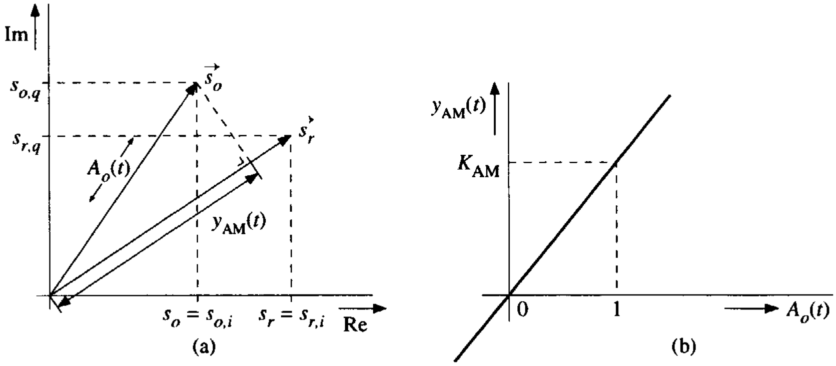


Figure 3.12: AM projection demodulation. a) phasor representation, b) demodulator transfer characteristic.

terms of the projection along the real and imaginary axis, this becomes

$$\vec{s}_o \cdot \vec{s}_r \overset{\text{def}}{=} s_{o,i}(t)s_{r,i}(t) + s_{o,q}(t)s_{r,q}(t). \tag{3.12}$$

The reference wave should be free of amplitude modulation, and synchronized to the wave $s_o(t)$, i.e.

$$s_r(t) = A\cos\Phi(t). \tag{3.13}$$

With this reference wave, the demodulator output signal from (3.12) equals

$$\begin{aligned} y_{\text{AM}} &= \vec{s}_o \cdot \vec{s}_r \\ &= A_o(t)\cos\Phi(t)A\cos\Phi(t) + A_o(t)\sin\Phi(t)A\sin\Phi(t) \\ &= AA_o(t). \end{aligned} \tag{3.14}$$

The transfer of this class of AM demodulators, depicted in figure 3.12b, is completely linear, even for negative values of $A_o(t)$, as opposed to the transfer of AM modulus demodulators (figure 3.8). Therefore, these demodulators allow a zero-valued offset component in the FM-AM converter output carrier amplitude, which is favorable for the DR.

From expression (3.14), it is observed that the demodulation algorithm consists of the following basic functions:

- generation of a synchronous reference wave $s_r(t)$;
- generation of waves in phase-quadrature with $s_o(t)$ and $s_r(t)$;
- multiplication;

- addition.

Again, considerable simplification of the algorithm is possible with the aid of a priori knowledge of the AM wave $s_o(t)$. The main simplifications are depicted in figure 3.13, together with the general projection demodulator architecture.

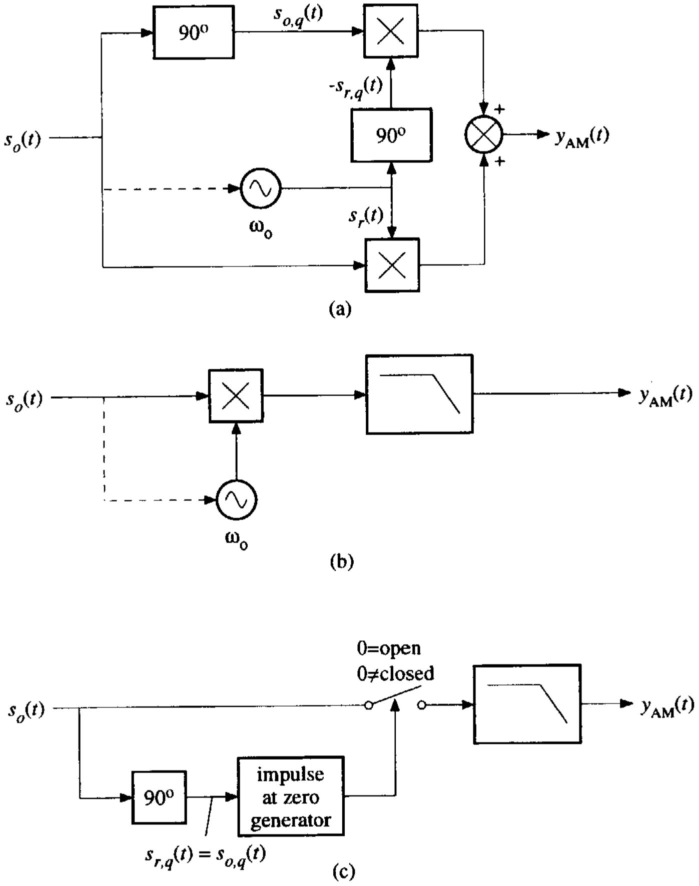


Figure 3.13: AM projection demodulator architectures obtained by algorithm simplification. a) general projection demodulator, b) synchronous detector, c) sampled detector.

Elimination of the Quadrature Generation and Addition

In the same way as in the AM-modulus demodulation algorithm, the quadrature carrier generation and addition can be replaced by low pass filtering, to remove double carrier frequency components of $s_o(t)s_r(t)$, when $s_o(t)$ is a narrow-band wave. The result of this simplification is the familiar synchronous detector of figure 3.13b.

Simplification by Sampling

Similar to the AM-modulus demodulation algorithm, the AM projection demodulation algorithm may be considerably simplified by application of sampling. As opposed to the asynchronous sampling AM-modulus demodulator, the sampling AM-projection demodulator synchronously determines the instant of the crossing of $\vec{s_o}$ through the real axis.

These crossing instant are obtained from the quadrature carrier $s_{o,q}(t)$, that crosses through zero at the instant that $s_o(t) = s_{o,i}(t)$ reaches its top. The resulting sampled FM projection detector is depicted in figure 3.13c.

3.4.6 FM Demodulation by AM-Modulus Detection

At this point in the development of the FM demodulator classification, it is possible to identify the complete class of FM-AM conversion FM demodulators, with the aid of the FM-AM conversion and AM demodulation principles considered in Section 3.4.1 through Section 3.4.4. One FM-AM conversion algorithm and two AM demodulation algorithms were found, resulting in two classes of FM-AM conversion FM demodulators.

This section discusses the first (sub-) class of FM-AM conversion FM demodulators, consisting of the demodulators that employ AM-modulus detection (linear rectification). First, their most important characteristics, partly derived in Chapter 4, are outlined. Subsequently, as an illustration, various types of demodulators encountered in literature are mapped on the classification.

Demodulator Characteristics

In the sequel, according to Shannon [12], the *Information Handling Capacity* (IHC), determined by the bandwidth and dynamic range (DR), is used as a criterion to judge the performance of the various types of demodulators. Therefore, we focus on those (implementation-independent) characteristics that limit the demodulator DR.

Amplitude Offset As explained in Section 3.4.4, distortion-free demodulation by AM-modulus detection is theoretically possible only for unipolar AM carrier amplitudes, corresponding to AM modulation indices smaller than unity. Thus, the offset component in the amplitude should always exceed the message signal. As discussed in Section 4.3, in general, this offset considerably reduces the FM-AM converter and FM demodulator DR. For example, for a sinusoidal message signal, the offset reduces the DR by at least 4.8 dB in comparison to the maximum DR, obtained for a zero-valued offset. Even worse, for Gaussian AM modulation, the reduction equals at least 12 dB.

With the aid of a balanced demodulator structure, so called "stagger-tuned" LC-tanks [21], it is possible to remove the offset component from the FM demodulator output signal. However, since the offset remains present in the FM-AM converter and AM demodulator, the same dynamic range reduction is observed.

Phase Selectivity Phase selectivity is a property that allows demodulators to discriminate between different directions in the phasor plane. For example, this property, enables detection of the tangential velocity component from the FM-AM converter output signal $s_o(t)$, as described by (3.5), in the presence of the undesired radial component.

AM-modulus demodulators are phase-inselective, and therefore unable to discriminate between the tangential and radial component in $s_o(t)$. As discussed in Section 4.3.1, the latter component may cause considerable distortion in the demodulator output signal, and therewith reduces the DR.

Noise As considered in Section 4.2, when the amplitude noise is not eliminated from the input FM wave, the amplitude offset introduces a considerable increase of the demodulator output noise level. Since such an offset is inevitable in the type of FM demodulators under consideration, a considerable reduction of the DR due to this effect has to be expected when amplitude noise is not suppressed.

Conclusion Based on the various short-comings outlined above, we must conclude that the performance of this type of FM demodulator is generally moderate. The main advantage of this oldest demodulator class, reported already in 1913 [22], is probably their simplicity.

Demodulators from Literature

In literature, a vast amount of this type of FM demodulators is known. Some interesting examples of them are discussed below.

Slope Detector

The Slope Detector, depicted in figure 3.14, consists of a detuned LRC-tank that is used as FM-AM converter (differentiator) and a peak detector, consisting of a diode and an RC low pass filter [11, 21, 23]. The balanced version, depicted in figure 3.14b, consists of two such structures, that eliminate the amplitude offset from the demodulator output. The resonant frequencies of both LRC tanks in this circuit are "stagger-tuned", such that a maximum linear slope, in order to minimize the distortion, is obtained in the region between both resonant frequencies, as depicted in figure 3.14c.

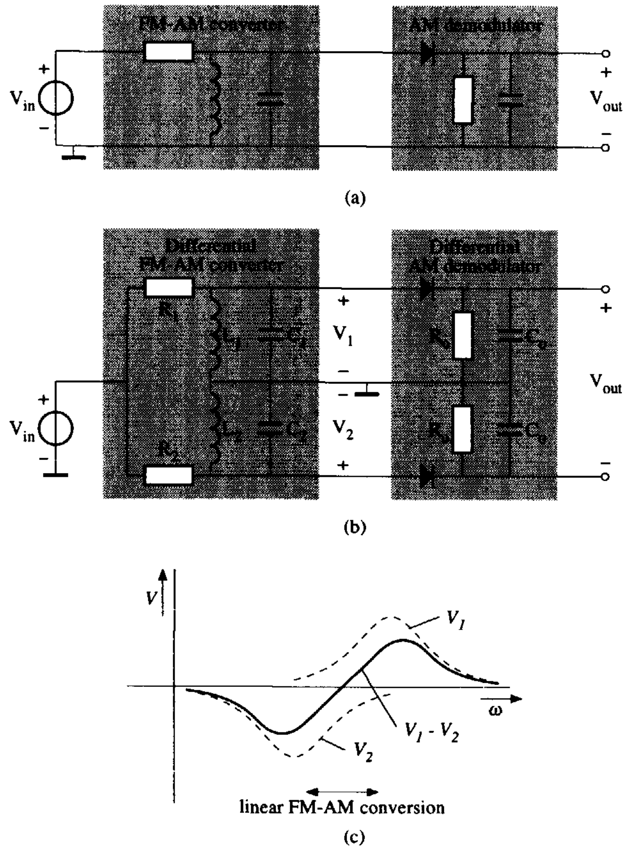


Figure 3.14: Slope detector a) single-ended, b) differential, c) transfer of the differential FM-AM converter.

This circuit is probably the oldest type of FM demodulator and was frequently used in FM radio receivers before world war II. Parasitic AM modulation and amplitude noise in the demodulator input signal were not suppressed by this circuit. Further, since the peak detector is incapable to suppress the radial component of the FM-AM converter output, the circuit demodulates AM waves as well [16].

Super Regenerative Detector

The super regenerative detector [24, 25], depicted in figure 3.15, is an improved version of the slope detector. This type of detector was was originally invented for the demodulation of AM waves. However, with some slight modifications, it appeared to be very suitable for FM demodulation as well. Its main advantages with respect to the slope detector is that it amplifies the received FM wave, which makes it far more sensitive than the slope detector. Its operation may be explained as follows. The FM-AM conversion is performed by the RLC-network in the input circuitry. Simultaneously, the positive feedback applied around this network by the hexode part of the electron tube, which acts as a negative resistor, introduces an oscillation, with an amplitude proportional to the instantaneous frequency of the FM wave (the message information). This amplitude is rectified by the hexode, i.e. AM-modulus detected, resulting in the baseband output signal at the anode of the hexode. However, once the oscillation is established, the RLC-hexode combination ceases 'listening' to the FM wave; a characteristic property of harmonic oscillations. Therefore, in order to continue demodulation of the FM wave, this oscillation has to be discontinued periodically. For that purpose, the RLC-network inside the amplifier and the triode part of the tube establish another, sustaining oscillation of at least twice the highest message frequency. This oscillation modulates the gain of the hexode, through the connection between a hexode grid and the triode grid, and therewith periodically reduces the positive loop gain below unity, which stops the oscillation.

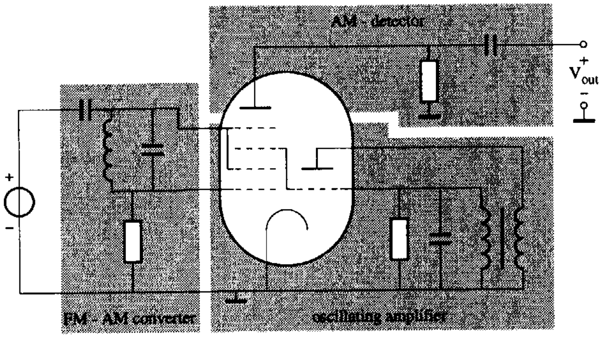


Figure 3.15: Super regenerative detector (biasing circuitry is omitted).

Finally, it is interesting to note that recently, the strange startup behavior of the hexode-oscillations, which result in an awkward kind of background 'noise', that could not be well explained by Armstrong [24], was identified to be chaotic [26].

Digital Square-Root Demodulator

A modern example of this type of FM demodulator is depicted in figure 3.16 [27]. This circuit is a fully digital implementation of the FM demodulation principle. Although digitally implemented, the circuit operates similar to its much

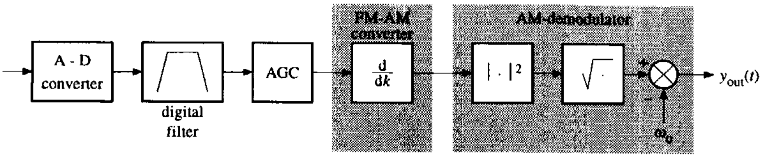


Figure 3.16: Digital Implementation of the FM-AM conversion demodulation principle followed by AM-modulus detection.

older analog equivalents and is subject to similar imperfections, inherent to the demodulation principle.

3.4.7 FM Demodulation by AM-Projection Detection

In this section we discuss the second subclass of FM-AM conversion demodulators, those employing AM-projection detection. We first summarize their characteristics, and subsequently discuss some example circuits from literature.

Demodulator Characteristics

The characteristics of this demodulator class, as far as their information handling capacity, and DR, is concerned, are as follows.

Amplitude Offset As opposed to AM-modulus demodulators, AM-projection demodulators allow a zero-valued offset in the FM-AM converter output carrier amplitude, due to their completely linear transfer. This property is favorable for the demodulator DR; it allows the upper bound on the DR to be reached.

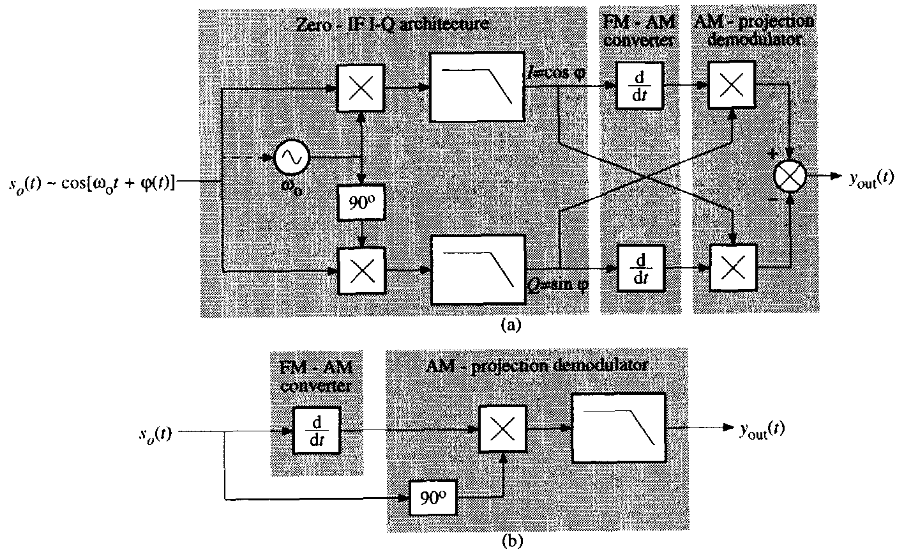
Phase Selectivity The AM projection demodulator is obviously phase-selective. Therefore, it is able to suppress the unwanted radial component in the FM to AM converter output signal, which is favorable for the DR.

Noise Since the amplitude offset may be nullified, the output noise level is theoretically able to reach the minimum possible level, set forward by the frequency noise contained in the input FM wave. Again, this is obviously favorable for the DR.

Conclusion The FM-AM demodulator, combined with AM-projection detection definitely outperforms the FM-AM demodulator with AM-modulus detection, and is able to reach the upper bound on the demodulator DR. This high-performance is attained at the price of an increased circuit/system complexity.

Demodulators from Literature

Since IC technology became widely available, this class of FM demodulators has gradually replaced the FM demodulator architectures based on AM-modulus detection discussed in Section 3.4.6, thanks to its performance, suitability for integration, and digital implementation [28, 29]. The main types of FM demodulators based on this demodulation principle are discussed below.



(a)

(b)

Figure 3.17: Mathematical demodulator a) balanced, b) single-ended.

Mathematical Demodulator

The mathematical or direct conversion demodulator, as depicted in figure 3.17a, converts the input FM wave to zero frequency, by means of a zero-IF I-Q architecture, subsequently converts it to an AM wave by means of differentiation, and finally demodulates this wave by AM projection detection. Due to the zero carrier frequency, the AM projection demodulator cannot be simplified to a synchronous demodulator that applies low-pass filtering; the FM bandwidth is much (infinitely) larger than the carrier frequency.

Single-ended Mathematical Demodulator

This type of FM demodulator, depicted in figure 3.17b, operates similarly as the balanced math demodulator, except for the fact that the FM carrier frequency is not converted to zero. Consequently, it allows simplification of the AM projection demodulator to the well-known synchronous demodulator.

It is obvious that, in this case, suppression of the offset in the FM-AM converter output carrier amplitude requires special measures. As explained in Section 4.2.1, a "band-pass differentiator" is needed for this purpose, instead of the usual differentiator.

3.5 Demodulation by Conversion to PM

This section investigates the principles of operation, and algorithms available for the implementation of FM demodulators based on FM-PM conversion, the second of the two classes identified in Section 3.3.2.

Similar to the previous section, the two sub-functions identified in these demodulators, FM-PM conversion and PM demodulation, are investigated separately. Considerations on the implementation of the resulting FM-PM demodulation algorithms are given in Chapter 4.

An outline is as follows. Section 3.5.1 considers the ideal FM-PM conversion function, while Section 3.5.2 through Section 3.5.5 explain the four different FM-PM conversion algorithms. Section 3.5.6 considers the ideal PM demodulation function, used in Section 3.5.7 to determine the PM demodulation algorithm. Section 3.5.8 through Section 3.5.11 compare three of the four types of FM-PM conversion FM demodulators with demodulators encountered in literature. The fourth type is considered in detail in Chapter 7.

3.5.1 Ideal FM-PM Conversion Function

The ideal FM-PM conversion function establishes a linear relation between the FM message signal $\dot{\varphi}(t)$, contained in the instantaneous *frequency* of the FM input wave, and the PM message signal $\Delta\varphi(t)$, contained in the instantaneous *phase* of the output wave, as depicted in figure 3.18. In this figure, $K_{\text{FM-PM}}$

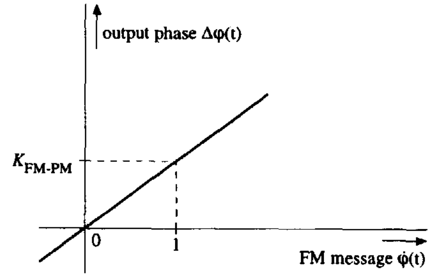


Figure 3.18: Transfer of the ideal FM-PM conversion function.

denotes the conversion gain.

In practice, since the FM modulation contained in the input wave cannot be eliminated, $\Delta\varphi(t)$ denotes the *phase difference* between the input FM wave and the FM-PM converter output wave, denoted by $s_o(t)$, instead of the absolute phase modulation level. Thus, the response of the ideal FM-PM converter to the input FM wave $s(t)$ from (2.1) may be expressed as

$$s_o(t) = A_o \cos\left[\omega_o t + \varphi(t) + \Delta\varphi(t)\right], \tag{3.15}$$

where, according to figure 3.18, $\Delta\varphi(t) = K_{\text{FM-PM}}\dot{\varphi}(t)$. Consequently, $s_o(t)$ is simultaneously modulated in phase and frequency by the message $\dot{\varphi}(t)$.

3.5.2 Algorithm based on a Fixed Time-Delay

This section investigates the first of the four different FM-PM conversion algorithms, available for the construction of FM-PM conversion FM demodulators.

The basic algorithm is derived with the aid of the quasi-stationary approximation. Subsequently, its phasor representation is considered. Finally, the corresponding demodulator architecture is discussed.

Quasi-Stationary Approximation

According to the quasi-stationary approximation, suppose that figure 3.18 represents the transfer from the spectral input frequency ω to the spectral phase difference $\Delta\Phi_o(\omega)$ between the FM-PM converter input and output wave.

In that case, FM-PM conversion corresponds to ideal, linear prediction with a prediction time τ_p, given by [30]

$$\tau_p = \frac{\partial\Delta\Phi_o(\omega)}{\partial\omega}. \tag{3.16}$$

Obviously, an ideal linear, predictor is non-causal and cannot be realized; such a system would generate a response to an input wave that arrives in the future.

Fortunately, when the slope of the linear phase characteristic is negative instead of positive, FM-PM conversion corresponds to a fixed time-delay (negative prediction time), which is causal and (approximately) realizable. The (group) delay time τ_d is therefore related to $\Phi_o(\omega)$ as [30]

$$\tau_d \overset{\text{def}}{=} -\frac{\partial\Delta\Phi_o(\omega)}{\partial\omega}. \tag{3.17}$$

By comparison of (3.17) with figure 3.18 is observed that the conversion gain for a linear delay equals $K_{\text{FM-PM}} = -\tau_d$.

Phasor Representation

The phasor representation of this FM-PM conversion algorithm is depicted in figure 3.19. In this figure $\Phi(t)$ denotes the instantaneous phase of the FM wave

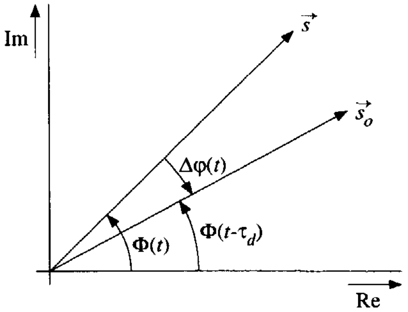


Figure 3.19: Phasor representation of FM-PM conversion by means of a fixed time-delay.

$s(t)$. The FM-PM converter output, represented by $\vec{s}_o$, tracks the input wave $\vec{s}$ at a fixed time-difference τ_d. However, due to the variable angular velocity of $\vec{s}$, as a result of the FM modulation, the corresponding phase difference $\Delta\varphi(t)$ is not constant, but, as shown below, approximately proportional to the instantaneous frequency of $s(t)$.

An expression for the phase difference $\Delta\varphi(t)$ in the time-delay FM-PM converter output is obtained as follows. The response of the converter to the FM wave $s(t)$ equals

$$s_o(t) = A_o \cos\left[\omega_o\left(t - \tau_d\right) + \varphi\left(t - \tau_d\right)\right]. \tag{3.18}$$

This wave is identical to the one described by (3.15). Therefore, equating the instantaneous phases of both waves yields

$$\Delta\varphi(t) = -\tau_d\omega_o - \varphi(t) + \varphi(t - \tau_d). \tag{3.19}$$

For small time delays, when $\varphi(t)$ and $\varphi(t-\tau_d)$ differ only slightly, (3.19) may be approximated as

$$\Delta\varphi(t) \approx -\tau_d\left[\omega_o + \dot{\varphi}(t)\right], \tag{3.20}$$

which is the expected result. Note that the factor $-\tau_d$ equals the previously obtained value for the conversion gain $K_{\text{FM-PM}}$ of this converter. Further, observe that, just as the 'standard' differentiator FM-AM converter, this FM-PM converter is unable to eliminate the carrier-offset ω_o. The consequences of this offset are, however, not as dramatic as in FM-AM conversion demodulators, for reasons discussed below.

Demodulator Architecture

The basic architecture of FM demodulators based on this FM-PM conversion algorithm, incorporating an ideal PM demodulator, is depicted in figure 3.20. In this architecture, the input FM wave is subjected to a delay τ_d, and subsequently supplied to the input of the (ideal) PM demodulator. Another copy of the input wave is supplied directly to this demodulator, and serves as reference wave.

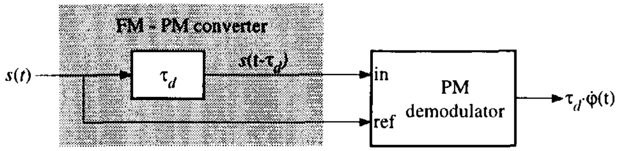


Figure 3.20: FM demodulator architecture based FM-PM conversion by a fixed time delay.

An important intrinsic property of this demodulator architecture is its ability to eliminate the offset term $\omega_o\tau_d$, observed in (3.20), with the aid of the periodicity in the FM demodulator transfer. This 'offset-cancellation' mechanism is illustrated by figure 3.21, which depicts the transfer characteristic of the PM demodulator. As a result of the periodicity, a consequence of the observation that PM demodulators cannot distinct phase differences larger than 2π[1], the demodulator output becomes zero for some nonzero values of the input phase difference. Therefore, when the phase offset $\omega_o\tau_d$ is chosen such that the PM demodulator response to it equals zero, as sketched in figure 3.21, the demodulator output (approximately) equals the FM message signal $-\tau_d\dot\varphi(t)$, without offset.

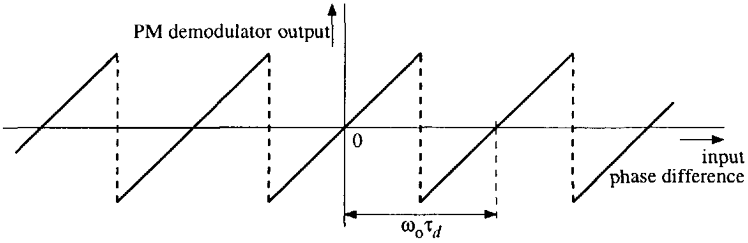


Figure 3.21: Elimination of the offset $\omega_o\tau_d$ with the aid of the periodic PM demodulator transfer.

[1] This range can be extended by application of a memory, but the periodicity remains present.

3.5.3 Algorithm based on a Fixed Phase Difference

This section investigates the second of the four FM-PM conversion algorithms.

First, the principles of this conversion algorithm are explained. Subsequently, the FM-PM converter transfer is determined with the aid of a phasor representation. Finally, the corresponding demodulator architecture is discussed.

Principles of the Algorithm

In essence, all FM-PM conversion algorithms somehow implement the definition formula that relates the instantaneous frequency and instantaneous phase of a carrier wave, i.e.

$$\dot{\Phi}(t) \overset{\text{def}}{=} \lim_{\tau \to 0} \frac{\Phi(t) - \Phi(t-\tau)}{\tau}. \tag{3.21}$$

In essence, this expression shows that an FM-PM conversion is established by differentiation of the carrier phase $\varphi(t)$.

The algorithm discussed in Section 3.5.2 implements an approximation of (3.21), given by

$$\tau_d \dot{\Phi}(t) \approx \Phi(t) - \Phi(t - \tau_d). \tag{3.22}$$

Thus, the lim-operation is omitted, the phase difference $\Phi(t) - \Phi(t - \tau_d)$ is detected by a PM demodulator, while a *fixed, finite time difference* $\tau = \tau_d$ is realized by means of a delay line. The approximation holds when τ_d is small, such that $\Phi(t)$ and $\Phi(t - \tau_d)$ differ only slightly.

The algorithm considered in this section adopts exactly the opposite approach. Instead of a fixed time difference τ_d, a *fixed, finite phase difference* $\Phi(t) - \Phi(t-\tau) = \Delta\Phi_o$ is realized, while the time difference τ is detected. Thus, this algorithm approximates (3.21) as

$$\dot{\Phi}(t) \approx \frac{\Delta\Phi_o}{\tau}. \tag{3.23}$$

The approximation holds as long as the message information contained in $\Phi(t) = \omega_o t + \varphi(t)$, i.e. $\varphi(t)$, differs only slightly when $\Phi(t)$ covers a phase difference $\Delta\Phi_o$ in a time interval τ. Thus, in other words, the carrier frequency ω_o should be much larger than the maximum value of $\dot{\varphi}(t)$.

Phasor Representation

The operation of the algorithm is illustrated by the the phasor diagram of figure 3.22. In this figure, the 'start'- and 'stop' phase Φ_{start} and Φ_{stop} define the bounds on a phasor-plane segment of angle $\Delta\Phi_o$, corresponding to the fixed

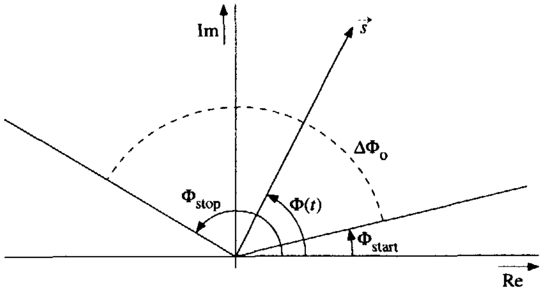


Figure 3.22: FM-PM conversion by means of a fixed phase difference.

phase difference in equation (3.23). The FM phasor $\vec{s}$, that rotates around the origin, covers this segment once per cycle. The time difference τ from (3.23), required to cover the segment, is measured by an internal clock inside the demodulator. This clock is started every time $\vec{s}$ enters the segment, i.e. when it instantaneous phase equals $\Phi(t) = \Phi_{\text{start}}$, and stopped when it leaves the segment, i.e. when $\Phi(t + \tau) = \Phi_{\text{stop}}$.

It is convenient to use a segment angle $\Delta\Phi_o$ that equals a multiple[2] of π. In that case, Φ_{start} and Φ_{stop} may both be chosen to coincide with the imaginary axis, i.e. with the *zero crossings* of the wave $s(t)$. A simple zero-crossing detector (a "binary PM demodulator") may be used to start and stop the clock.

An expression for the output signal of such a demodulator is obtained as follows. Assume that the segment size equals $\Delta\Phi_o = 2\pi$. Then, according to (3.15), the phase difference between the FM-PM converter input and output signal equals $\Delta\varphi(t) = 2\pi$. The phase of the FM-PM converter output wave at time $t + \tau$, the response to the FM input wave at time t, therefore equals $\Phi(t) + 2\pi$. Since, by definition, the input FM wave covers the phase segment $\Delta\Phi_o = 2\pi$ in a time interval of length τ, its phase at time $t+\tau$ equals $\Phi(t+\tau) = \Phi(t) + 2\pi$. By rewriting this expression, we obtain

$$2\pi = \omega_o\tau + \varphi(t + \tau) - \varphi(t). \tag{3.24}$$

If further τ is assumed small, such that $\varphi(t)$ and $\varphi(t + \tau)$ differ only slightly, then we obtain

$$\frac{2\pi}{\tau} \approx \omega_o + \dot{\varphi}(t). \tag{3.25}$$

This expression shows that, just as the previously discussed algorithm, this algorithm results in an offset at the FM-PM converter output.

[2] A segment size larger than 2π means that the clock measures the duration of several cycles of the input FM wave.

Demodulator Architecture

The FM demodulator architecture corresponding to this FM-PM conversion algorithm is depicted in figure 3.23. A simple phase-crossing detector (M-ary PM demodulator), usually a zero-crossing detector, determines the instants that the phase of the input FM wave enters and leaves the phase segment of size $\Delta\Phi_o$. Its output signal controls a pulse-generator, that implements the clock-function. Note that the FM-PM conversion operation is not concentrated in a small part of the architecture, as in the previously discussed algorithm, but is distributed over the entire system.

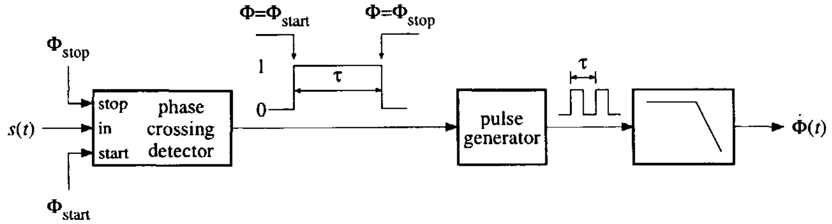


Figure 3.23: FM demodulator architecture based on FM-PM conversion by a fixed phase difference.

The pulse-generator generates a pulse of a *fixed* duration at every edge of the phase-crossing detector output signal. Since the time-interval τ between two such pulses is inversely proportional to the average frequency of the input wave, the average value of the pulse-train, obtained by low-pass filtering, is proportional to the 'instantaneous' frequency of the input FM wave, as long as this frequency changes only slightly during the time-interval τ.

As opposed to the algorithm discussed in Section 3.5.2, the algorithm discussed in this section is unable to eliminate the carrier offset ω_o. The transfer of a clock is non-periodic, which means that the offset cancellation technique discussed in Section 3.5.2 cannot be applied; as observed from (3.24), a zero offset is achieved only for $\tau = 0$. The only possibility is to subtract ω_o from the demodulator output signal, or, as used in [31], the average period time T from τ. However, this still requires the clock to measure τ, and therefore does not improve the demodulator DR significantly.

Finally, it should be remarked that the algorithm given by (3.25) is only the simplest possible; a zero-th order interpolation that yields a constant output signal between two zero-crossings of the input wave. More advanced schemes, i.e. first- and higher order interpolation may considerably reduce the distortion in the output signal [32, 33]. However, this becomes of interest only when the carrier frequency ω_o is in the same order of magnitude as the FM message signal.

3.5.4 Algorithm based on Phase Feedback

This section briefly outlines the third of the four FM-PM conversion algorithms. A detailed discussion of demodulators based on this algorithm is given in Chapter 7.

First, the principles of this algorithm are discussed subsequently, the corresponding demodulator architecture is outlined.

Principles of the Algorithm

The algorithm discussed in the two previous sections implement an approximation of the definition formula that relates the instantaneous frequency to the instantaneous phase of a carrier wave. The first employs a fixed time difference, while the second one employs a fixed phase difference.

As opposed to these approximative algorithms, the algorithm considered in this section implements the exact definition formula, i.e expression (3.21). The phase difference $\Phi(t) - \Phi(t - \tau)$ is measured and controlled to zero by means of a phase feedback loop.

In essence, this feedback performs the lim-operation in (3.21) and therewith automatically reduces the time difference τ to zero. The phase $\Phi(t)$ of the input FM wave $s(t)$ is compared with the instantaneous phase $\Phi(t - \tau)$ of a second FM wave, denoted by $s_o(t)$. This wave is reconstructed from the demodulator output signal by means of an FM modulator (controlled oscillator). The feedback mechanism attempts to reduce the phase difference between both waves to zero, such that in the ideal case, $s(t)$ and $s_o(t)$ possess the same instantaneous frequency. If this is achieved, the input signal of the FM modulator in the feedback path equals the demodulated FM message. An expression for this algorithm is derived in Section 7.1.

Phasor Representation

A phasor representation of the algorithm is depicted in figure 3.24. The phasor $\vec{s}$ represents the input FM wave $s(t)$, that possesses an instantaneous phase $\Phi(t)$, while the phasor $\vec{s}_o$ represents the reconstructed FM wave $s_o(t)$, with instantaneous phase $\Phi(t - \tau)$. The feedback loop measures the phase difference $\Phi(t) - \Phi(t - \tau)$, and advances/delays $\vec{s}_o$ with respect to $\vec{s}$, such that the difference approaches zero.

Demodulator Architecture

The FM demodulator architecture corresponding to this FM-PM conversion algorithm is depicted in figure 3.25. The phase detector in this figure detects the phase difference $\Phi(t) - \Phi(t - \tau)$ between the FM input wave $s(t)$ and its reconstruction $s_o(t)$, generated by the controlled oscillator in the feedback loop.

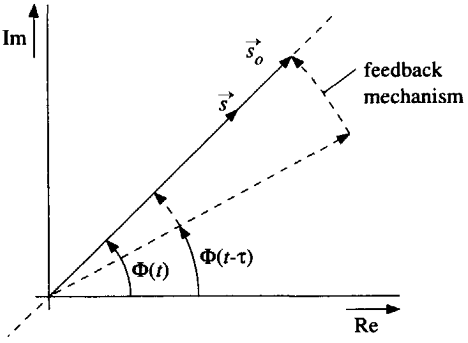


Figure 3.24: Phasor representation of the phase feedback algorithm.

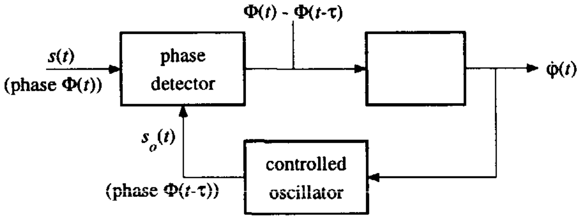


Figure 3.25: FM demodulator architecture based on FM-PM conversion by means of phase feedback.

When the loop 'locks' i.e. when $s_o(t)$ is a proper reconstruction of the input wave $s(t)$, the oscillator input wave corresponds to the demodulated FM message. The loop filter is usually included to improve the tracking properties of the loop, as discussed Chapter 7.

3.5.5 Algorithm based on Post-Detection Conversion

This section considers the last of the four FM-PM conversion algorithms. Opposed to the previously discussed algorithms, this algorithm is suited for narrow-band FM waves only, as explained below.

The algorithm considered in this section is a straight-forward implementation of the definition formula (3.21). It simply differentiates the phase of the FM wave, after its detection by a PM demodulator, as depicted in figure 3.26. The demodulator architecture therefore consists of a PM demodulator, followed by a phase-frequency converter (P-F), that transforms the detected instantaneous phase into a wave that is proportional to the instantaneous frequency of the input FM wave.

Unfortunately, this strategy is suited for narrow-band FM waves only, since

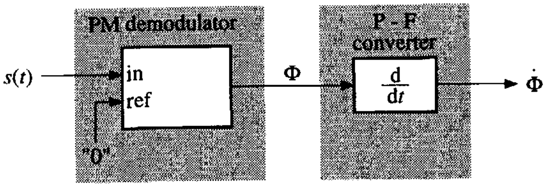


Figure 3.26: FM demodulator architecture based on post-detection FM-PM conversion.

the transfer of the PM demodulator possesses a limited domain, due to its periodicity, i.e. in-capability to detect phase differences larger than 2π. For proper demodulation, the phase of the FM wave should not exceed the bounds of this domain. Otherwise, as illustrated by figure 3.27, a type of distortion occurs that is similar to "cycle-slip" noise in phase feedback demodulators, considered in Chapter 7. When the PM demodulator runs out of its bounds, a phase step

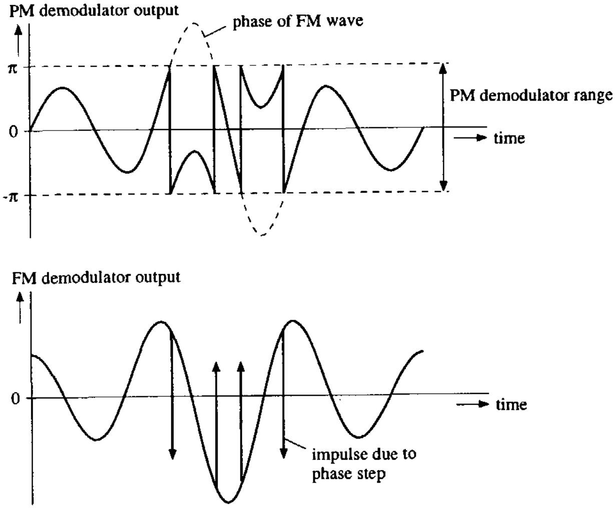


Figure 3.27: Generation of impulses in the FM demodulator output.

(of height 2π) is observed in its output signal. Due to the differentiation by the FM-PM converter, this results in an impulse (of area 2π) at the FM demodulator output.

Consequently, this algorithm requires that the integrated FM message $\varphi(t)$ fits within the domain of the PM demodulator. For instance, for a sinusoidal FM message signal and a PM demodulator with a domain of length 2π, follows that the maximum allowed FM modulation index equals π. Larger modulation indices inevitably result in spikes at the demodulator output. Further, notice that at low input CNR's, phase noise contained in the input FM wave cause the PM demodulator to run out of its bounds, resulting in spikes, even when the FM message wave fits within its domain.

3.5.6 Ideal PM Demodulation Function

As discussed previously, all demodulator architectures considered in Section 3.5.2 through Section 3.5.5 establish besides FM-PM conversion also PM demodulation, in order to recover the message information.

This section discusses the ideal PM demodulation function, the basis for the derivation of the PM demodulation algorithms, considered in Section 3.5.7.

The ideal PM demodulation function differs from the various ideal functions discussed previously, in the sense that it possesses two inputs, instead of one. Besides the input wave subjected to the demodulation, a (phase) reference wave is required. In the various phasor representations used in this chapter, we implicitly assumed a zero-valued reference phase: the instantaneous phase of a wave was defined as the angle between the phasor and the real axis.

The demodulator output signal is proportional to the phase difference between both inputs, as long as this difference is smaller than 2π. Phase differences larger than 2π cannot be detected instantaneously, but require a memory function. This is due to the fact that phase values separated by multiples of 2π possess exactly the same phasor representation. They correspond to the same position coordinates of the FM phasor tip. The ideal PM demodulation function is therefore periodic with period 2π. The resulting ideal transfer characteristic is depicted in figure 3.28.

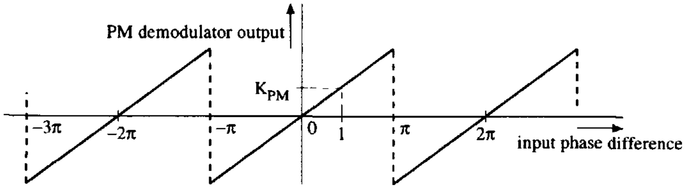


Figure 3.28: Ideal PM demodulation function.

3.5.7 PM Demodulation Algorithm

This section considers the PM demodulation algorithm available for the implementation of FM and PM demodulators. We first derive the general algorithm, valid in all possible circumstances, and subsequently consider its simplification.

General Algorithm

A PM demodulator determines the phase difference that exists between the two waves applied to its both inputs; the wave subjected to demodulation, denoted by $s(t)$, and a reference wave. The reference wave, denoted by $s_r(t)$, is defined by

$$s_r(t) = A_r \cos \Phi_r(t). \tag{3.26}$$

The PM demodulation algorithm is illustrated by figure 3.29. This figure depicts the phasor $\vec{s}$ of $s(t)$ and the phasor $\vec{s}_r$ of the reference wave $s_r(t)$. The PM demodulator output signal is proportional to the phase difference that exists between both phasors. When this phase difference is denoted by $\Delta\Phi_{\text{out}}$, then the output signal equals

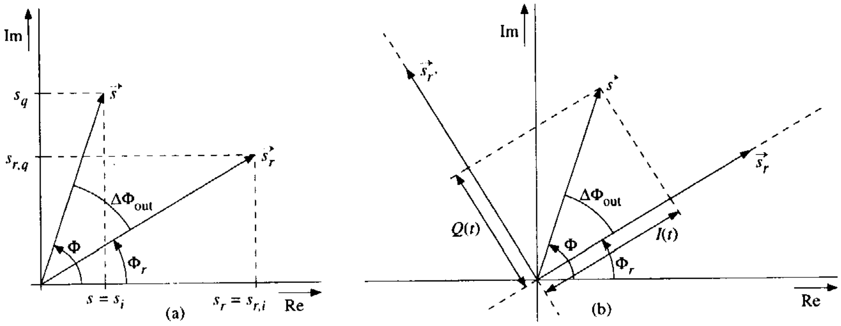


Figure 3.29: Phasor representation of PM demodulation. a) static frame of reference, b) dynamic frame of reference.

$$y_{\text{PM}}(t) = K_{\text{PM}} \Delta\Phi_{\text{out}}(t) = \Phi(t) - \Phi_r(t). \tag{3.27}$$

First, consider the case where the reference wave equals zero, i.e. $s_r(t) \equiv 0$. In that case, the demodulator output equals $\Delta\Phi_{\text{out}}(t) = \Phi(t)$. From a mathematical point of view, the phase $\Phi(t)$ equals the argument of the phasor $\vec{s}$, i.e.

$$\Phi(t) \overset{\text{def}}{=} \arg\{\vec{s}\} = \arg\{A(t) \exp[\mathrm{j}\Phi(t)]\}. \tag{3.28}$$

Expressed in terms of the projections on the real and imaginary axis, $s_i(t)$ and

$s_q(t)$ respectively, as depicted in figure 3.29a, this becomes

$$\Phi(t) = \begin{cases} \arctan\left[\frac{s_q(t)}{s_i(t)}\right] & s_i(t) \geq 0 \\ \arctan\left[\frac{s_q(t)}{s_i(t)}\right] + \pi & s_i(t) < 0, s_q(t) > 0 \\ \arctan\left[\frac{s_q(t)}{s_i(t)}\right] - \pi & s_i(t) < 0, s_q(t) \leq 0 \end{cases} . \tag{3.29}$$

In general, for a nonzero reference wave, the reference phasor $\vec{s}_r$ and the one orthogonal to it, $\vec{s}_{r'}$, define the frame of reference used by the PM demodulator. This frame of reference rotates around the origin with angular velocity $\dot{\Phi}_r(t)$, as opposed to the static frame of reference defined by the real and imaginary axis, as depicted in figure 3.29b.

For this situation, the phase difference $\Delta\Phi_{\text{out}}(t)$ should be expressed in terms of the coordinates $I(t)$ and $Q(t)$, the coordinates of $\vec{s}$ with respect to the axis through $\vec{s}_r$ and the axis through $\vec{s}_{r'}$ respectively, as depicted in figure 3.29b. By inspection, both components may be written as

$$\begin{aligned} I(t) &\overset{\text{def}}{=} \vec{s} \cdot \vec{s}_r \\ &= s_i(t)s_{r,i}(t) + s_q(t)s_{r,q}(t) \\ &= A(t)\cos\Phi(t)A_r\cos\Phi_r(t) + A(t)\sin\Phi(t)A_r\sin\Phi_r(t) \\ &= A(t)A_r\cos[\Phi(t) - \Phi_r(t)], \end{aligned} \tag{3.30}$$

$$\begin{aligned} Q(t) &\overset{\text{def}}{=} \vec{s} \cdot \vec{s}_{r'} \\ &= s_i(t)s_{r',i}(t) + s_q(t)s_{r',q}(t) \\ &= A(t)\sin\Phi(t)A_r\cos\Phi_r(t) - A(t)\cos\Phi(t)A_r\sin\Phi_r(t) \\ &= A(t)A_r\sin[\Phi(t) - \Phi_r(t)]. \end{aligned} \tag{3.31}$$

The phase difference between $\vec{s}$, and the phasor of the reference wave supplied to the demodulator, $\vec{s}_r$, can therefore be expressed as.

$$\Delta\Phi_{\text{out}}(t) = \begin{cases} \arctan\left[\frac{s_i(t)s_{r',i}(t) + s_q(t)s_{r',q}(t)}{s_i(t)s_{r,i}(t) + s_q(t)s_{r,q}(t)}\right] & I(t) \geq 0 \\ \arctan\left[\frac{s_i(t)s_{r',i}(t) + s_q(t)s_{r',q}(t)}{s_i(t)s_{r,i}(t) + s_q(t)s_{r,q}(t)}\right] + \pi & I(t) < 0, Q(t) > 0 \\ \arctan\left[\frac{s_i(t)s_{r',i}(t) + s_q(t)s_{r',q}(t)}{s_i(t)s_{r,i}(t) + s_q(t)s_{r,q}(t)}\right] - \pi & I(t) < 0, Q(t) \leq 0 \end{cases} . \tag{3.32}$$

This expression describes the general PM demodulation algorithm, valid in all circumstances. The following basic functions are identified from this expression:

1. quadrature wave generation for both $s(t)$ and $s_r(t)$;
2. arctan-function;
3. multiplication;
4. addition;
5. division;
6. "quadrant detection", that determines whether 0, π or $-\pi$ is added to the arctan-function.

Algorithm Simplification

Direct implementation of this algorithm, depicted in figure 3.30a, may be very well acceptable. However, implementations usually apply one or more of the following simplifications.

Elimination of the Addition

When the bandwidth of $s(t)$ and $s_r(t)$ is significantly smaller than their carrier frequency, $I(t)$ and $Q(t)$ can be obtained by one multiplication followed by low pass filtering, instead of two multiplications and a summation. The same type of simplification was already discussed in Section 3.4.7 in conjunction with AM demodulation. In AM demodulation, this simplification also eliminated the quadrature generation. However, as observed from (3.30) and (3.31), in PM demodulation, one of the two quadrature generation operations, the one for $s(t)$ or the one for $s_r(t)$ is eliminated, but not both. The resulting demodulator is depicted in figure 3.30b.

Elimination of the Arctan, Division and Quadrature Generation

In many cases, the phase difference $\Phi(t) - \Phi_r(t)$ is relatively small, i.e. considerably smaller than one radian. In that case, $I(t)$ and $Q(t)$ may be approximated as

$$I(t) \approx A(t)A_r, \tag{3.33}$$

$$Q(t) \approx A(t)A_r\left[\Phi(t) - \Phi_r(t)\right]. \tag{3.34}$$

Under the same conditions, the arctan-function may be approximated by its first-order Taylor term x. Therefore, the entire PM demodulation algorithm can in this case be reduced to the determination of $Q(t)$, optionally preceded by

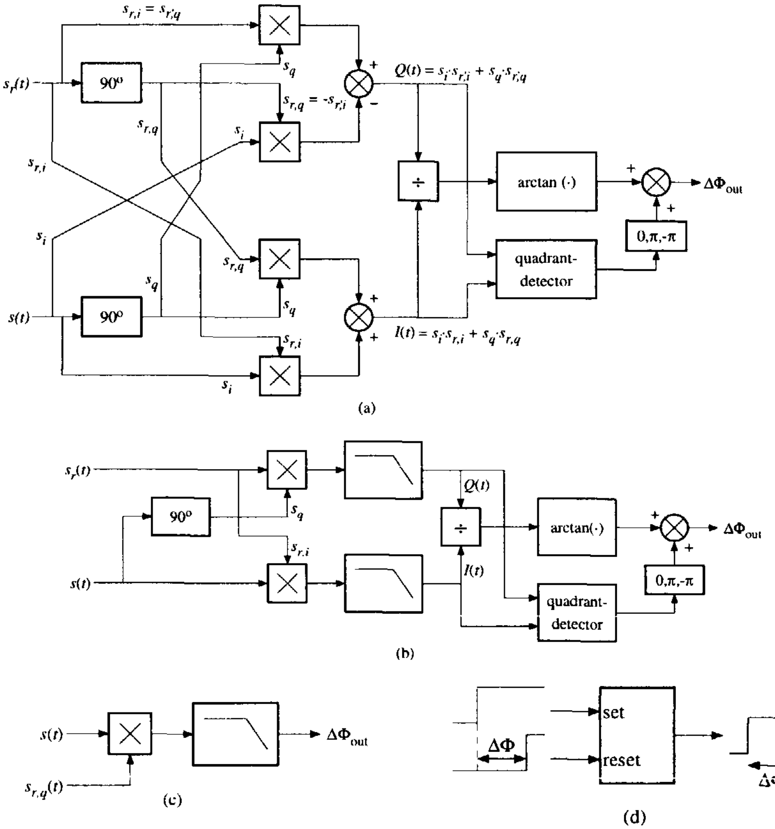


Figure 3.30: PM demodulation algorithm a) general algorithm, b) elimination of addition in I and Q, c) simple PM demodulator, d) sampling PM demodulator.

elimination of fluctuations in the FM carrier amplitude $A(t)$. A disadvantage of this type of PM demodulator is that its domain is at most of length π, instead of 2π, while its transfer is usually nonlinear for large phase differences. The resulting demodulator is depicted in figure 3.30c.

Simplification by Means of Sampling

When the bandwidth of the FM message signal is much smaller than the carrier frequency, the multiplication may be replaced by sampling, as shown in figure 3.30d. A clock is started at the time-instant that the reference wave crosses through some level, usually zero, and stopped when the other input wave crosses through zero. The phase difference is proportional to the measured time difference.

3.5.8 FM Demodulation by a Fixed Time Delay

The discussions in Section 3.5.1 through Section 3.5.7 showed that four different FM-PM conversion algorithms and one direct PM demodulation algorithm can be distinguished. Together, this yields four classes of FM demodulators based FM-PM conversion and direct PM demodulation.

This section discusses the first of the four sub-classes of FM-PM conversion FM demodulators, based on FM-PM conversion by means of a fixed time delay. First, their main characteristics, as obtained in Section 3.5.2 and Section 3.5.7 are outlined. Subsequently, examples found in literature are discussed.

Demodulator Characteristics

As far as the FM demodulator dynamic range is concerned, two main characteristics are of importance: the ability to suppress/prevent a carrier-induced phase offset, and the noise performance.

Phase Offset As explained in Section 3.5.2, the FM-PM conversion algorithm based on a fixed time-delay is able to suppress the carrier offset component observed in the instantaneous frequency of an FM wave. For this purpose, the time delay τ_d should be dimensioned such, that the PM demodulator response to a nonzero phase offset $\omega_o\tau_d$ equals zero. For a multiplier phase detector (PD), for example, this is the case when $\omega_o\tau_d = \pi/2$ (rad), while for a sampling PD, this is the case when $\omega_o\tau_d = 2\pi$ (rad).

Noise In Chapter 4 is shown that FM demodulators of the type discussed in this section perform the required quadratic noise shaping. As long as the phase offset is suppressed completely, no white noise floor is observed.

Conclusion Since FM demodulators based on FM-PM conversion by means of a fixed time-delay possess the ability to suppress the offset, and perform quadratic noise shaping, their DR is able to resemble the upper bound on the DR. Therefore, high-performance is feasible with these demodulators. The main problem in the design of these demodulators is the realization of a delay-element with a sufficiently large DR, as discussed in Chapter 4.

Demodulators from Literature

Numerous implementations of this type of FM demodulators are found in literature. Some interesting examples are discussed below.

Quadrature Demodulator

In this type of FM demodulator, which may be implemented by means of a single-ended or balanced structure, as depicted in figure 3.31, the PM demodulator is implemented by means of a multiplier, possibly preceded by limiters in order to eliminate AM modulation, or linearize its transfer (see Section 7.5.1). In the single-ended demodulator, this multiplier should be followed by a low pass filter in order to eliminate the double frequency terms. The PM demodulator in the balanced structure, which follows from figure 3.30a by application of the simplifications for small phase values $\Delta\Phi_{\text{out}}(t)$, double frequency terms are canceled. The balanced demodulator is therefore able to demodulate FM waves with zero carrier frequency, while the single-ended one is not. In both variants, the phase offset in the output signal due to the carrier frequency is eliminated when the input waves of the PM demodulator are phase in quadrature. This means that the time delay τ_d should be chosen such that

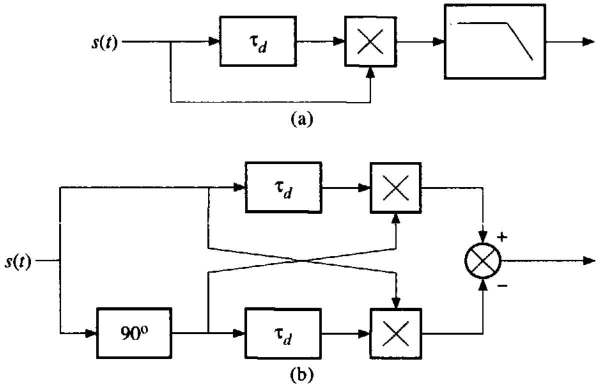


Figure 3.31: Quadrature demodulator. a) single-ended, b) balanced.

$$\tau_d = \frac{\pi}{2\omega_o} + k\pi, \tag{3.35}$$

where k denotes some integer number.

The various systems described in literature mainly differ in the way the delay-line is implemented. In [34] , a single tuned LC-circuit is used as delay line, a well-known implementation, and tuned by digital circuitry. In [35] an integrated delay line is realized by means of slow limiter stages, while in [36, 37] all-pass filters are used.

The φ-detector

This type of demodulator is similar to the quadrature demodulator, except for the fact that the PM demodulator is implemented by means of sampling. The detector output level becomes high when one of the input waves crosses through zero, and returns to zero again when the other wave crosses through zero. The duty-cycle, in this case the mean value of the detector output during a single period, is thus proportional to the phase difference between the input waves. Electron-tube implementations of this type of detector are found in [25, 38]. Modern implementations of this type is for example an S-R latch or an EX-OR port.

3.5.9 FM Demodulation by a Fixed Phase Difference

This section discusses the second sub-class of FM demodulators based on FM-PM conversion: those that establish the conversion with the aid of a fixed phase difference. First, the dynamic range capabilities of these demodulators are discussed. Subsequently, the examples encountered in literature are briefly discussed.

Demodulator Characteristics

With respect to the dynamic range and information handling capacity, the following can be remarked.

Phase Offset As discussed in Section 3.5.3, FM demodulators based on FM-PM conversion with the aid of a fixed phase difference, widely known as zero-crossing detectors, are unable to eliminate the carrier induced offset. This is due to the fact, that a clock, i.e. a 'time-detector', possesses a non-periodic transfer, as opposed to a phase-detector. Furthermore, the carrier frequency of the input FM wave cannot be zero-valued, but has to be chosen such, that the Nyquist rate for the samples of the instantaneous frequency is satisfied. A significant fraction of the demodulator DR is therefore spoiled to the non-informative offset.

Noise Due to the presence of a large offset component, the output noise level this type of FM demodulators is generally larger than the minimum possible level. As discussed in Chapter 4, it is the cause of a profound influence of internally generated noise on the output noise level.

Conclusion Based on the inevitable presence of a large phase offset, and the corresponding relatively large noise level, it is concluded that this type of FM demodulator cannot reach the upper bound on the demodulator DR, and is therefore unable to establish high-performance demodulation. The main advantages are its simplicity, suitability for integration, and quite good linearity.

Demodulators from Literature

Many variants of this type of FM demodulator have been developed in the past, see e.g. [32] for an overview. An more accurate clock-system was proposed that interpolates between several zero crossings of the FM wave [33]. Another example is used in [39], where the duty-cycle of the clock-system output wave is proportional to the instantaneous frequency of the FM wave.

3.5.10 FM Demodulation by Phase Feedback

This section discusses the third subclass of FM demodulators based on FM-PM conversion: phase feedback demodulators. A detailed discussion of this demodulator sub-class is given in Chapter 7. This section outlines their main characteristics.

Demodulator Characteristics

The dynamic range capabilities and information handling capacity of phase feedback demodulators are as follows.

Phase Offset From the discussion in Section 3.5.4, it follows that the FM-PM conversion algorithm is able to suppress the carrier-induced offset component. The objective of the loop is to drive the phase difference between the input FM wave, and the regenerated wave to zero. Consequently, when the oscillator in the feedback path, or some memory (filter/integrator) in the loop, supplies the offset component that is required for the carrier frequency of the regenerated wave, the offset component in the phase detector output, i.e. the *steady-state phase error* vanishes.

Noise In Chapter 7 is shown that phase feedback FM demodulators apply the required quadratic shaping to the input noise spectrum. Moreover, it is shown that these demodulators possess threshold extension capabilities.

Conclusion Phase feedback FM demodulators are capable to attain the upper bound on the FM demodulator DR, when the steady-state phase error equals zero. Therefore, high-performance demodulation can be established by these demodulators. Another advantage of these demodulators is their suitability for integration.

Demodulators from Literature

A vast amount of implementations of phase feedback demodulators have been developed and reported in literature. Some interesting examples are reported in [8, 9, 40]. In [40], a fully integratable, analog phase feedback demodulator for FM broadcast reception is described. The demodulators described in [8, 9] are digital implementations, that effectively apply some kind of sigma-delta modulation to the instantaneous frequency of the analog input FM wave. The output signal is a digital bit-stream, suitable for processing by a DSP.

3.5.11 FM Demodulation by Post Detection Conversion

This section discusses the last of the four sub-classes of FM demodulators based on FM-PM conversion: FM demodulators that employ post-detection conversion.

Demodulator Characteristics

As opposed to the three previously discussed FM demodulator sub-classes, FM demodulators based on post-detection conversion are unsuited for the demodulation of wideband FM waves, due to the limited phase-domain PM demodulators, in which demodulation without phase-jumps is possible. Therefore, the DR of these FM demodulators is besides by the supply currents and voltages, also limited by the range of the PM demodulator.

Phase Offset FM demodulators based on post-detection conversion are capable to suppress carrier induced phase offsets. In order to attain the upper bound on the demodulator DR, the PM demodulator should suppress any phase offset in its output signal. The frequency-offset in the FM demodulator output signal is suppressed when the differentiator transfer contains a zero at zero frequency.

Noise Thanks to the differentiator at the PM demodulator output, this type of FM demodulator applies quadratic shaping to the input noise spectrum. However, due to the limited range of PM demodulators, high modulation, and/or high noise introduce phase-jumps in the PM demodulator output signal, that, after differentiation, result in noise/distortion impulses at the FM demodulator output. Similar to so called "click noise", investigated in Chapter 5, this type of noise/distortion results in a white noise floor at the FM demodulator output, that may seriously deteriorate the demodulator DR.

Conclusion In theory, FM demodulators based on post-detection conversion are capable to attain the upper bound on the demodulator DR, as long as the phase noise and phase modulation in the input FM wave is small, compared to the range of the PM demodulator. In the presence of strong noise and/or large phase modulation, present in wideband FM waves, the performance rapidly degrades. True high-performance demodulation, in comparison to e.g. quadrature and phase feedback demodulators, is therefore infeasible with these demodulators.

Demodulators from Literature

As should be expected, the main concern in the implementation of post-detection conversion FM demodulators is the limited range of the PM demodulator. An attempt to increase this range is reported e.g. in [41]. The PM demodulator used in that reference is a digital implementation of the architecture depicted in figure 3.30b. The range extension is achieved by inclusion of a memory (counter) into the implementation of the quadrant detector, also called "jump detector". With the aid of the memory, this detector becomes capable to recognize and compensate for a large number of phase jumps; with n bits of memory, it is possible to detect and compensate for $2^n - 1$ phase jumps.

3.6 Demodulation by FM-PM-AM Conversion

In demodulators based on FM-PM-AM conversion, the actual demodulation operation is preceded by two conversion operations, that copy the message information from one carrier parameter to another. In Section 3.3 was discussed already that at most two such conversions may be performed on one and the same FM wave without loss of information. Moreover, it was discussed that the only valid scheme with two conversions is the one that converts the FM message information to phase information, subsequently converts this phase information into amplitude information and finally applies AM demodulation.

To date, this algorithm seems to be very inefficient, since two conversions are required, whereas the previous sections have shown that FM demodulation is possible already with only one conversion. However, in the 1940's and 1950's, this method of FM demodulation was an attractive alternative to FM demodulators based on FM-AM conversion and subsequent AM modulus detection. Tuned LC circuits instead of detuned ones could be used, that simultaneously participate in the demodulation process and perform IF filtering on the FM wave.

The only function performed in these FM demodulators that is not encountered in the other demodulator classes, studied in Section 3.4 and Section 3.5, is the PM-AM conversion operation. Therefore, this section focuses on that operation. First, the ideal PM-AM conversion function is discussed in Section 3.6.1, followed by a discussion of the PM-AM conversion algorithm in Section 3.6.2. Finally, the two well-known demodulators of this type, the Foster-Seeley detector and the ratio-detector, are discussed in Section 3.6.3.

3.6.1 Ideal PM-AM Conversion Function

The ideal PM-AM conversion function establishes a linear relation between the amplitude of the output wave and the phase difference between both its input waves. These input waves are derived from an FM input wave by means of an FM-PM converter. Thus, for an FM-PM converter output wave equal to $s_o(t) = A(t)\cos[\omega_o t + \varphi(t) + \Delta\varphi(t)]$, where $\Delta\varphi(t)$ denotes the phase difference between the FM-PM converter input and output, the PM-AM converter response becomes:

$$f_{\text{PM-AM}}\{A(t)\cos[\omega_o t + \varphi(t) + \Delta\varphi(t)]\} = \\ A_o\Delta\varphi(t)\cos[\omega_o t + \varphi(t) + \Delta\varphi(t)]. \quad (3.36)$$

The ideal PM-AM conversion transfer is depicted in figure 3.32.

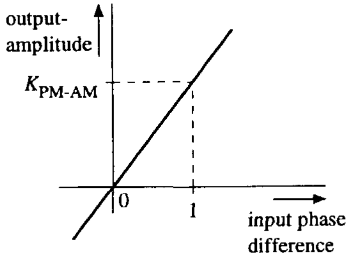


Figure 3.32: Ideal PM-AM conversion function.

3.6.2 PM-AM Conversion Algorithms

This section discusses the PM-AM conversion algorithm present in FM-PM-AM conversion FM demodulators. We first explain the principles of the algorithm by means of a discussion of its similarities and dissimilarities with the FM-AM and FM-PM conversion operations. Subsequently, it is shown that theoretically three different PM-AM conversion algorithms exist. Only one of them has been encountered in implementations of such FM demodulators. This algorithm is therefore elaborated in more detail.

Principles of the Algorithms

As discussed in Section 3.4 and Section 3.5, FM-AM can FM-PM conversion operations somehow implement a differentiation to time. The FM-AM conversion algorithm implements this differentiation directly, while FM-PM conversion algorithms implement time- differentiation of the carrier phase. As a result, most of these algorithms contain special kinds of linear filtering. The required type of filtering followed from a quasi-stationary interpretation of the ideal conversion functions.

The PM-AM conversion operation cannot be implemented by linear filtering, since it is not based on some differentiation to time. This is also observed from the fact that a quasi-stationary interpretation of the ideal PM-AM transfer, depicted in figure 3.32 does not represent the spectral amplitude or phase characteristic of a linear filter. Instead, it relates the spectral amplitude and phase to each-other, with the frequency as implicit parameter. This relation is, however, not very useful.

Insight into the principles of the PM-AM conversion algorithm is gained by an investigation of the FM-AM conversion algorithm, that copies the message information to the same carrier parameter: the carrier amplitude. From a mathematical point of view, that algorithm is based on the chain-rule of differentiation to time; the converter output equals the derivative of the FM wave to its instantaneous phase, times the time-derivative of this phase.

The PM-AM conversion algorithm may be based on the chain-rule in a similar way. In this case, however, differentiation should not be applied to time, but to a parameter that satisfies the following conditions:

- it is an implicit parameter of the phase difference $\Delta\varphi(t)$;
- the derivative of $\Delta\varphi(t)$ to this parameter is proportional to $\Delta\varphi(t)$.

From these conditions follows that the conversion gain $K_{\text{FM-PM}}$ of the FM-PM converter, positioned in front of the PM-AM converter, is the parameter that satisfies these conditions. Since, according to the discussion in Section 3.5, the phase difference $\Delta\varphi(t)$ introduced by the FM-PM converter this proportional to this parameter, and can be written as

$$\Delta\varphi(t) = K_{\text{FM-PM}}\dot{\varphi}(t), \tag{3.37}$$

the PM-AM converter output signal $s_o(t)$, obtained by differentiation to $K_{\text{FM-PM}}$, becomes

$$\begin{aligned} f_{\text{PM-AM}}\left[s_o(t)\right] &= \frac{\partial s_o(t)}{\partial K_{\text{FM-PM}}} \\ &= \frac{\partial A(t)\cos\Phi_o(t)}{\partial\Phi_o(t)}\,\frac{\partial\Phi_o(t)}{\partial K_{\text{FM-PM}}} \\ &= -A(t)\dot{\varphi}(t)\sin\left[\omega_o t + \varphi(t) + \Delta\varphi(t)\right]. \end{aligned} \tag{3.38}$$

A necessary requirement for proper PM-AM conversion, reflected by this expression, is that the FM-PM conversion operation should be applied directly to the input FM wave, prior to any demodulation operation. Otherwise, direct PM demodulation instead of indirect PM demodulation is established. From

Section 3.5 is noticed, that this requirement is satisfied only by FM-PM conversion based on a fixed, finite time delay τ_d. Every FM-PM-AM conversion FM demodulator therefore necessarily contains an implementation of that type of FM-PM conversion.

Derivation of the Algorithms

Similar to the procedure followed in Section 3.5, the three different PM-AM conversion algorithms follow by application of the definition formula for the differentiation in (3.38). If the FM-PM converter output signal is denoted by $s_o\left(t, K_{\text{FM-PM}}\right)$, which explicitly states its dependence on $K_{\text{FM-PM}}$, the PM-AM conversion algorithm may be expressed as

$$\frac{\Delta\varphi(t)}{K_{\text{FM-PM}}} s_o\left(t, K_{\text{FM-PM}}\right) = \lim_{\Delta K \to 0} \frac{s_o\left(t, K_{\text{FM-PM}} + \Delta K\right) - s_o\left(t, K_{\text{FM-PM}}\right)}{\Delta K}. \tag{3.39}$$

Thus, this expression shows that PM-AM conversion basically consists of a subtraction of the output signals of two FM-PM converters with a slightly different gain.

This algorithm can theoretically be implemented in three different ways:

- by a fixed, finite gain-difference ΔK;
- by a fixed, finite difference between both waves in the numerator;
- by application of adaptive feedback.

The first approach corresponds to subtraction of the outputs of two FM-PM converters, with different conversion gains. This is similar to FM-PM conversion by means of a fixed time-delay.

The second approach, similar to FM-PM conversion by means of a fixed phase difference, corresponds to sweeping the conversion gain of one FM-PM converter, until the difference between both FM-PM converter outputs reaches a fixed "threshold level". The PM-AM converter output is proportional to the ratio of this difference and the gain-difference that establishes it.

The third approach, similar to phase feedback in FM-PM converters, establishes an adaptive feedback loop around two FM-PM converters, that controls the gain-difference to zero. An essential part in this scheme, similar to the integration to time performed by the oscillator in phase feedback demodulators (see Chapter 7), is the rather awkward integration of the PM-AM converter output wave to the FM-PM conversion gain.

Only the first of these three algorithms has been applied in FM demodulator implementations. Therefore, it is interesting to consider this algorithm in more detail. The latter two are merely theoretical possibilities, that are not elaborated in detail in this section.

PM-AM Conversion by a Fixed Gain Difference

The architecture corresponding to this PM-AM conversion algorithm is depicted in figure 3.33. Figure 3.33a depicts the general architecture, that is obtained from expression (3.39). However, since only the gain difference of both FM-PM converters is of interest, it is allowed to simplify this architecture by setting one of the two conversion gains to zero. Obviously, a zero-valued FM-PM conversion gain corresponds to linear amplification of the input FM wave. Further, as discussed at the beginning of this section, the FM-PM converter is necessarily implemented by means of a delay-line. Figure 3.33b depicts the resulting simplified architecture.

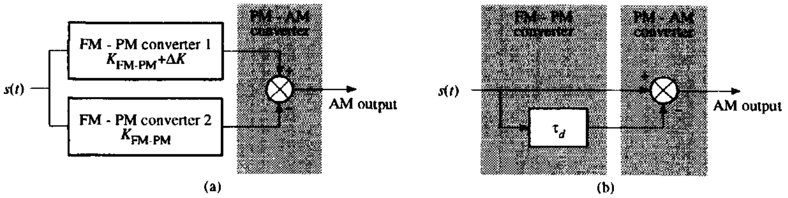


Figure 3.33: Architecture of the PM-AM converter. a) general architecture, b) simplified architecture.

The demodulators discussed in Section 3.6.3 use a balanced version of this architecture, that contains one PM-AM converter with a positive conversion gain and one with a negative conversion gain, obtained by replacement of the subtraction in figure 3.33 by an addition. This balanced double conversion demodulator architecture is depicted in figure 3.34.

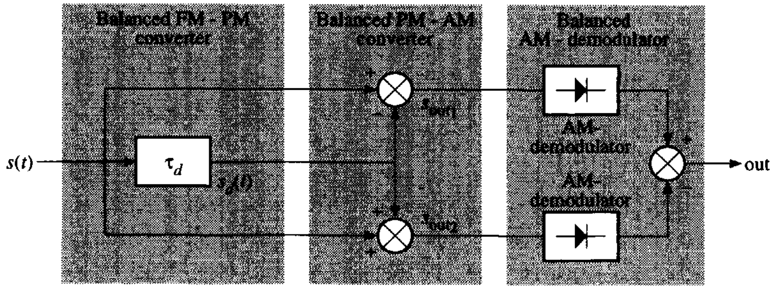

Balanced FM - PM converter
Balanced PM - AM converter
Balanced AM - demodulator
s(t)
τ_d
$s_d(t)$
+
−
s_{out1}
+
+
s_{out2}
AM-demodulator
AM-demodulator
+
−
out

Figure 3.34: Balanced double conversion demodulator architecture.

The operation of this demodulator is well explained by means of the phasor representation of the PM-AM converter output waves $s_{\mathrm{out},1}(t)$ and $s_{\mathrm{out},2}(t)$, depicted in figure 3.35. In this phasor representation, it is assumed that the

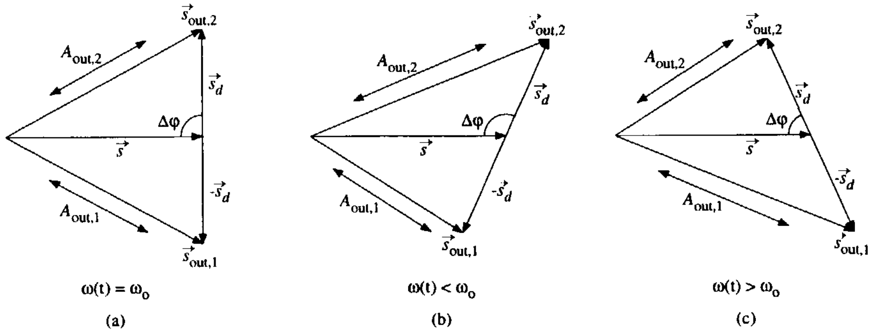


Figure 3.35: Phasor representation of the PM-AM converter output waves, for three different values of the instantaneous frequency $\omega(t) = \omega_o + \dot{\varphi}(t)$ of the FM wave. a) $\omega(t) = \omega_o$, b) $\omega(t) < \omega_o$, c) $\omega(t) > \omega_o$.

FM wave at the delay-line output, denoted by $s_d(t)$, is shifted with respect to $s(t)$ by $-\omega_o\tau_d = \pi/2$, such that $\vec{s}$ and $\vec{s}_d$ are orthogonal in the absence of modulation, as depicted in figure 3.35a. The phasors corresponding to the PM-AM converter output waves, $\vec{s}_{\mathrm{out},1}$ and $\vec{s}_{\mathrm{out},2}$, equal the vector subtraction and addition of $\vec{s}$ and $\vec{s}_d$ respectively. The balanced AM demodulator detects the difference between the length of these phasors, denoted by $A_{\mathrm{out},1}$ and $A_{\mathrm{out},2}$.

We apply the quasi-stationary approximation in order to explain the three situations in figure 3.35. In figure 3.35a, $\dot{\varphi}(t) = 0$, and consequently, the 'instantaneous' frequency of $s(t)$ equals the carrier frequency ω_o. Consequently, $\vec{s}$ and $\vec{s}_d$ are in quadrature, i.e. the angle $\Delta\varphi = \pi/2$, resulting in a zero demodulator output signal since $A_{\mathrm{out},1} = A_{\mathrm{out},2}$. When $\dot{\varphi}(t) < 0$, as in figure 3.35b, the input carrier frequency $\omega(t)$ is smaller than ω_o, such that $\omega(t) - \omega_o < 0$, resulting in a phase angle $\Delta\varphi = \omega(t)\tau_d < \pi/2$, $A_{\mathrm{out},1} < A_{\mathrm{out},2}$. Consequently, the demodulator output signal is negative in this case. Oppositely, when $\omega(t) > \omega_o$ the demodulator output signal becomes positive.

An expression of the demodulator output signal is obtained as follows. The amplitudes $|A_{\mathrm{out},1}|$ and $|A_{\mathrm{out},2}|$ follow by elaboration of $s(t) \pm s_d(t)$ as

$$A_{\mathrm{out},1,2}(t) = A(t)\sqrt{2\left\{1 \pm \cos\left[\varphi(t) - \varphi(t-\tau_d) + \omega_o\tau_d\right]\right\}}. \tag{3.40}$$

This expression shows that the maximum PM-AM conversion gain is obtained when $\omega_o\tau_d = \pi/2$, as expected from figure 3.35. In that case, the cos-function is replaced by a sin-function. Further, if τ_d is such that $\varphi(t)$ and $\varphi(t-\tau_d)$ differ only slightly, then

$$A_{\mathrm{out},1,2}(t) \approx A(t)\sqrt{2}\left[1 \pm \frac{\tau_d}{2}\dot{\varphi}(t)\right]. \tag{3.41}$$

The AM modulation index of the output wave is usually much smaller than unity, which allows the use of AM-modulus detection in the subsequent AM demodulation.

3.6.3 Examples of FM-PM-AM Conversion Demodulators

In this section we discuss to well-known FM demodulator circuits that employ the FM-PM-AM conversion FM demodulation principle outlined in the previous section.

The Foster-Seeley Detector

The Foster-Seeley detector [21, 25, 42, 43], is depicted in figure 3.36. This circuit consists of two inductively coupled LC circuits. The capacitor C_1 couples both circuits in such a way that the voltages across them are in quadrature at the resonant frequency, for both LC-tanks equal to the FM carrier frequency ω_o. In figure 3.35, $\vec{s}$ corresponds to the voltage across the primary LC circuit, while $\vec{s}_d$, the delayed FM wave, corresponds to the voltage across the secondary LC-tank. The coil L_1 establishes the appropriate DC reference level. Finally, the balanced peak detector determines the difference between the amplitudes $A_{\text{out},1}(t)$ and $A_{\text{out},2}(t)$. It is clear that this circuit is very sensitive to variations in the amplitude of the input FM wave; any fluctuations in this amplitude result in multiplicative errors in the demodulator output signal. An important advantage of this type of FM demodulators is the fact that the center frequency

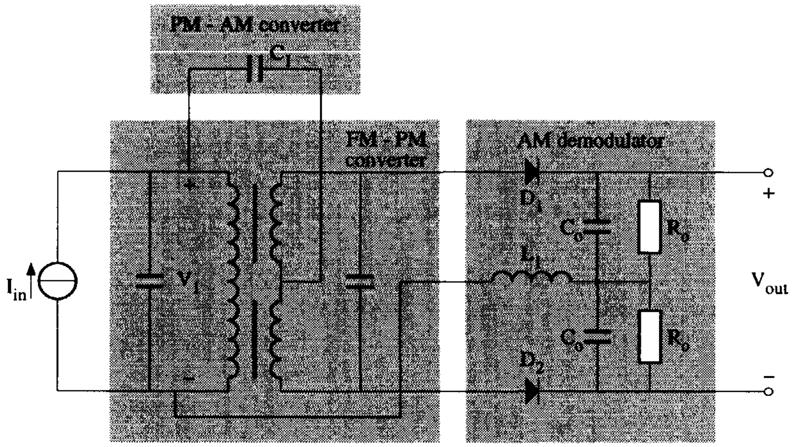


Figure 3.36: Foster-Seeley FM detector.

of the LC circuits coincides with the carrier frequency. Consequently, both LC circuits operate as IF filter on the input FM wave.

The Ratio Detector

The ratio detector [10, 21, 25, 43, 44] has for a long time been one of the most popular FM demodulators. The advantage of this demodulator in comparison to the Foster-Seeley detector is its ability to suppresses fluctuations and noise in the amplitude of the input FM wave. The Foster-Seeley detector lacks such a function. Due to this function, the ratiodetector outperforms the Foster-Seeley detector in the presence of noise.

A simplified schematic of the ratio detector is depicted in figure 3.37. The amplitude limiting function is established by the resistors R_o, that, as opposed to the Foster-Seeley detector of figure 3.36, are *not* connected to the tap of both capacitors C_o, an additional capacitor C_2, and a *reversed* diode D_2. Due to the reversal of D_2, the voltage across C_2 corresponds to the *addition* of the envelopes $A_{\text{out},1}$ and $A_{\text{out},2}$, instead of their difference, as in the Foster-Seeley detector. Both these envelopes are proportional to the fluctuating input FM carrier amplitude. Their addition depends only on the fluctuations on the input carrier amplitude, but *not* on the FM-PM-AM converted message information. Therefore, if C_2 is chosen sufficiently large, the filtering applied by combination $2R_o$, C_2 suppresses all fluctuations in the carrier amplitude, such that the voltage across C_2 becomes independent of these fluctuations.

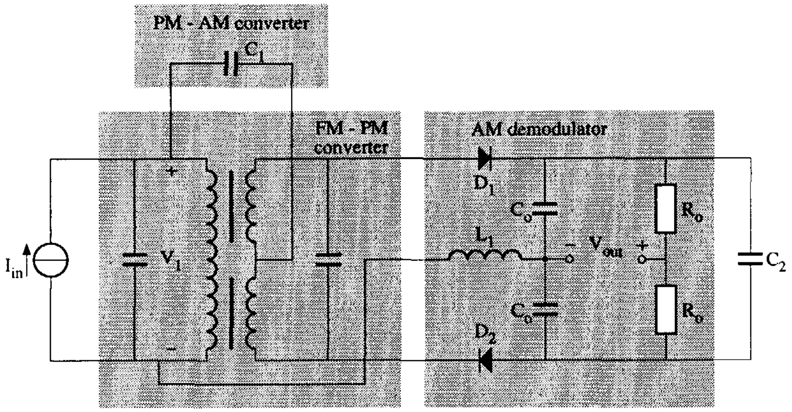


Figure 3.37: Ratio-Detector.

The demodulator output signal equals the voltage difference between the tap of the capacitors C_o, and the tap of the resistors R_o. The voltage across each of the resistors R_o equals half the voltage over the capacitor C_2, which corresponds to the average value of $A_{\text{out},1}+A_{\text{out},2}$ (see figure 3.35). The voltages across the capacitors C_o correspond to $A_{\text{out},1}$ and $A_{\text{out},2}$ respectively. They are proportional to the FM-PM-AM converted message information, but do *not* contain fluctuations due to the input carrier amplitude; these are eliminated by C_2. The output voltage is therefore proportional to $A_{\text{out},1} - A_{\text{out},2}$. Thus, in fact, the output is determined by the *ratio* of the voltages across the capacitors C_o, which explains the name of the circuit.

3.7 Conclusions

A classification of FM demodulation principles is an indispensable instrument in a structured approach towards FM demodulator design, that allows introduction of hierarchy into the design procedure. Moreover, it allows deliberate selection of the most suitable demodulator architecture in a very early stage of this procedure.

The classification developed in this chapter is schematically depicted in figure 3.38.

Physical FM demodulators cannot detect the instantaneous frequency of FM waves directly, since it is not associated to the energy of the wave. Instead, a conversion of the FM message information to the carrier amplitude (FM-AM) or the carrier phase (FM-PM) is required, followed by AM or PM demodulation.

Conversion to amplitude (FM-AM) is established by differentiation of the FM wave to time. Subsequently, the modulus of the FM/AM wave or its projection on a reference should be detected. Demodulators equipped with AM projection detection outperform those equipped with AM modulus detection, since this method of AM demodulation allows suppression of the offset component due to the carrier frequency. Since AM modulus detection does not allow AM modulation indices larger than unity, at least half of the dynamic range of FM demodulators equipped with such AM demodulators is spoiled to a non-informative component.

Conversion to phase (FM-PM) corresponds to differentiation of the instantaneous FM carrier phase, which can be established in four different ways. The PM demodulation can be established directly, by detection of the carrier phase, or indirectly, by conversion of the information contained in the carrier phase to the carrier amplitude, and subsequent AM demodulation.

Two of the four subclasses of FM demodulators based on FM-PM conversion and subsequent direct PM demodulation, the one based on a fixed time-delay and the one based on phase feedback, outperform the other two, based on a fixed time difference and post-detection conversion respectively, since they allow suppression of the carrier induced offset. With the FM-PM conversion principles employed by these demodulators, the maximum possible demodulator DR can be attained, i.e. the same DR as with FM demodulators based on FM-AM

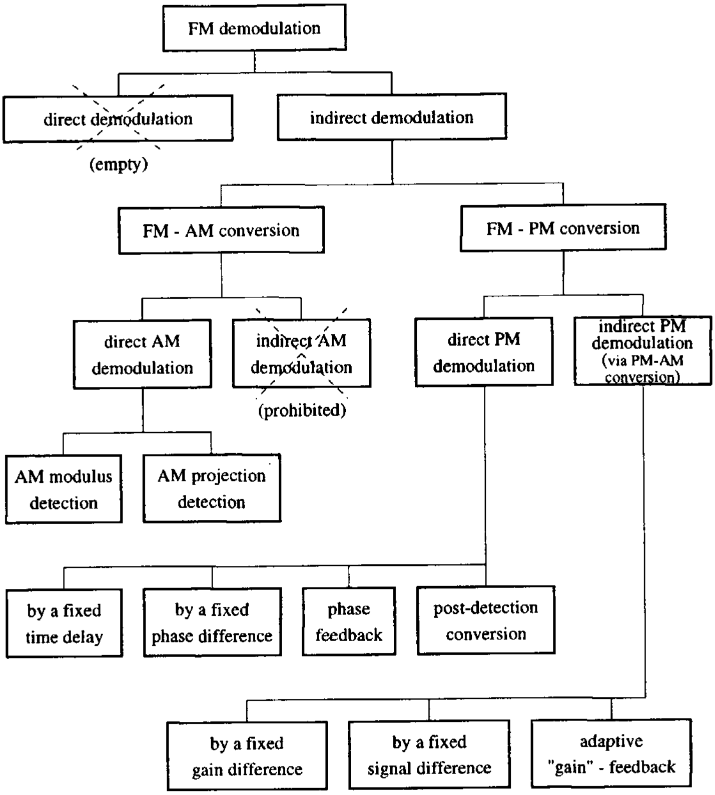


Figure 3.38: Classification of FM demodulation principles developed in this chapter.

conversion and subsequent AM projection detection.

Three subclasses of FM demodulation principles based on FM-PM conversion, followed by indirect PM demodulation, i.e. FM-PM-AM conversion demodulators, exist. However, in practice, only one of them has been implemented in demodulator circuits. Demodulators based on FM-PM-AM conversion necessarily establish FM-PM conversion with the aid of a fixed time-delay, and AM modulus demodulation. Therefore, their performance is moderate; the upper bound on the demodulator DR cannot be attained with these demodulators.

All demodulators that were encountered in literature could be classified according to this classification.

References

[1] E.H. Nordholt, *Design of High-Performance Negative Feedback Amplifiers,* Else-

vier, Amsterdam, 1983.

[2] C. J. M. Verhoeven, *First Order Oscillators*, PhD thesis, Delft University of Technology, Delft, The Netherlands, 1990.

[3] C. A. M. Boon, *Design of High-Performance Negative Feedback Oscillators*, PhD thesis, Delft University of Technology, Delft, The Netherlands, 1989.

[4] Gert Groenewold, *Optimal Dynamic Range Integrated Continuous-Time Filters*, PhD thesis, Delft University of Technology, Delft, The Netherlands, 1992.

[5] A. van Staveren, *Structured Electronic Design of High-Performance Low-Voltage Low-Power References*, PhD thesis, Delft University of Technology, Delft, The Netherlands, 1997.

[6] W.A. Serdijn, C.J.M. Verhoeven, and A.H.M. van Roermund, Eds., *Analog IC Techniques for Low-Voltage Low-Power Electronics*, Delft University Press, Delft, 1995.

[7] C.J.M. Verhoeven, A. van Staveren, and G.L.E. Monna, "Structured electronic design, negative-feedback amplifiers", To be published.

[8] R. Douglas Beards and Miles A. Copeland, "An oversampling delta-sigma frequency discriminator", *IEEE Transactions on Circuits and Systems-II*, vol. 41, no. 1, pp. 26–32, Jan. 1994.

[9] Ian Galton, "Analog-input digital phase-locked loops for precise frequency and phase demodulation", *IEEE Transactions on Circuits and Systems-II*, vol. 42, pp. 621–630, Oct. 1995.

[10] Edwin H. Armstrong, "A study of the operating characteristics of the ratio detector and its place in radio history", *Proceedings of the Radio Club of America*, vol. 64, no. 3, pp. 217–232, Nov. 1990, Reprint of vol. 25, no. 3, 1948.

[11] H. Meinke und F.W. Gundlach, *Taschenbuch der Hochfrequenztechnik*, Springer-Verlag, Berlin, 1956.

[12] C.E. Shannon, "A mathematical theory of communication", *The Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, 1948.

[13] Marcelo Alonso and Edward J. Finn, *Fundamental University Physics volume III, quantum and statistical physics*, Addison-Wesley Publishing Company, Reading, Massachusetts, 1968.

[14] L. Solymar and D. Walsh, *Lectures on the Electrical Properties of Materials, fourth edition*, Oxford University Press, Oxford, 1990.

[15] Edwin H. Armstrong, "A method of reducing disturbances in radio signaling by a system of frequency modulation", *Proceedings of the IRE*, , no. 5, pp. 689–741, May 1936.

[16] Edwin H. Armstrong, "Evolution of frequency modulation", *Proceedings of the Radio Club of America*, vol. 64, no. 3, pp. 179–188, Nov. 1990, Reprint of dec. 1940, pp.485-494, embracing the substance of lectures presented before the *American Institute of Electrical Engineers*.

[17] Wilbur B. Davenport and William L. Root, *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill Book Company, New York, 1958.

[18] A. Bruce Carlson, *Communication Systems, An introduction to Signals and noise in Electrical Communication,* McGraw-Hill International Editions, Singapore, 1986.

[19] G.L.E. Monna, *Design of Low-Voltage Integrated Filter-Mixer Systems,* PhD thesis, Delft University of Technology, Delft, The Netherlands, 1996.

[20] Eric A. M. Klumperink, Carlo T. Klein, Bas Rüggeberg, and Ed J. M. van Tuijl, "AM suppression with low AM to PM conversion with the aid of a variable-gain amplifier", *IEEE Journal of Solid State Circuits,* vol. 31, no. 5, pp. 625–633, May 1996.

[21] Samuel Seeley, *Electron-Tube Circuits, second edition,* McGraw-Hill Book Company, New York, 1958.

[22] J. Zenneck, *Lehrbuch der drahtlosen Telegraphy,* Enke, Stuttgart, 2nd edition, 1913.

[23] Lawrence B. Arguimbau, *Vacuum-Tube Circuits,* John Wiley and Sons, London, 1948.

[24] Edwin H. Armstrong, "Some recent developments of regenerative circuits", *Proceedings of the IRE,* vol. 10, no. 8, pp. 244–260, 1922.

[25] L. Ratheiser, *Rundfunk-Röhren, Eigenschaften und Anwendung der neuen UKW Röhren,* Regelien, Berlin, 1951.

[26] Domine M. W. Leenaerts, "Chaotic behaviour in super regenerative detectors", *IEEE Transactions on Circuits and Systems-I,* vol. 43, no. 3, pp. 169–176, Mar. 1996.

[27] Heinz Göckler, "Verfahren zur Demodulation von frequenzmodulierten Signalen", Offenlegungsschrift DE 43 10 462 A1, Deutsches Patentamt, 1994.

[28] Bang-Sup Song and In Seop Lee, "A digital FM demodulator for FM, TV and wireless", *IEEE Transactions on Circuits and Systems-II,* vol. vol. 42, no. 12, pp. 821–825, Dec. 1995.

[29] John F. Wilson, Richard Youell, Tony H. Richards, Gwilym Luff, and Ralf Pilaski, "A single-chip VHF and UHF receiver for radio paging", *IEEE Journal of Solid State Circuits,* vol. 26, no. 12, pp. 1944–1950, Dec. 1991.

[30] Anatol I. Zverev, *Handbook of Filter Synthesis,* John Wiley and Sons, New York, 1967.

[31] Timo Rahkonen, Kari Kananen, and Juha Kostamovaara, "A digital FM demodulator chip based on measurement of IF-signal's period", in *Proceedings of the European Solid-State Circuits Conference,* xxxx, Sept. xxx, pp. 102–105.

[32] Edouard Labin, "Theory of frequency counting and its application to the detection of frequency-modulated waves", *Proceedings of the IRE,* vol. 36, no. 7, pp. 828–839, July 1948.

[33] R.G. Wiley, "Approximate FM demodulation using zero crossings", *IEEE Transactions on Communications,* vol. 29, no. 7, pp. 1061–1065, July 1981.

[34] Rick W. Miller and Daniel M. Hutchinson, "Bus aligned quadrature FM detector", U.S. Patent 5,596,298, January 21 1997.

[35] L.P. de Jong, E.H. Nordholt, and C.M.C.J. Hooghiemstra, "High-performance integrated receiver for optical fiber transmission of wideband FM video signals", *IEEE Transactions on Consumer Electronics*, vol. 33, no. 3, pp. 473–480, Aug. 1987.

[36] D. Kasperkovitz, "An integrated FM receiver", *Microelectronics Reliability*, vol. 21, no. 2, pp. 183–189, 1981.

[37] W.G. Kasperkovitz, "FM receivers for mono and stereo on a single chip", *Philips Technical Review*, vol. 41, no. 6, pp. 169–182, 1983–1984.

[38] J.L.H. Jonker en A.J.W.M. Overbeek, "De ϕ-detector, een detectorbuis voor frequentiemodulatie (in dutch)", *Philips Technisch Tijdschrift*, vol. 11, no. 2, pp. 33–64, Februari 1949.

[39] Bang-Sup Song and Jeffrey R. Barner, "A CMOS double-heterodyne FM receiver", *IEEE Journal of Solid State Circuits*, vol. 21, no. 6, pp. 916–923, Dec. 1986.

[40] A. Sempel and H. van Nieuwenburg, "A fully-integrated HIFI PLL FM-demodulator", in *Dig. IEEE International Solid State Circuits Conference*, Feb. 1990, pp. 102–103.

[41] Noël Boutin, "An arctangent type wideband PM/FM demodulator with improved performance", *IEEE Transactions on Consumer Electronics*, vol. 38, no. 1, pp. 5–9, Feb. 1992.

[42] D.E. Foster and S. Seeley, "Automatic tuning, simplified circuits, and design practice", *Proceedings of the IRE*, vol. 25, no. 3, pp. 289–313, Mar. 1937.

[43] Austin V. Eastman, *Fundamentals of Vacuum Tubes, second edition*, McGraw-Hill Book Company, New York, 1941.

[44] Stuart W. Seeley and Jack Avins, "The ratio detector", *RCA Review*, vol. 8, pp. 201–236, 1947.

Chapter 4

FM Demodulator Design

The investigation of FM demodulation principles and algorithms in Chapter 3 showed, that nearly all types of FM demodulators consist of combinations of four classes of subsystems:

- FM-AM converters;
- AM demodulators;
- FM-PM converters;
- PM demodulators.

The various algorithms available for the implementation of each of these subsystems were derived.

This chapter discusses the design of these four classes of subsystems encountered in FM demodulators. Starting from the algorithms available for the implementation of these subsystems, derived in Chapter 3, the influence of the technology-independent design aspects on the performance of the FM demodulator as a whole is evaluated. Design rules for the optimization of this performance, expressed in terms of distortion, noise, and dynamic range are derived.

The design of PM-AM converters, contained in FM-PM-AM conversion demodulators, is disregarded in this chapter for two reasons. In the first place, FM-PM-AM conversion demodulators have currently lost their attractiveness, especially in IC technology, due to the presence of two conversion operations. Secondly, since PM-AM conversion simply consist of an addition, there isn't much to say about its design.

Due to the nonlinear nature of FM demodulation, the formulation of a proper definition of the FM demodulator dynamic range is a nontrivial problem. A completely satisfactory definition for nonlinear systems has not yet been obtained, despite some preliminary attempts to find one [1, 2]. In the context of this chapter, a suitable measure for this dynamic range is given by the maximum Signal-to-Noise-and-Distortion ratio (SNDR) that can exist at the demodulator output, for a given input carrier level and input CNR. This measure is used for the derivation of the various design rules. In order to clarify the discussion, the maximization of the FM demodulator SNDR is treated in detail for analog, continuous-time systems, intended for processing of amplitude- and time-continuous signals. Analogous discussions hold for the processing of signals from the other three domains, discussed in Section 3.2.

An outline is as follows. Section 4.1 discusses the dependence of the FM demodulator dynamic range on the characteristics of the subsystems contained in it. Subsequently, design rules for these subsystems, with the objective to maximize the FM demodulator DR, are derived in Section 4.2 through Section 4.5. The conclusions are given in Section 4.6.

4.1 FM Demodulator Dynamic Range

Maximization of the FM demodulator performance, represented by the demodulator dynamic range (DR), should be established by proper design of the subsystems contained in the demodulator. This section identifies the various characteristics of the subsystems that determine the FM demodulator DR. The design of these subsystem characteristics is considered in subsequent sections.

As mentioned in the introduction, the maximum SNDR that can exist in the demodulator is used as measure for the demodulator DR. The influence of non-idealities in the subsystems on this SNDR is illustrated by figure 4.1. The maximum SNDR that can exist in an 'ideal' FM demodulator, containing ideal

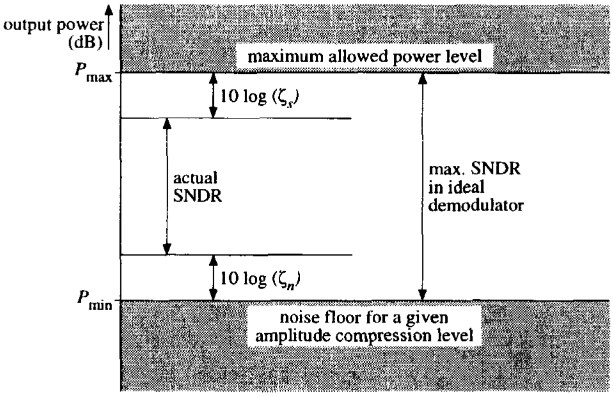

output power (dB)
maximum allowed power level
P_{max}
$10 \log (\zeta_s)$
actual SNDR
max. SNDR in ideal demodulator
$10 \log (\zeta_n)$
P_{min}
noise floor for a given amplitude compression level

Figure 4.1: Deterioration of the FM demodulator SNDR due to non-idealities in the demodulator subsystems.

subsystems, is represented by the distance between the maximum power level P_{max}, and the minimum power level P_{min}. The maximum power level represents the fact that the signal swings in electronic demodulators are generally bounded by the supply currents and voltages. The minimum power level represents the noise power generated at the output of the ideal demodulator, in response to amplitude and frequency noise contained in the input FM wave. For a fair comparison between the ideal demodulator and a non-ideal demodulator, it is assumed that the same level of compression is applied to the input carrier amplitude of both FM demodulators, that, as will be explained in Section 5.4.2, determines the contribution of the amplitude noise.

The parameters ξ_s and ξ_n represent the deterioration of the maximum power level and minimum power level respectively due to non-idealities in the demodulator subsystems. Their dependence on the characteristics of the FM-AM converter, or FM-PM converter, and the AM demodulator, or PM demodulator, contained in the FM demodulator architecture, is discussed in Section 4.1.1 and Section 4.1.2 respectively.

4.1.1 Dependence on the FM-AM/FM-PM Converter

This section considers the reduction of the FM demodulator SNDR due to nonidealities in FM-AM converters and FM-PM converters. First, the reduction of the maximum signal power level P_{max}, represented by ξ_s is considered. Subsequently, the increment of the minimum required signal power level P_{min}, represented by ξ_n, is discussed.

Deterioration of the Maximum Signal Power Level

The maximum allowed demodulator output signal power is generally smaller than P_{max} due to the presence of an 'offset' component. This offset, denoted by ω_{offs}, is the converter response to the carrier frequency ω_o, i.e. the DC-component of the instantaneous frequency, of the FM wave. In FM-AM converters, it results in an output carrier amplitude given by

$$A_o(t) = AK_{\text{FM-AM}}\left[\omega_{\text{offs}} + \dot{\varphi}(t)\right], \tag{4.1}$$

while in FM-PM converters, it results in an instantaneous phase difference between the input and output carrier wave, equal to

$$\Delta\Phi_o(t) = K_{\text{FM-PM}}\left[\omega_{\text{offs}} + \dot{\varphi}(t)\right]. \tag{4.2}$$

In a properly configured FM demodulator, the power contents of the converter output signal, and the demodulator output signal should be as large as possible, but is not allowed to exceed P_{max}. Therefore, when $\Delta\omega$ denotes the RMS frequency deviation of the FM message signal $\dot{\varphi}(t)$, and K_{dem} denotes the combination of the FM-AM/FM-PM converter and AM/PM demodulator conversion gain, it follows from (4.1) and (4.2) that

$$\begin{aligned} P_{\text{max}} &\geq K_{\text{dem}}^2(\Delta\omega)^2\left[1+\left(\frac{\omega_{\text{offs}}}{\Delta\omega}\right)^2\right] \\ &= K_{\text{dem}}^2(\Delta\omega)^2\xi_s. \end{aligned} \tag{4.3}$$

From this expression, it follows that large offsets leave only a small fraction of the demodulator DR to the power contents of the message $\dot{\varphi}(t)$, while the largest part of the DR is spoiled to the (non-informative) carrier-induced amplitude/phase offset.

As an illustration of the DR deterioration due to a carrier-induced offset, consider an FM demodulator for FM broadcast reception, that expects input waves with a center frequency of $\omega_o/2\pi = 10.7$ MHz, and a maximum frequency deviation of 75 kHz. Even when the RMS deviation, which in general is considerably smaller than the maximum deviation, equals $\Delta\omega = 75$ kHz, the offset would reduce the demodulator DR by 43 dB (!) when no offset reduction techniques are applied, i.e. when $\omega_{\text{offs}} = \omega_o$. Of course, such a reduction is totally unacceptable, and should be prevented by suppression of the amplitude/phase offset, as will be discussed in Section 4.2.1.

Deterioration of the Minimum Signal Power Level

The minimum demodulator output signal power level that still results in an intelligible output signal is generally larger than the minimum level P_{min} in the ideal demodulator. The following converter non-idealities contribute to this increase, represented by the factor ξ_n:

- distortion and interference due to the converter frequency transfer;
- generation of a carrier-induced offset ω_{offs};
- non-ideal shaping of the input noise;
- internal noise generation.

The distortion and interference can be minimized by proper design of the FM-AM and FM-PM converter frequency transfer function, as discussed in Section 4.2.2 and Section 4.4.2.

Besides a reduction of the maximum power level, the carrier-induced amplitude/phase offset ω_{offs} also increases the contribution of the noise contained in the input carrier amplitude to the demodulator output, as will be explained in Section 5.5. Consequently, when only part of the noise is eliminated from the input carrier amplitude, i.e. when *finite* or *no* amplitude compression is applied,

ω_{offs} should be minimized in order to avoid a (further) increase of the output noise level.

The quadratic shaping applied by an ideal FM demodulator to the frequency noise of the input FM wave constitutes the SNR improvement of FM transmission, in comparison to AM transmission, as explained in Section 2.3.2. The shaping established by FM-AM and FM-PM converters should therefore also be truly quadratic, as considered in Section 4.2.3 and Section 4.4.3.

The level of internally generated noise, by components inside the converters, depends on the applied technology. However, as shown in Section 4.2.3 and Section 4.4.3, its effect on the demodulator output noise is also highly dependent on its position inside the demodulator architecture.

4.1.2 Dependence on the AM/PM Demodulator

This section considers the reduction of the FM demodulator SNDR due to non-idealities in AM demodulators and PM demodulators. First, the reduction of the maximum signal power level is discussed. Subsequently, the increase of the minimum signal power level is considered.

Deterioration of the Maximum Signal Power Level

The AM and PM demodulators contained in the FM demodulator architecture do not deteriorate the demodulator SNDR by generation of offset components, but define a lower bound on the offset level that is required for proper demodulation. Thus, even when the converter possesses the ability, complete suppression of the offset is not allowed when the AM or PM demodulator defines a nonzero lower bound on the offset.

The minimum required amplitude/phase offsets for the various types of AM and PM demodulators are considered in Section 4.3 and Section 4.5 respectively.

Deterioration of the Minimum Signal Power Level

The minimum required signal power level is generally increased by the following non-idealities in AM and PM demodulators:

- the absence of phase selectivity;
- distortion;
- a nonzero lower bound on the carrier-induced offset;
- internal noise generation.

In essence, phase selectivity is the ability to distinguish different directions in the phasor plane, and allows suppression of non-ideal components that point in directions different from the requested component, containing the required message information. Obviously, this non-ideality, considered in Section 4.3, is only encountered in AM demodulators; PM demodulators are phase selective by definition.

The distortion introduced into the demodulator output signal is highly dependent on the applied type of AM/PM demodulator. Further, also the possibilities to minimize the distortion differ considerably among the various AM and PM demodulator types, as considered in Section 4.3 and Section 4.5.

As discussed previously, carrier induced offsets increase the contribution of amplitude noise to the demodulator output noise level. A lower bound on this offset, set forward by the AM or PM demodulator, is therefore also a lower bound on the deterioration of the FM demodulator SNDR.

The contribution of internally generated noise to the FM demodulator output noise level is highly dependent on the applied type of FM demodulator architecture. In some architectures, the internal noise results in a deteriorative white noise floor, while in others it results only in quadratic shaped noise.

4.2 FM-AM Converter Design

FM-AM converters constitute the operation of the class of FM demodulators discussed in Section 3.4. It was shown there, that the FM-AM conversion algorithm consists of differentiation to time, i.e. a special kind of linear filtering.

This section considers the design of FM-AM converters. Various design rules are derived, that aim at maximization of the FM demodulator SNDR.

As discussed in Section 4.1.1, the following FM-AM converter characteristics have a major influence on the FM demodulator SNDR:

- generation of an offset in the output carrier amplitude;
- distortion and interference in the frequency transfer;
- non-ideal noise shaping;
- internal noise generation.

Design rules for minimization of the offset component are considered in Section 4.2.1. Subsequently, design rules for the minimization of the distortion and interference are discussed in Section 4.2.2. Finally, design rules for minimization of the demodulator output noise level are considered in Section 4.2.3.

4.2.1 Offset Minimization

Section 3.4.2 noticed already, that a straight-forward implementation of the FM-AM conversion algorithm, differentiation to time, generally results in a large carrier-induced offset component in the FM-AM converter output carrier amplitude. As discussed in Section 4.1, such an offset considerably reduces the FM demodulator DR, and should therefore be eliminated.

This section discusses the possibilities to minimize carrier-induced offset in the FM demodulator output signal, by proper design of the FM-AM converter.

In essence, the offset in the amplitude of the converter output wave $s_o(t)$ is the result of a nonzero converter response to spectral components located at the carrier frequency ω_o of the input FM wave $s(t)$. Since FM-AM conversion is established by linear filtering of the FM wave, minimization, or even elimination, of this offset is achieved when the zero in the Laplace-domain converter transfer, a differentiator, coincides with the carrier frequency, i.e. when it is positioned at $s = \pm j\omega_o$. As illustrated by figure 4.2, this may be accomplished in two fundamentally different ways:

- translation of the FM wave's center-frequency to the differentiator-zero;
- translation of the differentiator-zero to the FM wave's center-frequency.

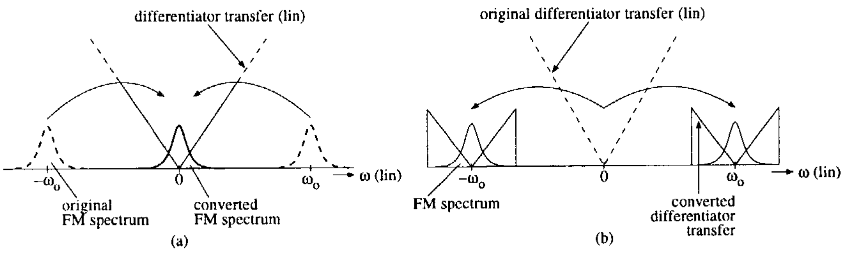


Figure 4.2: Offset elimination by a) translation of the FM center-frequency to the differentiator-zero ("zero IF"), b) translation of the differentiator-zero to the FM center-frequency ("band-pass differentiation").

These possibilities are separately discussed below.

Zero-IF Architecture

The first approach results in a so called "zero-IF" architecture, depicted in figure 4.3. This architecture converts the input FM wave to a zero center-frequency by means of a down-conversion mixer, and subsequently performs the FM-AM conversion by means of a differentiator, that contains a zero at $s = 0$. According to Section 3.4.7, AM demodulation of a wave with a zero-valued carrier frequency requires two such waves in phase quadrature, (I-Q), which explains the configuration in figure 4.3.

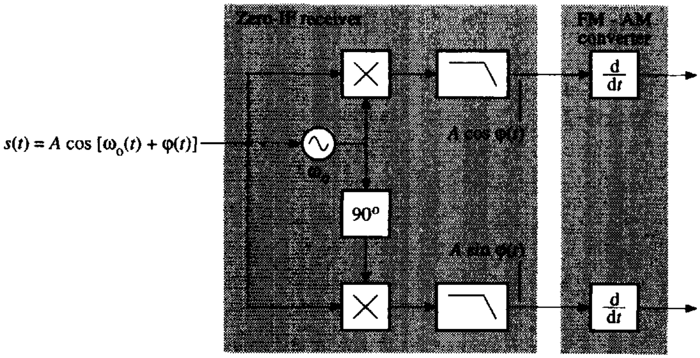


Figure 4.3: Minimization of the offset by means of a zero-IF architecture.

Band-Pass FM-AM Conversion

The second approach transforms the 'low-pass' differentiation established by the FM-AM converter into a "band-pass differentiation". This transformation is established by the so called 'biquadratic' transform [3], that replaces the complex frequency s in the FM-AM converter Laplace-domain transfer function with

$$s \rightarrow \frac{\omega_o^2}{s}, \tag{4.4}$$

where ω_o equals the carrier frequency of the input FM wave. An advantageous property of this transformation is its *orthogonality* with respect to distortion minimization; if the 'low-pass' FM-AM converter transfer, denoted by $H_{\text{diff,lp}}(s)$, yields the lowest possible level of distortion in the demodulator output signal of all 'low-pass' FM-AM converters, the corresponding 'band-pass' FM-AM converter, denoted by $H_{\text{diff,bp}}(s)$, yields the lowest possible distortion level of all band-pass FM-AM converters.

For example, if the low-pass FM-AM conversion is implemented by a 'low-pass' differentiator transfer $H_{\text{diff,lp}}(s)$, given by

$$H_{\text{diff,lp}}(s) = \frac{\omega_r s}{s^2 + s\frac{\omega_r}{Q} + \omega_r^2}, \tag{4.5}$$

where ω_r denotes the resonant frequency and Q the quality factor, the corresponding band-pass differentiator transfer, with zeros at $s = \pm j\omega_o$, equals

$$\begin{aligned} H_{\text{diff,bp}}(s) &= H_{\text{diff,lp}}\left(s + \frac{\omega_o^2}{s}\right) \\ &= \frac{\omega_r s\left(s^2 + \omega_o^2\right)}{s^4 + \frac{\omega_r}{Q}s^3 + \left(\omega_r^2 + 2\omega_o^2\right)s^2 + \frac{\omega_r\omega_o^2}{Q}s + \omega_o^4}. \end{aligned} \tag{4.6}$$

In early FM demodulators, this bandpass converter transfer was approximated by subtraction of the output signal of two slightly detuned, so called "stagger-tuned" LC-tanks [4], in order to eliminate the offset from the demodulator output and minimize the distortion, as mentioned in Section 3.4.6. The demodulator DR is however not improved by stagger-tuning, since the offset cancellation is positioned at the AM-demodulator output.

4.2.2 Distortion Minimization

This section discusses the minimization of the distortion and interference, caused by undesired components in the output signal, in the FM demodulator output signal, by proper design of the FM-AM converter frequency transfer. Since offset minimization by means of 'band-pass' FM-AM conversion is orthogonal to distortion minimization, as discussed in Section 4.2.1, it is sufficient to consider only distortion and interference minimization in 'low-pass' FM-AM converters.

The Laplace-domain transfer of a non-ideal FM-AM converter generally differs from the ideal transfer, denoted by $H_{\mathrm{FM-AM,id}}(s)$, equal to

$$H_{\mathrm{FM-AM,id}}(s) = K_{\mathrm{FM-AM}}s, \tag{4.7}$$

in the following aspects:

- the zero generally possess a non-zero real part;
- the transfer contains poles, and, possibly, additional zeros.

The types of distortion in the FM demodulator output signal due to each of these imperfections, and the measures required to minimize this distortion, are separately considered below.

Zero with a Non-zero Real Part

Unwanted components in the FM-AM converter output signal are basically due to deviations of the zero in the FM-AM converter transfer from its theoretical, ideal position. As illustrated by figure 4.4, two different types of deviations can be distinguished, that result in different unwanted components:

- deviations in the imaginary part of the zero;
- deviations in the real part of the zero.

From the discussions in Section 3.4.2 and Section 4.2.1 follows that a deviation of the *imaginary* part of the zero from the FM carrier frequency ω_o yields a carrier-induced offset in the *tangential* component of the FM-AM converter output wave $s_o(t)$, and, consequently also in the FM demodulator output signal. Thus,

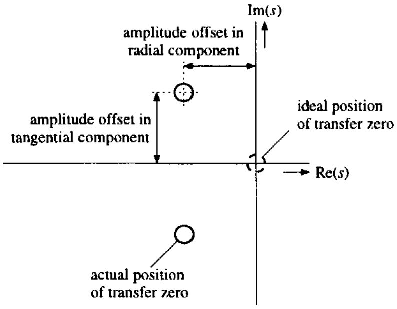

Im(s)
amplitude offset in radial component
amplitude offset in tangential component
ideal position of transfer zero
Re(s)
actual position of transfer zero

Figure 4.4: Deviations of the zero in the FM-AM converter transfer from its ideal position.

in order to suppress this offset, the zero in the transfer of low-pass FM-AM converters should possess a zero-valued imaginary part.

A deviation of the *real* part of the zero from the ideal value Re$(s) = 0$ yields an amplitude offset component in the *radial* component of the converter output wave, instead of in the tangential component. As shown below, this 'offset' complicates the elimination of the radial component from the demodulator output signal. A zero-valued real part is therefore favorable.

The cause of the amplitude offset in the radial component can be explained as follows. Suppose that the FM-AM converter transfer is given by

$$H_{\mathrm{FM-AM}}(s) = K_{\mathrm{FM-AM}}(s + z), \tag{4.8}$$

i.e. a parallel connection of an ideal differentiator, and an amplifier/attenuator with transfer z, followed by an amplifier of gain $K_{\mathrm{FM-AM}}$. Then, its response to the FM wave $s(t) = A(t)\cos[\omega_o t + \varphi(t)]$ equals

$$\begin{aligned} s_o(t) &= K_{\mathrm{FM-AM}}\frac{\mathrm{d}s(t)}{\mathrm{d}t} + K_{\mathrm{FM-AM}}zs(t) \\ &= K_{\mathrm{FM-AM}}\left[\omega_o + \dot{\varphi}(t)\right]u_{\mathrm{tan}}(t) \\ &\quad + K_{\mathrm{FM-AM}}\left[\dot{A}(t) + zA(t)\right]u_{\mathrm{rad}}(t). \end{aligned} \tag{4.9}$$

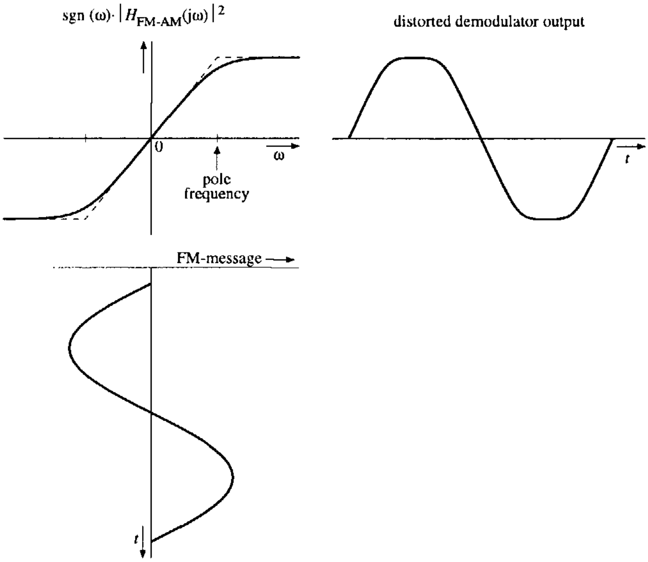
The required tangential component remains unaffected, but the radial component contains an additional term $zA(t)$, as a result of the direct feed-through. As a result of this component, the radial component can no longer be eliminated by suppression of the fluctuations in the amplitude $A(t)$ prior to FM-AM conversion (see Section 3.4.2).

Consequently, when a nonzero real part cannot be avoided, the AM demodulator itself should suppress the radial component in $s_o(t)$, in order to avoid interference and/or distortion in the FM demodulator output signal. This topic is considered in Section 4.3.

Poles and Additional Zeros

Poles and additional zeros in the FM-AM converter transfer distort the message signal at the FM demodulator output in two different ways. In the first place, they limit the converter bandwidth, resulting in narrow-band filtering distortion. Secondly, they introduce curvatures in the frequency transfer.

Distortion Modeling Qualitative understanding of these distortion mechanisms is conveniently obtained with the aid of the quasi-stationary approximation (see Chapter 2). As an illustration, figure 4.5 depicts the response of a single-pole FM-AM converter to the FM message modulation. The spectral frequency is replaced by the instantaneous frequency of the FM wave, while the converter transfer is inverted for negative frequencies, where the time-domain converter output becomes negative. For small values of the FM modulation, the transfer is essentially linear, and approaches the ideal differentiator. For large values of the modulation, however, considerable distortion is introduced, due to the curvature, and eventually the 'saturation' of the transfer beyond the frequency of the pole.



sgn (ω)· |$H_{FM\text{-}AM}(j\omega)$|^2
distorted demodulator output
0
ω
pole frequency
t
FM-message
t

Figure 4.5: Distortion of the FM message information due to curvature and a finite bandwidth of the FM-AM converter transfer.

Positions of the Poles and Zeros for Minimal Distortion Minimum distortion is introduced into the FM demodulator output signal when the converter transfer $H_{\text{FM-AM}}(s)$ approaches the ideal converter transfer $H_{\text{FM-AM,id}}(s)$ as close as possible over the entire bandwidth of the input FM wave. The FM-AM converter transfer can generally be written as the product of the ideal transfer $H_{\text{FM-AM,id}}(s)$ from (4.7), and a low-pass filter (LPF) $H_{\text{LPF}}(s)$, that represents the non-idealities,

$$H_{\text{FM-AM}}(s) = K_{\text{FM-AM}} s H_{\text{LPF}}(s). \tag{4.10}$$

Distortion minimization corresponds to proper design of the LPF transfer function $H_{\text{LPF}}(s)$, such that:

- the bandwidth is sufficiently large;
- the transfer is as flat as possible in the pass band.

Generally, low-distortion FM-AM conversion requires a bandwidth that is somewhat larger than the Carson bandwidth (2.8), which is a slightly too optimistic estimation of the FM transmission bandwidth. The required bandwidth depends on the allowed level of distortion, and should be obtained e.g. by means of simulations. A much larger bandwidth is generally unfavorable with respect to noise, unwanted harmonics of multipliers, etc..

To fulfill the second requirement, $H_{\text{LPF}}(s)$ should belong to the class of Maximum Flat Magnitude (MFM) filters [3]. The characteristic property of these filters is that their derivative, a measure for the "flatness" of the transfer, at the center frequency $\omega = \omega_c$ of the filter vanishes, i.e.

$$\left.\frac{\mathrm{d}\,|H_{\text{MFM}}(\mathrm{j}\omega)|}{\mathrm{d}\omega}\right|_{\omega=\omega_c} = 0. \tag{4.11}$$

Hence, $H_{\text{LPF}}(s)$ is an all-pole low-pass filter, with $\omega_c = 0$, a Butterworth transfer, which is the all-pole MFM transfer, with a sufficiently large bandwidth is the best ('optimal') choice. In any case, minimum distortion is achieved when the filter center frequency ω_c coincides with the position of the zero in the converter transfer.

Distortion Simulations on a Second-Order FM-AM Converter The latter conclusion is confirmed by a simulation results on a second-order FM-AM converter, with a transfer-function given by (4.5). The numerator of this transfer corresponds to the ideal converter transfer, while the denominator represents the non-ideal LPF.

A Butterworth transfer is obtained when the quality factor equals $Q = 1/\sqrt{2} \approx 0.7$. Figure 4.6 depicts the time-domain transfer of this converter, from instantaneous frequency of the input FM wave to the time-domain output amplitude, for various values of Q. This figure clearly shows that the best linearity is indeed obtained for $Q = 0.7$, i.e. a Butterworth transfer.

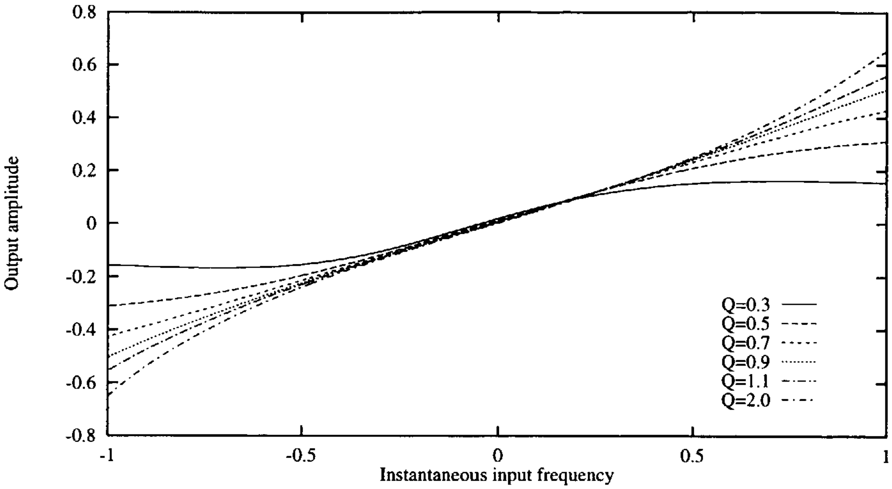


Figure 4.6: Time-domain FM-AM converter transfer for various values of Q.

4.2.3 Noise Minimization

This section considers the minimization of the FM demodulator output noise power by proper design of the FM-AM converter transfer, and realization of suitable trade-offs between sources of noise inside the converter.

As discussed in Section 4.1, the noise performance of FM-AM converters is determined by two characteristics:

- the shaping applied to the frequency noise;
- internal noise generation.

Minimization of the output noise due to these characteristics is discussed below.

Noise Shaping

As discussed previously, the SNR improvement of FM transmission in comparison to AM transmission relies on quadratic shaping of the frequency noise spectrum. In FM demodulators based on FM-AM conversion, this shaping is performed by the FM-AM converter.

The shaping operation is illustrated by figure 4.7, that depicts the low-pass equivalent model of the FM-AM converter, i.e. a 'low-pass' differentiator, and the AM demodulator, represented by its conversion gain K_{AM}. The input of the model consists of the message phase $\varphi(t)$ and the phase noise $\theta(t) \approx n_{s,q}(t)/A$, as explained in Section 2.3.2. The power-density spectrum of the frequency

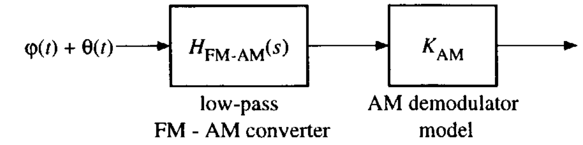


Figure 4.7: Low-pass equivalent demodulator model.

noise observed at the demodulator output, denoted by $S_{\dot{\theta}}(\omega)$, may therefore be expressed as

$$S_{\dot{\theta}}(\omega) = \left(\frac{K_{\text{AM}}}{A}\right)^2 |H_{\text{FM-AM}}(\mathrm{j}\omega)|^2 \, S_n(\omega), \tag{4.12}$$

where $S_n(\omega)$ denotes the spectrum of $n_{s,q}(t)$. This expression shows that only a true differentiator, with a transfer function $H_{\text{FM-AM}}(\mathrm{j}\omega) = \mathrm{j}\omega$, yields the noise shaping that is required to attain the FM transmission improvement, as discussed in Section 2.3.2.

From (4.12) is also observed, that an integrator, as used in combination with a differentiator in [5], is a profoundly bad implementation of the FM-AM converter, as far as the noise, and distortion, performance is concerned. Although essentially correct FM demodulation is possible with such a converter, the noise is shaped in an incorrect way. This conclusion is illustrated by figure 4.8. The

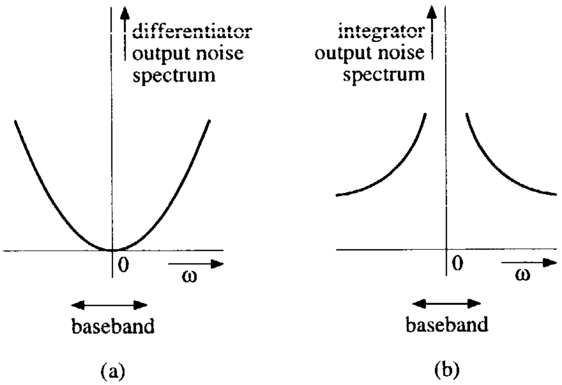


Figure 4.8: Shaping of the demodulator output noise spectrum. a) differentiator as FM-AM converter, b) integrator as FM-AM converter.

differentiator in figure 4.8a suppresses the largest part of the input noise at low frequencies, i.e. inside the baseband, at the expense of a noise enhancement at high frequencies, above the highest message frequency, and therefore maximizes the output SNR. The integrator however enhances the part of the noise located at low frequencies, i.e. inside the baseband, while it suppresses noise at high frequencies. Therefore, it deteriorates the SNR.

Internal Noise Generation

The contribution of the internal noise, generated by the electronic components inside the FM-AM converter, to the demodulator output noise spectrum can be estimated with the aid of the demodulator model depicted in figure 4.9. The noise process $n_1(t)$ represents the equivalent input noise, due to all internal noise contributions that are shaped by the FM-AM converter transfer, and are *uncorrelated* with any noise contributions that are not shaped. The noise process $n_2(t)$ represents equivalent input noise, due to shaped noise contributions that are *correlated* to contributions that are not shaped, while $n_3(t)$ represents all uncorrelated contributions that are not shaped by the transfer.

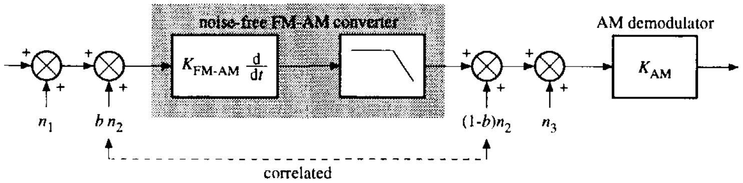

noise-free FM-AM converter
AM demodulator
$K_{\text{FM-AM}} \frac{\mathrm{d}}{\mathrm{d}t}$
K_{AM}
+
+
+
+
+
+
+
+
n_1
$b\,n_2$
$(1-b)n_2$
n_3
correlated

Figure 4.9: Contribution of the internal noise to the FM-demodulator output noise.

The contribution of these sources, which are generally Gaussian, to the noise power density spectrum at the demodulator output can be expressed as

$$S_{n,\text{out}}(\omega) = K^2_{\text{FM-AM}}\omega^2\frac{N_1}{2\pi} + K^2_{\text{FM-AM}}b^2\omega^2\frac{N_2}{2\pi} + (1-b)^2\frac{N_2}{2\pi} + \frac{N_3}{2\pi}, \tag{4.13}$$

where N_1 through N_3 represent the spectral densities of n_1, n_2 and n_3, while the conversion-gain K_{AM} equals unity. This expression contains some interesting conclusions on FM-AM converter design.

In the first place, (4.13) shows that the conversion-gain $K_{\text{FM-AM}}$ should be realized at the converter input, and made as large as possible, in order to nullify the relative noise contribution of the sources $(1-b)n_2$ and n_3. This follows directly from Friis' formula [6, 7]. An upper bound on $K_{\text{FM-AM}}$ is set forward by the maximum allowed power-level at the converter output (see Section 4.1).

Secondly, it should be noted that n_1 and bn_2 are transferred to the output according to the FM transmission scheme, while $(1-b)n_2$ and n_3 are transferred according to the AM transmission scheme. The latter two sources therefore result in a white noise floor, that deteriorates the output SNR. Therefore, whenever $K_{\text{FM-AM}}$ cannot be made sufficiently large to nullify the relative contribution of $(1-b)n_2$ and n_3, a trade-off should be established that increases

n_1 and bn_2 in favor of a decrease of $(1-b)n_2$ and n_3. In this respect, the optimal value of the parameter b, that minimizes the contribution of n_2 inside the baseband ($\omega \in [-W, W]$), equals

$$b_{\text{opt}} = \frac{3}{K^2_{\text{FM-AM}} W^2 + 3}. \tag{4.14}$$

Finally, since n_1 and bn_2 contribute to the frequency noise as well as to the amplitude noise, it is extremely important that the carrier-induced offset component ω_{offs} is minimized, in order to avoid a huge noise floor at the output (see Section 5.5). This is due to the fact that the contribution to the amplitude noise cannot be eliminated by means of amplitude compression; the compressor is placed in front of the FM-AM converter (see Chapter 5).

4.3 AM Demodulator Design

The AM demodulator succeeds the FM-AM converter in FM demodulators based on FM-AM conversion. According to Section 4.1, the two main characteristics of AM demodulators, that are decisive for the FM demodulator DR are phase selectivity, and the noise behavior. Therefore, these characteristics are analyzed in further detail in this section.

Section 4.3.1 considers phase selectivity as means to suppress unwanted components in the FM-AM converter output signal. Section 4.3.2 considers the noise performance of AM demodulators. Both sections compare the performance of the two AM demodulation algorithms discussed in Chapter 3: AM modulus detection and AM projection detection.

4.3.1 Suppression of Non-idealities by Phase Selectivity

Phase selectivity is the ability to distinguish carrier waves on the basis of their instantaneous phase. In a phasor-plane representation, phase selectivity represents the ability to discriminate between phasors that point in different directions.

In AM demodulators, phase selectivity is an important characteristic, due to the observation in Section 3.4 and Section 4.2, that the FM-AM converter output signal generally consists of two orthogonal components: a radial component and a tangential component. Only the tangential component contains the FM message information. In order to avoid interference/distortion in the FM demodulator output signal, the AM demodulator should distinguish the tangential component from the radial component.

As discussed in Section 3.4.2, suppression of fluctuations in the FM-AM converter input carrier amplitude eliminates the radial component from the converter output signal. However, Section 4.2.1 showed, that when the zero in its transfer is not exactly matched with the carrier frequency of the converter input wave, a residual radial component remains, which can only be suppressed by the AM demodulator.

The ability of both types of AM demodulators to suppress the radial component of the FM-AM converter output wave is considered below.

AM Modulus Detection

The class of AM demodulators based on AM modulus detection, discussed in Section 3.4.4, does not possess phase selectivity, and is therefore unable to distinguish between the radial and tangential component of the FM-AM converter output wave. Consequently realization of a large FM demodulator dynamic range with the aid of this type of AM demodulator puts severe requirements on the FM-AM converter transfer.

The reason for the absence of phase selectivity in AM modulus detectors is the absence of a reference wave, i.e. a phase reference, in the AM modulus demodulation algorithm; the demodulator simply determines the length of the phasor $\vec{s}_o$. In the presence of a nonzero radial component in the converter output wave $s_o(t)$, the FM demodulator output signal obtained with this type of AM demodulator becomes

$$\begin{aligned} y_{\mathrm{AM,modulus}}(t) &= |v(t)| = \sqrt{v_{\mathrm{tan}}^2(t) + v_{\mathrm{rad}}^2(t)} \\ &= \sqrt{A^2(t)\left[\omega_o + \dot{\varphi}(t)\right]^2 + \left[\dot{A}(t) + zA(t)\right]^2}, \end{aligned} \tag{4.15}$$

where $v_{\mathrm{tan}}(t)$ is given by (3.6) and $v_{\mathrm{rad}}(t)$, that includes the effect of a non-ideal zero in the transfer, is adopted from (4.9).

This expression demonstrates that the radial component $v_{\mathrm{rad}}(t)$ causes distortion in the FM demodulator output signal, and therewith reduces the FM demodulator DR. Application of this type of AM demodulator therefore requires suppression of fluctuations in the amplitude $A(t)$, and an ideal zero in the FM-AM converter transfer.

AM Projection Detection

The class of AM demodulators based on AM projection detection does possess phase selectivity, and is therefore able to eliminate the radial component from the FM-AM converter output wave. Consequently, application of this type of AM demodulator alleviates the requirements on the FM-AM converter transfer.

The origin of the phase selective behavior of these demodulators is the presence of a (phase) reference wave in their demodulation algorithm; the output signal equals the projection of $\vec{s}_o$ on the phasor of the reference wave $s_r(t)$. Thus, when the reference wave is synchronized to the tangential component of $\vec{s}_o$, the orthogonal, radial component is suppressed. This is also observed from the following expression for the output signal in the presence of a radial component:

$$\begin{aligned}\vec{s}_o \cdot \vec{s}_r =& s_{o,i}(t)s_{r,i}(t) + s_{o,q}(t)s_{r,q}(t)\\ =& v_i(t)s_{r,i}(t) + v_q(t)s_{r,q}(t)\\ =& \{v_{\text{rad}}(t)\cos\Phi(t) - v_{\text{tan}}(t)\sin\Phi(t)\}\, A\cos\Phi(t) +\\ & \{v_{\text{rad}}(t)\sin\Phi(t) + v_{\text{tan}}(t)\cos\Phi(t)\}\, A\sin\Phi(t)\\ =& Av_{\text{tan}}(t).\end{aligned} \tag{4.16}$$

As expected, the radial component $v_{\text{rad}}(t)$ is completely suppressed, as long as perfect synchronization between the reference wave and the tangential component of $s_o(t)$ exists. When the synchronization is not ideal, a small radial component remains.

The filtering applied by the FM-AM converter generally distorts the instantaneous phase of $s_o(t)$, and introduces a finite delay. In order to attain proper synchronization, the reference wave used by the AM projection demodulator should generally be low-pass filtered, by $H_{\text{LPF}}(s)$ from (4.10), i.e. the FM-AM converter transfer without the differentiator zero.

4.3.2 Noise Minimization

The AM demodulator has a decisive influence on the noise performance of the FM demodulator.

In the first place, the type of demodulator determines the minimum magnitude of the carrier-induced offset in the FM-AM converter output carrier amplitude, that still allows correct AM demodulation. As discussed in Section 4.1, this offset severely limits the FM demodulator DR; it reduces the maximum allowed output signal power, and, when noise in the input carrier amplitude is only partially suppressed, also increases the noise floor.

Section 3.4.4 noticed already, that AM modulus detection requires an AM modulation index smaller than unity, which means that the offset should exceed the FM message signal. In some types of AM modulus detectors, e.g. the peak detector, the requirements are even worse; these detectors are able to handle AM modulation that are much smaller than unity only. Moreover, due to their nonlinear transfer, i.e. the modulus operation, these demodulators generally show a threshold in their output SNR characteristic [7, 8], which deteriorates the demodulator DR even further. Consequently, AM modulus demodulators are not suited for realization of FM demodulators with a large DR.

The performance achieved with of AM projection demodulators is considerably better than the maximum possible performance attained with AM modulus demodulators. These demodulators allow a zero-valued carrier-induced offset, i.e. an (approximately) infinite AM modulation index. Therefore, no reduction of the maximum signal power, nor an increase of the noise level due to the offset occurs in these demodulators. Consequently, these demodulators are able to attain the maximum possible demodulator SNR and DR. If the most general AM projection detection algorithm is used, a zero-valued input carrier frequency is allowed, and zero-valued carrier offset can be achieved with a low-pass FM-AM converter. If a simplified algorithm is used, e.g. synchronous detection, a nonzero carrier frequency is required, and a zero-valued carrier offset should be achieved with the aid of a band-pass FM-AM converter.

4.4 FM-PM Converter Design

FM-PM converters constitute the operation of the class of FM demodulators discussed in Section 3.5. Four different FM-PM conversion algorithms were obtained.

This section considers the design of FM-PM converters, resulting in design rules for maximization of the converter DR, and FM demodulator DR. Similar to Section 4.2 on FM-AM converter design, the discussion on FM-PM converter design in this section concentrates on maximization of the FM demodulator SNDR through minimization of the carrier-induced phase offset, minimization of the distortion, and minimization of the output noise level.

Much of the material discussed in Section 4.2 in conjunction with FM-AM converters is also applicable to FM-PM converter design. Therefore, this section concentrates on the differences between FM-AM and FM-PM converter design, and the differences among the various FM-PM conversion algorithms.

An outline is as follows. Section 4.4.1 considers the minimization of the carrier-induced phase offset. Design rules for minimization of the distortion are considered in Section 4.4.2. Finally, minimization of the demodulator output noise is discussed in Section 4.4.3.

4.4.1 Offset Minimization

In Section 3.5 was shown, that similar to FM-AM conversion, FM-PM conversion algorithms generally result in a carrier-induced phase offset in the demodulator output signal, that deteriorates the DR. Therefore, also in FM-PM converter design, minimization of the offset is required.

This section discusses the possibilities to minimize the carrier-induced phase offset in each of the four different FM demodulator architectures based on FM-PM conversion, that were derived in Section 3.5.

Conversion based on a Fixed Time-Delay

In FM demodulators based on FM-PM conversion by means of a fixed time-delay, discussed in Section 3.5.2, both offset reduction techniques of Section 4.2.1, i.e. a zero-IF architecture and bandpass FM-PM conversion, can be applied in order to suppress the carrier-induced phase offset, introduced by the converter.

However, instead of the offset introduced by the FM-PM converter (time-delay) alone, the total offset of the converter and the built-in phase shift of the succeeding PM demodulator has to be nullified. When a nonzero built-in phase shift is present in the PM demodulator, such as the 90^o phase shift in a multiplier phase detector with sinusoidal inputs (see Chapter 7), the techniques of Section 4.2.1 *do not* suppress the phase offset in the demodulator output signal completely: the built-in phase offset of the PM demodulator remains.

The built-in phase offset of the PM demodulator can be eliminated in two different ways:

- application of an extra phase shifter that cancels the offset;
- cancellation of the PM demodulator offset by the converter offset $\omega_o\tau_d$.

The disadvantage of the first approach is, that it is difficult to realize a phase shift *without* introduction of a delay. This can generally be realized only for particular phase shifts, such as 90^o, e.g. by means of a zero-IF architecture. Application of this approach to the balanced quadrature demodulator of figure 3.31b, preceded by a zero-IF architecture, can be used to eliminate the built-in offset of the PM demodulator.

The disadvantage of the second approach is, that it generally yields an FM-PM converter with a distortion that is larger than the minimum possible distortion, as discussed in Section 4.4.2.

Conversion based on a Fixed Phase Difference

In FM demodulators based on FM-PM conversion by means of a fixed phase difference, discussed in Section 3.5.3, the carrier-induced phase offset cannot be eliminated. This is due to the sampling of the carrier phase performed by these demodulators. The sampling rate should at least equal twice the bandwidth of the FM message signal. Therefore, when one sample per carrier cycle is obtained, the carrier frequency should at least equal twice the message bandwidth W.

Conversion based on Phase Feedback

In FM demodulators based on phase feedback, the carrier-induced phase offset, usually called the *steady-state phase error* in these systems, should be eliminated by proper design of the phase detector and the loop filter. This subject is considered in Chapter 7.

Post-Detection Conversion

In FM demodulators based on post-detection phase-frequency conversion, no other means than application of an extra phase shifter in front of the PM demodulator is available to suppress the phase offset in the PM demodulator output signal. Furthermore, in order to avoid excessive cycle-slipping, the extra phase shift should be dimensioned such, that on average, the instantaneous phase of the input FM wave is positioned at the center of the PM demodulator characteristic, i.e. removed as far as possible from the bounds, that are represented in figure 3.27 by the dashed lines at $\pm\pi$.

4.4.2 Distortion Minimization

This section considers minimization of the distortion in the FM demodulator output signal by proper design of the FM-PM converter.

In FM demodulators based on FM-PM conversion by means of a fixed time-delay, this corresponds to proper design of the converter frequency transfer, as will be discussed below. In FM demodulators based on FM-PM conversion by means of a fixed phase difference, i.e. zero-crossing detection, the distortion can be reduced, if necessary, by higher-order interpolation between the samples, as mentioned already in Section 3.5.3. In phase feedback demodulators, the distortion is generally minimized when the tuning range of the oscillator in the feedback path is as linear as possible over the range of interest, and the the loopgain is sufficiently large. Finally, the distortion in post-detection conversion demodulators is essentially minimized by a maximally linear PM demodulator characteristic.

The sequel of this section is confined to distortion minimization in FM demodulators based on FM-PM conversion by means of a fixed time delay. Distortion minimization in other types of demodulators is relatively straight forward.

Section 3.5.2 showed already that FM-PM conversion on the basis of a fixed time-delay is associated to the spectral phase characteristic of the FM-PM converter transfer. Therefore, minimization of the distortion requires proper design of this phase characteristic. However, in addition, proper design of the amplitude characteristic of the FM-PM converter is generally required in order to minimize undesired, i.e. "parasitic", FM-AM conversion. This is especially important in FM receivers that do not apply amplitude compression, i.e. limiting, to the input FM wave. The design of both characteristics is discussed below.

Design of the Phase Characteristic

It follows from the the quasi-stationary approximation and the discussion in Section 3.5.2, that the distortion in the FM demodulator output signal is minimized when the spectral phase characteristic of the FM-PM converter is maximally linear. A maximally linear phase characteristic corresponds to a maximally flat group delay (MFD) [3].

Implementation of a maximum flat group delay in electronic systems is a difficult matter, especially when an IC realization is required. Generally, the best linearity is obtained with distributed element filters such as transmission lines. Integratable delay lines with a linear phase characteristic can be constructed with the aid of Surface Acoustic Wave (SAW) technology, as used in [9]. However, SAW filters require special IC processing steps that are often not acceptable.

When transmission lines or SAW filters cannot be used, one has to resort to lumped element filters for realization of the time delay. A maximally flat group delay is established with such filters, when the (all-pole) filter transfer is of the Bessel/Thomson type [3]. A characteristic property of these filters is, that the second-order derivative of their spectral phase characteristic vanishes at the center-frequency $\omega = \omega_c$,

$$\left.\frac{\partial^2 \Phi_{\text{Bessel}}(\omega)}{\partial \omega^2}\right|_{\omega=\omega_c} = 0. \tag{4.17}$$

Obviously, for low-pass filters $\omega_c = 0$. Of course, the linearity improves with increasing order of the filter. Further, if a band-pass converter is required, the frequency $\omega = 0$ should be transformed to the FM carrier frequency ω_o (see Section 4.2.1).

The of a Bessel filter as the optimal FM-PM converter transfer for a given converter bandwidth is confirmed by simulations on the second-order low-pass filter given by

$$H_{\text{FM-PM}}(s) = \frac{\omega_r^2}{s^2 + \frac{\omega_r}{Q}s + \omega_r^2}. \tag{4.18}$$

This filter possesses a Bessel characteristic when the quality factor equals

$$Q_{\text{bessel}} = \frac{1}{\sqrt{3}} \approx 0.6. \tag{4.19}$$

The group delay at $\omega = 0$, obtained by differentiation of the phase characteristic of (4.18), equals

$$\tau_d = \frac{1}{Q\omega_r}. \tag{4.20}$$

Further, similar to FM-AM converters, the bandwidth of the FM-PM converter should at least comply with Carson's bandwidth formula.

Figure 4.10 shows the simulated transfer from instantaneous input frequency of the input FM wave to the instantaneous phase difference between the FM-PM converter input and output wave. Maximum linearity is indeed attained when $Q = 0.6$, instead of $Q = 0.7$ found for FM-AM converters in Section 4.2.2.

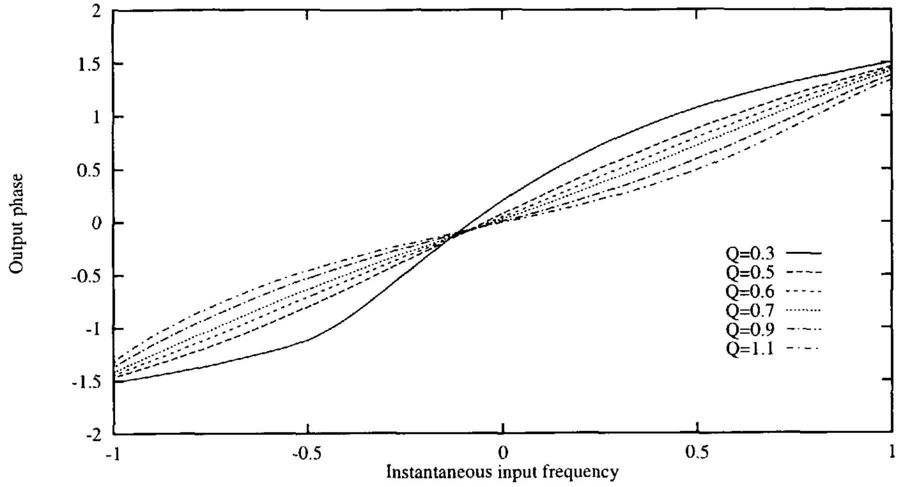


Figure 4.10: Time-domain FM to PM converter transfer for various values of Q.

Design of the Amplitude Characteristic

An ideal delay element is not subject to unwanted FM-AM conversion, since its spectral amplitude characteristic is constant; the conversion gain $K_{\text{FM-AM}} \equiv 0$. In practice however, delay lines, and, in particular those implemented by a Bessel filter, are subject to FM-AM conversion, since their amplitude characteristic shows some curvature.

A maximum flat magnitude (MFM) characteristic would minimize the FM-AM conversion. However, in low-pass and bandpass filters, MFM and MFD characteristics are conflicting requirements [3]. This is illustrated already by the examples given for a second-order FM-AM and FM-PM converter; the optimum value obtained for the quality factor Q was different.

Besides application of limiting/compression of the FM-PM converter output carrier amplitude, undesired FM-AM conversion can be avoided by application of a Bessel all-pass filter, that, by convention, possesses a constant magnitude characteristic. The spectral phase characteristic equals twice the phase characteristic of the corresponding low pass/bandpass Bessel filter. When $D(s)$

denotes the denominator polynomial of a Bessel low-pass filter, the transfer function of the corresponding all-pass filter equals [10]

$$H_{\text{all-pass}}(s) \overset{\text{def}}{=} K\frac{D(-s)}{D(s)}. \tag{4.21}$$

Thus, the right-half plane zeros of this filter transfer are located at the same frequency as the left-half plane poles.

4.4.3 Noise Minimization

This section considers the minimization of the FM demodulator output noise level by proper design of the FM-PM converter transfer. The discussion is confined to the noise behavior of FM demodulators based on FM-PM conversion by means of a fixed time-delay, which are potentially able to establish high performance demodulation. The noise performance of FM demodulators based on FM-PM conversion by means of a fixed phase difference, and FM demodulators based on post-detection conversion is not considered, since it follows in a straight-forward fashion from the discussion in this section, Section 4.2.3, and Section 5.5. A discussion of the noise performance of phase feedback demodulators is postponed to Chapter 7, since it differs significantly from the noise behavior of all other demodulator types.

First, the noise shaping by the FM-PM converter, implemented by a fixed time-delay, is considered. Subsequently, the increase of the output noise level due to internally generated noise is discussed.

Noise Shaping

Although established in a slightly different way, the FM-PM converter applies the same type of quadratic noise shaping to the phase/frequency noise of the input wave as the FM-AM converter. This is seen as follows.

If $\Delta\Phi_n(t)$ denotes the phase/frequency noise observed at the output of the FM demodulator of figure 3.20, and $\theta(t) \approx n_{s,q}(t)/A$ represents the phase noise in the input FM wave, then

$$\Delta\Phi_n(t) = \theta(t) - \theta(t - \tau_d), \tag{4.22}$$

where τ_d denotes the time-delay realized by the FM-PM converter. The power spectral density of $\Delta\Phi_n(t)$ may be expressed in terms of $S_\theta(\omega)$, the spectral density of $\theta(t)$, as

$$\begin{aligned}
S_{\Delta\Phi_n}(\omega) &= \left|1 - \exp(\mathrm{j}\omega\tau_d)\right|^2 S_\theta(\omega) \\
&= 4\sin^2\left(\frac{\omega\tau_d}{2}\right) S_\theta(\omega) \\
&= 2\left[1 - \cos(\omega\tau_d)\right] S_\theta(\omega) \\
&\approx (\omega\tau_d)^2 S_\theta(\omega) \\
&\approx \frac{\tau_d^2}{A^2}\omega^2 S_n(\omega),
\end{aligned} \tag{4.23}$$

where the first approximation holds as long as $\omega\tau_d \ll 1$. Since τ_d is related to the carrier frequency, as discussed in Section 3.5.2, this is usually the case inside the baseband.

Expression (4.23) demonstrates that the required quadratic noise shaping is established as a result of the correlation between both FM waves (the wave subjected to demodulation, and the reference wave) at the PM demodulator input. At low frequencies, the correlation is almost unity, resulting in a very small phase difference. At high frequencies, the correlation gradually decreases, resulting in a larger phase difference.

Internal Noise Generation

The discussion on internal noise of FM-AM converters showed that part of this noise is shaped by the converter transfer, i.e. transferred to the output according to the FM transmission scheme, while the other part is not shaped, and transferred according to the AM transmission scheme.

In FM-PM converters (a delay-line), the internally generated noise is not shaped by the converter transfer, but transferred to the FM demodulator output according to the PM transmission scheme. Therefore this noise inevitably results in a deteriorative white noise floor at the demodulator output. This behavior can be explained by the observation that the shaping is established on the basis of the correlation between the input signal of the FM-PM converter and its output signal, as discussed previously. The noise produced inside the converter is uncorrelated with the input signal of the delay-line, and therefore cannot be shaped with the aid of the correlation between the converter input and output signal.

The only means to minimize the contribution of the internal noise to the demodulator output noise, besides minimization of the intensity of the noise sources, is to maximize the amplification applied to the FM wave prior to FM-PM conversion.

4.5 PM Demodulator Design

PM demodulators perform the actual demodulation of the FM wave in FM demodulators based on FM-PM conversion. As discussed in Section 3.5, the position and type of the PM demodulator included in the FM demodulator architecture differs considerably among the four FM demodulation algorithms based on FM-PM conversion.

This section briefly outlines the main characteristics of PM demodulators, that affect the FM demodulator SNDR. A slightly more detailed discussion of phase detector design is contained in Chapter 7, in conjunction with phase feedback demodulators.

Section 4.5.1 considers minimization of the distortion in PM demodulators, while Section 4.5.2 discusses the noise performance.

4.5.1 Distortion Minimization

At an architectural level of consideration, distortion in PM demodulators is due to simplification of the rather complex general PM demodulator structure, depicted in figure 3.30a. For example, distortion is introduced into the transfer of the multiplier PM demodulator of figure 3.30c, due to its sinusoidal transfer.

A frequently encountered technique to linearize the transfer of the PM demodulator depicted in figure 3.30c, is to change the shape of the waves that enter the multiplier in this architecture from sinusoids to square waves, by insertion of hard-limiters at both demodulator inputs (the signal input and the reference input). In that case, the demodulator nonlinearity changes from a sinusoid to a triangular wave, which is exactly linear in bounded intervals.

Another technique that linearizes the transfer is replacement of the multiplier by a sampler, resulting in a sawtooth demodulator characteristic.

4.5.2 Noise Minimization

The influence of PM demodulators on the noise performance of FM demodulators based on FM-PM conversion is similar to the influence of AM demodulators in FM demodulators based on FM-AM conversion.

In the first place, in order to prevent a significant increase of the FM demodulator output noise floor (and a decrease of the SNDR), the built-in phase offset present in many types of PM demodulators has to be canceled by an opposite phase shift realized by the FM-PM converter, or by an additional phase shifter. As opposed to some types of AM demodulators, PM demodulators do generally not require a non-zero offset component in order to operate properly.

Secondly, similar to AM demodulators, noise generated internally in PM demodulators results in a white noise floor at the PM demodulator output. Therefore, especially in FM demodulators based on FM-PM conversion with the aid of a fixed time-delay, where the PM demodulator output equals the FM demodulator output, it is essential to minimize the internally generated noise, in order to prevent deterioration of the output SNDR. In such demodulators, no shaping is applied to this noise. Oppositely, in FM demodulators based on post-detection conversion, the internal noise of the PM demodulator is shaped by the phase-frequency converter, and in general does not significantly deteriorate the SNDR.

4.6 Conclusions

This chapter considered the design of the four types of subsystems encountered in FM demodulators.

It was shown that a carrier-induced offset in the instantaneous frequency of the FM demodulator output, FM-AM converter output carrier amplitude, and FM-PM converter output carrier phase may considerably reduce the demodulator Dynamic Range (DR). It reduces the maximum allowed signal power, and, when amplitude noise is not or only partly eliminated, also increases the noise floor.

Minimization/ elimination of the carrier-induced offset is possible in two fundamentally different ways; by application of a zero-IF architecture and a low-pass FM-AM or FM-PM converter, or by application of a band-pass FM-AM or FM-PM converter. When a nonzero built-in phase offset exists in the PM demodulator, contained in an FM-PM conversion FM demodulator, an extra phase shifter is generally required, or the periodicity of the PM demodulator transfer has to be exploited in order to suppress the offset.

In FM-AM conversion demodulators, the minimum allowed level of the offset is determined by the applied type of AM demodulator. AM modulus demodulators do not allow a zero-valued offset, as opposed to AM projection demodulators, and are therefore unsuited for realization of high-DR FM demodulators.

The distortion introduced by low-pass FM-AM converters is minimized when their frequency characteristic equals an ideal differentiator, cascaded by a low-pass filter of the Maximum Flat Magnitude (MFM) type. Further, the bandwidth of this filter should be sufficiently large to accommodate the FM wave. For band-pass FM-AM converters, related to low-pass converters by the "biquadratic transformation", the same conclusions hold.

The distortion introduced by FM-PM converters is minimized when their transfer is of the Maximum Flat Delay type (MFD). Further, their amplitude characteristic should be as flat as possible in order to avoid undesired FM-AM conversion. This can be achieved by application of an all-pass MFD filter.

Noise generated internally in the input circuitry of FM-AM converters is transferred to the demodulator output by means of the FM transmission scheme, and therefore results in a slightly increased quadratic output noise spectrum.

Noise generated in the output circuitry, however, is transferred by means of the AM transmission scheme, and results in a white noise floor that considerably deteriorates the output SNR. The influence of this noise should be minimized, possibly at the expense of larger noise generation at the input. The FM-AM conversion gain should therefore be realized at the input, and made as large as possible.

Noise generated internally in FM-PM converters is always transferred to the output by means of the PM transmission scheme, and therefore results in a deteriorative white noise floor. The influence of this noise can be minimized only by sufficient amplification of the input FM wave, and minimization of the noise generation itself.

Finally, it was shown that the non-informative radial component of the FM-AM converter output wave cannot be eliminated by limiting of the input FM wave when the zero in the converter transfer does not match the carrier frequency. In that case, the AM demodulator should suppress this component. It was shown that only AM projection demodulators possess the ability to suppress this component.

References

[1] Yannis P. Tsividis, "Externally linear, time-invariant systems and their application to companding signal processors", *IEEE Transactions on Circuits and Systems-II*, vol. 44, no. 2, pp. 65–85, Feb. 1997.

[2] Wouter A. Serdijn, Michiel H.L. Kouwenhoven, Jan Mulder, and Arthur H.M. van Roermund, "Design of high dynamic range fully integratable translinear filters", Accepted for publication in *Analog Integrated Circuits and Signal Processing.*

[3] Anatol I. Zverev, *Handbook of Filter Synthesis*, John Wiley and Sons, New York, 1967.

[4] Samuel Seeley, *Electron-Tube Circuits, second edition*, McGraw-Hill Book Company, New York, 1958.

[5] Jacob Klapper and Edward J.A. Kratt, III, "A new family of low-delay FM detectors", *IEEE Transactions on Communications*, vol. 27, no. 2, pp. 419–429, Feb. 1979.

[6] E.H. Nordholt, *Design of High-Performance Negative Feedback Amplifiers*, Elsevier, Amsterdam, 1983.

[7] A. Bruce Carlson, *Communication Systems, An introduction to Signals and noise in Electrical Communication*, McGraw-Hill International Editions, Singapore, 1986.

[8] Wilbur B. Davenport Jr. and William L. Root, *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill Book Company, New York, 1958.

[9] Paul T. M. van Zeijl, *Fundamental Aspects and Design of an FM Upconversion Receiver Front-End with On-Chip SAW Filters*, PhD thesis, Delft University of Technology, 1990.

[10] Herman J. Blinchikoff and Anatol I. Zverev, *Filtering in the Time and Frequency Domains*, John Wiley and Sons, New York, 1976.

Chapter 5

FM Receiver Design

The discussion on FM demodulation principles in Chapter 3, and their implementation in Chapter 4, implicitly assumed an FM demodulator input signal that consists of a single noise-free FM wave.

The FM *receiver* however, which 'embeds' the demodulator, is usually confronted with a crowded band of such FM carriers, with different intensities, separated in frequency by Frequency Division Multiplexing (FDM) and mutilated with noise and interference. In order to establish reliable reconstruction of the FM message signal, the receiver should therefore regenerate a *demodulator* input signal from the received band of FM carriers, that closely resembles a single-noise free FM wave. In fact, it should provide the embedding that maximizes the demodulator performance.

Besides pre-processing of the demodulator input signal, FM receivers generally also perform some post-processing to improve the quality of the demodulator output signal. The algorithms of these pre- and post-processing operations necessarily require information of the requested FM wave's characteristics, in order to distinct and extract it from noise, interference and other carrier waves. This information may be acquired in two different ways:

- by inclusion of a priori information in the receiver architecture;
- by extraction from the receiver input signal during reception.

Pre- and post-demodulation processing functions implemented according to the first approach are entirely based on the *expected* characteristics of the FM wave. The required processing is therefore established only when the *actual* characteristics of the FM wave, observed during reception, resemble the expected ones.

Processing functions implemented according to the second approach introduce (adaptive) control schemes into the FM receiver architecture, that adjust its behavior at the basis of the detected signal characteristics. However, since these control schemes themselves generally need to be supplied with some a priori information, the adjustments are mostly limited to fine-tuning of a coarse receiver behavior, that is determined with the aid of a priori information. For example, a phase locked loop contains a control scheme that adjusts the frequency of the internal reference, i.e. the controlled oscillator, to the carrier frequency of the received carrier wave. The free running frequency of the oscillator, a coarse a priori estimation of the carrier frequency, is included into the PLL architecture design, while the phase-lock mechanism performs the fine-tuning on the basis of the detected signal.

The intended improvement of the demodulator input and output signal is realized only as long as the a priori information, used by the various pre- and post-demodulation processing functions, is reliable. When e.g. noise and interference introduce discrepancies between the actual and the expected characteristics of the FM wave, the a priori information becomes invalid. In that case, pre- and post-processing is likely to introduce performance degradation, instead of improvement, since it is 'deceived' of the actual characteristics of the FM wave. Therefore, performance improvement, by inclusion of (additional) a priori information, is usually obtained at the expense of performance degradation, whenever this information becomes invalid.

A design strategy for FM demodulators and receivers should therefore be aware of the *trade-offs* introduced by the various pre- and post-processing functions.

This chapter investigates the possibilities for inclusion of pre- and post-processing into the FM receiver architecture, and *qualitatively* analyzes the performance improvement/degradation it effectuates. A *quantitative* elaboration of the most important types of processing is the subject of Chapter 6 to Chapter 8.

The generalized FM receiver architecture depicted in figure 5.1, visualizes the pre- and post-processing functions considered in this chapter. Besides reception and initial amplification, these functions may be arranged in six different classes, that are separately discussed throughout the chapter. Except for FM demodulation, the main receiver function, all these functions are optional.

Section 5.1 outlines the three types of preprocessing that may be used to extract the required FM wave. These types are separately considered in detail in Section 5.2 through Section 5.4. The most important FM demodulator characteristics that determine the output signal quality are summarized in Section 5.5. Improvement of this signal by means of post-demodulation processing is discussed in Section 5.6. Finally, improvement techniques based on adaptive feedback/feed forward control schemes, including frequency feedback, are discussed in Section 5.7. Section 5.8 presents the conclusions.

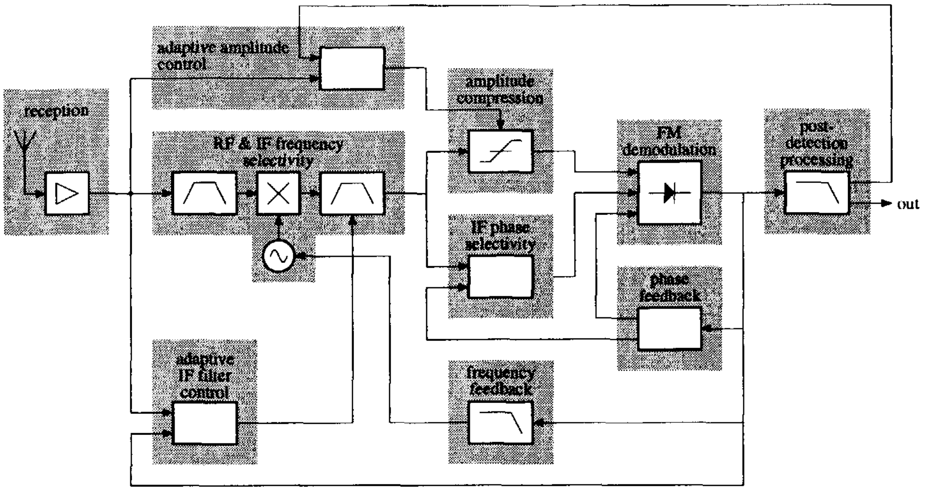

adaptive amplitude control
reception
RF & IF frequency selectivity
adaptive IF filter control
amplitude compression
IF phase selectivity
frequency feedback
FM demodulation
phase feedback
post-detection processing
out

Figure 5.1: General FM receiver architecture.

5.1 Pre-Detection Processing

The introduction of this chapter discussed already that besides demodulation, an FM receiver usually contains some pre- and post-processing in order to establish reliable reconstruction of the transmitted intelligence.

This section gives an overview of the three distinguishable types of pre-demodulation processing functions, and outlines their main characteristics. Each of these functions is separately analyzed in Section 5.2 through Section 5.4.

Section 5.1.1 relates the three types of pre-processing to each other, while Section 5.1.2 through 5.1.4 outline their main characteristics.

5.1.1 Extraction of the Requested FM Wave

Since frequency modulation belongs to the class of FDM transmission schemes, as discussed in Chapter 2, the signal detected by an FM receiver usually consists of a large number of carrier waves, (equally) spaced over the assigned frequency band. Familiar examples of such systems are found in radio broadcasting, satellite communications and (wireless) telephony.

Reconstruction of the requested intelligence requires extraction of the corresponding FM carrier from the receiver input signal, prior to demodulation. This extraction operation requires a priori knowledge of the FM wave's characteristics. Since such a wave is characterized by three parameters, frequency, phase and amplitude, such an extraction may be accomplished by a combination of the following three separations:

- separation in spectral frequency (frequency domain);
- separation in phase (time domain);
- separation in amplitude (amplitude domain).

The principles of these separations are separately considered below.

5.1.2 Separation in Frequency

Separation in the frequency domain allows discrimination between signals that occupy non-coincident frequency bands. It is usually implemented by means of linear filtering.

The required a priori information consists of the edge-frequencies, or, alternatively, the center-frequency and bandwidth of the frequency range occupied by the signal of interest. All signals located *outside* this range should be suppressed, while the one(s) *inside* should remain unaffected by this operation.

Although linear filtering implements the requested separation, it is simultaneously one of the causes of distortion in the reconstructed FM intelligence, observed at the demodulator output. Especially in case of narrow-band filtering, considerable distortion is observed. This *narrow-band filtering distortion* is due to the nonlinear nature of frequency modulation. A filter that performs *linear* operations on the FM carrier wave, simultaneously performs *nonlinear* operations on the intelligence contained in the instantaneous frequency, as a result of the FM scheme: *linear* distortion of an FM wave results in *nonlinear* distortion of the FM message signal contained in that wave. Estimation of this distortion and filter design for minimum distortion is considered in Section 5.2.

5.1.3 Separation in Phase

Separation of signals located in coincident frequency ranges, e.g. the requested FM wave mutilated by *co-channel interference*, cannot be established by means of frequency selectivity. Instead, separation in phase, established by phase selectivity, should be established.

Especially in crowded frequency bands, such as those assigned to mobile telephony and satellite communications, co-channel interference seriously deteriorates the demodulator output signal quality, by virtue of the well-known capture effect. Due to this effect, see e.g. [1], the instantaneous frequency of the composite wave, i.e. the addition of the requested FM wave and the co-channel interferer, essentially copies the intelligence contained in the instantaneous frequency of the strongest wave, while the intelligence of the other(s) is suppressed. This yields especially annoying results when both waves are of comparable strength. In that case, the demodulator output randomly switches between the requested intelligence and the intelligence contained in the interferer. The deterioration of the demodulator output signal due to this effect is, among others, analyzed in [2, 3].

Separation in phase allows discrimination between carrier waves that occupy the same frequency range, but possess different instantaneous phases. Applied in combination with an FM demodulator that intrinsically contains phase selectivity, such as phase feedback demodulators, this approach may be used to suppress co-channel interference. A receiver architecture that establishes suppression of co-channel interference is discussed in in Section 5.3.

5.1.4 Separation in Amplitude

Separation in amplitude, established by compression of the modulation contained in the carrier amplitude, allows discrimination between weak and strong signals with the same frequency and phase. In FM demodulators, amplitude compression, usually implemented by means of a (hard-) limiter, has found wide-spread application as means to suppress amplitude modulation introduced by (weak) additive noise.

At high input CNRs, this operation establishes a considerable improvement of the demodulator output SNR, by application of the a priori knowledge that only the instantaneous phase/frequency of the FM wave, which is orthogonal/independent of the carrier amplitude, contains the intelligence. Despite this fact, it is observed from Chapter 3 that the output signal of many types of FM demodulators depends on the demodulator input carrier amplitude as well. Noise and interference contained in this amplitude therefore penetrates the output signal, and reduces the output SNR. Eventually, this noise reduces the performance of the entire FM transmission system to a level comparable with, or even below that of AM.

At low CNRs however, amplitude compression deteriorates the demodulator output signal by means of so called "click noise", discussed in Section 5.4, as a result of the fact that the applied a priori knowledge is no longer valid. At these CNRs, the instantaneous frequency of the noisy FM wave is no longer a reliable representative of the transmitted intelligence. Further, it is no longer true that the carrier amplitude contains no information; it contains information about the reliability of the intelligence in the instantaneous frequency. Amplitude compression deprives the demodulator from the possibility to distinguish between reliable and unreliable FM waves, required to prevent severe deterioration of the output signal. Instead, it forces the demodulator to treat all FM waves in the same way.

A trade-off between continuous noise and click noise, in order to reduce the deterioration of the output signal, may be established by application of finite compression to the demodulator input wave. The details of this trade-off mechanism are considered in Section 5.4.

5.2 Pre-Detection Frequency Selectivity

Separation in the frequency domain by application of pre-detection, i.e. RF and IF, frequency selectivity is essential for reliable reconstruction of the transmitted intelligence in almost every FM receiver. Generally, there are three reasons for the inclusion of RF and IF frequency selectivity into the receiver architecture:

- selection of the requested FM wave from the received frequency band;
- suppression of out-of-band noise;
- prevention of aliasing in subsequent nonlinear operations, e.g. limiting.

In all of these functions, an as small as possible bandwidth, and an as steep as possible transfer function, of the filter(s) is required, in order to establish a maximum suppression of interference, noise and aliasing noise/distortion respectively. This maximum suppression results in an as large as possible demodulator input CNR, and therewith establishes an as low as possible receiver threshold.

However, since narrow-band filtering distorts the intelligence contained in FM waves, maximum suppression does generally not correspond to a maximum DR of the demodulator output signal. Instead, the filter bandwidth and transfer function should be a *trade-off* between a maximum *demodulator input CNR* and minimum *distortion* of the intelligence.

This section is concerned with the modeling of this distortion, and its implications on the design of RF and IF filters for minimum distortion of the FM intelligence.

Accurate modeling of distortion introduced by narrow-band filtering is extremely difficult, and has been the subject of study from the 1930's on. The results encountered in literature seem to be confined to very wideband FM waves, with large frequency deviation ratios, and narrow-band waves, with very small deviation ratios. Both types of waves require different modeling, and are therefore considered in separate sections. Section 5.2.1 considers the distortion in wideband waves, while Section 5.2.2 considers the distortion in narrow-band waves. Both sections focus on qualitative understanding of the distortion, and its implications on RF and IF filter design.

5.2.1 Distortion in Wideband FM Waves

This section investigates the distortion introduced in wideband FM wave by means of linear filtering. We first discuss the two distortion mechanisms and their cause. Subsequently, their implications on the design of low-distortion FM receivers are considered.

Distortion Modeling

Linear-filtering distortion in wideband FM waves is conveniently modeled by means of the quasi-stationary approximation. In such waves, the bandwidth of the intelligence is much smaller than the transmission bandwidth, which allows them to be modeled as sinusoid that slowly swings along the FM transmission bandwidth. The same approach was applied already in Chapter 4, to the design of low-distortion FM-AM and FM-PM converters.

The linear-filtering distortion observed in the demodulator output signal is the resultant of two different distortion mechanisms:

- distortion due to a nonlinear spectral phase characteristic;
- 'parasitic' FM-AM conversion;

The first type of distortion is due to the effect that in practical RF/IF filters, the various spectral frequencies in the FM wave are subjected to unequal time-delay's. This effect may be viewed as a kind of 'dispersion', similar to wavelength dispersion in e.g. optical fibers. In bandpass-filters (BPF), this type of distortion results in a kind of 'clipping' of the FM message wave, as schematically depicted in figure 5.2.

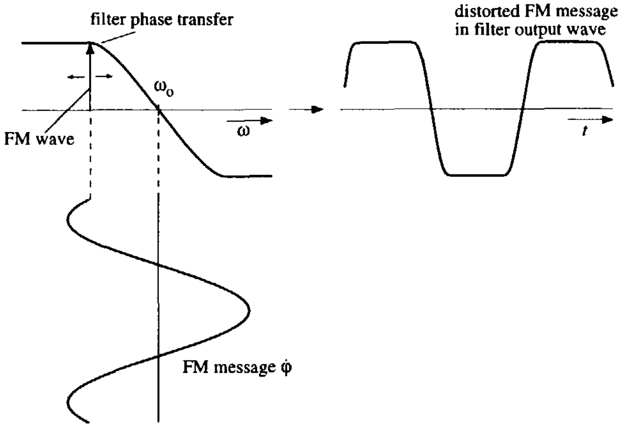


Figure 5.2: Phase distortion in band-pass filters.

The same type of distortion was considered already in Section 4.4, in conjunction with low-distortion FM-PM converter design. It was concluded there, that minimum distortion, corresponding to a maximally linear phase characteristic, is attained with so called Maximum Flat Delay (MFD) filters. An all-pole MFD transfer is (approximately) realized by Bessel-Thompson filters [4]. For such filters, the second derivative of the phase characteristic at the filter's center frequency equals zero.

The second distortion mechanism corrupts the intelligence when the demodulator output signal depends on the input carrier envelope. As observed from Chapter 3, in the absence of amplitude compression, this is the case for many FM demodulator types. The FM-AM conversion is due to a nonzero slope of the filter's spectral magnitude characteristic, as depicted in figure 5.3.

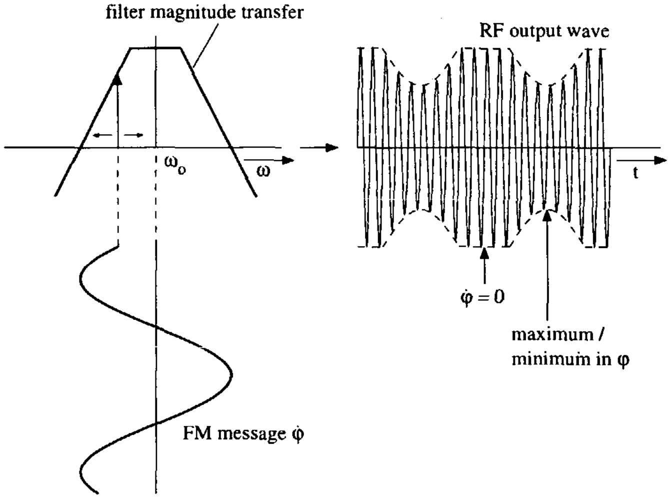


Figure 5.3: "Parasitic" FM-AM conversion in band-pass filters.

Since the output signal of most types of FM-demodulators depends on the FM receiver input carrier amplitude, when no or finite amplitude compression is applied to their input signal, the amplitude fluctuations caused by this FM-AM conversion generally penetrate the output signal.

Section 4.2 discussed this type of distortion already in conjunction with the design of low-distortion FM-AM converters. It was concluded there that 'parasitic' FM-AM conversion is minimized by Maximum Flat Magnitude (MFM) filters. An all-pole MFM transfer is realized by Butterworth filters.

Basically the same conclusions were obtained by Roder [5] in 1937, and confirmed by his measurements on a single-tuned LRC-circuit and two coupled RLC-circuits.

Low-Distortion Design

It is clear from the previous analysis, that simultaneous minimization of both types of distortion results in conflicting requirements on the filter transfer; phase distortion and amplitude distortion cannot be minimized simultaneously by the same band-pass filter transfer, since MFD and MFM band-pass transfers are different.

In fact, this means that insufficient degrees of freedom are left by the filter transfer for an optimal design. According to the orthogonalization principle (see Section 3.1.3), distortion minimization should therefore be achieved by inclusion of an additional function into the receiver architecture. One of both distortion types should be minimized by this function, while the other one is minimized by optimization of the filter transfer. This approach yields two possible solutions of the distortion minimization problem, as discussed below.

MFD Filter and Amplitude Compressor The first solution applies an MFD filter in order to minimize the phase-distortion, and an "amplitude compressor", such as a limiter or a fast AGC, in order to remove the amplitude modulation introduced by parasitic FM-AM conversion. High order MFD filters yield a very linear phase characteristic, resulting in low phase-distortion. However, compared to e.g. MFM filters, the noise bandwidth of MFD filters is relatively large, which is unfavorable for the demodulator input CNR, and the threshold level. Further, the magnitude characteristic of high order MFD's is considerably steeper than the slope of low order MFD's, resulting in considerable FM-AM conversion. At high CNRs, this AM modulation is eliminated by the amplitude compressor. At low CNRs however, this modulation may somewhat decrease the average compressor input carrier power level (see figure 5.3), which reduces the effective input CNR and thereby increases the threshold level.

Consequently, a *trade-off* between the demodulator DR at high CNRs (above threshold) and the threshold level exists.

MFM Filter and Phase Equalizer The second solution applies an MFM filter in order to minimize the FM-AM conversion, and a separate (all-pass) equalizing filter in order to cancel the phase-distortion. Again, high order filters result in negligible FM-AM conversion, but at the same time cause considerable phase distortion. An equalizing all-pass filter may somewhat reduce this distortion, but exact cancellation is generally impossible. Therefore, the resulting distortion in the demodulator output signal is generally larger than the distortion level attained with the first solution. At the same time however, since the noise bandwidth of MFM filters is smaller than the noise bandwidth of MFD filters of the same order and the same bandwidth, the demodulator input CNR is larger than in the first configuration, which means that the threshold level will be smaller. Again, this shows the existence of a trade-off between the the demodulator DR and its threshold level.

5.2.2 Distortion in Narrow-Band FM Waves

This section considers the modeling of linear-filtering distortion in narrow-band FM waves, and its implications on the design of low-distortion FM receivers. Modeling of the distortion in these waves is of interest, e.g. in the design of frequency feedback receivers, considered in Chapter 8.

Distortion Modeling

The distortion introduced by linear-filtering in narrow-band FM waves cannot be modeled properly with the aid of the quasi-stationary approach, since the transmission bandwidth and the bandwidth of the message signal are in the same order of magnitude. In order to attain insight into the distortion mechanism, one has to resort to approximate analytic models, that exploit the narrow-band property. Such models have been obtained only for two specialized cases: modulation that resembles Gaussian noise [6], and modulation that resembles a sinusoid, described e.g. in [7, 8]. Further, all these models consider only phase-distortion; AM modulation due to parasitic FM-AM conversion is assumed to be eliminated by a hard-limiter. The sequel of this section is confined to the model for Gaussian noise modulation, since that model is best suited to gain insight into the distortion mechanism. The model for sinusoidal modulation is merely a straight-forward elaboration of the well known Bessel-function expansion for the FM spectrum, as described e.g. in [1].

A useful approximate model for linear-filtering distortion in narrow-band waves with Gaussian modulation, that are passed through band-pass filter with a symmetrical characteristic is developed in [6]. The advantage of this model in comparison to other models, as discussed e.g. in [5], is that it separates the phase/frequency of the filtered output wave into components with a clear physical meaning. Measurements [9] indicate that the model is useful for small and moderate modulation indices, corresponding to frequency deviations of at most the same order of magnitude as the signal, but loses accuracy for large indices. We briefly outline the model and consider its implications on FM receiver design.

The model distinguishes three uncorrelated components in the power density spectrum of the FM demodulator response to the filtered FM wave:

- a signal component;
- a cross-term between the signal and the distortion;
- a component that consists of distortion only.

For weak distortion the first two terms dominate, while the third one is relatively small. Therefore, the latter component is neglected in the sequel.

Signal Component The signal component represents the demodulator response to the intelligence, in the absence of distortion. This component may be expressed as follows.

Let $S_{\dot{\varphi}}(\omega)$ denote the power density spectrum of the FM message, $S_{\dot{\Phi}}(\omega)$ the power density spectrum of the FM message in the filtered FM wave, and $H_{\mathrm{RF}}(\mathrm{j}\omega)$ the transfer of the bandpass filter with center frequency ω_o. The normalized low pass equivalent of this filter transfer is formally defined as

$$\Gamma_{\mathrm{IF}}(\mathrm{j}\omega) \overset{\mathrm{def}}{=} \frac{H_{\mathrm{RF}}(\mathrm{j}\omega + \mathrm{j}\omega_o)}{H_{\mathrm{RF}}(\mathrm{j}\omega_o)} u(\omega + \omega_o), \tag{5.1}$$

where $u(x)$ denotes the Heaviside step function ($u(x) = 1$ for $x > 0$, $u(0) = 1/2$, and $u(x) = 0$ for $x < 0$). In practice, however, it is more convenient to approximate $\Gamma_{\mathrm{IF}}(\mathrm{j}\omega)$ by the low pass filter that is related to the band-pass filter $H_{\mathrm{RF}}(\mathrm{j}\omega)$ by means of the biquadratic transformation [10], as explained in Section 4.2.1.

It follows [6], that the signal component contained in the power density spectrum $S_{\dot{\Phi}}(\omega)$ equals the linearly filtered message spectrum $S_{\dot{\varphi}}(\omega)$. Thus, when this component is denoted by $S^L_{\dot{\Phi}}(\omega)$, we may write

$$S^L_{\dot{\Phi}}(\omega) = |\Gamma_{\mathrm{IF}}(\mathrm{j}\omega)|^2 \, S_{\dot{\varphi}}(\omega). \tag{5.2}$$

Thus apparently, the spectrum of the message contained in the instantaneous *phase/frequency* of the FM wave is filtered by the low-pass equivalent of the spectral *magnitude* characteristic. This rather remarkable result may be explained as follows. For narrow-band FM waves, the instantaneous message phase $\varphi(t)$ is generally much smaller than one radian. Consequently, such waves may be approximated as

$$\begin{aligned} s(t) &= A\cos\left[\omega_o t + \varphi(t)\right] \\ &= A\cos\varphi(t)\cos\left[\omega_o t + \varphi(t)\right] - A\sin\varphi(t)\sin\left[\omega_o t + \varphi(t)\right] \\ &\approx A\cos\omega_o t - A\varphi(t)\sin\omega_o t. \end{aligned} \tag{5.3}$$

Since this is essentially the same expression as for an AM wave, except for the fact that the modulation is in quadrature with the carrier instead of in-phase,[1] linear filtering of such a wave has essentially the same effect as filtering of an AM wave.

[1] The behavior expressed by (5.2) and (5.3) is also the explanation for the fact that the phase noise spectrum in harmonic oscillators decreases with 20 dB per decade [11, 12]; in that case, white noise is filtered by the second-order frequency selectivity (resonator) in the oscillator.

Cross Term It follows [6], that the cross-term between signal and distortion, $S^C_{\dot{\Phi}}(\omega)$, is given by

$$S^C_{\dot{\Phi}}(\omega) = 2S_{\dot{\varphi}}(\omega)\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{S_{\dot{\varphi}}(y)}{y^2}\mathrm{Re}\left[\Gamma_{\mathrm{IF}}(\mathrm{j}\omega)\Gamma_{\mathrm{IF}}(\mathrm{j}y)\Gamma_{\mathrm{IF}}(-\mathrm{j}y-\mathrm{j}\omega)\right]\mathrm{d}y$$
$$- 2S_{\dot{\varphi}}(\omega)\left|\Gamma_{\mathrm{IF}}(\mathrm{j}\omega)\right|^2\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{S_{\dot{\varphi}}(y)}{y^2}\left|\Gamma_{\mathrm{IF}}(\mathrm{j}y)\right|^2\mathrm{d}y. \quad (5.4)$$

Note that the integral in the second terms equals the power contents of the linearly filtered instantaneous phase, denoted by $\Phi_L(t)$, of which the power density spectrum is given by (5.2). This indicates that the cross-term is roughly proportional to the output signal power.

Implications on Low-Distortion Design

Two important conclusions for the design of frequency selectivity in receivers for narrow-band FM can be drawn from the previously discussed model.

In the first place, since the message signal is filtered by the magnitude characteristic of the RF/IF filter, the high-frequency spectral contents of the message, and also of the output noise, is usually somewhat suppressed due to the finite roll-off of the filter. Since this suppression is known from the filter transfer, compensation by means of an equalizing filter at the demodulator output can be applied, when necessary. However, usually, such compensation is not required.

The second conclusion follows from figure 5.4, that depicts Signal-to-Distortion Ratio (SDR), i.e. the ratio of the power contained in (5.2) and (5.4), for a fourth-order band-pass filter (represented by its second-order low-pass equivalent), as function of the quality factor Q of the low pass equivalent filter. The bandwidth of the filter was selected according to Carson's rule, which is usually a slightly too small estimation of the actual bandwidth of the FM wave [1, 8]. This is also reflected by the relatively low SDR values in figure 5.4. The FM message signal was assigned a rectangular spectrum, and a frequency deviation ratio $\Delta\omega/W = 0.5$, where W denotes the message bandwidth, corresponding to a narrow-band wave.

The SDR shows a steep optimum for $Q = 0.57$, corresponding to a Bessel filter, with a maximally linear phase characteristic. This observation agrees with the conclusions obtained for wideband waves with the aid of the quasi-stationary approximation, discussed in Section 5.2.1. Apparently, although that approximation is formally invalid in this case, it still seems to yield valuable information, as far as distortion minimization is concerned.

The maximum in the SDR around $Q = 1.3$ is probably due to the 'peaking' of the transfer, that increases the spectral contents of the message at the band edge, while it does not (yet) result in excessive distortion. The optimum around $Q = 0.6$ is theoretically the most proper choice, since it does not result in

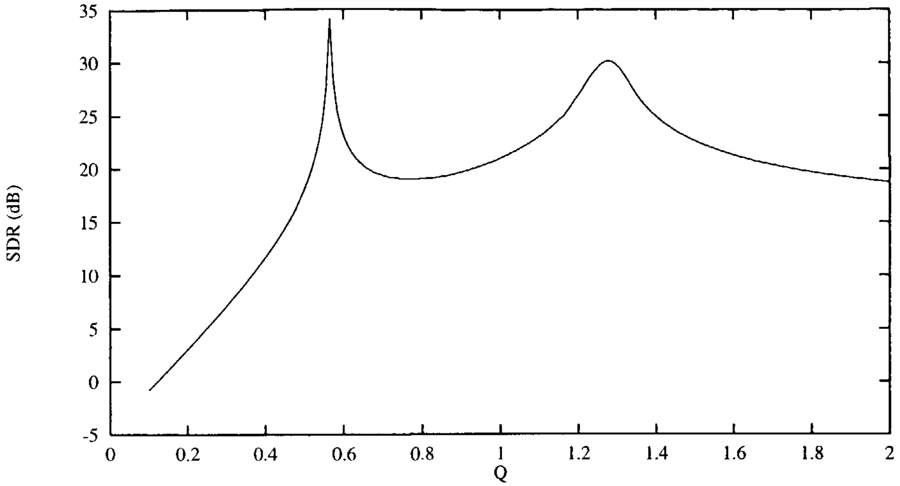


Figure 5.4: Signal-to-Distortion Ratio as function of the quality-factor Q of the low-pass equivalent.

a 'peaked' message (and noise) spectrum. In practice, the optimum around $Q = 1.3$ has the advantage to be relatively shallow, and therefore far more insensitive to tolerances in the filter transfer than the optimum around $Q = 0.6$.

5.3 Pre-Detection Phase Selectivity

Frequency selectivity is able to suppress the main portion of noise and interference in the receiver input signal, as far as it is located outside the channel occupied by the requested FM wave. Interference located in the same channel as the requested wave, so called co-channel interference, cannot be eliminated in this way. This requires the use of 'phase selectivity', i.e. the capability to distinguish between the instantaneous phase of signals with the same frequency.

This section considers a type of receiver architectures intended for suppression of co-channel interference, by application of phase selectivity. First, the principles of these architectures are outlined. Subsequently, some implementations encountered in literature are considered.

5.3.1 Principles of Operation

The architectures capable of co-channel interference suppression reported in literature are based on the cancellation scheme depicted in figure 5.5. It is assumed that the signal at the IF filter output equals the addition of two FM

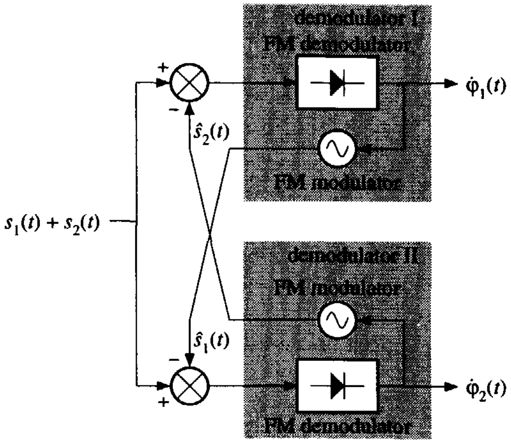


Figure 5.5: Suppression of co-channel interference by application of phase selectivity.

waves, denoted by $s_1(t)$ and $s_2(t)$ respectively, with approximately the same carrier frequency.

The input signal of the demodulators I and II in this figure equals the subtraction of the IF filter output signal, and a reconstruction, symbolized by the ˆ-sign, of one of the two FM waves. The reconstruction is obtained by re-modulation of the output signal of the other demodulator. When both reconstructions resemble the original FM waves, this subtraction yields the other FM wave, which can subsequently be demodulated.

Proper operation of this scheme requires phase selectivity, since the demodulators I and II should keep track of the instantaneous phase of the input FM wave, and produce a wave that is exactly in anti-phase with it. Further, it is also required that the amplitude of the reconstruction closely matches the original amplitude, in order to establish effective cancellation.

5.3.2 Implementations

In [13], an architecture of the type depicted in figure 5.5 was obtained by application of estimation theory. In short, such theories, see e.g. [14, 15], consider the received FM wave, mutilated by noise and interference, as a stochastic process, of which a certain parameter, the message information, has to be obtained. With the aid of the probability density of the message, the noise and interference, an estimator is developed that reconstructs the message information and thereby minimizes the reconstruction errors due to noise and interference with respect to a predefined criterion. A Maximum A Posteriori (MAP) estimate, which in most cases equals the Minimum Mean Squared Error (MMSE) estimate, results in a recursive estimation scheme that obtains the new estimate from the previous estimate and the input signal. Phase Lock Loop (PLL) structures are considered as the (approximate) physical realizations of such estimators.

The cross-coupled PLL architecture realized (and measured) according to this estimation algorithm is depicted in figure 5.5. One of both PLL's, i.e. phase selective FM demodulators, in this figure locks on the strongest FM wave, which is possible due to the fact that the capture effect suppresses the weakest. The controlled oscillator in this PLL produces a reconstruction that has a predefined phase relation, determined by the nature of the phase detector, with the original FM wave. With the aid of a phase shifter, this wave is subtracted from the input of the second PLL, which thereby becomes able to demodulate the weakest FM wave of the two.

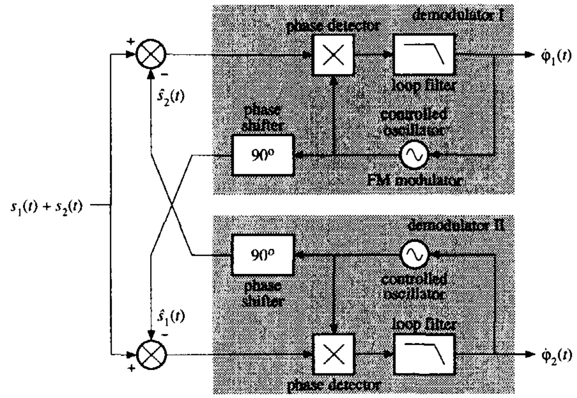


Figure 5.6: Cross-coupled PLL FM demodulator reported in [13], capable of co-channel interference suppression.

The disadvantages of this structure are that in the first place, it cannot be predicted in advance which of the two PLL's locks on the strongest carrier wave, unless both PLL's possess different closed-loop transfers. Further, it is usually unclear which of the two waves, the requested one or the interferer, is actually the the strongest one. In order to solve this problem, the received FM waves should contain a unique identifier, that describes its origin. Secondly, in order to establish cancellation of the interference with the aid of a reconstructed carrier, their amplitudes should match very accurately. The architecture in figure 5.6 does not provide means to establish this automatically. In [16], amplitude detectors are added to the architecture for this purpose, while in [17], the subtraction at the PLL inputs is replaced by a controllable notch filter, that suppresses the interferer. Finally, since all schemes are based on the phase lock principle, they inherently suffer from the disadvantages of a PLL architecture (see Chapter 7), such as cycle-slipping and loss of lock phenomena.

5.4 Pre-Detection Amplitude Selectivity

Section 5.4 mentioned already that considerable improvement of the FM demodulator output SNR may be achieved by application of *infinite* compression to the input carrier amplitude, i.e. by pre-detection amplitude selectivity. Further, it was mentioned that this improvement is established at the expense of the generation of click noise, that deteriorates the output signal quality at low CNRs. Especially in FM receivers intended for the reception of audible intelligence, this type of noise yields extremely annoying results.

This section qualitatively investigates the mechanisms that constitute the SNR improvement at high input CNRs, and the deteriorative click noise at low CNRs. It is shown that, in order to improve the intelligibility of the output signal at low CNRs, a *trade-off* between click noise and continuous noise may be established by application of *finite* compression to the demodulator input wave, instead of the usually applied infinite compression. This improvement technique is particularly suited for applications where operation above threshold cannot be guaranteed, such as car radio. Other improvement techniques, elaborated in Section 5.7, Chapter 7 and Chapter 8 are less suited for such applications. They shift the threshold to a lower input CNR, but at the same time increase the steepness of the threshold curve, resulting in increased "aggressiveness" of the threshold.

An outline of this section is as follows. Section 5.4.1 introduces the general model, used to describe arbitrary types of amplitude compression systems. Section 5.4.2 considers the SNR improvement mechanism observed for high input CNRs. Section 5.4.3 explains the click noise generation mechanism in FM receivers with infinite compression, Subsequently, Section 5.4.4 considers the extension to click noise in receivers with finite compression. Finally, Section 5.4.5 considers a type of noise that is encountered in the output signal of FM receivers with finite compression. The quantitative modeling of all mechanisms described in this section is considered in Chapter 6.

5.4.1 General Amplitude Compressor Model

The models developed in the sequel of this thesis for the FM demodulator output noise require a description of the amplitude compression operation, applied to the demodulator input carrier envelope. This section describes the general model, valid for arbitrary, instantaneously reacting amplitude compressors, such as limiters and fast AGC's.

We first consider the representation of the compressor output wave, and subsequently discuss the corresponding description of the demodulator output signal.

Amplitude Compressor Transfer

The general model for the amplitude compressor is illustrated by figure 5.7. The compressor input signal equals the addition of the noise free FM wave $s(t)$ from (2.1), and additive Gaussian noise $n(t)$, described in Section 2.3. According to Chapter 2, the composite noisy FM wave $r(t)$ may be written in polar format as

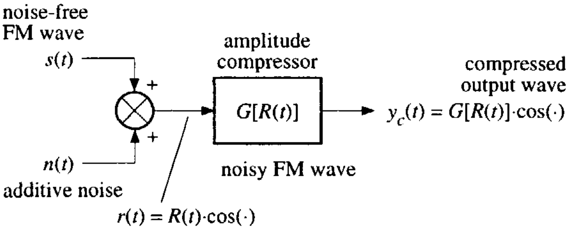


Figure 5.7: General model of an amplitude compressor.

$$s(t) + n(t) \overset{\text{def}}{=} r(t) = R(t)\cos\left[\omega_o t + \varphi(t) + \theta(t)\right], \tag{5.5}$$

where the noisy amplitude $R(t)$ and phase noise $\theta(t)$ are described by (2.16) and (2.17) respectively.

The ideal amplitude compressor transforms the input carrier amplitude $R(t)$ into the output amplitude $G\left[R(t)\right]$, while it passes the carrier phase. The output signal thus becomes

$$y_c(t) = G\left[R(t)\right]\cos\left[\omega_o t + \varphi(t) + \theta(t)\right]. \tag{5.6}$$

In practice, amplitude compressors, as e.g. limiters and AGC's, do affect the phase of the output wave by parasitic AM to PM conversion of noise, that corrupts the FM message signal, as a result of their finite response time [18]. At this level of consideration however, such effects are ignored. In (5.6), $G[\ldots]$ denotes the transfer function of the compressor, or, as e.g. in case of limiters, its first harmonic response[2].

Demodulator Output Signal

A generalized expression of the demodulator output signal, contains the transfer functions G_1 and G_2 of two mutually independent amplitude compressors:

$$y_{\text{dem}}(t) = G_1\left[R(t)\right]G_2\left[R(t)\right]\left[\dot{\varphi}(t) + \dot{\theta}(t)\right]. \tag{5.7}$$

[2] In Chapter 6 is shown that it's often advantageous to use only the first harmonic of the output, since this yields the highest output SNR.

This expression covers the various types of demodulators discussed in Chapter 3 in the following way. With respect to their dependence on $R(t)$, three classes of FM demodulator responses are distinguished:

- the response is independent of $R(t)$;
- the response is proportional to $R(t)$;
- the response is proportional to $R^2(t)$.

The first response is exhibited by demodulators based on zero-crossing detection (FM-PM conversion with a fixed phase difference). They intrinsically contain an amplitude compression mechanism that removes all amplitude noise. Additional compression by means of a separate amplitude compressor in front of such demodulators therefore makes no sense. Their response is represented by (5.7) when the compressor transfers G_1 and G_2 equal a proportionality constant. Consequently, this type of demodulators leaves no degrees of freedom to optimize their sensitivity and response by proper design of the amplitude compressor transfer

The second response is exhibited by FM demodulators that employ an AM modulus demodulator, which does not require a reference wave for demodulation. For these demodulators, the transfer G_2 (or G_1) equals a constant, while G_1 (or G_2) corresponds to an optional compressor that precedes the demodulator. The demodulator transfer may therefore be optimized by proper design of this transfer.

The third response is exhibited by all FM demodulators that employ a reference wave, e.g. those equipped with an AM projection detector or a quadrature phase detector. In such demodulators, G_1 represents the compressor that processes the FM wave subjected to the demodulation, while G_2 processes the reference wave required by the AM/PM detector, that is eventually derived from the input FM wave. Thus, in these demodulators, both G_1 and G_2 may be used to optimize the demodulator performance.

5.4.2 SNR Improvement Above Threshold

It was mentioned previously that amplitude compression provides an improvement of the demodulator output SNR at high input CNRs. This section investigates the mechanism of this improvement, as function of the amplitude compressor 'transfer function' $G(R)$.

First, a time-domain representation of the improvement mechanism is discussed. Subsequently, the corresponding demodulator output noise spectrum and output SNR are considered.

Improvement Mechanism

The SNR improvement mechanism established by compression of the noisy demodulator input carrier amplitude is based on the compressor behavior depicted in figure 5.8. The noisy input wave $r(t)$, corresponding to the phasor $\vec{r}$, is the resultant of the noise-free FM wave $s(t)$, corresponding to the phasor $\vec{s}$, and the noise $n(t)$, represented by $\vec{n}$. At high CNRs, i.e. when $|\vec{n}| \ll |\vec{s}|$, the noise components in-phase with $s(t)$, represented by $n_{s,i}(t)$, represents the amplitude noise in the input carrier, while the component in quadrature with $s(t)$, represented by $n_{s,q}(t)$, represents the phase noise (see Chapter 2).

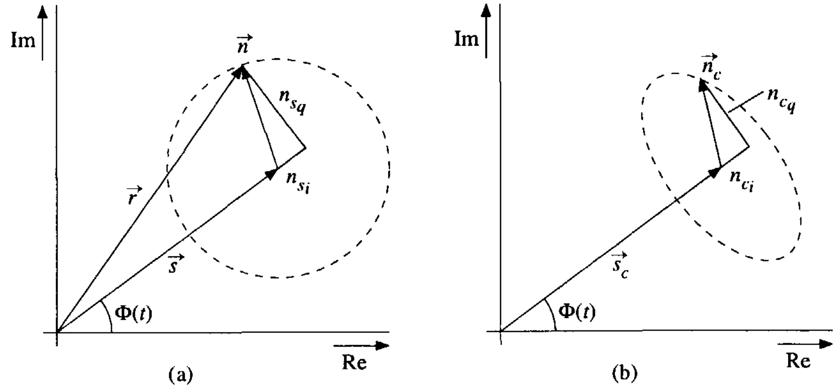


Figure 5.8: Amplitude compressor operation on a noisy input wave. a) compressor input signal, b) compressor output signal.

In the phasor representation of the compressor input wave, depicted in figure 5.8a, the noise components $n_{s,i}(t)$ and $n_{s,q}(t)$, in-phase and in quadrature with $s(t)$ respectively, are of the same magnitude (on average), which means that the tip of the noise phasor $\vec{n}$ describes a *circular* path around the tip of $\vec{s}$.

In the phasor representation of the compressor output wave, depicted in figure 5.8b, the tip of the noise phasor $\vec{n}_c$ describes an *elliptic* path around the tip of $\vec{s}$, instead of the circular path described by the input noise. Due to the amplitude compression, the in-phase noise component, representing the amplitude noise, is (partly) suppressed, while the quadrature noise, representing the phase noise, remains essentially unaffected. Thus, small signals (and noise) in-phase with the large FM wave $\vec{s}$ are compressed, while those in-quadrature with $\vec{s}$ remain unaffected.

A mathematical description of the compressor action can be obtained by means of a small-signal approximation, which exploits the property that the noise is small compared to the FM wave at high input CNR‘s. This approximation, elaborated in Section 6.3, considers the input noise $n(t)$ as a small ‘signal’, that is superimposed on a time-dependent 'bias', the wave $s(t)$. With the aid of expression (5.7), which describes the demodulator output signal in terms of the amplitude $R(t)$ and the frequency noise $\dot{\theta}(t)$, expression (2.16) and (2.19), which express $R(t)$ and $\dot{\theta}(t)$ in terms of the noise components $n_{s,i}(t)$, $n_{s,q}(t)$, and their derivatives, the demodulator output signal can be expanded into a Taylor series to the noise components. This series expansion yields two interesting conclusions, concerning the transfer of small noise from the amplitude compressor input, to the FM demodulator output signal (see Section 6.3). It is shown that, in terms of the compressor transfer $G(R) = G_1(R)G_2(R)$ (see equation (5.7)),

- the amplitude noise is transferred by the *small-signal transfer*;
- the phase/frequency noise is transferred by the *large signal transfer.*

This follows also directly from expression (5.7). The amplitude noise represents (small) deviations of the actual carrier amplitude $R(t)$ from the noise free amplitude A, and is therefore transferred by the small signal transfer. The instantaneous frequency is multiplied by the input carrier amplitude, and is therefore transferred by the compressor large signal transfer.

Consequently, amplitude compression, and (partial) suppression of the noise is established by the fact that the amplitude noise $n_{s,i}(t)$ is transferred by a different transfer than the phase noise $n_{s,q}(t)$. The achieved level of compression may therefore be expressed in terms of a *compression factor*, equal to the ratio of both transfers. In the sequel, however, it appears to be more convenient to use the inverse compression factor, instead of the compression factor itself. This factor is defined as follows.

Definition 1 *The inverse, first-order amplitude compression factor, denoted by $C_{n,1}(A)$, equals the ratio of the amplitude compressor small-signal transfer $\left.\frac{\partial G(R)}{\partial R}\right|_{R=A}$ and the compressor large signal transfer $G(R)/R|_{R=A}$,*

$$C_{n,1}(A) \overset{def}{=} \left.\frac{R}{G(R)}\frac{\partial G(R)}{\partial R}\right|_{R=A}, \qquad (5.8)$$

where A denotes the amplitude of the FM wave $s(t)$.

This definition is consistent with the strict definition of the compression factor, i.e. $1/C_{n,1}(A)$, used for AGC's (see e.g. [19]).

Thus, according to definition 1, compression is established when $C_{n,1}(A) < 1$. Furthermore, *infinite* compression is established when the inverse compression factor equals zero, i.e. $C_{n,1}(A) = 0$. In that case, the in-phase component of the compressor output noise vanishes completely.

Output Noise Spectrum

The effect of amplitude compression on the power spectral density of the demodulator output noise is as follows. Generally, if the spectrum of the input noise $n(t)$ is flat over the entire FM bandwidth, the demodulator output noise power spectral density (PSD) consists of two mutually uncorrelated components:

- a parabolic shaped component, representing the frequency noise $\dot{\theta}(t)$;
- a white component, representing the contribution of the amplitude noise.

Both components are sketched in figure 5.9. The parabolic shape of the first component, a result of the differentiation to time from phase noise to frequency noise, is responsible for the improved transmission capabilities of FM, in comparison to AM. The parabolic shaping moves the largest portion of the input noise power from low frequencies, where the message signal resides, to high frequencies, where it is easily eliminated by a low-pass filter.

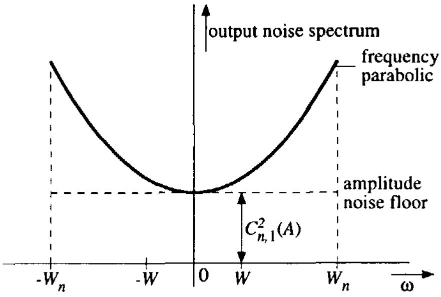


Figure 5.9: Demodulator output noise spectrum above threshold.

The white component, which is proportional to the squared inverse compression factor $C^2_{n,1}(A)$, introduces a considerable amount of noise at baseband frequencies, and therefore deteriorates the output SNR. This component should thus be minimized by choosing $C_{n,1}(A)$ as small as possible, which is effectuated by application of a maximum level of compression (in the ideal case infinite) to the input carrier amplitude.

Output SNR

The demodulator output SNR corresponding to the noise spectrum of figure 5.9 may be calculated as (see Chapter 6)

$$\mathrm{SNR}_{\mathrm{out}} = \frac{3p\frac{W_n}{W}\left(\frac{\Delta\omega}{W}\right)^2}{1 + 3C^2(A)\left(\frac{\Delta\omega}{W}\right)^2}, \tag{5.9}$$

where W_n denotes the FM transmission bandwidth, W the message bandwidth, $\Delta\omega$ the RMS frequency deviation and p the input CNR. As discussed below, this expression hides some interesting conclusions concerning the output SNR of the three demodulator classes distinguished in Section 5.4.1.

Output Signal Independent of Carrier Amplitude When $G(R)$ is a proportionality constant independent of $R(t)$, the inverse compression factor $C_{n,1}(A)$ equals zero, and consequently infinite compression is established. In that case, all amplitude noise is suppressed, the white noise floor vanishes, and the output SNR (5.9) attains the maximum value given by (2.21).

Output Signal Proportional to Carrier Amplitude When $G(R)$ is proportional to $R(t)$, the inverse compression factor $C_{n,1}(A) = 1$. In these systems, the amplitude noise deteriorates the performance of the FM system to the level of an AM transmission system. As observed from (5.9), the output SNR approaches the SNR of a Double Side Band (DSB) demodulator with transmission bandwidth W_n, and message bandwidth W [1].

Output Signal Proportional to Squared Carrier Amplitude When $G(R)$ is proportional to $R^2(t)$, it follows from (5.8), that the inverse compression factor $C_{n,1}(A) = 2$. The output SNR of the FM demodulator is at least 6 dB below the level achieved with a comparable DSB system, due to the fact that both the demodulated wave and the reference wave used during demodulation contribute (mutually correlated) amplitude noise to the output signal.

5.4.3 Click Noise Generated by Infinite Compression

It is well known that the threshold effect in FM demodulators, i.e. the phenomenon that the demodulator output SNR decreases faster than predicted by (5.9) at low CNRs, is a direct consequence of the generation of impulsive noise, usually called "click noise" [20]. This noise consists of discrete pulses of high energy, with a rate that rapidly increases when the input CNR decreases.

The first attempts to model the FM threshold effect [21–26], based on Rice's work concerning the statistical properties of random noise [27–29], resulted in rather complicated mathematical descriptions. Despite their scientific relevance, their value in engineering practice is limited, due to the mathematical complexity, and the absence of the perceptive notion of click noise. Further, these models consider some special types of amplitude compressors only, and do not hold for arbitrary amplitude compressor transfers.

The first suitable engineering models of the threshold effect, that explicitly include the notion of click noise, were independently developed by Cohn [30] and Rice [20]. These models consider demodulators with infinite amplitude compression only, i.e. elimination of all amplitude noise.

This section discusses the principles of the click noise model, and the mechanism that constitutes the generation of this noise, for FM receivers with infinite compression. Section 5.4.4 considers the extension of this model to arbitrary types of compression.

First, the click noise generation mechanism is outlined. Subsequently, the basics of the click noise model are discussed. Finally, expressions for the demodulator output click rate and power spectral density are given for important types of FM message signals.

Click Noise Mechanism

Click noise is basically due to the introduction of zero-crossings by the input noise $n(t)$ into the composite amplitude compressor input wave $r(t)$.

The effect of the noise-induced zero crossings on the output signal of the amplitude compressor, which, without loss of generality, will be modeled by an ideal hard-limiter, and the output signal of the demodulator is illustrated by figure 5.10. Figure 5.10a depicts *minus* the input wave $s(t)$, the noise $n(t)$, and the corresponding composite compressor input wave $r(t)$. Besides shifting the zero-crossings introduced by $s(t)$, resulting in continuous phase/frequency noise at the demodulator output, the noise introduces additional zero crossings, or cancels crossings introduced by $s(t)$, when its intensity temporarily exceeds $-s(t)$. Each time this occurs, two zero crossings are introduced into $r(t)$, or, alternatively, one is canceled. Around the threshold, it is on average, still very unlikely that $n(t)$ exceeds $-s(t)$ for a long time. Therefore, both induced zero-crossings follow shortly after each other.

The compressor response, depicted in figure 5.10b, equals the polarity of the composite input wave $r(t)$. Therefore, each time the noise introduces introduces/cancels zero-crossings in $r(t)$, this response rapidly gains/losses one cycle with respect to the noise free FM wave $s(t)$. During such an event, the phase difference between $r(t)$ and $s(t)$, i.e. the phase noise $\theta(t)$, increases/decreases nearly stepwise by an amount 2π. Consequently, its derivative, the frequency noise $\dot{\theta}(t)$, shows an impulse of area 2π, i.e. a click. In FM receivers with infinite compression, the noise observed at the demodulator output is proportional to $\dot{\theta}(t)$, and therefore shows the same impulse, as depicted in figure 5.10c.

A phasor representation of the click noise mechanism is depicted in figure 5.11. In this figure, generation of a click corresponds to an encirclement of the noise phasor $n(t)$ around the origin. During such an event, the phase noise $\theta(t)$ increases/decreases by 2π, resulting in an impulse of area 2π in $\dot{\theta}(t)$.

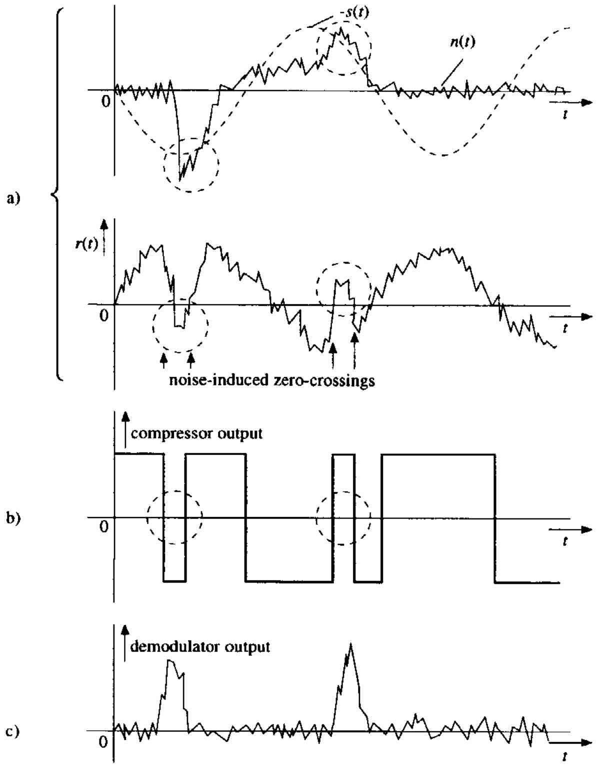


Figure 5.10: Origin of click noise. a) compressor (limiter) input signal, b) output signal, c) demodulator output signal.

Click Noise Model

The approximate models developed in [20, 30] start from the description of the FM demodulator output signal in the presence of infinite compression. In that case, G_1 and G_2 in (5.7) are proportionality constants, resulting in an output signal that is directly proportional to the instantaneous frequency of the input FM wave $r(t)$, i.e.

$$y_{\text{dem}}(t) = G_o \left[\dot{\varphi}(t) + \dot{\theta}(t)\right] . \tag{5.10}$$

Based on this expression, the following approximations are applied.

In the first place, the frequency noise $\dot{\theta}(t)$ is decomposed into two components that are considered mutually independent:

- continuous frequency noise, that behaves as described in Section 5.4.2;

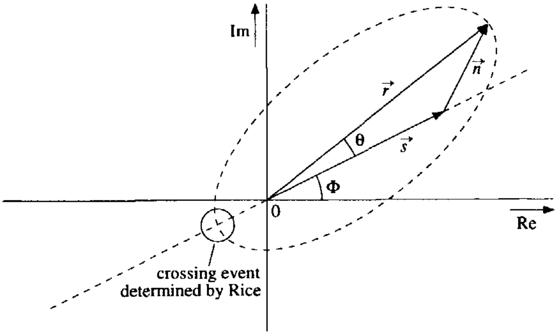


Figure 5.11: Phasor representation of the click noise mechanism.

- impulsive noise, consisting of a stochastic train of pulses.

According to a theoretical study in [31], the assumption on the approximate independence of both components is justified, at least in the threshold region, i.e. input CNRs typically around 10 dB. This can be explained by the observation that click noise completely dominates the demodulator output during a noise impulse, and therewith temporarily suppresses the FM message signal and the continuous noise. In this view, clicks and the continuous demodulator output components do not coexist on the same instant, and are therefore uncorrelated.

Secondly, the click pulses are not described instantaneously, i.e. by the pulse intensity as function of time, but by two stochastic averages:

- the area of a single click pulse;

- their average rate of occurrence.

The shape of the click pulses is approximated by a Dirac-impulse, due to their short duration. This averaged description is the main strength of the click noise model that, in comparison with previous threshold models, yields a significant simplification.

In a receiver with infinite compression, the average click pulse area is a rather trivial parameter, since the area of all such pulses in $\dot{\theta}(t)$ equals, by definition, exactly 2π. In receivers with finite compression however, considered in the next section, this is no longer the case.

The average rate of clicks is entirely determined by the properties of the compressor input signal $r(t) = s(t) + n(t)$, and is therefore, at least in a first approximation, independent of the amplitude compressor transfer $G(R)$. In order to calculate these rates, it is assumed in [20, 30] that clicks are mutually independent, which means that they can be modeled as Poisson processes. This is allowed, since the click rate around the threshold is still very low; in FM receivers for audio broadcasting, the click rate at the threshold is in the order of one per second [20]. Further, the encirclement performed by $\vec{n}$ during a click is modeled as a crossing event. It is assumed that the encirclement is completed (with a probability of unity) when $\vec{n}$ crosses through $-\vec{s}$ (see figure 5.11), i.e. when the phase noise $\theta(t) \mod 2\pi$ exceeds π. The rate of these crossings is subsequently determined exactly.

Power Spectral Density and Click Rates

The double-sided power spectral density of click noise $\dot{\theta}_{\text{click}}(t)$, is essentially white within the baseband, due to the impulsive, and consequently wideband, nature of the click pulses. As discussed in Section 5.4.2, a white noise floor results in considerable deterioration of the demodulator output SNR. When N_+ and N_- denote the rates of anti-clockwise and clockwise origin encirclements performed by $\vec{n}$ respectively, it can be shown, that the double-sided click noise spectral density, denoted by $S_{\dot{\theta}_{\text{click}}}(\omega)$, equals the total click rate times the squared frequency spectrum of a single click pulse [20], i.e.

$$S_{\dot{\theta}_{\text{click}}}(\omega) = 4\pi^2 \left(N_+ + N_-\right), \tag{5.11}$$

were the factor $4\pi^2$ represents the squared click pulse area.

The click rates N_+, N_- have been calculated in literature for a large number of different situations, including the presence of oscillator phase noise in $r(t)$ [32], and co-channel interference [2]. The results derived in [20] for the cases of interest in the sequel of this thesis are briefly outlined below.

Unmodulated Carrier The case when the FM wave $s(t)$ is an unmodulated carrier is of considerable theoretical interest, since it allows comparisons between the performance of various types of FM demodulators, with minimum complexity of the models. For this case, it can be shown, that N_+ and N_- can be obtained exactly, i.e. without approximation, as

$$N_+ = N_- = \frac{r}{2}\left[1 - \operatorname{erf}\left(\sqrt{p}\right)\right], \tag{5.12}$$

where r denotes the so called *radius of gyration*, defined as

$$r \overset{\text{def}}{=} \frac{1}{2\pi}\sqrt{\frac{\int_{-\infty}^{\infty} \omega^2 S_n(\omega)\mathrm{d}\omega}{\int_{-\infty}^{\infty} S_n(\omega)\mathrm{d}\omega}}, \tag{5.13}$$

where $S_n(\omega)$ denotes the power density spectrum of $n_{s,i}(t)$ and $n_{s,q}(t)$. This parameter represents the total, average number of zero-crossings that occurs in $n_{s,i}(t)$ and $n_{s,q}(t)$ per unit time, which follows from (5.12) by setting $p = 0$.

Gaussian Modulation For additive Gaussian noise and Gaussian message signal $\dot{\varphi}(t)$ with an RMS frequency deviation equal to $\Delta\omega$, N_+ and N_- can be shown to satisfy [20]

$$N_+ = N_- \approx r\exp(-p)\sqrt{\frac{1+p\left(\frac{\Delta\omega}{2\pi r}\right)^2}{4\pi p}}. \tag{5.14}$$

Sinusoidal Modulation For sinusoidal modulation with maximum frequency deviation $\Delta\omega$, the click rates can be shown to be

$$N_+ = N_- \approx \frac{\Delta\omega}{2\pi^2}\exp(-p) + \frac{r\exp(-p)}{\sqrt{4\pi p}}\exp\left[-\frac{p}{2}\left(\frac{\Delta\omega}{2\pi r}\right)^2\right]\mathrm{I}_0\left[\frac{p}{2}\left(\frac{\Delta\omega}{2\pi r}\right)^2\right] \tag{5.15}$$

is obtained. Here $\mathrm{I}_0(.)$ denotes the zero-th order modified Bessel function of the first kind. Although the appearance of the expressions (5.12), (5.14), and (5.15) is quite different, it follows in all three the cases that the click rates increase approximately exponential for decreasing input CNRs. Therefore, when the click noise dominates, the relation between the demodulator output SNR and the input CNR becomes approximately exponential.

5.4.4 Click Noise Generated by Finite Compression

It was previously mentioned in this chapter that application of finite compression instead of the usual infinite compression to the input carrier amplitude establishes a trade-off between click noise and continuous noise. This section explains the mechanism of this trade-off.

A problem that arises in the discussion on receivers with finite compression is that the click noise models encountered in literature are confined to receivers with *infinite* compression. A click noise model for the case of *finite* compression has, as far as known, not been developed. Therefore, such a model is developed in this thesis. This section outlines the principles of the newly developed model, while Section 6.5 develops the mathematical formulation.

First, the mechanism of the trade-off between click noise and continuous noise is discussed. Subsequently, the requirements on the compressor transfer in order to establish click noise suppression are considered. Finally, the principles of the extended click noise model are discussed.

Trade-off Mechanism

According to (5.7), the output noise of FM receivers with finite amplitude compression is besides by the frequency noise $\dot{\theta}(t)$, also determined by the noise in the carrier amplitude $R(t)$. This differs from receivers with infinite compression, where the latter contribution is completely suppressed.

The level of click noise observed at the demodulator output in such receivers is therefore the resultant of two processes:

- generation of click pulses in the frequency noise $\dot{\theta}(t)$;
- correlation between $\dot{\theta}(t)$ and the amplitude $R(t)$.

The level of click noise that is generated in $\dot{\theta}(t)$ is not affected by the amplitude compressor or the demodulator, as long as no phase- or frequency feedback is applied. As discussed in Section 5.4.3, this process is entirely determined by the input waves $s(t)$ and $n(t)$. This is also reflected by expression (5.7); the compressor does not affect $\dot{\theta}(t)$ itself, but only its contribution to the demodulator output signal.

The correlation between $\dot{\theta}(t)$ and $R(t)$ provides a means to suppress click noise in the demodulator output signal, and is the basis of the trade-off between click noise and continuous noise. A click in the frequency noise $\dot{\theta}(t)$ is accompanied by a fade in the carrier amplitude $R(t)$, as discussed below. A strong dependence on the input carrier amplitude, determined by the compressor transfer $G\left[R(t)\right]$, causes the output signal (partly) to fade during the click, resulting in click noise reduction. However, at the same time, it also enables amplitude noise to penetrate the output signal, which considerably increases the level of continuous output noise, as was discussed in Section 5.4.2.

Figure 5.12, that visualizes the demodulator response described by (5.7), illustrates the click suppression mechanism. The frequency noise $\dot{\theta}(t)$ represents

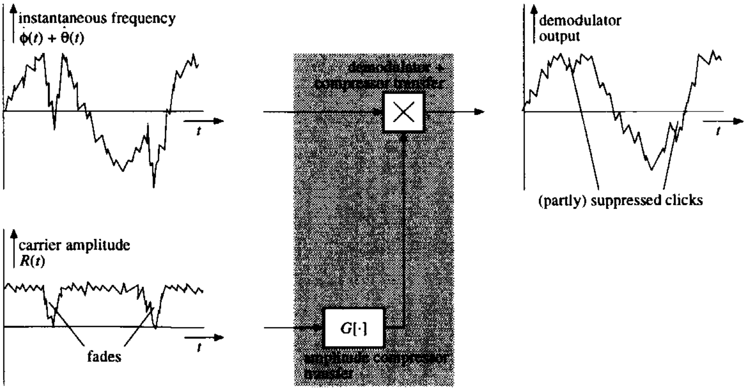

instantaneous frequency
$\dot{\phi}(t) + \dot{\theta}(t)$
t
carrier amplitude
$R(t)$
fades
t
demodulator + compressor transfer
$G[\cdot]$
amplitude compressor transfer
demodulator output
t
(partly) suppressed clicks

Figure 5.12: Partial suppression of click noise in the presence of finite compression, as a result of correlation between $\dot{\theta}(t)$ and $R(t)$.

the output noise of a receiver with infinite compression. When the amplitude compressor transfer passes the fades in $R(t)$, which is possible only with finite compression, the click and the fade are 'multiplied', resulting in partial click noise suppression. Thus, in fact, finite amplitude compression partly prohibits propagation of clicks, generated in $\dot{\theta}(t)$, to the demodulator output.

Noise Fades in the Carrier Amplitude

The origin of the fading mechanism in $R(t)$ follows from the phasor diagram depicted in figure 5.13, representing the input signal $r(t) = s(t) + n(t)$ during a click in $\dot{\theta}(t)$. For CNRs above and around the threshold, clicks are very rare and of very short duration. The corresponding excursions of $\vec{n}$, depicted by curve (a) in figure 5.13, therefore posses an as small as possible length, and closely encircle the origin. The length of the phasor $\vec{r}$, the amplitude $R(t)$, thereby temporarily becomes very small, resulting in a fade at the time instant of the click. At low CNRs, i.e. below the threshold, the duration of the clicks gradually increases, as a result of the increased probability that the noise $n(t)$ will exceed the FM wave $s(t)$. As shown by curve (b) in figure 5.13, the increased click duration corresponds to a larger radius of the origin encirclements of $\vec{n}$. As a result, the fades in the carrier amplitude $R(t)$ become shallower, but of longer duration [33–35].

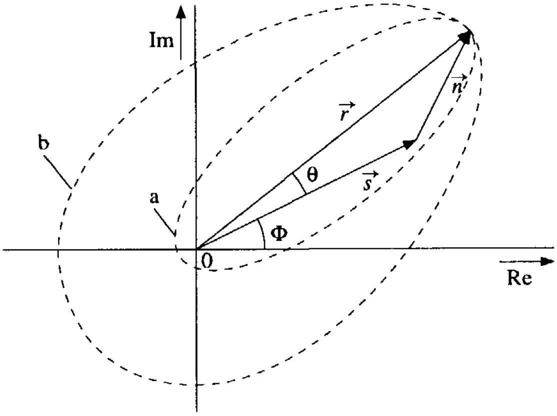


Figure 5.13: Fading in the envelope $R(t)$. Click trajectory at high input CNR a), low input CNR b).

Requirements on the Compressor Transfer

The shape of the amplitude compressor transfer $G[R(t)]$ required to suppress the click noise in the demodulator output signal follows from the expression for

$\dot{\theta}(t)$, obtained by combination of (2.19) and (2.16):

$$\dot{\theta}(t) = \frac{\dot{n}_{s,q}(t)\,[A + n_{s,i}(t)] - \dot{n}_{s,i}(t)n_{s,q}(t)}{R^2(t)}. \tag{5.16}$$

The numerator in this expression equals the product of Gaussian noise processes, and is therefore of a continuous nature. The impulsive nature of click noise in $\dot{\theta}(t)$ is therefore mainly due to the fades in $R(t)$ during the origin encirclements of $\vec{n}$; when $R(t)$ becomes very small for a short time, a steep pulse is observed in (5.16).

Thus, as observed from (5.7) and (5.16), in order to eliminate all clicks from the demodulator output the amplitude compressor transfer $G\,[R(t)]$ should be proportional to $R^2(t)$, at least at low CNRs. This is realized e.g. when no amplitude compression is applied at all; in that case, both G_1 and G_2 in (5.7) are proportional to $R(t)$.

Extension of the Click Noise Model for Infinite Compression

The click noise models [20, 30] described in the previous section, may be extended to arbitrary types of compression in the following way.

As discussed before, these models describe click noise in terms of the average click rate, and the click pulse area. The average click rate is not affected by amplitude compression, since it is entirely determined by the input wave $s(t)$ and noise $n(t)$.

The click pulse area however does depend on the compressor transfer $G\,[R(t)]$, due to the correlation between $R(t)$ and $\dot{\theta}(t)$. When finite compression is applied, this area no longer equals 2π, as in the case of full normalization, but becomes dependent on $R(t)$, i.e. on the shape of the encirclement produced by $\vec{n}$.

As shown in Section 6.5, the known click models can therefore be extended to include finite amplitude compression by replacement of the fixed pulse area 2π in case of infinite compression, with the average area of the pulses observed at the demodulator output in case of an arbitrary type of compression.

5.4.5 Second-Order Noise

This section considers a type of noise, called "second-order noise" in this thesis, which is characteristic for FM receivers with finite compression. In fact, in such receivers, this continuous noise component is part of the trade-off between click noise and continuous noise.

We first explain the impact of second-order noise on the demodulator output SNR. Subsequently, an heuristic explanation for its origin is given on the hand of an example. Finally, its relation with the amplitude compressor transfer is considered. A detailed quantitative analysis of second-order noise is given in Section 6.4.

Impact on the Receiver Output SNR

The impact of second-order noise on the output SNR is illustrated by figure 5.14, that sketches a typical threshold curve, i.e. output SNR versus input CNR, of an FM receiver with finite compression. The threshold curve of a receiver with infinite compression is included as reference. At high input CNRs, the dominant noise contribution is due to the first-order continuous noise, discussed in Section 5.4.2, resulting in an output SNR that increases proportional (10 dB per decade) to the input CNR, as described by (5.9).

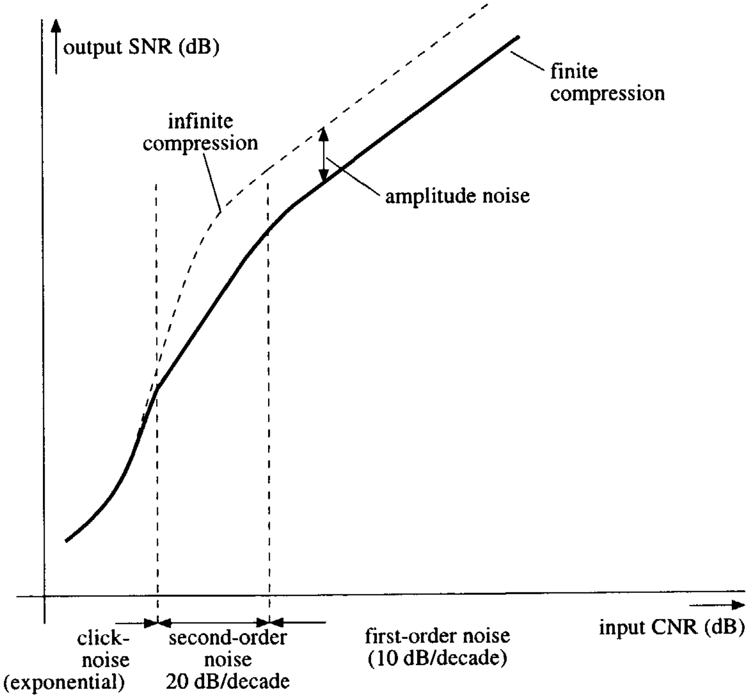


Figure 5.14: Typical threshold curve of an FM demodulator with finite amplitude compression.

At very low input CNRs, a very steep curve is observed, as a result of click noise. There, the output SNR decreases exponentially for decreasing input CNRs, as a result of the exponential increase of the click rate (see Section 5.4.3).

The second-order noise dominates the intermediate region of the threshold curve, between the click noise and the first-order noise. This noise depends on the squared input noise $n(t)$, and therefore (asymptotically) increases the output SNR by 20 dB per decade. In receivers with infinite compression, or those without any kind of compression at all (zero compression), second-order noise becomes noticeable only below input CNRs of about 0 dB. In case of finite compression however, it is already noticeable at CNRs of 10 – 15 dB. Thus, at the expense of a decreased output SNR, at high and intermediate input CNRs, in comparison to infinite compression, finite compression establishes a 'smoothed' threshold and a reduced level of click noise, instead of the rather abrupt threshold and high click noise level obtained with infinite compression.

Origin of Second-Order Noise

The increased level of second-order noise in receivers with finite compression may be explained in an heuristic way by the presence of "noise-induced modulation" in the transfer of amplitude noise to the demodulator output.

A clear example of this effect is observed in receivers that establish finite compression by means of a soft-limiter [36–38], as illustrated by figure 5.15. Figure 5.15a depicts the compressor (soft-limiter) response to the sinusoidal, noisy input FM wave $r(t)$. Figure 5.15b depicts the corresponding small-signal transfer as a function of time. When the input wave is clipped, and the limiter saturates, this transfer equals zero, while it equals the limiter gain during linear

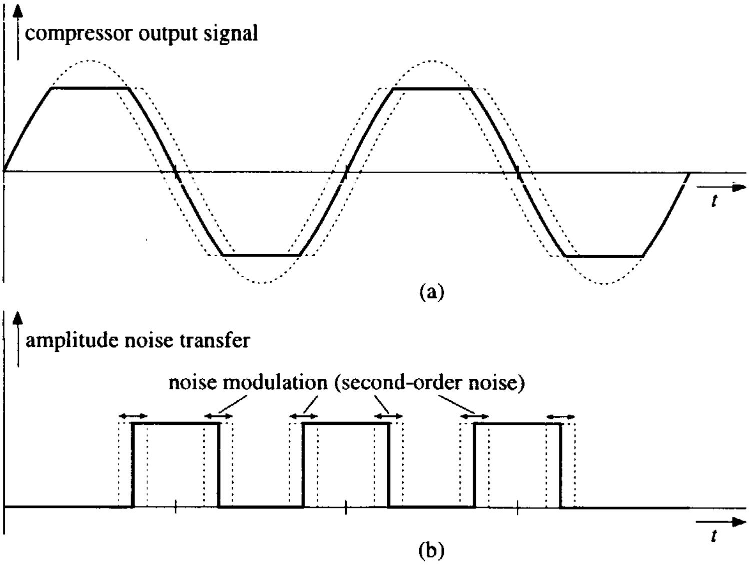


Figure 5.15: Second-order noise in soft-limiting FM receivers. a) compressor (soft-limiter) output signal, b) corresponding small-signal amplitude noise transfer.

operation. The Fourier coefficient corresponding to the fundamental frequency of this transfer equals the transfer $G(A)C_{n,1}(A)$ of the amplitude (in-phase) noise from compressor input to the demodulator output (see Chapter 6).

The modulation effect observed in figure 5.15b, the cause of the significant amount of second-order noise, is explained as follows. The calculation of the first-order demodulator output noise, discussed in Section 5.4.2 and Section 6.3, assumes that the input noise $n(t)$ is very small compared with the FM wave $s(t)$, such that $s(t)$ determines the value of the compressor transfer $G(.)$, and the inverse compression factor $C_{n,1}(.)$. In figure 5.15b, the small-signal transfer corresponding to the soft-limiter response in the absence of noise, i.e. $s(t)$ alone, is represented by the drawn curve. In reality, however, the *combination* of $s(t)$ and $n(t)$ determines the value of both transfers, since the compressor input signal equals $r(t) = s(t) + n(t)$ instead of the noise-free wave $s(t)$. Therefore, the actual positions of the gain pulse edges in figure 5.15, represented by the dashed curves, that correspond to the value of the input wave $s(t) + n(t)$ where the limiter is just driven into saturation, slightly differ from the positions in the absence of noise.

Consequently, the pulse-edge positions are modulated by the noise. The influence of this modulation is especially significant for low limiter gains, since in that case the boundary between saturation and linear operation is positioned close to the top of the input wave, where its slope is small. This illustrates the fact that the level of second order noise generated at the demodulator output, as a result of the influence of the input noise $n(t)$, highly depends on the characteristics of the compressor transfer, as discussed below.

Dependence on the Compressor Transfer

Expressed in a mathematically formal way, second-order noise equals the second-order derivative to the input noise, of the demodulator response to the noisy FM wave $r(t)$. Equivalently, it equals the first-order derivative of the amplitude noise transfer, and therefore represents the variations in this transfer ("modulation") as a result of the input noise.

An advanced analysis in Section 6.4 shows that the level of second-order noise observed at the demodulator output is determined by the first-order inverse compression factor $C_{n,1}(A)$, that also determines the level of first-order noise, and the second-order inverse compression factor $C_{n,2}(A)$, that is defined as follows.

Definition 2 *The inverse second-order compression factor, denoted by $C_{n,2}(A)$, corresponding to the amplitude compressor transfer $G(R)$, is defined as*

$$C_{n,2}(A) \overset{def}{=} \frac{R^2}{G(R)} \frac{\partial^2 G(R)}{\partial R^2}\bigg|_{R=A} . \tag{5.17}$$

Further, the analysis shows that a minimum amount of second-order noise is observed when $C_{n,1}(A) = C_{n,2}(A)$, i.e. when both compression factors are equal. This condition is satisfied by receivers with infinite compression, where $C_{n,1}(A) = C_{n,2}(A) \equiv 0$, and receivers without compression, where $G(A) = A^2$ and $C_{n,1}(A) = C_{n,2}(A) \equiv 2$. In receivers with finite compression, however, the condition is generally *not* satisfied, which explains the increased level of second-order noise.

5.5 FM Demodulation

Since the main task of the FM receiver is to provide a suitable embedding for the FM demodulator, it should not be a surprise that decisions made during the design of the FM demodulator, i.e. the implementation of the FM demodulation function, will generally have a profound influence on the receiver architecture. Besides the FM receiver performance the FM demodulator also determines the degrees of freedom left to the receiver to adjust/improve this performance.

Implementation of the demodulation function was considered in detail already in Chapter 3 and Chapter 4. In this section, we summarize the impact of two main FM demodulator characteristics on the performance, and degrees of freedom left in the design of the receiver architecture:

- the presence of intrinsic amplitude compression;
- the presence of a frequency offset in the demodulator output signal.

Both these demodulator characteristics are conveniently described by the following general expression for the demodulator output signal:

$$y_{\mathrm{dem}}(t) = G_1\left[R(t)\right] G_2\left[R(t)\right] \left[\dot{\varphi}(t) + \dot{\theta}(t) + \omega_{\mathrm{offs}}\right], \tag{5.18}$$

which differs from expression (5.7) only by the frequency offset ω_{offs}. This expression will be used in the sequel to demonstrate the impact of both properties.

5.5.1 Intrinsic Amplitude Compression

As concluded in Section 5.4.1, the amount of freedom in the design of the amplitude compressor transfers G_1 and G_2 is severely limited by certain types of FM demodulators. In this respect, three classes of demodulators where distinguished:

- demodulators that allow free design of both G_1 and G_2;
- demodulators that allow free design of G_1 *or* G_2;
- demodulators that fix both transfers.

The first transfer-type can be established by the following classes of FM demodulators:

- FM-AM conversion demodulators with AM projection detection;
- FM-PM conversion demodulators, based on a fixed time-delay;
- post-detection conversion demodulators.

These demodulators somehow employ a reference wave, besides the FM wave subjected to the demodulation. As an example, figure 5.16 depicts a demodulator based on FM to AM conversion that establishes different transfers G_1 and G_2. It consists of two cross-coupled balanced math-demodulators, of the type depicted in figure 3.17. The cross coupling is required in this case in order to eliminate undesired cross-terms, corresponding to the time-derivative of $G_{1,2}\left[R(t)\right]$.

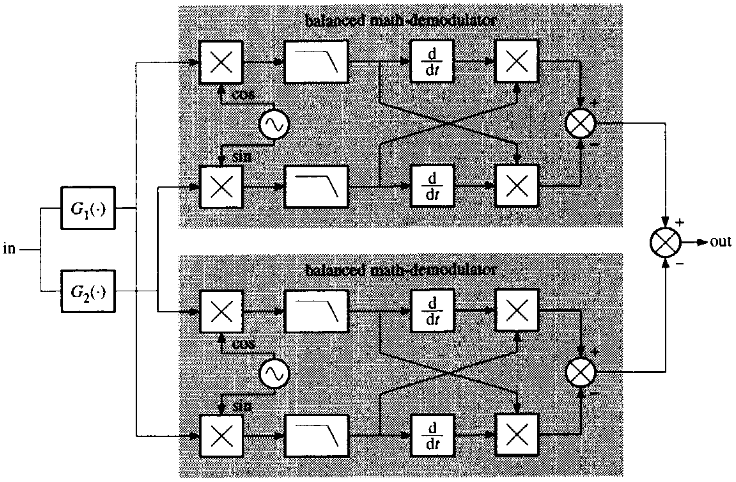


Figure 5.16: FM demodulator with different transfers G_1 and G_2.

The second transfer-type can, besides by the previously mentioned demodulators, be established by the following demodulator classes, that do not employ a reference wave:

- FM-AM conversion demodulators with AM modulus detection;
- FM-PM-AM conversion demodulators.

The third transfer-type, essentially uses only the information contained in the zero crossings of the FM wave, and is realized e.g. by a zero-crossings detector (FM-PM conversion based on a fixed phase-difference).

5.5.2 Frequency Offsets

It was noticed already throughout Chapter 3 and Chapterc:FMdesign that an offset term in the demodulator output signal adversely influences the demodulator dynamic range.

At the upper side of the dynamic range, the offset limits the maximum (distortion free) swing of the output signal, while it simultaneously increases the output noise level at the bottom side when no or only finite amplitude compression is performed. The latter statement follows by application of the theory discussed in Section 5.4.2, and the analysis to be discussed in Section 6.3, with the message signal $\dot{\varphi}(t)$ replaced by $\dot{\varphi}(t) + \omega_{\text{offs}}$. Inspection of (5.18) shows already that any residual amplitude noise in the output signal of the compressors G_1 and G_2 multiplies the offset ω_{offs}, and thus increases the highly undesirable white noise floor at the demodulator output.

The demodulator output SNR above threshold in the presence of ω_{offs}, may be shown to equal

$$\text{SNR}_{\text{out}} = \frac{3p\left(\frac{W_n}{W}\right)\left(\frac{\Delta\omega}{W}\right)^2}{1 + 3C_{n,1}^2(A)\left(\frac{\Delta\omega}{W}\right)^2\left[1 + \left(\frac{\omega_{\text{offs}}}{\Delta\omega}\right)^2\right]}. \tag{5.19}$$

It is clearly shown by this expression that the offset should be kept small compared to the RMS frequency deviation $\Delta\omega$ of the message. Otherwise, it significantly increases the level of the white noise floor, which deteriorates the demodulator output SNR.

5.6 Post Detection Processing

Apart from pre-demodulation signal processing, the performance of the demodulator output signal may be significantly improved by application of post demodulation processing. In general, post demodulation processing establishes a reduction of the demodulator output noise, and possibly of distortion. In this respect, two classes of post processing can be distinguished, that realize different types of noise reductions:

- reduction of continuous demodulator output noise and interference;
- reduction of click noise.

The continuous demodulator output noise and interference is usually reduced by application of linear filtering, including de-emphasis, as discussed in Section 5.6.1.

Reduction of click noise has been attempted by means of click detection, i.e. recognition of clicks from the demodulator output signal, and subsequent elimination. It is an alternative to application of finite compression to the demodulator input wave. As discussed in Section 5.4, finite compression reduces the level of click noise at the expense of increased level of continuous demodulator output noise. Click elimination on the other hand attempts to reduce the level of click noise without such a noise increased, i.e. without establishing a trade-off. This may heuristically explain the limited success of click elimination in practice, which so far has been able to reduce the FM demodulator threshold by a few dB's only. The subject of click detection is considered in Section 5.6.2, while the associated problem of click elimination is considered in Section 5.6.3.

5.6.1 Baseband Frequency Selectivity

Frequency selectivity may be used to improve the demodulator output signal in two different ways:

- prevention of aliasing by high-frequency noise in subsequent systems;
- de-emphasis.

Prevention of Aliasing

For reduction of out of band noise and interference, the baseband (low-pass) filter should possess an asymptotic decay of at least second-order, as illustrated in figure 5.17. Due to the intrinsic differentiation of FM demodulation, the frequency noise at the demodulator output increases with 20 dB per decade, assuming the spectrum of the demodulator input noise is flat. A first order filter would stop this increment outside the message bandwidth, by transforming the

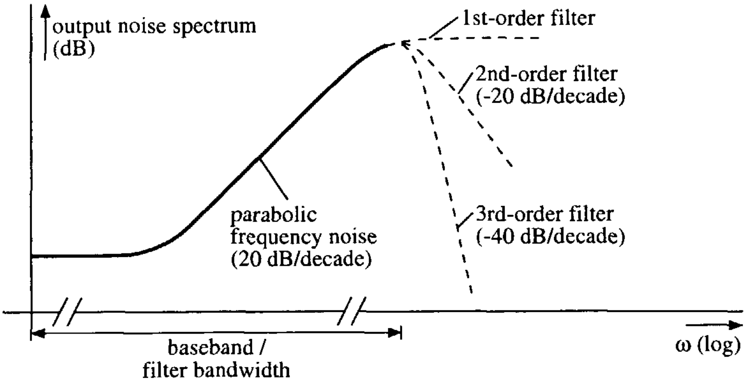

output noise spectrum (dB)
1st-order filter
2nd-order filter (-20 dB/decade)
parabolic frequency noise (20 dB/decade)
3rd-order filter (-40 dB/decade)
baseband / filter bandwidth
ω (log)

Figure 5.17: Reduction of out of band noise and interference by means of baseband filtering.

'triangular' out of band frequency noise into white noise, but does not eliminate it. Second and higher order filters cause a decay of the noise spectral density outside the message bandwidth, and thus eliminate the main portion of the out of band noise and interference.

De-emphasis

De-emphasis techniques suppress the asymptotic 20 dB per decade increase of the frequency noise at the top of the baseband, by application of a priori information on the shape of the pre-distorted message spectrum, as depicted in figure 5.18. A (standardized) pre-emphasis filter in the FM transmitter predistorts the message signal by enhancing its high frequency spectral contents, before it is modulated on a carrier wave. Basically, the differentiating character of this filter causes the high frequency contents of the message to modulate the carrier phase (PM), while the low frequency contents modulates the carrier frequency (FM). The de-emphasis filter at the output of the demodulator in the FM receiver restores the original message spectrum and transforms the triangular noise into white noise by suppression of the high frequency contents, using the knowledge of the pre-emphasis filter transfer.

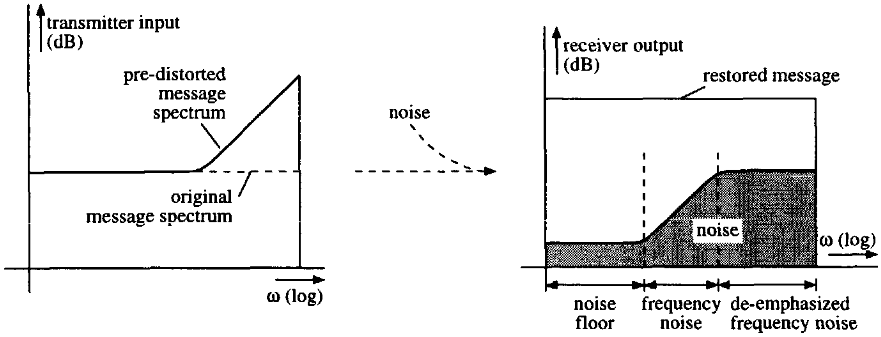


Figure 5.18: Principle of noise reduction by pre- and de-emphasis.

The penalty paid for the application of pre-emphasis is an increased FM demodulator threshold. The pre-emphasis filter enhances the high frequency contents of the message and thereby increases the RMS frequency deviation of the corresponding FM wave. According to Carson's bandwidth formula, expression (2.8), this causes an increase of the FM wave's bandwidth [1]. Consequently, in the presence of noise, the FM receiver input CNR is decreased, and the receiver threshold is shifted towards a lower intensity of the input noise spectral density, corresponding to a higher input CNR with respect to the original bandwidth of the wave in the absence of pre-emphasis.

5.6.2 Click Detection

The concept of click detection is based on the idea that the distinctive, impulsive shape of click pulses should allow their recognition, and subsequent elimination, from the demodulator output signal. The feasibility of this idea is intuitively enhanced by the fact that humans are annoyed by click noise, and thus able to distinct clicks from the message signal and continuous noise. By automation of click recognition and subsequent elimination, it might therefore be possible to shift the threshold towards a lower input CNR, and therewith achieve a significant improvement of the receiver output signal quality.

Automated detection of clicks should at least use information about the shape of click pulses. Several studies have been devised to obtain detailed information about this click pulse shape [39–41]. Their results basically confirm Rice's very simple click model, at least around the demodulator threshold.

A serious problem in click detection, that limits the attainable performance improvement, is the fact that clicks are not the only cause of impulses in the demodulator output signal [42–44]. So called "doublets" produce sharp peaks in the demodulator output signal as well, but do not contribute to the click noise or the FM threshold effect initiated by clicks. The difference between clicks and doublets may be understood from figure 5.19. A click in the instantaneous frequency noise $\dot{\theta}(t)$ corresponds to a trajectory of the noise phasor $\vec{n}$ that encircles the origin, i.e. curve (a) in figure 5.19, and thereby increases the phase noise $\theta(t)$ by an amount 2π. The impulsive shape such a click observed at input CNRs around the threshold, is due to the fact that the carrier amplitude $R(t)$ temporarily becomes very small during the encirclement: the impulse is

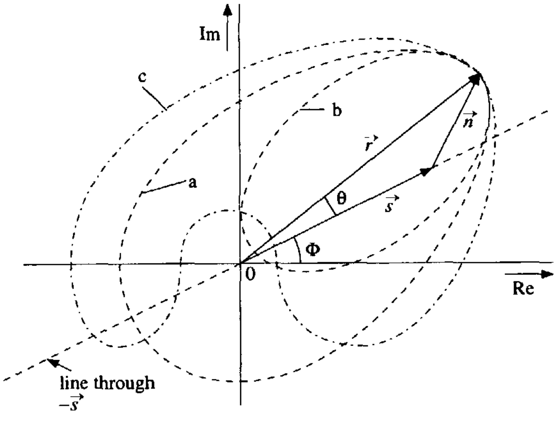


Figure 5.19: Causes of impulses in the demodulator output signal a) a click, b) a doublet, c) false click.

inversely proportional to the squared envelope $R^2(t)$. Since the net area of a click pulse, i.e. the 'DC-component' of its spectrum, equals 2π, the spectral density of click noise contains a significant amount of energy at low frequencies, i.e. inside the message bandwidth. This low frequency spectral contents causes serious deterioration of the demodulator output signal.

A doublet corresponds to a trajectory of $\vec{n}$ that closely approaches the origin of the phasor plane, curve (b) in figure 5.19, resulting in a small value of $R(t)$ and thus a peak in $\dot{\theta}(t)$, but does not encircle it. The net increment of the phase noise $\theta(t)$ during a doublet, the net area of the corresponding doublet-impulse in the instantaneous frequency noise $\dot{\theta}(t)$, therefore equals zero. In fact, a doublet pulse consists of two pulses of equal area and opposite sign, that follow immediately one after each other. The low frequency spectral contents of doublet pulses is therefore insignificant and does hardly affect the message information. The main portion of the doublet-pulse energy is concentrated at relatively high frequencies, outside the message bandwidth, and may easily be removed by means of low-pass filtering.

A special type of doublets, so called "false clicks", correspond to the noise trajectory described by curve (c) in figure 5.19. Although this curve doesn't encircle the origin, and therefore does not contribute to the click noise, it does cross the vector $-\vec{s}$ used by the Rician click model. Consequently, the Rician click model considers this type of doublets as clicks! Fortunately, false clicks do hardly violate the validity of the click model, since they are very rare around the threshold, as a result of the fact that the length of the noise trajectory required for them is relatively large (and thus unlikely to occur). Deep below the threshold however, the probability of false clicks rapidly increases, which causes the Rician click noise model slightly to overestimate the click rate. This inaccuracy of the click model and the limited validity of the Poisson model for the click instants has been used by some authors as explanation for the limited success of click elimination [31, 45]. In [43], however, it was shown that this inaccuracy is not essential in the click detection problem, and hardly limits the attainable threshold extension, i.e. decrement of the input CNR where the threshold occurs; click detectors do not use these properties.

The fundamental problem that complicates the discrimination between clicks and doublets is the fact that the noise free FM wave $s(t)$ is not known by the FM receiver. Only the noisy FM wave $r(t)$ is observed. Therefore, the trajectories (a), (b) and (c) depicted in figure 5.19 cannot be observed directly by the receiver, but have to be derived, as far as possible, from information supplied by the carrier envelope $R(t)$ and the instantaneous frequency $\dot{\varphi}(t)+\dot{\theta}(t)$. In [43, 44], it is shown that reliable click detection, and elimination without introduction of distortion, based on the information contained in the *instantaneous* values of $R(t)$ and $\dot{\varphi}(t) + \dot{\theta}(t)$ is impossible; it follows, that the optimum click detector, supplied with a distortion-free click eliminator that cancels click pulses with pulses of the same area and opposite sign, supplied only with this information should decide to consider any impulse in the demodulator output signal as a doublet, and pass it to the receiver output. Reliable click detection therefore requires memory, in order to use knowledge about the behavior of the envelope and frequency in the (near) past. For example, one might think of a scheme that investigates the spectrum of the peaks in $\dot{\varphi}(t) + \dot{\theta}(t)$, in order to find out if a significant low frequency spectral contents is present (this property distinguishes clicks from doublets).

Due to its complexity, click detection is currently mainly of theoretical interest and not of significant practical interest. Threshold extensions achieved by practical click detectors and theoretical predictions of the attainable threshold extension reported so far have not exceeded 6 dB.

5.6.3 Click Elimination

Click elimination is the operation that removes pulses from the demodulator output signal, once the click detector has decided to consider a particular pulse as a click.

Click elimination algorithms can be divided into two classes [43, 44]:

- elimination by means of cancellation;
- elimination by means of interpolation.

Elimination by Cancellation

Click eliminators based on cancellation exploit the fact that the area of a click pulse equals 2π. These systems add a pulse of 2π and sign opposite to the click pulse to the demodulator output signal shortly after the occurrence of a click. In this way, the click is transformed into some kind of a doublet, with zero net area and a correspondingly insignificant low frequency spectral contents. The advantage of this type of click elimination is that it does not introduce any distortion into the demodulator output signal.

On the other hand however, an extremely reliable click detector, which is has not been found yet, is required, since the compensation pulse produced by the eliminator generates an artificial click at the demodulator output for every "false alarm", i.e. every doublet that is erroneously considered as a click.

Elimination by Interpolation

Click eliminators based on interpolation suppress the demodulator output signal during a click, similar to the muting mechanism in FM demodulators equipped with partial amplitude normalization. Between the start and end of the click pulse, a zero-th order, first-order or even higher order interpolation of the message signal is constructed. Obviously, this type of click eliminator does introduce distortion into the demodulator output, since not only the click, but also the message signal is suppressed during the click. Higher order interpolation may reduce the distortion to an acceptable level.

On the other hand, this scheme is far less susceptible to unreliable click detection. Any false alarm of the click detector causes muting of the demodulator output signal, but does not cause the generation of an artificial click. The selection of a suitable click eliminator is thus associated to a trade-off between requirements on the reliability of the click detector and the amount of distortion introduced by the eliminator into the demodulator output signal.

Further it should be noted that application of finite amplitude compression to the demodulator input wave complicates click elimination by means of cancellation, due to the fact that the area of click pulses in the demodulator output become dependent on the value of the carrier envelope $R(t)$ in that case. Finite compression does not affect click elimination by means of the interpolation, as long as its influence on the reliability of the click detector is negligible, since this approach does not use information about the area of click pulses.

5.7 Adaptive Signal Processing

The performance of FM receivers may be improved significantly by application of adaptive control, possibly by means of feedback, to the various sub-systems. These types of processing use additional a priori information, obtained from the expected properties of the FM wave, and a posteriori information, obtained from the demodulator output signal (adaptive feedback), or the receiver input signal (feed-forward). Obviously, performance improvement is achieved as long as the information is reliable. Unreliable information will somehow result in performance deterioration instead of improvement.

This section outlines the principles of the main types adaptive processing that facilitate improvement of the demodulator performance. Section 5.7.1 considers the improvement facilitated by frequency feedback, i.e. adaptive control of the local oscillator frequency. Section 5.7.2 considers adaption of the RF frequency selectivity, while Section 5.7.3 considers adaptive amplitude compression. Phase feedback is not considered in this section, since it cannot be considered as a performance improvement technique that is applicable to arbitrary types of FM demodulators; it constitutes a particular FM demodulation algorithm.

5.7.1 Frequency Feedback

Frequency feedback is a means to reduce the threshold CNR of an FM demodulator, which exploits the property of (wideband) FM waves that the transmission bandwidth is (considerably) larger than the bandwidth of the intelligence. Alternatively, it may be used to reduce the distortion in the demodulator output signal. Thus, in fact, it provides a trade-off between the demodulator threshold CNR and the distortion in the output signal above the threshold. As explained below, this is essentially the same trade-off as the one encountered in the design of the RF/IF selectivity, considered in Section 5.2.

Principle of Operation

The principle of FM frequency feedback is depicted in figure 5.20. An FM demodulator is enclosed by a feedback loop that reconstructs the (wideband) FM wave at the receiver input from the demodulated FM message at the receiver output by means of an FM modulator. The instantaneous frequency of this regenerated wave $r_r(t)$ is subtracted from that of the received wave $r(t)$. When the demodulator output signal is a reliable copy of the original message information, which is the case above threshold, this subtraction results in an FM wave, denoted by $r_l(t)$, with a considerably reduced RMS frequency deviation.

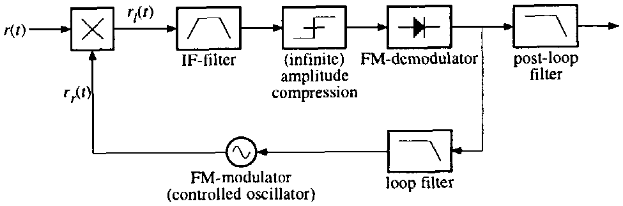


Figure 5.20: Basic FM frequency feedback receiver architecture.

According to Carson's bandwidth formula (2.8), this means that the bandwidth of $r_l(t)$ is significantly smaller than the bandwidth of $r(t)$. Thus, the feedback loop transforms the wideband FM wave $r(t)$ into a narrow-band FM wave $r_l(t)$. This feature is the basis for the threshold/distortion reduction capabilities of frequency feedback receivers.

Threshold Reduction

Reduction of the threshold CNR, i.e. "threshold extension", is achieved by application of narrow band filtering to the compressed wave $r_l(t)$.

Assumed that the receiver input noise is relatively wideband, i.e. of roughly the same bandwidth as the received wideband FM wave $r(t)$, narrow band filtering considerably reduces the noise power level at the input of the FM demodulator. The signal power level is not affected by this filtering, since the narrow-band wave $r_l(t)$ fits entirely within the filter bandwidth. This means that the FM demodulator experiences a larger input CNR than the CNR observed at the receiver input, i.e. in front of the feedback loop.

Notice that this type of threshold reduction reduces the amount of click noise generated in the frequency noise $\dot{\theta}(t)$ itself, instead of affecting the propagation of clicks from the frequency noise to the demodulator output, which is the approach followed by finite compression and click detection/elimination.

At a first glance, one should expect that the FM receiver threshold level is reduced by a factor equal to the ratio of the noise bandwidth of the receiver input noise power, and the noise bandwidth of the IF-filter inside the feedback loop. The actual threshold extension however is smaller than this estimation, due to noise that is fed back from the FM demodulator output to the input, as discussed in detail in Chapter 8.

Further, it should be noted that frequency feedback does not improve the demodulator output SNR above threshold. The same SNR is attained without feedback; only the threshold is reduced. This is a consequence of the fact that feedback reduces the FM message signal and the frequency noise contained in the input FM wave in equal proportions.

The penalty paid for the extension is a steeper, and consequently more "aggressive" threshold behavior, as discussed in Chapter 8.

Distortion Reduction

Instead of threshold reduction, distortion reduction is achieved when the narrow-band FM wave $r_l(t)$ is filtered by a relatively wideband IF filter, and the FM modulator in the feedback path provides a better linear transfer than the FM demodulator.

As a result of the feedback mechanism, the linearity of the feedback demodulator approaches that of the FM modulator in the feedback path if the loop gain is sufficiently large. The wide bandwidth of the IF filter reduces the level of filtering distortion, at the expense of a smaller threshold reduction. This is exactly the same trade-off as discussed in Section 5.2.

5.7.2 Adaption of the RF/IF Frequency Selectivity

Adaption of the RF/IF frequency selectivity, usually implemented by one or several filters, possibly interconnected by mixers, may be employed to improve the rejection of (out-band) noise and interference in the FM demodulator input signal.

As far as this rejection is concerned, basically two different types of selectivity parameters may be controlled adaptively:

- parameters that determine the center frequency;
- parameters that affect the bandwidth of the selectivity.

For convenience, we assume in the sequel of this section that the RF/IF selectivity is realized by a single bandpass filter, although this is never-since a fundamental restriction to the adaption schemes.

Adaption of the Center Frequency

The center-frequency is adaptively controlled in so called *dynamic tracking filters*. The basic architecture of such FM receivers is depicted in figure 5.21. In these receivers, adaptive tuning of the (usually symmetrical) bandpass filter's center-frequency to the instantaneous frequency of the received FM wave, by means of the feedback loop, allows a filter bandwidth that is considerably smaller than the bandwidth of the FM wave itself, without introduction of excessive distortion. This smaller bandwidth increases the FM demodulator input CNR over the FM receiver input CNR and thus reduces the threshold of the FM receiver as a whole. This is the same threshold reduction mechanism, as the one established by means of frequency feedback, discussed in Section 5.7.1.

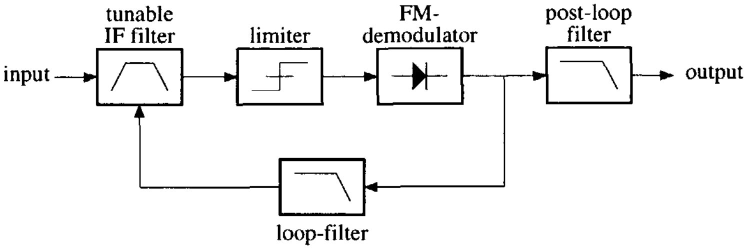


Figure 5.21: Dynamic tracking filter FM receiver.

Its operation may be heuristically understood from figure 5.22, where the quasi-stationary approximation, that represents the FM wave as an impulse that moves along the transmission bandwidth, is applied to model the received FM wave. If the center-frequency of the bandpass filter is fixed to a certain value (no adaption), its bandwidth should be sufficiently large to accommodate all possible positions of the impulse in the spectrum, as depicted in figure 5.22a.

However, when the center-frequency tracks the instantaneous frequency of the FM wave (adaption), as in figure 5.22b, a considerably smaller bandwidth is allowed; only the moving impulse in the FM wave's spectrum should be positioned within the filter bandwidth. It has been recognized that, despite their different implementation, the threshold extension mechanism and the threshold behavior of dynamic tracking filters is basically equal to those exhibited by frequency feedback receivers [46–48].

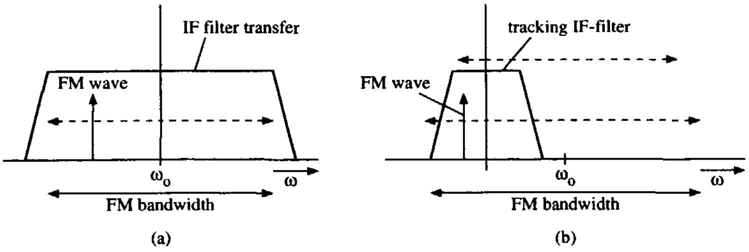


Figure 5.22: Filtering of the FM demodulator input wave a) filter with fixed center-frequency, b) filter with adaptively tuned center-frequency.

The tracking filter basically compresses the received FM wave into a smaller bandwidth by reducing its frequency deviation, like the frequency feedback receiver. This is also observed from figure 5.22b. Here, the spectrum of the compressed FM demodulator input wave $r_l(t)$ is represented by the perturbations of the FM-impulse relative to the center-frequency of the moving IF filter. Due to the tracking mechanism, the average magnitude of these perturbations, i.e. the average frequency deviation, is considerably less than those observed in the input wave, represented by the perturbations of the impulse relative to the center-frequency ω_o.

Adaption of the Bandwidth

Adaption of the filter bandwidth may be used to improve adjacent channel rejection, at the expense of an increased level of narrow-band filtering distortion observed at the demodulator output.

Such schemes reduce the bandwidth of the IF filter when an interfering signal, located at an adjacent channel, exceeds a certain level. In [48, 49] such a scheme is included into a dynamic tracking filter FM receiver by clipping the tuning signal of the tracking filter when it exceeds a certain level. This clipping level is adjusted as function of the interference level in the adjacent channel. As a result, the IF filter tracks the FM wave only in some fraction of FM bandwidth around the center-frequency ω_o, which is basically equivalent, probably except for the threshold behavior, to reducing the bandwidth of a non-adaptive IF filter.

In principle, the information contained in both the demodulator input and output signal can be used to establish the adaptive control scheme. Each of the two inputs to the "adaptive IF filter control" block in figure 5.1 represents one of these possibilities.

5.7.3 Adaptive Amplitude Compression

The trade-off between click noise and continuous noise in the demodulator output signal is besides its dependency on the amplitude compressor characteristic/gain also a function of the receiver input CNR. Optimization of this trade-off to some predefined criterion, over a certain range of input CNRs, can therefore generally not be achieved with an invariant compressor transfer characteristic. Instead, adaption of this characteristic to the receiver input CNR is required.

At high CNRs a high level of compression of is favorable, as a result of the very small amount of click noise observed here, whereas at low CNRs a smaller level of compression is generally favorable, due to the dominance of click noise here.

The information required to control the amplitude compressor, i.e. the demodulator output SNR and input CNR, may be attained in two fundamentally different ways, that are both sketched in figure 5.1:

- by detection of the noise level at the receiver input;
- by detection of the amount of clicks at the demodulator output.

The input noise may be determined e.g. from an adjacent channel.

The click detection approach requires regular transmission of a deterministic test signal, since reliable click detection is impossible if the message signal is unknown (see Section 5.6.2). In existing transmission schemes, such as FM radio broadcasting, this approach is not an option. In newly defined systems however, it may probably be applicable.

5.8 Conclusions

This chapter considered the design of FM receivers, analyzed the various types of processing that can be included into the receiver architecture, and discussed the trade-offs involved with each type of processing.

The main objective of FM receiver design is to maximize the performance of the FM demodulator included in the receiver architecture, by means of various types of pre- and post-demodulation processing. These pre- and post-processing necessarily use a priori or a posteriori information of the characteristics of the FM wave to be demodulated. The processing generally results in an improved performance as long as the applied information is valid. When the information becomes invalid, however, e.g. due to noise, pre- and post processing may eventually degrade the receiver performance; every improvement is somehow paid by deterioration when the improvement mechanism fails.

The general objective of pre-demodulation processing is extraction of the required FM wave from the receiver input signal. Three types of separation operations are available: separation in frequency, phase and amplitude.

Separation in frequency, by means of linear RF/IF filtering, allows discrimination between signals that occupy non-coincident frequency bands. The linear filtering applied to the FM wave was shown to result in nonlinear distortion of the FM message signal. Minimal distortion is attained with Maximum Flat Delay filters, e.g. Bessel filters. The trade-off involved in the design of linear filtering in FM receivers is the one between the demodulator threshold, output SNR at one side, and the distortion in the output signal at the other side.

Separation in phase, by means of cross-coupled PLL structures, allows discrimination between signals with coincident frequency bands, but different instantaneous phases, e.g. co-channel interference. A phase selective FM demodulator, i.e. a phase feedback demodulator, is indispensable for this separation. The disadvantages of phase feedback demodulators, such as the potential danger of loss of lock, are thus inevitable in such structures.

Separation in amplitude, by means of amplitude compression, allows discrimination between signals with the same frequency and phase, but different intensities, e.g. the FM wave and additive noise. Amplitude compression establishes a trade-off between the output SNR observed above threshold, and the "steepness" or "aggressiveness" of the FM threshold at low input CNRs, caused by the generation of impulsive click noise. Infinite compression achieves the highest output SNR above threshold, but yields a maximum level of click noise at low input CNRs. Finite compression establishes a trade-off between both types of noise.

Two characteristics of the FM demodulator architecture are decisive for the performance of the entire FM receiver. In the first place, the degrees of freedom left by the demodulator architecture for optimization of the performance through amplitude compression differ considerably among the various demodulator classes. Secondly, frequency offset components in the demodulator output signal, that cannot be avoided in all architectures, deteriorate the performance.

The general objective of post-demodulation processing is reduction of the FM demodulator output noise. Base-band filtering, including de-emphasis, reduces the continuous demodulator output noise, while click detection and elimination reduces the click noise. Since reliable click detection often requires knowledge of the message signal that is not available, this type of processing should (currently) be considered to be of marginal practical interest.

Frequency feedback and adaptive control of the IF center-frequency, as in dynamic tracking filters, establish a trade-off between reduction of the demodulator threshold, and reduction of the distortion at the demodulator output. Adaption of the IF filter bandwidth establishes a trade-off between distortion and suppression of interference in adjacent channels. Finally, adaption of the amplitude compressor transfer allows optimization of the trade-off between click noise and continuous noise to the input CNR/output SNR.

References

[1] A. Bruce Carlson, *Communication Systems, An introduction to Signals and noise in Electrical Communication*, McGraw-Hill Book Company, Singapore, 1986.

[2] C. Chayavadhanangkur and J.H. Park, "Analysis of FM systems with co-channel interference using a click model", *IEEE Transactions on Communications*, vol. 24, no. 8, pp. 903–910, Aug. 1976.

[3] Toshio Mizuno and Osamu Shimbo, "Response of an FM discriminator in the presence of noise and cochannel interference", *IEEE Transactions on Communications*, vol. 42, no. 11, pp. 3003–3009, Nov. 1994.

[4] Herman J. Blinchikoff and Anatol I. Zverev, *Filtering in the Time and Frequency Domains*, John Wiley and Sons, New York, 1976.

[5] Hans Roder, "Effects of tuned circuits upon a frequency modulated signal", *Proceedings of the IRE*, vol. 25, no. 12, pp. 1617–1647, Dec. 1937.

[6] Edward Bedrosian and Stephen O. Rice, "Distortion and crosstalk of linearly filtered, angle-modulated signals", *Proceedings of the IEEE*, vol. 56, no. 1, pp. 2–13, Jan. 1968.

[7] John R. Carson and Thornton C. Fry, "Variable frequency electric circuit theory with applications to the theory of frequency-modulawtion", *The Bell System Technical Journal*, vol. 16, no. 10, pp. 513–540, Oct. 1937.

[8] F.G.M. Bax, *Analysis of the FM Receiver with Frequency Feedback*, PhD thesis, Catholic University of Nijmegen, Nijmegen, The Netherlands, Oct. 1970.

[9] J.H. Roberts, "FM distortion: A comparison of theory and measurement", *Proceedings of the IEEE*, vol. 57, no. 4, pp. 728–732, Apr. 1969.

[10] Anatol I. Zverev, *Handbook of Filter Synthesis*, John Wiley and Sons, New York, 1967.

[11] C. A. M. Boon, *Design of High-Performance Negative Feedback Oscillators*, PhD thesis, Delft University of Technology, 1989.

[12] A. van Staveren, *Structured Electronic Design of High-Performance Low-Voltage Low-Power References*, PhD thesis, Delft University of Technology, 1997.

[13] Tippure S. Sundresh, Frank A. Cassara, and Harry Schachter, "Maximum a posteriori estimator for suppression of interchannel interference in FM receivers", *IEEE Transactions on Communications*, vol. 25, no. 12, pp. 1480–1485, Dec. 1977.

[14] Harry L. van Trees, *Detection, Estimation, and Modulation Theory-Part I*, John Wiley and Sons, New York, 1968.

[15] Harry L. van Trees, *Detection, Estimation, and Modulation Theory-Part II*, John Wiley and Sons, New York, 1971.

[16] Frank A. Cassara, Harry Schachter, and Gerald H. Simowitz, "Acquisition behaviour of the cross-coupled phase-locked loop FM demodulator", *IEEE Transactions on Communications*, vol. 28, no. 6, pp. 897–904, June 1980.

[17] David A. Rich, Steven Bo, and Frank A. Cassara, “Cochannel FM interference suppression using adaptive notch filters”, *IEEE Transactions on Communications*, vol. 42, no. 7, pp. 2384–2389, July 1994.

[18] Eric A. M. Klumperink, Carlo T. Klein, Bas Rüggeberg, and Ed J. M. van Tuijl, “AM suppression with low AM to PM conversion with the aid of a variable-gain amplifier”, *IEEE Journal of Solid State Circuits*, vol. 31, no. 5, pp. 625–633, May 1996.

[19] Wouter A. Serdijn, *The design of Low-Voltage Low-Power Analog Integrated Circuits and their application in Hearing Instruments*, PhD thesis, Delft University of Technology, 1994.

[20] S. O. Rice, “Noise in FM receivers”, in *Proceedings of the Symposium on Time Series Analysis, Brown University, 1962.* 1963, pp. 395–422, M.Rosenblatt Ed., John Wiley and Sons, New York.

[21] F. L. H. M. Stumpers, “Theory of frequency-modulation noise”, *Proceedings of the IRE*, vol. 36, no. 9, pp. 1081–1092, Sept. 1948.

[22] David Middleton, “On theoretical signal-to-noise ratios in F-M receivers: A comparison with amplitude modulation”, *Journal of Applied Physics*, vol. 20, pp. 334–351, Apr. 1949.

[23] David Middleton, “The spectrum of frequency-modulated waves after reception in random noise-I”, *Quarterly of Applied Mathematics*, vol. vol. VII, no. 2, pp. 129–174, July 1949.

[24] David Middleton, “The spectrum of frequency-modulated waves after reception in random noise-II”, *Quarterly of Applied Mathematics*, vol. vol. VIII, no. 1, pp. 59–80, Apr. 1950.

[25] Nelson M. Blachman, “The demodulation of an F-M carrier and random noise by a limiter and discriminator”, *Journal of Applied Physics*, vol. 20, pp. 38–47, Jan. 1949.

[26] Nelson M. Blachman, “The demodulation of a frequency-modulated carrier and random noise by a discriminator”, *Journal of Applied Physics*, vol. 20, pp. 976–983, Oct. 1949.

[27] S. O. Rice, “Mathematical analysis of random noise-I”, *The Bell System Technical Journal*, vol. 23, pp. 282–332, 1944.

[28] S. O. Rice, “Mathematical analysis of random noise-II”, *The Bell System Technical Journal*, vol. 24, pp. 46–156, 1945.

[29] S. O. Rice, “Statistical properties of a sine wave plus random noise”, *The Bell System Technical Journal*, vol. 27, pp. 109–157, 1948.

[30] John Cohn, “A new approach to the analysis of FM threshold reception”, in *Proceedings of the National Electronics Conference*, 1956, pp. 221–236.

[31] Davras Yavuz and Donald T. Hess, “FM noise and clicks”, *IEEE Transactions on Communication Technology*, vol. COM-17, no. 6, pp. 648–653, Dec. 1969.

[32] J.E. Mazo and Shlomo Shamai (Shitz), "Theory of FM clicks with brownian motion phase noise", *IEEE Transactions on Information Theory*, vol. 38, no. 7, pp. 1022–1030, July 1990.

[33] S.O. Rice, "Distribution of the duration of fades in radio transmission: Gaussian noise model", *The Bell System Technical Journal*, vol. 37, no. 3, pp. 581–635, May 1957.

[34] A.J. Rainal, "Theoretical duration and amplitude of an FM click", *IEEE Transactions on Information Theory*, vol. 26, no. 3, pp. 369–372, May 1980.

[35] E. Bozzoni, G. Marchetti, U. Mengali, and F. Russo, "Probability density of the click duration in an ideal FM discriminator", *IEEE Transactions on Aerospace and Electronic Systems*, vol. 6, pp. 249–252, Mar. 1970.

[36] M.H.L. Kouwenhoven, C.J.M. Verhoeven, and A.H.M. van Roermund, "Frequency demodulator threshold minimization by application of soft-limiters", in *Proceedings of the ProRISC/IEEE Workshop on Circuits, Systems and Signal Processing*, Mierlo, November 27 - 28, 1996, pp. 195 – 200.

[37] M.H.L. Kouwenhoven, C.J.M. Verhoeven, and A.H.M. van Roermund, "A new simple design model for FM demodulators using soft-limiters for click noise suppression", in *Proceedings of the IEEE International Symposium on Circuits and Systems*, Hong Kong, June 9-12, 1997, vol. 1, pp. 265–268.

[38] M.H.L. Kouwenhoven, C.J.M. Verhoeven, and A.H.M. van Roermund, "A new design model for click noise and continuous noise in soft-limiting FM demodulators", in *Proceedings of the 13-th European Conference on Circuit Theory and Design*, Budapest, August 30 - September 3, 1997, vol. 3, pp. 1387–1392.

[39] A. J. Rainal, "Power spectrum of FM clicks", *IEEE Transactions on Information Theory*, vol. 30, no. 1, pp. 122–124, Jan. 1984.

[40] George Lindgren, "On the shape and duration of FM-clicks", *IEEE Transactions on Information Theory*, vol. 29, no. 4, pp. 536–543, July 1983.

[41] George Lindgren, "Shape and duration of clicks in modulated FM transmission", *IEEE Transactions on Information Theory*, vol. 30, no. 5, pp. 728–735, Sept. 1984.

[42] A.J. Rainal, "Optimal detection of FM clicks", *IEEE Transactions on Information Theory*, vol. 28, no. 6, pp. 971–973, Nov. 1982.

[43] M. Polacek, S. Shamai (Shitz), and I. Bar-David, "On FM threshold extension by click noise elimination", *IEEE Transactions on Communications*, vol. 36, no. 3, pp. 375–380, Mar. 1988.

[44] Israel Bar-David and Shlomo Shamai (Shitz), "On the Rice model of noise in FM receivers", *IEEE Transactions on Information Theory*, vol. 34, no. 6, pp. 1406–1419, Nov. 1988.

[45] Davraz Yavuz and Donald T. Hess, "False clicks in FM detection", *IEEE Transactions on Communication Technology*, vol. 18, no. 6, pp. 751–756, Dec. 1970.

[46] J.H. Roberts, "Frequency-feedback receiver as a low-threshold demodulator in FM FDM satellite systems", *Proceedings of the IEE*, vol. 115, no. 11, pp. 1607–1618, Nov. 1968.

[47] J.H. Roberts, "Dynamic tracking filter as a low-threshold demodulator in FM FDM satellite systems", *Proceedings of the IEE*, vol. 115, no. 11, pp. 1597–1606, Nov. 1968.

[48] W. Bijker and W.G. Kasperkovitz, "A top-down design methodology applied to a fully integrated adaptive FM IF system with improved selectivity", in *Proceedings of the European Solid-State Circuits Conference*, Sevilla, Sept. 1993, pp. 53–56.

[49] W. Bijker and W.G. Kasperkovitz, "FM receiver with dynamic intermediate frequency (IF) filter tuning", U.S. Patent 5,404,589, Apr. 1995.

Chapter 6

Amplitude Compression

Section 5.4 showed that compression of the demodulator input carrier amplitude generally improves the output SNR by several tens of dBs at high input CNRs. This improvement is not free. It is paid for at low input CNRs by the generation of click noise that introduces the threshold effect. Due to the concentration of its energy in separate short time slots, this type of noise causes great perceptive annoyance to humans.

The amplitude compressor should be designed such that it realizes a suitable trade-off between the SNR improvement at high CNRs, and click noise generation at low CNRs. Obviously, the choice of this trade-off is entirely dependent on the application. For instance, in wired FM transmission systems, the CNR of the input wave is generally sufficiently large to guarantee demodulator operation above threshold. Such systems should therefore apply infinite amplitude compression in order to establish a maximum SNR improvement. In many types of wireless transmission systems however, such as mobile telephony and car radio, demodulator operation above threshold cannot be guaranteed in many circumstances, even when "threshold extending" demodulators such as PLLs and frequency feedback receivers, discussed in Chapter 7 and 8, are applied. In such applications, some kind of finite compression is generally a better alternative.

Although intuitively obvious, the trade-off associated with finite amplitude compression has never been fully explored in literature. Very little is known other than a few theoretical studies in the late 1940s [1–4]. These investigations did not address the aforementioned trade-off at all due to the fact that suitable mathematical descriptions for click noise were not known until Cohn's publication in 1956 [5], and became widely known only after Rice's publication in 1963 [6].

This chapter develops a general model for the output signal and noise of FM demodulators with an arbitrary type of amplitude compression applied to their input signal. The model is valid above and around the threshold, and incorporates a description of the trade-off between the SNR improvement at high CNRs and the generation of click noise at low CNRs.

An overview of the chapter is as follows. Section 6.1 shows how the various types of amplitude compressors are included in the general model. Section 6.2 outlines the principles of the model and its elaboration. Sections 6.3, 6.4 and 6.5 develop the descriptions for the first and second-order noise, and the click noise respectively. An expression for the demodulator output SNR, and its dependence on the applied type of amplitude compression is considered in Section 6.6. With the aid of the previously developed theory, this section also derives an expression for the optimum amplitude compressor transfer, that maximizes the output SNR. Verification of the model by simulation and measurement is discussed in Section 6.7. The conclusions are given in Section 6.8.

6.1 Amplitude Compressor Modeling

Throughout this chapter and the previous chapter, the amplitude compressor is modeled by its 'transfer' $G(A)$ that describes the relation between the compressor input and output carrier amplitude.

This section considers the derivation of this transfer and the most important properties of amplitude compressor behavior in the presence of noise.

Section 6.1.1 describes the two distinguishable types of amplitude compressors. The quite remarkable behavior of one of both types in the presence of noise, of which the hard-limiter is an important example, are considered in sections 6.1.2 and 6.1.3 respectively. Finally, Section 6.1.4 derives the transfer $G(A)$ of a soft-limiter that is used in the simulations and measurements described in Section 6.7.

6.1.1 Amplitude Compressor Types

Generally, two different classes of amplitude compressors can be distinguished that differ in their use of a priori knowledge of the FM wave characteristics, and in the carrier parameters used to establish the compression. In this respect, we distinguish between application of compression:

- directly to the RF/IF FM wave $r(t) = R(t)\cos[\omega_o t + \varphi(t) + \theta(t)]$;
- to the (LF) carrier amplitude $R(t)$, obtained by AM demodulation.

For both types, separately discussed below, the discussion is limited to the case where the response of the compressor is approximately instantaneous. Only in such cases is it possible to model the compressor by means of an instantaneous transfer $G[R(t)]$. The transfer of non-instantaneous compressors is described by means of a (generally nonlinear) differential/difference equation, resulting in additional cross correlations between the amplitude and frequency noise in the demodulator output.

Compression of the RF/IF Carrier Wave

This class of amplitude compressors directly applies a (nonlinear) compression operation to the FM input carrier intensity $r(t)$, as depicted in figure 6.1, *without explicit detection of the carrier amplitude.* Amplitude compression is therefore established implicitly, through the relation between the carrier intensity $r(t)$ and the carrier amplitude $R(t)$.

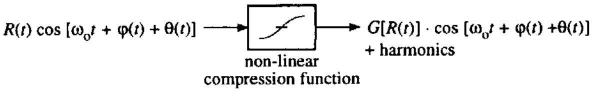

$R(t)\cos[\omega_0 t + \varphi(t) + \theta(t)]$
$G[R(t)]\cdot\cos[\omega_0 t + \varphi(t) + \theta(t)]$
+ harmonics
non-linear
compression function

Figure 6.1: Amplitude compression by nonlinear processing of FM wave intensity.

This simple type of amplitude compressor architecture has two significant disadvantages. In the first place, the output signal of such compressors contains also components at the harmonics, besides a component at the input carrier frequency itself. As discussed in Section 6.1.3, these harmonics basically result in a waste of signal power since it is unfavorable for the output CNR to incorporate them into the compressor output signal. Secondly, instantaneous operation is established only when the compressor bandwidth is at least as large as the input carrier frequency. This is a severe requirement, since the bandwidth of the FM wave and carrier amplitude that should actually be processed is usually considerably smaller than the carrier frequency.

The transfer $G[R(t)]$ for this type of compressor corresponds to the Fourier coefficient of the fundamental frequency, or one of the harmonics. The other harmonics should generally be suppressed, usually by means of a filter, since they merely reduce the compressor output CNR, as considered in Section 6.1.3.

Computation of $G[R(t)]$ requires some additional provisions in comparison with the standard Fourier series expansion. Due to the presence of FM modulation, the input carrier wave $r(t)$ from (2.15) is *not* generally periodic in time, while a Fourier series exists only for periodic signals. However, by rewriting $r(t)$ as

$$r(t) = R(t)\cos\left[\omega_o t + \varphi(t) + \theta(t)\right] = R(t)\cos\Theta(t), \tag{6.1}$$

it is seen that it is periodic in $\Theta(t)$. The time t is included here as an implicit parameter. Thus, when the nonlinear compressor operation is denoted by

$f_{\text{comp}}(.)$, the Fourier series of its response to $r(t)$ may be written as

$$f_{\text{comp}}\left[R(t)\cos\Theta(t)\right] = \sum_{k=0}^{\infty} a_k\left[R(t)\right]\cos k\Theta(t), \tag{6.2}$$

where the coefficients $a_k\left[R(t)\right]$ are given by

$$a_k\left[R(t)\right] \overset{\text{def}}{=} \frac{\varepsilon_k}{2\pi}\int_{-\pi}^{\pi} f_{\text{comp}}\left[R(t)\cos\Theta\right]\cos k\Theta \mathrm{d}\Theta. \tag{6.3}$$

The so called "Neumann factor" ε_k in this expression is defined as [7]

$$\varepsilon_k = \begin{cases} 1 & k = 0, \\ 2 & k \neq 0. \end{cases} \tag{6.4}$$

The amplitude compressor transfer $G\left[R(t)\right]$ used by the demodulator model of equation (5.7) equals one of the coefficients $a_k\left[R(t)\right]$.

Compression of the Carrier Amplitude

This class of amplitude compressor, usually implemented by AGCs, *explicitly* detects the carrier amplitude $R(t)$ by means of an AM demodulator, prior to the nonlinear compression operation, and re-modulates the input carrier wave with the compressed amplitude by means of an AM modulator.

As depicted in figure 6.2, a feed-forward and a feedback variant can be distinguished that differ by the position of the AM detector (positioned at the input and output respectively). The feed-forward architecture in figure 6.2a detects the input carrier amplitude, applies the required compression operation and re-modulates the input wave such that the output amplitude equals the output of the compression function. For this purpose, the output of the compression function should be divided by the original amplitude $R(t)$, prior to re-modulation.

The feed-back architecture in figure 6.2b detects the output carrier amplitude, and compares it with a reference amplitude. This reference is derived from the input carrier amplitude by a separate AM demodulator and the nonlinear compression operation. The error between the output amplitude and the reference approaches zero for large loop gains.

At the cost of an increased complexity, true carrier amplitude compression has two main advantages over intensity compression, discussed previously. In the first place, no signal power is wasted in harmonics since the processing applied to the carrier wave is essentially linear. Secondly, the compressor bandwidth should only exceed the *bandwidth* of the FM wave, instead of its *carrier frequency.* This property has been exploited in [8] to reduce the AM-PM conversion in

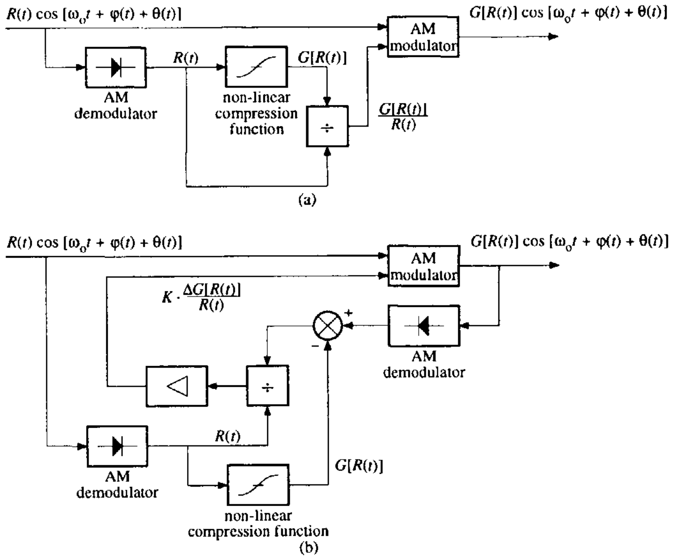

R(t) cos [ω0t + φ(t) + θ(t)]
AM modulator
G[R(t)] cos [ω0t + φ(t) + θ(t)]
AM demodulator
R(t)
G[R(t)]
non-linear compression function
÷
G[R(t)] / R(t)
(a)
R(t) cos [ω0t + φ(t) + θ(t)]
AM modulator
G[R(t)] cos [ω0t + φ(t) + θ(t)]
K · ΔG[R(t)] / R(t)
+
−
AM demodulator
÷
R(t)
G[R(t)]
AM demodulator
non-linear compression function
(b)

Figure 6.2: Amplitude compressor architectures a) feed-forward, b) feedback.

amplitude compressors that results from bandwidth limitations. Usually, AGCs are used to suppress relatively slow variations in the carrier amplitude, while rapid variations are passed unaltered [9, 10]. In an FM demodulator however, both slow and rapid variations should be suppressed, which means that the compressor bandwidth should at least equal the bandwidth of the FM wave.

The discussion in sections 6.1.2 through 6.1.4 is concerned with the non-idealities in the behavior of amplitude compressors based on RF/IF intensity compression only. Since such non-idealities are not, or at least far less important in compressors based on true carrier amplitude compression, this type of compressor is not considered in further detail.

6.1.2 Carrier Suppression in RF/IF Carrier Compressors

Carrier suppression, or signal suppression, is the effect that the compressor output carrier/signal component decreases disproportionately at low input CNRs, such that the compressor output CNR decreases faster than the input CNR. This effect is, in essence, noticeable only in compressors based on RF/IF carrier intensity compression with a discontinuous transfer, such as hard-limiters. In such systems, it reduces the demodulator output SNR [11], and is also responsible for transfer degeneration at low CNRs in hard-limiter based phase detectors, as discussed in Section 7.5.1.

This section explains the cause of the carrier suppression effect in hard-limiters that constitute the operation of many amplitude compressors based on RF/IF carrier intensity compression. An expression for the limiter output signal, subjected to the suppression, is determined and compared with measurement results.

Cause of Carrier Suppression

Carrier suppression is basically due to the effect that noise at the input of discontinuous nonlinearities as hard-limiters linearizes the nonlinear transfer. This linearization causes the harmonics in the output signal to decrease faster than the fundamental frequency, which reduces the gain and the compressor transfer $G(A)$. In A-D converters, this linearization effect is intentionally used to increase the resolution by a technique called "dithering" [12–16].

The origin of the linearization effect can be explained from the behavior of the harmonics $\cos k\,[\omega_o t + \varphi(t) + \theta(t)]$ in the Fourier series expansion of the compressor output signal, given by (6.2). These harmonics are visualized by the phasor representation of figure 6.3. Due to the phase noise $\theta(t)$, generated in response to the input noise $n(t)$, the phasors $\vec{s}_k$ in the output signal deviate from the noise free phase $k\omega_o t + k\varphi(t)$, represented by the horizontal axis in figure 6.3. From a statistical point of view, the compressor output noise therefore represents the uncertainty introduced into the direction of the phasors $\vec{s}_k$ by the input noise $n(t)$. The output carrier/signal component corresponds to the components of $\vec{s}_k$ that point in their average direction, which usually corresponds to the noise-free phase $k\omega_o t + k\varphi(t)$. When the noise induced deviation increases, this component, equal to the average value of $\cos k\theta(t)$, rapidly decreases resulting in carrier suppression.

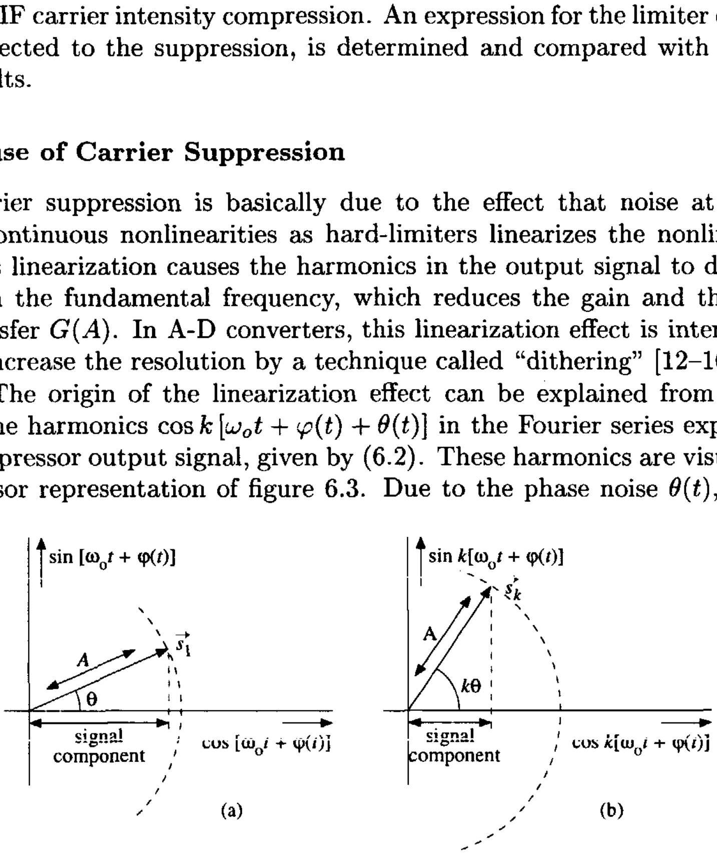


Figure 6.3: Phasor representation of the components in the compressor output signal. a) fundamental harmonic, b) k-th harmonic.

As observed from figure 6.3, for a certain value of the phase noise $\theta(t)$, the phase deviation of k-th harmonic is k times as large as the phase deviation of the fundamental frequency component. Therefore, the carrier/signal component around the k-th harmonic decreases considerably faster than the carrier component located at the fundamental frequency.

Calculation of the Carrier Suppression

An expression for the limiter output carrier/signal component as a function of the input CNR p, which includes a description of the signal suppression effect, is obtained as follows.

According to [17], the previously discussed definition of the output signal component, denoted by $s_o(t)$, corresponds to the expected value of the limiter output for a given input wave $s(t)$. For the Gaussian input noise $n(t)$, with variance σ_n^2, this component equals

$$s_o(t) \overset{\text{def}}{=} \mathrm{E}\left\{\mathrm{sgn}[s(t)+n(t)]|s(t)\right\} = \mathrm{erf}\left[\frac{\sqrt{p}}{A}s(t)\right], \tag{6.5}$$

where sgn(.) denotes the hard-limiter transfer and A denotes the amplitude of $s(t)$. This transfer is plotted in figure 6.4 for various limiter input CNRs as a function of the normalized input signal $s(t)/A = \cos\Phi(t)$. This figure shows that the transfer is considerably linearized at low CNRs, as a result of the fast decay of the harmonics.

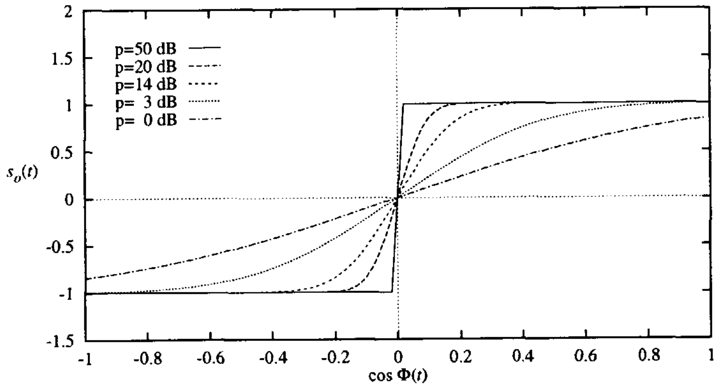


Figure 6.4: Transfer of the normalized limiter input signal $s(t)/A = \cos\Phi(t)$ to output carrier/signal component $s_o(t)$, as a function of the input CNR p.

As elaborated in Appendix A, the Fourier coefficients a_{2k+1} of (6.5), that represent the carrier component located at the $(2k+1)$-th harmonic (the even harmonics equal zero), can be expressed as

$$a_{2k+1} = (-1)^k \frac{2}{2k+1}\sqrt{\frac{p}{\pi}}\exp\left(-\frac{p}{2}\right)\left[\mathrm{I}_k\left(\frac{p}{2}\right)+\mathrm{I}_{k+1}\left(\frac{p}{2}\right)\right], \tag{6.6}$$

where $I_k(.)$ denotes the k-th order Modified Bessel Function of the first kind. At high input CNRs these coefficients approach the coefficients of a square wave, $a_{2k+1} = 4/[\pi(2k+1)]$, due to the fact that the output noise component becomes negligible, and the total limiter output power is contained in the output carrier/signal component. At low input CNRs, the carrier component around the k-th harmonic becomes proportional to p^k. Therefore, the harmonics decrease considerably faster than the fundamental frequency component for decreasing input CNRs.

Measurements

The Fourier coefficients (6.6) were measured as function of the input CNR with the aid of the limiter circuit depicted in figure 6.5. The limiter input was supplied with a sinusoid and additive Gaussian noise. The power contained in carrier component of the limiter output, $a_{2k+1}^2/2$, was determined for various input CNRs according to (6.5) by time-averaging of the output spectrum with the aid of a spectrum analyzer. The measurement results for the fundamental frequency, the third, fifth and seventh harmonic are depicted in figures 6.6 through 6.9, together with the computed curves obtained from (6.6). The measurements and computations match satisfactorily. The discrepancies in the higher harmonics are due to the finite transition gain of the circuit in figure 6.5, whereas the calculations assume an infinite transition gain. Another source of inaccuracies is the very low level of the high order coefficients, which is in the same order of magnitude as the noise floor.

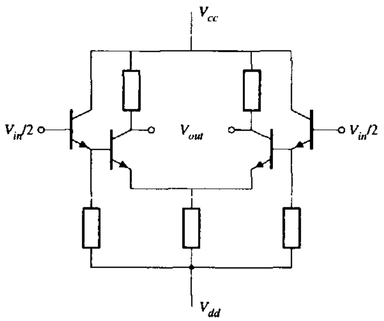


Figure 6.5: Limiter circuit used for the measurement of the coefficients a_k.

As far as the fundamental frequency is concerned, figure 6.6 shows that the carrier suppression is only about 2 dB for CNRs down to 0 dB, and is therefore of minor importance. The harmonics however decay considerably faster, which confirms the linearization of the transfer by the noise. This indicates that it is advantageous to use only the fundamental frequency for demodulation, while

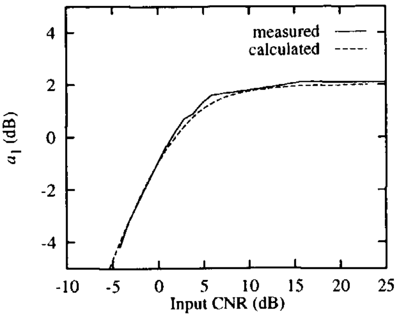


Figure 6.6: Measured and calculated curves for a_1.

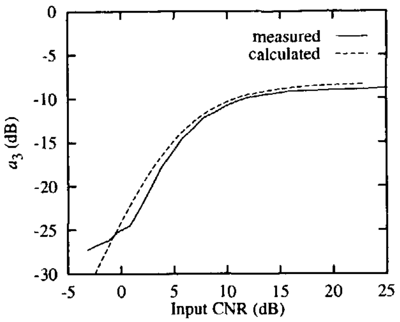


Figure 6.7: Measured and calculated curves for a_3.

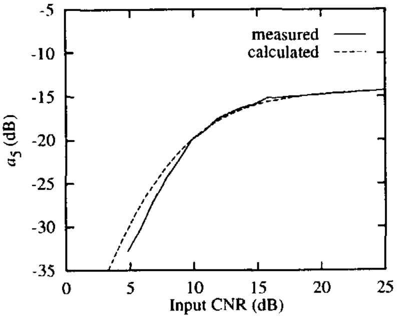


Figure 6.8: Measured and calculated curves for a_5.

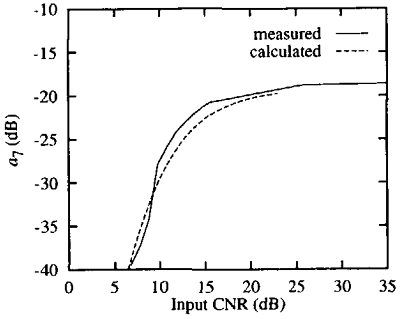


Figure 6.9: Measured and calculated curves for a_7.

the harmonics are suppressed.

6.1.3 Output CNR of RF/IF Carrier Compressors

The compressor output CNR, which equals the FM demodulator input CNR, is an important parameter in FM demodulator design since it determines the demodulator output SNR. Therefore, it is desirable to establish a compressor output CNR that is as large as possible for a given receiver input CNR p.

Section 6.1.1 stated that amplitude compressors based on RF/IF carrier intensity compression should suppress the harmonics in the output signal in order to attain an output CNR that is as large as possible. This section shows the validity of that statement by an investigation of the output CNR of a hard-limiter. The output signal of compressors based on true carrier amplitude compression automatically complies with this statement since it (ideally) does not contain harmonics.

First, an intuitive explanation for the improvement of the output CNR by suppression of harmonics in the output signal is given. Subsequently, a quantitative measure for the CNR improvement is derived. Finally, the limitations on the validity of the analysis are considered.

Output CNR Improvement by Suppression of Harmonics

The possibility of improving the compressor output CNR by suppression of the harmonics stems from the redundancy in the compressor output signal, described by (6.2). Each harmonic contains the same message information, but they are of different qualities.

The CNRs around the harmonics, i.e. the ratio of the carrier component and the noise power located at the individual harmonics, is considerably smaller than the CNR around the fundamental frequency. Inclusion of harmonics in the output signal therefore results in a faster increase of the total noise level than of the total signal level.

These conclusions can be easily derived from the Fourier series (6.2) when the limiter input CNR is assumed to be high, say 10 dB or more, and the phase noise of the k-th harmonic, $k\theta$, is assumed to be considerably smaller than one radian. In that case, the factor $\cos k\Phi(t)$, contained in the k-th harmonic of the limiter output, can be written as

$$\begin{aligned} \cos k\Phi(t) &= \cos k\left[\omega_o t + \varphi(t) + \theta(t)\right] \\ &= \cos k\theta(t) \cos k\left[\omega_o t + \varphi(t)\right] - \sin k\theta(t) \sin k\left[\omega_o t + \varphi(t)\right] \\ &\approx \cos k\left[\omega_o t + \varphi(t)\right] - k\theta(t) \sin k\left[\omega_o t + \varphi(t)\right]. \end{aligned} \qquad (6.7)$$

The first term in the final approximation, $\cos k\left[\omega_o t + \varphi(t)\right]$, represents the noise-free FM carrier at the k-th harmonic of the amplitude compressor output. The second term, $k\theta(t) \sin k\left[\omega_o t + \varphi(t)\right]$, represents the noise observed at the k-th harmonic of the compressor output, as long as $k\theta(t) \ll 1$ (rad).

Thus, this expression shows that the (phase) noise power level in $\cos k\Phi(t)$ increases proportionally to the square of the index k, while the carrier power level is equal for all k. Consequently, the CNR around the k-th harmonic of the compressor output decreases proportionally to k^2. This agrees with the observation in Section 6.1.2 that the uncertainty introduced into the k-th harmonic component by the phase noise $\theta(t)$ increases with k. For very large values of k, however, the uncertainty increase slows down and (6.7) becomes invalid.

According to (6.2), the k-th harmonic component of the square-wave limiter output signal equals the product of $\cos k\Phi(t)$ and the Fourier-coefficient a_k. Since the coefficients a_k of a square-wave are inversely proportional to k, the signal power and the noise power contained in $\cos k\Phi(t)$ are both multiplied by

$1/k^2$. Consequently, the (noise-free) carrier power located at the k-th harmonic of the limiter output signal decreases inversely proportionally to k^2, while the noise level is independent of k. This is schematically depicted in figure 6.10. This

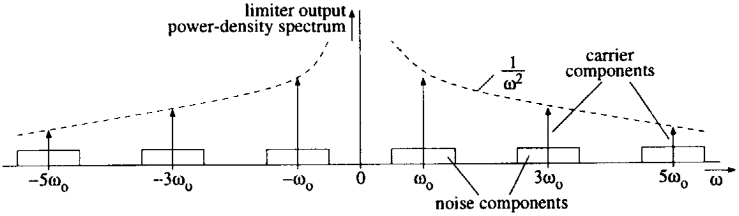


Figure 6.10: Limiter output carrier and noise components.

figure clearly shows that the fundamental frequency possesses the largest CNR. Including harmonics into the output signal decreases the output CNR, since the total noise power increases faster than the total signal power. The precise input CNR experienced by the FM demodulator depends on the type of processing applied to the limiter output wave prior to demodulation, as discussed below.

Calculation of the CNR Improvement

Determination of the CNR improvement established by suppression of the harmonics in the limiter output signal, requires a proper definition of this CNR. In turn, such a definition requires

- identification of 'signal' and 'noise' components at the harmonics;
- determination of the correlation between the harmonics.

The signal components were already identified in Section 6.1.2. The noise power located around each harmonic follows from the property of hard-limiters that the limiter output square-wave, expressed as a function of the instantaneous phase of the input wave $\Phi(t)$ with the time included as an implicit parameter, remains unchanged for all limiter input CNRs. Consequently, the Fourier-coefficients of this square-wave, and the corresponding distribution of the total demodulator output power contained in its harmonic components $4/[\pi(2k+1)]\cos(2k+1)\Phi$, remains unchanged for all CNRs. The noise power located around the k-th harmonic therefore equals the difference between the total power around the k-th harmonic, i.e. $8/[\pi(k)]^2$, and the power contained in the signal-component around the k-th harmonic, which equals $a_k^2/2$ in terms of the Fourier coefficients a_k from (6.6).

The correlation between the contributions of the harmonics to the demodulator input signal is determined by the type of processing applied to the limiter output prior to demodulation. As an upper- and lower bound, we consider the compressor output CNR for the case when no correlation is present and the case when full correlation is present. In other cases, when there is partial correlation, the compressor output CNR will generally attain a value between the CNRs obtained for both these limiting cases.

Uncorrelated Addition An uncorrelated addition of the harmonic components, i.e. an addition of signal and noise *power* components, is established for example by a quadrature demodulator consisting of an 'ideal' all-pass filter of an infinite bandwidth, realizing the time-delay, and a multiplier phase-detector. The response of such a demodulator to a hard-limited, noisy FM square wave equals the *addition of* the *carrier power* and the *noise power* located at all harmonics. This is an uncorrelated addition, since it applies to signal and noise *power* components.

Mathematically, the effective compressor output CNR, or, equivalently, demodulator input CNR that results from uncorrelated addition, denoted by CNR_{uc}, equals the total compressor output signal power divided by the total compressor output noise power, i.e.

$$\text{CNR}_{\text{uc}} = \frac{\sum_{k=0}^{\infty} a_{2k+1}^2}{\sum_{k=0}^{\infty} \frac{16}{\pi^2(2k+1)^2} - a_{2k+1}^2}, \tag{6.8}$$

where a_{2k+1} is given by (6.6).

Fully Correlated Addition The other, merely theoretical possibility of a fully correlated addition, occurs when the compressor output signal component equals the sum of the Fourier coefficients a_k and the noise equals the correlated addition of the noise located around all harmonics. The corresponding effective input CNR, denoted by CNR_{fc}, equals

$$\text{CNR}_{\text{fc}} = \left(\frac{\sum_{k=0}^{\infty} |a_{2k+1}|}{\sum_{k=0}^{\infty} \sqrt{\frac{16}{\pi^2(2k+1)^2} - a_{2k+1}^2}} \right)^2 . \tag{6.9}$$

Notice that this expression is equivalent to (6.8) when only the fundamental frequency is used and all other components are suppressed.

Comparison Figure 6.11 depicts the ratio of the compressor output CNR and the compressor/receiver input CNR for a compressor that uses only the fundamental frequency component, and compressors that additionally use the third harmonic in a fully correlated and an uncorrelated fashion. Figure 6.12 depicts the same ratio for compressors that use all harmonics in a fully correlated

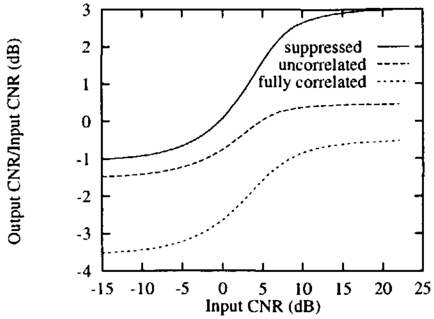


Figure 6.11: CNR improvement for inclusion of the third harmonic.

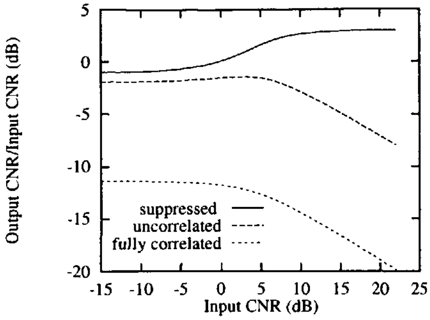


Figure 6.12: CNR improvement for inclusion of all harmonics.

and an uncorrelated fashion. These figures show that correlated addition of the harmonics yields the worst possible output CNR, and should therefore be avoided. Uncorrelated addition yields a higher output CNR, but is still worse than the output CNR obtained by suppression of all harmonics. The curve for the latter case, identical to the one obtained in [18], yields an output CNR that is 3 dB higher than the input CNR at low noise levels due to the fact that the amplitude noise (in-phase noise) is suppressed completely. Finally, it is observed that the deterioration of the output CNR increases when more harmonics are included.

The conclusion is, therefore, that only the fundamental frequency component should be used for demodulation. In many types of quadrature demodulators, for example, this is established automatically when the time delay is realized by a high-Q bandpass filter [19].

Validity of the Analysis

The result obtained in this section should be interpreted with caution. This is due to the fact that the analysis accounts only for noise located at the input of the limiter and not for noise produced by the limiter and demodulator circuits themselves. When the circuit noise becomes dominant, it might be more convenient to include (some of) the harmonics in the output signal. This maximizes the slope of the output wave at the zero crossings and in this way decreases the contribution of circuit noise to the carrier phase noise.

Further, the signal and noise components around the various harmonics can be combined in a variety of different ways, besides addition, that yield somewhat different results. An example is the transfer degradation in phase detectors, considered in Section 7.5.1.

6.1.4 Soft-Limiter Transfer

As an example of the computation of the amplitude compressor transfer $G(R)$, this section determines the first harmonic response of the soft-limiter depicted in figure 6.13a. This limiter, an extremely simplified model of an electronic soft-limiter, is frequently used in subsequent sections of this chapter to supply the theory with numerical results.

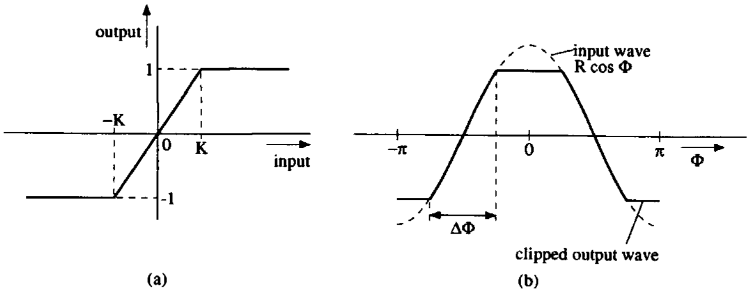


Figure 6.13: Soft-limiter transfer characteristic.

Figure 6.13b depicts the soft-limiter response to the sinusoidal wave $R\cos\Phi$, where the time t is included in $R(t)$ and $\Phi(t)$ as an implicit parameter. Five different regions can be identified in the output signal when the phase increases through the interval $[-\pi, \pi]$; three saturation regions and two linear transition regions of width $\Delta\Phi$. Inspection of (6.3) for the coefficient of the first harmonic shows that the contributions of the saturation regions cancel each other out. This was to be expected due to the anti-symmetry of the limiter transfer and the zero-mean of the input wave. Integration over the two linear transition regions then yields for the amplitude transfer $G(R)$

$$G_{\rm sl}\left[R(t)\right] \overset{\rm def}{=} a_1\left[R(t)\right] =$$

$$\frac{2}{\pi}\int_{\frac{\pi}{2}-\frac{\Delta\Phi}{2}}^{\frac{\pi}{2}+\frac{\Delta\Phi}{2}} \frac{R}{K}\cos^2\Phi d\Phi = \frac{R\Delta\Phi}{\pi K}\left(1+\frac{\sin\Delta\Phi}{\Delta\Phi}\right). \quad (6.10)$$

By inspection it is found that the phase interval $\Delta\Phi = 2\arcsin\left(\frac{K}{R}\right)$. Substitution into (6.10) then yields the following "amplitude compression" transfer for the soft-limiter

$$G_{\rm sl}\left[R(t)\right] =$$

$$\begin{cases} \frac{R(t)}{K}, & |R(t)| \leq K \\ \frac{2}{\pi}\left\{\frac{R(t)}{K}\arcsin\left[\frac{K}{R(t)}\right] + \sqrt{1-\left[\frac{K}{R(t)}\right]^2}\right\}, & |R(t)| > K. \end{cases} \quad (6.11)$$

This transfer is plotted in figure 6.14 as a function of the normalized amplitude $R(t)/K$. The transfer is linear for small values of the amplitude, while it saturates and suppresses amplitude noise for large values.

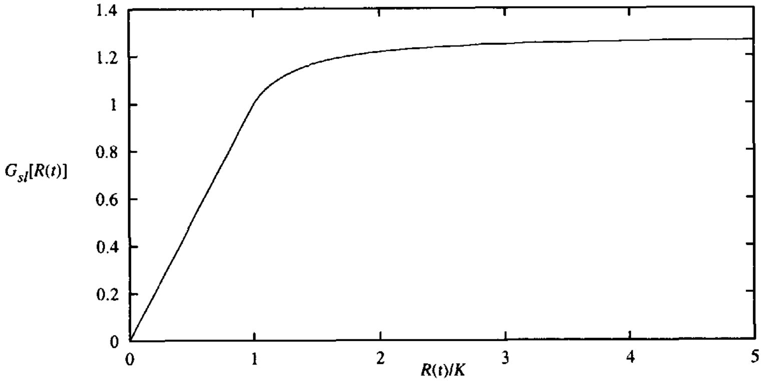


Figure 6.14: First harmonic response of the soft limiter as a function of the normalized input amplitude $R(t)/K$.

The the discussion of the simulation and measurement results given in Section 6.7, frequently uses the inverse first-order compression factor $C_{n,1}(A)$ of an FM demodulator that applies soft-limiting to both the input FM wave $s(t)$, and to a reference wave derived from $s(t)$ that is used during demodulation. Thus, according to (5.7), both the compressor transfers $G_1\left[R(t)\right]$ and $G_2\left[R(t)\right]$ equal the soft-limiter transfer $G_{\text{sl}}\left[R(t)\right]$, such that $G\left[R(t)\right] = G_1\left[R(t)\right] G_2\left[R(t)\right] = G_{\text{sl}}^2\left[R(t)\right]$. The inverse first-order compression transfer corresponding to this transfer equals

$$C_{n,1,\text{sl}}(x) = \begin{cases} 2, & x \geq 1, \\ 2\dfrac{\arcsin^2(x)+x^2\left(x^2-1\right)}{\arcsin^2(x)+2x\sqrt{1-x^2}\arcsin(x)-x^2\left(x^2-1\right)}, & x < 1, \end{cases} \tag{6.12}$$

where $x = K/A$ denotes the inverse of the so called limiter over-drive factor [20].

6.2 Approach to Output Noise Calculation

This section outlines the principles and the approach that is followed in sections 6.3 and 6.4 in order to calculate the first-order and second-order continuous demodulator output noise power spectral density and SNR. A slightly more comprehensive approach is required for the calculation of the click noise, as discussed in Section 6.5.

The general block schematic of the FM receiver architecture that is considered throughout this chapter, including the various signals inside the receiver, is depicted in figure 6.15. The amplitude compressor and FM demodulator in this figure are arranged in such a way that the FM demodulator response $y_{\text{dem}}(t)$ to the noisy compressor input FM wave $r(t)$ becomes

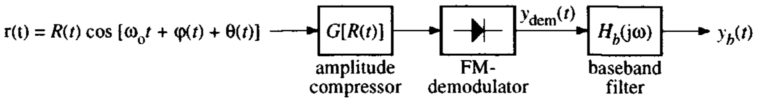


Figure 6.15: Position of the various signals in the FM receiver.

$$y_{\text{dem}}(t) = G\left[R(t)\right]\left[\dot{\varphi}(t) + \dot{\theta}(t)\right]. \tag{6.13}$$

The compressor transfer $G(.)$ in this expression equals the product of two compression functions, $G_1(.)G_2(.)$, in expression (5.7), that correspond to the compression applied to the FM wave subjected to demodulation, and the reference wave respectively. The transfer $H_b(\mathrm{j}\omega)$ represents the baseband (low pass) filter at the FM demodulator output.

In order to obtain the FM receiver output SNR as a function of the receiver input CNR, the statistical properties of the receiver output signal have to be expressed in terms of the (known) statistical properties of the receiver input signal, i.e. the statistics of the message $\dot{\varphi}(t)$ and the noise $n(t)$. As schematically depicted in figure 6.16, this calculation basically requires four steps. The procedure starts from expression (6.13). In the first step of the calculation, discussed in Section 6.2.1, the demodulator output signal $y_{\text{dem}}(t)$ from (6.13) is expressed in terms of the message signal $\dot{\varphi}(t)$ and the in-phase and quadrature components $n_{s,i}(t)$ and $n_{s,q}(t)$ of the input noise $n(t)$. Further, the first and second-order noise components, corresponding to the first and second-order terms of a Taylor series of the output signal are identified.

In the second step, discussed in Section 6.2.2, the autocorrelation function of the output signal is expressed in terms of known correlation functions of the input signals $\dot{\varphi}(t)$, $n_{s,i}(t)$ and $n_{s,q}(t)$.

Subsequently, in the third step, the autocorrelation function of $y_{\text{dem}}(t)$ is used to obtain the power spectral density, as discussed in Section 6.2.3.

Finally, this spectral density is used to obtain the power contents of the receiver output signal $y_b(t)$ and the output SNR, as described in Section 6.2.4.

6.2.1 Time-Domain Expression for the Output Signal

This section discusses the first step in the calculation of the receiver output SNR: determination of an expression for the demodulator output signal $y_{\text{dem}}(t)$

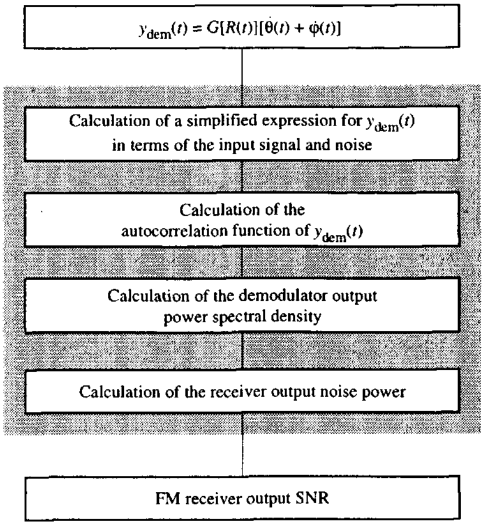

$y_{\rm dem}(t) = G[R(t)][\dot{\theta}(t) + \dot{\varphi}(t)]$
Calculation of a simplified expression for $y_{\rm dem}(t)$ in terms of the input signal and noise
Calculation of the autocorrelation function of $y_{\rm dem}(t)$
Calculation of the demodulator output power spectral density
Calculation of the receiver output noise power
FM receiver output SNR

Figure 6.16: Outline of the procedure required to calculate the receiver output SNR as a function of the input CNR.

that is suitable for calculation of the autocorrelation function. As outlined in figure 6.17, the derivation of such an expression for $y_{\rm dem}(t)$ consists of two steps. First the general expression for $y_{\rm dem}(t)$ in terms of the FM message signal $\dot{\varphi}(t)$, the noise components $n_{s,i}(t)$, $n_{s,q}(t)$, and derivatives has to be determined. Subsequently, since this general expression is too complex for direct use, (general) expressions for the first-order and second-order demodulator output noise components have to be derived by means of a Taylor series. The detailed calculation of the first-order and second-order terms is considered in sections 6.3.1 and 6.4.1 respectively.

General Expression

A general expression for $y_{\rm dem}(t)$ in terms of $\dot{\varphi}(t)$ and the components of $n(t)$ is obtained by substitution of expression (2.16) and (2.19), that express $R(t)$ and $\dot{\theta}(t)$ in terms of $n_{s,i}(t)$ and $n_{s,q}(t)$, into (6.13). The result of this substitution is a rather complicated nonlinear expression. In [1–4, 21] the demodulator output noise power spectral density was calculated exactly from this expression by means of the so called "transform method" [22, 23] for a few special cases of $G(R)$. However, the result is an extremely complicated expression that includes a badly converging series.

The approach followed in this chapter is in a way similar to the transform method, but circumvents badly converging series expansions. Further reduction

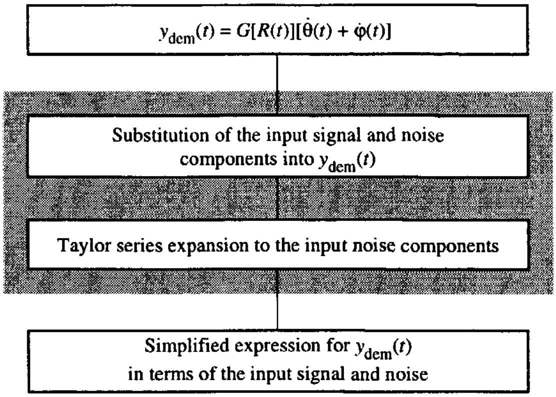

$y_{\text{dem}}(t) = G[R(t)][\dot{\theta}(t) + \dot{\varphi}(t)]$
Substitution of the input signal and noise components into $y_{\text{dem}}(t)$
Taylor series expansion to the input noise components
Simplified expression for $y_{\text{dem}}(t)$ in terms of the input signal and noise

Figure 6.17: Outline of the derivation of a suitable expression for $y_{\text{dem}}(t)$.

of the complexity is possible due to the fact that the main interest is in CNRs of 0 dB and above, which allows all but a few dominant noise contributions to be ignored: the first-order and second-order noise components.

Taylor Series

For input CNRs of 0 dB and larger, the input noise and its components can be considered to be small compared to the FM wave $s(t)$. A proper approximation of the continuous demodulator output noise is then obtained by expansion of $y_{\text{dem}}(t)$ into a Taylor series to the four noise components $n_1 = n_{s,i}(t)$, $n_2 = n_{s,q}(t)$, $n_3 = \dot{n}_{s,i}(t)$ and $n_4 = \dot{n}_{s,q}(t)$. By writing these four components as a noise vector $\underline{n}$, given by

$$\underline{n} = (n_1, n_2, n_3, n_4) = (n_{s,i}, n_{s,q}, \dot{n}_{s,i}, \dot{n}_{s,q}) . \tag{6.14}$$

the four dimensional Taylor series of $y_{\text{dem}}(t)$ to $\underline{n}$ becomes

$$y_{\text{dem}}(t) = y_{\text{dem}}\left[\underline{n}(t)\right] = \sum_{k=0}^{\infty} \frac{1}{k!} \left(\underline{n} \cdot \nabla\right)^k y_{\text{dem}}\left(\underline{u}\right)\Big|_{\underline{u}=\underline{0}} , \tag{6.15}$$

where ∇ is the vectorial differentiation operator

$$\nabla = \left(\frac{\partial}{\partial u_1}, \frac{\partial}{\partial u_2}, \frac{\partial}{\partial u_3}, \frac{\partial}{\partial u_4}\right) . \tag{6.16}$$

The first-order output noise terms, denoted by $y_{\text{dem},1}(t)$, follow from (6.15) for $k = 1$, i.e.

$$y_{\text{dem},1}(t) = \underline{n} \cdot \nabla y_{\text{dem}}\left(\underline{u}\right)\big|_{\underline{u}=\underline{0}} . \tag{6.17}$$

The second-order output noise terms, denoted by $y_{\text{dem},2}(t)$, follow from (6.15) for $k = 2$, i.e.

$$y_{\text{dem},2}(t) = \left(\underline{n} \cdot \nabla\right)^2 y_{\text{dem}}\left(\underline{u}\right)\Big|_{\underline{u}=\underline{0}} . \tag{6.18}$$

Both expressions are expanded in further detail in sections 6.3.1 and 6.4.1 respectively.

The noise power contained in the k-th order terms of (6.15) decreases inversely proportional, to the k-th power of the input CNR p. Therefore, terms of the third order and higher are hardly of interest above $p = 0$ dB and are consequently ignored. The series expansion is therefore terminated after the terms for $k = 2$.

6.2.2 Autocorrelation Function

The second, and computationally most comprehensive step is the determination of the autocorrelation function of the demodulator output signal expressed in terms of the correlation of the message signal $\dot{\varphi}(t)$, and the input noise components $n_{s,i}(t)$ and $n_{s,q}(t)$. The general derivation of such expressions is outlined in this section. The actual computation of the first-order and second-order correlation functions is considered in sections 6.3.2 and 6.4.2 respectively.

According to the definition, this correlation equals

$$R_{\text{dem}}(\tau) \overset{\text{def}}{=} \mathrm{E}\left[y_{\text{dem}}(t+\tau)y_{\text{dem}}(t)\right]_{n_1,n_2,n_3,n_4} . \tag{6.19}$$

where the expectations over n_1 through n_4 are denoted explicitly as a subscript.

As illustrated by figure 6.18, the calculation of the autocorrelation function $R_{\text{dem}}(\tau)$ of the demodulator output signal consists of three steps. First, the noise components $n_{s,i}(t)$ and $n_{s,q}(t)$ have to be substituted according to (2.13) and (2.14) with their truly Gaussian counterparts $n_i(t)$ and $n_q(t)$. Subsequently, this substitution allows the use of some very useful theorems for Gaussian noise, discussed below, in order to determine the correlation functions of the Gaussian components in the output signal. Finally, the correlation functions of the remaining components of the output, those corresponding to the FM message $\dot{\varphi}(t)$, are determined.

Simplification by Theorems for Gaussian Noise

The auto correlation $R_{\text{dem}}(\tau)$ contains many products of Gaussian noise processes. The following two theorems, adopted from [7], therefore allow considerable simplification of the calculations.

Theorem 1 *If $n_1, \ldots, n_{2k+1}$ denote an odd number of zero-mean Gaussian random variables, not necessarily independent, then*

$$\mathrm{E}\left(n_1 \ldots n_{2k+1}\right) = 0. \tag{6.20}$$

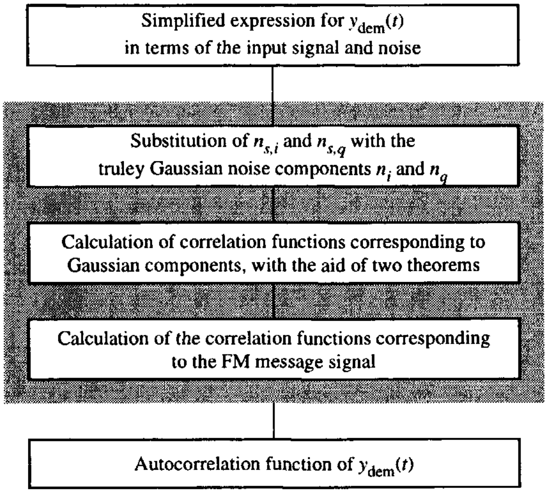

Simplified expression for $y_{\mathrm{dem}}(t)$ in terms of the input signal and noise
Substitution of $n_{s,i}$ and $n_{s,q}$ with the truley Gaussian noise components n_i and n_q
Calculation of correlation functions corresponding to Gaussian components, with the aid of two theorems
Calculation of the correlation functions corresponding to the FM message signal
Autocorrelation function of $y_{\mathrm{dem}}(t)$

Figure 6.18: Calculation of the autocorrelation function of the demodulator output signal.

Theorem 2 *If $n_1, \ldots, n_{2k}$ denote an even number of zero-mean Gaussian random variables, not necessarily uncorrelated, then*

$$\mathrm{E}(n_1 \ldots n_{2k}) = \sum_{\text{all pairs}} \prod_{\substack{i \neq j \\ 1 \leq i,j \leq 2k}}^{k} \mathrm{E}(n_i n_j). \tag{6.21}$$

Thus, for instance, when $k = 2$ theorem 2 states that

$$\begin{aligned} \mathrm{E}(n_1 n_2 n_3 n_4) &= \mathrm{E}(n_1 n_2)\,\mathrm{E}(n_3 n_4) \\ &+ \mathrm{E}(n_1 n_3)\,\mathrm{E}(n_2 n_4) \\ &+ \mathrm{E}(n_1 n_4)\,\mathrm{E}(n_2 n_3). \end{aligned} \tag{6.22}$$

The difference between the results for products of odd numbers and even numbers of Gaussian random variables can be explained by the fact that expansion of a product of an odd number of variables yields terms that consist of (several) factors $\mathrm{E}(n_i n_j)$, and one factor $\mathrm{E}(n_i)$. The latter expectation equals zero since all variables n_i are assumed to be zero-mean. Consequently, for an odd number of variables, all terms in the expansion of their product equal zero. Expression (6.21) shows that the expectation of products of Gaussian random variables can be expressed in terms of the (cross-)correlations of all possible permutations of pairs of these variables.

Application to the FM Demodulator Output Signal

The two theorems save a lot of dull and elaborate calculus, with an inherently high probability of mistakes. However, it should be stated explicitly that they are only valid for true Gaussian random variables. In the presence of modulation, the noise components $n_{s,i}(t)$, $n_{s,q}(t)$ and their derivatives are not truly Gaussian since they depend on $\varphi(t)$. Therefore, in order to be allowed to use theorem 1 and theorem 2, we first rewrite the demodulator output signal in terms of truly Gaussian components. Subsequently, the correlation functions of the remaining components are determined.

In order to attain an expression for $y_{\text{dem}}(t)$ that allows application of theorem 1 and 2, the noise components $n_{s,i}(t)$, $n_{s,q}(t)$ and their derivatives have to be expressed in terms of the noise components $n_i(t)$, $n_q(t)$, the components of $\vec{n}$ on the real and imaginary axis, $\varphi(t)$, and their derivatives, according to expression (2.13) and (2.14). The noise processes $n_i(t)$, $n_q(t)$, and their derivatives are truly Gaussian whenever $n(t)$ is, since they are independent of $\varphi(t)$.

Further, the calculation of the autocorrelation function $R_{\text{dem}}(\tau)$ should be split into two consecutive phases, as expressed by

$$\begin{aligned} R_{dem}(\tau) &\overset{\text{def}}{=} \mathrm{E}\left[y_{\text{dem}}(t)y_{\text{dem}}(t+\tau)\right]_{n_i,n_q,\dot{n}_i,\dot{n}_q,\varphi,\dot{\varphi}} \\ &= \mathrm{E}\left\{\mathrm{E}\left[y_{\text{dem}}(t)y_{\text{dem}}(t+\tau)\right|\varphi,\dot{\varphi}]_{n_i,n_q,\dot{n}_i,\dot{n}_q}\right\}_{\varphi,\dot{\varphi}}. \end{aligned} \tag{6.23}$$

The inner expectation in this expression can be determined with the aforementioned theorems. This expectation represents correlation functions of the noise $n(t)$, in which $\varphi(t)$ and $\dot{\varphi}(t)$ are considered as 'deterministic signals'. In that case, the argument of the expectation contains only Gaussian random variables, and no other random variables, for which the theorems hold. The result consists of correlation functions of the Gaussian noise components, multiplied by functions of $\varphi(t)$ and $\dot{\varphi}(t)$.

Subsequently, the outer expectation is calculated over the generally non-Gaussian random variables φ and $\dot{\varphi}$. Due to the Taylor series expansion of $y_{\text{dem}}(t)$, the resulting expression will consist of products of correlation functions that solely depend on $n(t)$ (determined in the first step), and correlation functions that solely depend on $\dot{\varphi}(t)$ and $\varphi(t)$.

In Section 6.3.2 and Section 6.4.2, this approach is applied to determine the autocorrelation functions of the first and second-order noise respectively, and their cross-correlation. The autocorrelation of $y_{\text{dem}}(t)$ equals the addition of these correlation functions.

6.2.3 Power Spectral Density

The third step is the determination of the power spectral density. According to the Wiener-Khintchine theorem [24, 25], the power spectral density of the demodulator output signal, denoted by $S_{\mathrm{dem}}(\omega)$, equals the Fourier transform of the autocorrelation function,

$$S_{\mathrm{dem}}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R_{\mathrm{dem}}(\tau) \exp(-\mathrm{j}\omega\tau) \mathrm{d}\tau, \tag{6.24}$$

where ω represents the radial frequency in (rad/s). The autocorrelation function $R_{\mathrm{dem}}(\tau)$ generally consists of products of one correlation function that depends only on the input noise $n(t)$, and another correlation function that depends only on the modulation $\varphi(t)$. The demodulator output spectral density therefore consists of convolutions of the spectra of these components.

The calculation of $S_{\mathrm{dem}}(\omega)$ therefore generally consists of two steps, as illustrated by figure 6.19. First, the products of autocorrelation functions are transformed into convolutions of the corresponding spectra. Subsequently, these convolutions are calculated.

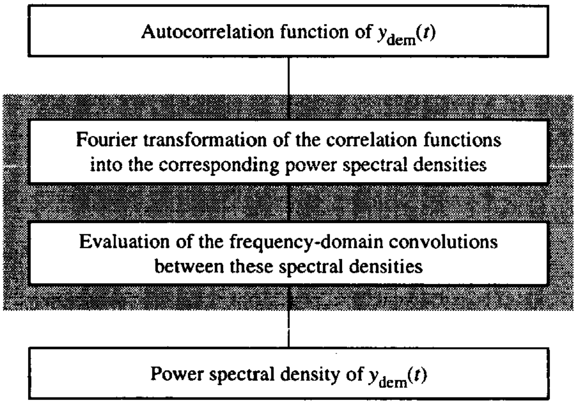

Autocorrelation function of $y_{\mathrm{dem}}(t)$
Fourier transformation of the correlation functions into the corresponding power spectral densities
Evaluation of the frequency-domain convolutions between these spectral densities
Power spectral density of $y_{\mathrm{dem}}(t)$

Figure 6.19: Calculation of the demodulator output power density spectrum.

The spectra corresponding to the autocorrelation functions of the components that depend only on $\varphi(t)$ and $\dot{\varphi}(t)$ may generally be obtained with the aid the quasi-stationary approximation, explained in Section 2.2.2. In addition, for wideband FM waves, the effect of the modulation on the demodulator output noise is usually negligible, which allows further simplification.

6.2.4 Baseband Filter Output Signal and Noise Power

The final step in the calculation of the output SNR is the determination of the signal and noise power observed at the output of the baseband filter. This section outlines the general derivation of an expression for the output noise power that is valid for arbitrary transfer characteristics of the RF/IF filter, the baseband output filter, and the amplitude compressor.

Formally, the total receiver output signal and noise power, denoted by P_b, is obtained by integration of the demodulator output spectral density, weighted by the baseband filter curve $H_b(\mathrm{j}\omega)$, as

$$P_b = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{\mathrm{dem}}(\omega)\,|H_b(\mathrm{j}\omega)|^2\,\mathrm{d}\omega. \tag{6.25}$$

The contribution of the noise to this output power depends on the shape of the IF/RF filter transfer and the baseband filter transfer. The dependency on the baseband filter transfer is demonstrated by (6.25), while the dependency on the IF filter transfer is contained in the spectral density $S_{\mathrm{dem}}(\omega)$. Thus, in order to obtain a generally valid, explicit expression for the receiver output noise power, the shape of the various filter transfers has to be included explicitly by means of some 'shape parameters'. In this way, the entire expression for the receiver output noise power and output SNR can be expressed in terms of a set of shaping parameters, the input CNR, and the parameters of the input FM wave.

Thus, as illustrated by figure 6.20, the calculation of the output noise power generally consists of two steps. First, the various shaping parameters are determined. Subsequently, the noise power is expressed in terms of these shaping parameters.

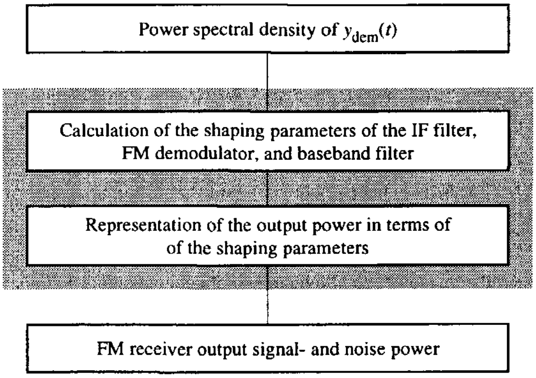

Power spectral density of $y_{\mathrm{dem}}(t)$
Calculation of the shaping parameters of the IF filter, FM demodulator, and baseband filter
Representation of the output power in terms of of the shaping parameters
FM receiver output signal- and noise power

Figure 6.20: Calculation of the receiver output noise power.

In subsequent sections it is shown that, with respect to the shaping introduced by the IF/RF filter, the baseband filter and the FM demodulator, six different components can generally be distinguished in the demodulator output noise power spectral density:

- first-order amplitude noise;

- first-order frequency noise;
- second-order amplitude noise;
- second-order frequency noise;
- a second-order noise×noise component;
- click noise.

The first and second component represent the first-order signal×noise intermodulation, the third and fourth represent the second-order signal×noise intermodulation, and the fifth represents the second-order demodulator response to the noise alone.

The contribution of each of these components to the receiver output noise, observed at the baseband filter output, is described by separate shaping parameter(s), as schematically represented by figure 6.21. Expressions for the various

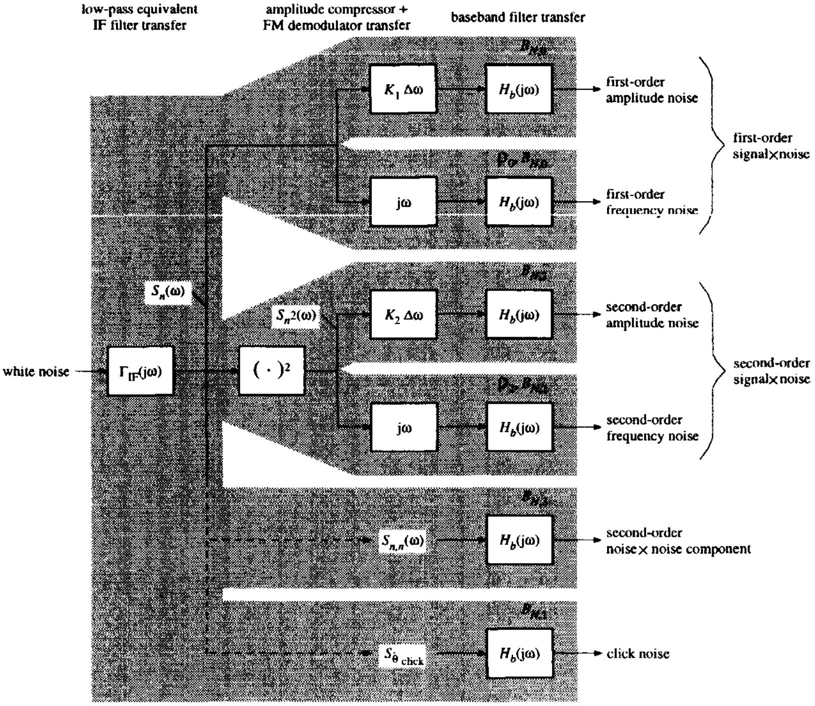


Figure 6.21: Schematic representation of the various noise components and their shaping parameters.

shaping parameters, in terms of the IF filter, baseband filter and demodulator transfer are derived below.

First-Order Noise

The first-order output noise, characterized by a power content that is inversely proportional to the receiver input CNR, generally consists of two contributions: continuous amplitude noise and continuous frequency noise. The amplitude noise is subjected to 'white' shaping, i.e. transferred by a frequency-independent gain, while the frequency noise is subjected to quadratic shaping by the demodulator.

Amplitude Noise The amplitude noise spectrum observed at the demodulator output is proportional to the spectral density $S_n(\omega)$ of the low-pass equivalent input noise processes $n_i(t)$ and $n_q(t)$, defined in Section 2.3. As illustrated by figure 6.21, this power density spectrum is obtained when zero-mean white Gaussian noise is filtered by the low-pass equivalent of the IF/RF filter, denoted by $\Gamma_{\text{IF}}(\mathrm{j}\omega)$.

The contribution of this noise component to the receiver output noise power is therefore proportional to the double-sided noise bandwidth of $\Gamma(\mathrm{j}\omega)H_b(\mathrm{j}\omega)$, which denotes the cascade of the low-pass equivalent IF filter and the baseband filter. This bandwidth, denoted by $B_{N,0}$, is defined as

$$\begin{aligned} B_{N,0} &\overset{\text{def}}{=} \frac{1}{2\pi S_n(0)\left|H_b(0)\right|^2}\int_{-\infty}^{\infty} S_n(\omega)\left|H_b(\mathrm{j}\omega)\right|^2 \mathrm{d}\omega \\ &= \frac{1}{2\pi\left|\Gamma_{\text{IF}}(0)H_b(0)\right|^2}\int_{-\infty}^{\infty} \left|\Gamma_{\text{IF}}(\mathrm{j}\omega)H_b(\mathrm{j}\omega)\right|^2 \mathrm{d}\omega, \end{aligned} \tag{6.26}$$

with unit (Hz). The first-order amplitude noise power at the receiver output, denoted by $P_{n,\text{ampl},1}$, is therefore proportional to

$$\begin{aligned} P_{n,\text{ampl},1} &\propto \frac{1}{2\pi}\int_{-\infty}^{\infty} S_n(\omega)\left|H_b(\mathrm{j}\omega)\right|^2 \mathrm{d}\omega \\ &= B_{N,0}S_n(0)\left|H_b(0)\right|^2. \end{aligned} \tag{6.27}$$

Frequency Noise The first-order frequency noise is shaped by the quadratic frequency transfer ω^2 of differentiation with respect to time. The demodulator emphasizes the high frequency spectral contents of this noise, while it reduces the low-frequency contents.

According to figure 6.21, the contribution of the frequency noise to the receiver output noise power is proportional to the double-sided 'noise bandwidth' of the cascade $\mathrm{j}\omega\Gamma_{\text{IF}}(\mathrm{j}\omega)H_b(\mathrm{j}\omega)$. The effect of the quadratic shaping applied by the demodulator on the frequency noise power at the receiver output is conveniently described by the *radius of gyration* of the FM receiver. This parameter, denoted by ρ_0, is defined in analogy with the radius of gyration of the IF/RF filter output noise r, defined by (5.13), as

$$\rho_0 \overset{\text{def}}{=} \frac{1}{2\pi}\sqrt{\frac{\int_{-\infty}^{\infty} \omega^2 S_n(\omega)\,|H_b(\mathrm{j}\omega)|^2\,\mathrm{d}\omega}{\int_{-\infty}^{\infty} S_n(\omega)\,|H_b(\mathrm{j}\omega)|^2\,\mathrm{d}\omega}}, \tag{6.28}$$

with unit (Hz). The first-order frequency noise power at the receiver output, denoted by $P_{n,\text{freq},1}$, is therefore proportional to

$$\begin{aligned} P_{n,\text{freq},1} &\propto \frac{1}{2\pi}\int_{-\infty}^{\infty} \omega^2 S_n(\omega)\,|H_b(\mathrm{j}\omega)|^2\,\mathrm{d}\omega \\ &= (2\pi\rho_0)^2\, B_{N,0} S_n(0)\,|H_b(0)|^2\,. \end{aligned} \tag{6.29}$$

Thus, according to this expression, $(2\pi\rho_0)^2$ denotes the ratio of the frequency noise power and the amplitude noise power at the receiver output.

Second-Order Noise

The second-order demodulator output noise, characterized by a power contents that increases inversely proportional, to the square of the input CNR, consists of three components: a 'white' shaped amplitude noise component, a 'quadratic' shaped frequency noise component that together represent the second-order signal×noise intermodulation, and a 'white' shaped noise×noise component, which is the demodulator response to the input noise alone.

Amplitude Noise Similar to the first-order amplitude noise, the second-order amplitude noise is transferred to the demodulator output by means of a frequency-independent 'gain'.

The contribution of this noise component to the output noise power is therefore also described by a noise bandwidth $B_{N,2}$, which is defined as

$$B_{N,2} \overset{\text{def}}{=} \frac{1}{2\pi S_{n^2}(0)\,|H_b(0)|^2}\int_{-\infty}^{\infty} S_{n^2}(\omega)\,|H_b(\mathrm{j}\omega)|^2\,\mathrm{d}\omega. \tag{6.30}$$

As explained in Section 6.4, the spectrum $S_{n^2}(\omega)$ in this expression equals the convolution of the low-pass equivalent input noise spectrum $S_n(\omega)$ with itself, i.e.

$$S_{n^2}(\omega) = S_n(\omega) * S_n(\omega). \tag{6.31}$$

In essence, this spectrum corresponds to the squared low-pass equivalent demodulator input noise components $n_i^2(t)$ and $n_q^2(t)$.

The second-order amplitude noise power at the receiver output, denoted by $P_{n,\mathrm{ampl},2}$, is therefore proportional to

$$\begin{aligned} P_{n,\mathrm{ampl},2} &\propto \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{n^2}(\omega)\,|H_b(\mathrm{j}\omega)|^2\,\mathrm{d}\omega \\ &= B_{N,2}S_{n^2}(0)\,|H_b(0)|^2\,. \end{aligned} \tag{6.32}$$

Frequency Noise Similar to the first-order frequency noise, the second-order frequency noise is subjected to quadratic shaping by the FM demodulator. Therefore, its contribution to the receiver output noise power is described by two parameters: the noise bandwidth $B_{N,2}$ from (6.32), and a radius of gyration ρ_2, that is defined as

$$\rho_2 \overset{\mathrm{def}}{=} \frac{1}{2\pi}\sqrt{\frac{\int_{-\infty}^{\infty}\omega^2 S_{n^2}(\omega)\,|H_b(\mathrm{j}\omega)|^2\,\mathrm{d}\omega}{\int_{-\infty}^{\infty} S_{n^2}(\omega)\,|H_b(\mathrm{j}\omega)|^2\,\mathrm{d}\omega}}, \tag{6.33}$$

with unit (Hz).

Consequently, the second-order receiver output frequency noise power, denoted by $P_{n,\mathrm{freq},2}$, is proportional to

$$\begin{aligned} P_{n,\mathrm{freq},2} &\propto \frac{1}{2\pi}\int_{-\infty}^{\infty} \omega^2 S_{n^2}(\omega)\,|H_b(\mathrm{j}\omega)|^2\,\mathrm{d}\omega \\ &= (2\pi\rho_2)^2\,B_{N,2}S_{n^2}(0)\,|H_b(0)|^2\,. \end{aligned} \tag{6.34}$$

Thus, according to this expression, $(2\pi\rho_2)^2$ denotes the ratio of second-order frequency noise power and the second-order amplitude noise power at the receiver output.

Noise×Noise Component The last second-order noise component represents the response of the FM demodulator to the input noise alone. In essence, it is the response to a mixture of amplitude noise and frequency noise.

As explained in Section 6.4.3, the power density spectrum of this component, denoted by $S_{n,n}(\omega)$, can be expressed in terms of the input noise spectrum $S_n(\omega)$ as

$$S_{n,n}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} y(2y-\omega)S_n(\omega-y)S_n(y)\mathrm{d}y. \tag{6.35}$$

Its contribution to the receiver output noise power is therefore described by the noise bandwidth $B_{N,3}$, defined as

$$B_{N,3} \overset{\mathrm{def}}{=} \frac{1}{2\pi S_{n,n}(0)\,|H_b(0)|^2}\int_{-\infty}^{\infty} S_{n,n}(\omega)\,|H_b(\mathrm{j}\omega)|^2\,\mathrm{d}\omega. \tag{6.36}$$

Therefore, the resulting noise×noise receiver output noise power, denoted by $P_{n,n,2}$, is proportional to

$$\begin{aligned} P_{n,n,2} &\propto \frac{1}{2\pi}\int_{-\infty}^{\infty} \omega^2 S_{n,n}(\omega)\,|H_b(\mathrm{j}\omega)|^2\,\mathrm{d}\omega \\ &= B_{N,3}S_{n,n}(0)\,|H_b(0)|^2\,. \end{aligned} \tag{6.37}$$

Click Noise

Click noise is not transferred from the receiver input to the output, but generated at the FM demodulator output. Its contribution to the output noise power is therefore shaped only by the baseband filter, and not by the RF/IF filter transfer, as depicted in figure 6.21. However, the click *rate* does depend on the IF/RF filter transfer shape, by means of the *radius of gyration*, as discussed in Section 5.4.3.

Due to its approximately white spectrum, the contribution of click noise to the output noise power is represented by the double-sided noise bandwidth of the baseband filter, given by

$$B_{N,1} \overset{\mathrm{def}}{=} \frac{1}{2\pi\,|H_b(0)|^2}\int_{-\infty}^{\infty} |H_b(\mathrm{j}\omega)|^2\,\mathrm{d}\omega, \tag{6.38}$$

with unit (Hz).

The click noise power observed at the receiver output, P_{click}, is therefore proportional to

$$\begin{aligned} P_{\mathrm{click}} &\propto \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{\dot{\theta}_{\mathrm{click}}}\,|H_b(\mathrm{j}\omega)|^2\,\mathrm{d}\omega \\ &= B_{N,1}S_{\dot{\theta}_{\mathrm{click}}}(0)\,|H_b(0)|^2\,. \end{aligned} \tag{6.39}$$

Total Output Noise Power

The total receiver output noise power is a linear combination of the previously described contributions, where the amplitude compressor transfer, represented by $C_{n,1}(A)$ and $C_{n,2}(A)$, serves as a weighting factor. Further, the receiver output noise power is, of course, dependent on the input CNR p. In terms of the double-sided noise bandwidth of the input noise, denoted by B, which is determined by the IF filter, i.e.

$$\begin{aligned} B_{N,\mathrm{IF}} &= \frac{1}{2\pi S_n(0)}\int_{-\infty}^{\infty} S_n(\omega)\mathrm{d}\omega \\ &= \frac{1}{2\pi\Gamma_{\mathrm{IF}}(0)}\int_{-\infty}^{\infty} \Gamma_{\mathrm{IF}}(\omega)\mathrm{d}\omega, \end{aligned} \tag{6.40}$$

shaping parameter	rectangular baseband filter	CCIR 468 baseband filter (kHz)
$B_{N,0}$	$\frac{W}{\pi}$	6.43
$B_{N,1}$	$\frac{W}{\pi}$	6.43
$B_{N,2}$	$\frac{W}{\pi}\left(1-\frac{W}{2W_n}\right)$	17.58
$B_{N,3}$	$\frac{W}{\pi}\left[1-\left(\frac{3W}{2W_n}\right)+\left(\frac{W}{W_n}\right)^2-\left(\frac{W}{4W_n}\right)^3\right]$	10.58
ρ_0	$\frac{W}{2\pi\sqrt{3}}$	6.99
ρ_2	$\frac{W}{2\pi}\sqrt{\frac{1-(3W/4W_n)}{3-(3W/2W_n)}}$	12.61

Table 6.1: Noise shaping parameters for a rectangular IF filter, and a rectangular/CCIR 468 baseband filter.

this CNR can be expressed as

$$p = \frac{A^2}{2S_n(0)B_{\mathrm{N,IF}}}, \tag{6.41}$$

where A denotes the receiver input carrier amplitude.

Table 6.1 lists the values of the parameters $B_{N,0}$ through ρ_2 for two different types of baseband output filters: a rectangular filter with bandwidth W, and a filter that complies with the CCIR 468 standard. The frequency transfer of the latter filter is depicted in figure 6.22. The IF filter is assumed to possess a rectangular frequency transfer, and a bandwidth W_n that is considerably larger than the bandwidth of the baseband filters.

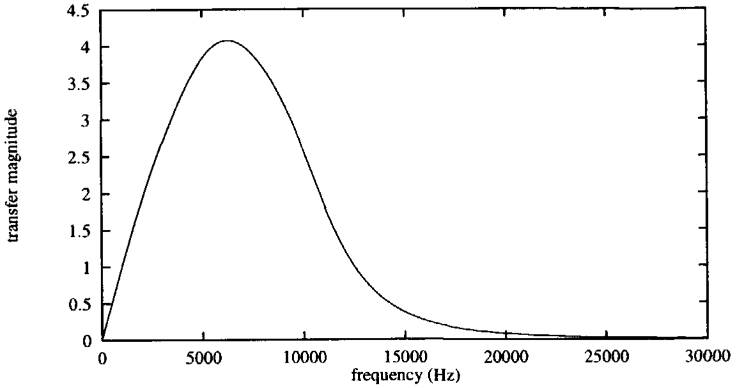


Figure 6.22: Frequency transfer of a CCIR 468 baseband filter.

6.3 First-Order Continuous Noise

The first-order continuous components dominate the demodulator output signal at high CNRs, as discussed in Section 5.4.2.

This section considers the derivation of these noise components, with the amplitude compressor transfer as parameter. Section 6.3.1 discusses their identification from the demodulator output signal. Section 6.3.2 discusses their autocorrelation function, required for the power density spectrum considered in Section 6.3.3. Section 6.3.4 derives an expression of the first-order noise power located at baseband frequencies, while Section 6.3.5 discusses the implications of this expression on the design of the amplitude compressor.

6.3.1 Time Domain First-Order Noise

As outlined in Section 6.2.1, a time domain expression for the first-order output noise components is obtained by a four-dimensional first-order Taylor expansion of $y_{\rm dem}(t)$, given by (6.15). From that expression, the first-order noise components, denoted by $y_{\rm dem,1}(t)$, are obtained for $k = 1$ as

$$y_{\rm dem,1}(t) = n_{s,i}\frac{\partial y_{\rm dem}}{\partial n_{s,i}} + n_{s,q}\frac{\partial y_{\rm dem}}{\partial n_{s,q}} + \dot{n}_{s,i}\frac{\partial y_{\rm dem}}{\partial \dot{n}_{s,i}} + \dot{n}_{s,q}\frac{\partial y_{\rm dem}}{\partial \dot{n}_{s,q}}. \qquad (6.42)$$

Evaluation of the four partial derivatives shows that only $n_{s,i}(t)$ and $\dot{n}_{s,q}(t)$ yield a nonzero contribution to the output signal. The reason for this is that $n_{s,i}(t)$ represents the first-order term of the amplitude $R(t)$, i.e. the amount of amplitude noise that penetrates the output, while $\dot{n}_{s,q}(t)$ represents the first-order frequency noise, as discussed in Section 2.3.2.

By substitution of the first-order inverse suppression factor $C_{n,1}(A)$, defined by (5.8), the final expression for the first-order output noise becomes

$$y_{\text{dem},1}(t) = \frac{G(A)}{A}\left[C_{n,1}(A)\dot{\varphi}(t)n_{s,i}(t) + \dot{n}_{s,q}(t)\right]. \tag{6.43}$$

This expression demonstrates the observation in Section 5.4.2 that the amplitude noise is transferred to the output by the compressor small-signal transfer, while the phase/frequency noise is transferred by the large signal transfer.

6.3.2 Autocorrelation Function of the First-Order Noise

The contribution of the first-order noise $y_{\text{dem},1}(t)$ to the demodulator output noise spectrum and the receiver output noise power can be obtained from its auto correlation, as discussed in Section 6.2. The cross correlation of the first-order noise with the FM message signal, which is a product of one Gaussian noise component with the FM message, and the cross-correlation with the second-order noise, which is a product of three Gaussian noise components, do not contribute to the noise spectrum and the output noise power. This follows directly from theorem 1: both cross correlations consist of products of an *odd* number of zero-mean Gaussian random variables of which, according to the theorem, the expectation equals zero.

From expression (6.43) it follows that the autocorrelation function of the first-order noise generally consists of three contributions:

- autocorrelation of the frequency noise $\dot{n}_{s,q}$;
- autocorrelation of the amplitude noise contribution $\dot{\varphi}n_{s,i}$;
- cross-correlation between both components.

As shown in Appendix B, application of theorem 2 to the definition formula (6.19), with $y_{\text{dem}}(t)$ replaced by $y_{\text{dem},1}(t)$, yields the following expression for the autocorrelation of the output signal $y_{\text{dem},1}(t)$:

$$\begin{aligned}R_{\text{dem},1}(\tau) = \left[\frac{G(A)}{A}\right]^2 \Bigg\{ & R_{\dot{n}\dot{n}}(\tau)\mathrm{E}\left[\cos\left(\varphi_1 - \varphi_2\right)\right] \\ & + \left[1 - C_{n,1}(A)\right] R_{\dot{n}n}(\tau)\mathrm{E}\left[(\dot{\varphi}_1 + \dot{\varphi}_2)\sin\left(\varphi_1 - \varphi_2\right)\right] \\ & + \left[1 - C_{n,1}(A)\right]^2 R_{nn}(\tau)\mathrm{E}\left[\dot{\varphi}_1\dot{\varphi}_2\cos\left(\varphi_1 - \varphi_2\right)\right]\Bigg\},\end{aligned} \tag{6.44}$$

where $R_{nn}(\tau)$ denotes the autocorrelation of the noise components $n_i(t)$ and $n_q(t)$ (see Section 2.3.1), $R_{\dot{n}n}(\tau)$ the cross-correlation between $n_i(t)$, or $n_q(t)$, and its time-derivative, and $R_{\dot{n}}(\tau)$ the autocorrelation of the time-derivative.

Further, the expectations of $\varphi(t)$ and $\dot{\varphi}(t)$, where φ_1 denotes $\varphi(t)$ at $t = t_1$, and φ_2 on $t = t_2$, represent the modulation of the noise spectrum by the message.

It is important to notice that the autocorrelation function of the first order noise consists of the same components for all types of amplitude compressors. The compression factor $C_{n,1}(A)$ serves as a weighting factor that determines only the contribution of these components to the demodulator output noise. The first term in (6.44) basically represents the frequency noise, while the last term is mainly due to the amplitude noise. The middle term represents their cross-correlation. This can be observed from their weighting factors. The contribution of the amplitude noise to $y_{\text{dem},1}(t)$ is proportional to the inverse first-order compression factor $C_{n,1}(A)$, while the frequency noise contribution is independent of this factor. Consequently, the autocorrelation function of the frequency noise contribution is independent of $C_{n,1}(A)$, the autocorrelation function of the amplitude noise contribution is proportional to $C^2_{n,1}(A)$, while their cross correlation is proportional to $C_{n,1}(A)$.

Expression (6.44) agrees with the output noise autocorrelation functions that were derived in [26] for three specific types of amplitude compression:

- infinite compression, where $C_{n,1}(A) \equiv 0$;
- finite compression, where $G(R) = R$ and $C_{n,1}(A) = 1$;
- no compression, where $G(R) = R^2$ and $C_{n,1}(A) = 2$.

In all these cases, (6.44) matches the results obtained in [26].

6.3.3 Power Spectral Density of the First-Order Noise

The power spectral density of the first-order output noise is obtained by Fourier transformation of the autocorrelation function $R_{\text{dem},1}(\tau)$. We first consider the general expression for this spectrum, and subsequently discuss a simplified expression for wideband FM.

General Expression

Expression (6.44) for the autocorrelation function consists of three terms, that are each a product of two autocorrelation functions. One correlation function depends entirely on the input noise $n(t)$, while the other one depends entirely on the message signals $\varphi(t)$ and $\dot{\varphi}(t)$.

In the corresponding spectrum, these products of correlation functions transform into a convolution of the corresponding spectra. The spectrum of the correlation function that depends on the noise is a function of the spectral density $S_n(\omega)$, of the low-pass equivalent noise processes $n_i(t)$ and $n_q(t)$. The spectrum of the correlation function that depends on $\varphi(t)$ and $\dot{\varphi}(t)$ can be expressed in terms of the probability density of the instantaneous message frequency $\dot{\varphi}(t)$, denoted by $p_{\dot{\varphi}}(\omega)$ with the aidof the quasi-stationary approximation.

As shown in Appendix C, the final result for the first order output noise spectrum, obtained by a proper rearrangement of the various spectral components, equals

$$\begin{aligned} S_{\mathrm{dem},1}(\omega) = \left[\frac{G(A)}{A}\right]^2 \bigg[\omega^2 \int_{-\infty}^{\infty} S_n(\omega - y) p_{\dot{\varphi}}(y) \mathrm{d}y \\ - 2\omega C_{n,1}(A) \int_{-\infty}^{\infty} y S_n(\omega - y) p_{\dot{\varphi}}(y) \mathrm{d}y \\ + C_{n,1}^2(A) \int_{-\infty}^{\infty} y^2 S_n(\omega - y) p_{\dot{\varphi}}(y) \mathrm{d}y \bigg]. \quad (6.45) \end{aligned}$$

This expression clearly shows the effect of (finite) compression on the demodulator output noise. The first term, a parabolic shaped spectrum, corresponds to the frequency noise. Its contribution to the output spectrum is therefore independent of the inverse compression factor $C_{n,1}(A)$. The last term represents a noise floor that is due to the amplitude noise $n_{s,i}(t)$. Therefore this term is weighted by $C_{n,1}^2(A)$. The second term represents a cross-term between the frequency noise and the amplitude noise, and is therefore proportional to $C_{n,1}(A)$. This term mainly affects the spectrum at the top of the FM bandwidth, i.e. usually far above the highest message frequency, and is therefore often negligible.

As an illustration, figure 6.23 depicts the spectrum $S_{\mathrm{dem},1}(\omega)$ on a logarithmic scale for an FM demodulator that employs a soft-limiter as amplitude compressor, for various values of the limiter gain. The compressor transfer $G(A) = G_{\mathrm{sl}}^2(A)$ is chosen according to (6.11), the message signal consists of Gaussian noise with frequency deviation ratio $\Delta\omega/W = 5$. The input noise $n(t)$ is assumed to be white over the entire FM transmission bandwidth. The parameter $x = K/A$ denotes the ratio of the limiter's linear region and the carrier amplitude, i.e. the inverse of the "over-drive factor" [20]. Further, the spectral frequency is normalized to the RMS frequency deviation.

For high limiter gain values, i.e. x close to zero, the output basically consists of the parabolic spectrum, represented by an asymptote that increases by 20 dB per decade. For decreasing limiter gain values, i.e. x in the vicinity of unity, the white component due to the amplitude noise gradually increases the noise level inside the baseband, located around zero frequency. The decay of the spectral density at $\omega/\Delta\omega = 1$ for low limiter gain values is due to the second term in (6.45).

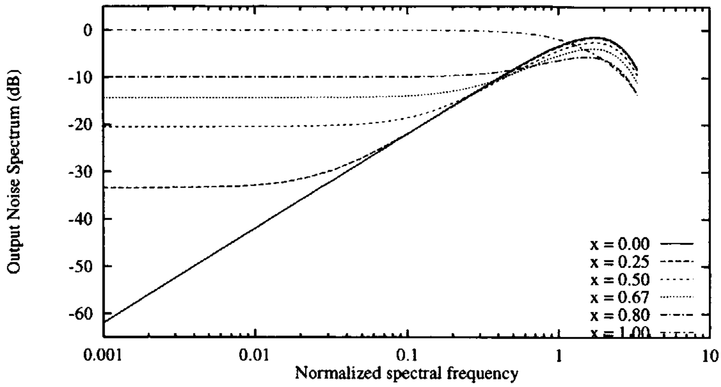


Figure 6.23: First-order output noise spectrum of a soft-limiting FM demodulator for various values of the limiter overdrive $x = K/A$.

Spectrum for Wideband FM

For wideband FM waves, and in many cases also for narrow-band waves, expression (6.45) can be simplified significantly by application of the knowledge that the bandwidth of the message signal $\dot{\varphi}(t)$ is considerably smaller than the FM transmission bandwidth.

The modulation of the noise spectrum introduced by $\varphi(t)$ and $\dot{\varphi}(t)$ mainly affects the demodulator output noise spectrum around its cut-off frequency (see figure 2.5). As shown in Appendix C, $S_{\text{dem},1}(\omega)$ can therefore be approximated as

$$S_{\text{dem},1}(\omega) \approx \left[\frac{G(A)}{A}\right]^2 \left[\omega^2 + (\Delta\omega)^2 C_{n,1}^2(A)\right] S_n(\omega), \tag{6.46}$$

where $\Delta\omega$ denotes the RMS frequency deviation. This expression basically ignores the correlation between the frequency noise and the amplitude noise, the second term in (6.45); it roughly equals the addition of the frequency noise spectrum and the amplitude noise spectrum. The term ω^2 represents the quadratic shaping that is applied to the frequency noise, while $(\Delta\omega)^2 C_{n,1}^2(A)$ represents the white noise floor due to the amplitude noise. The factor $(\Delta\omega)^2$ represents the power contents of the FM message signal $\dot{\varphi}(t)$ contained in the amplitude noise contribution $\dot{\varphi}(t) n_{s,i}(t)$.

6.3.4 Output Power due to First-Order Noise

This section determines the first-order noise power observed at the FM receiver output, i.e. at the output of the baseband filter in terms of the parameters defined in Section 6.2.4. The resulting expression is used as basis of the discussion on amplitude compressor design in Section 6.3.5.

In order to gain insight into the various mechanisms involved in the generation of the output noise power, it is assumed throughout this section that the simplified expression (6.46) for the demodulator output noise spectrum is valid.

The first-order receiver output noise power, consisting of an amplitude noise and a frequency noise contribution, denoted by $P_{\text{dem},1}$, can be obtained by application of the Wiener-Khintchine theorem to the demodulator output noise spectrum $S_{\text{dem},1}(\omega)$. By substitution of the receiver noise bandwidth $B_{N,0}$, the corresponding radius of gyration ρ_0, and (6.41) for the input CNR, this yields

$$P_{\text{dem},1} = (2\pi\rho_0)^2 \, \frac{G^2(A)}{p} \, \frac{B_{N,0}}{2B_{N,\text{IF}}} \, |H_b(0)|^2 \left[1 + \left(\frac{\Delta\omega}{2\pi\rho_0} \right)^2 C_{n,1}^2(A) \right] . \qquad (6.47)$$

The second term inside the brackets in this expression represents the contribution the amplitude noise; it is proportional to $C_{n,1}^2(A)$. Further, observe that the noise power is proportional to the ratio of the baseband filter and IF filter noise bandwidths; this represents the trade-off realized by the FM scheme between the receiver output SNR and the transmission bandwidth.

6.3.5 Implications on Amplitude Compressor Design

Expression (6.47) for the first-order receiver output noise power hides a design rule for the minimum level of compression, required for a receiver output SNR that is close to the theoretical maximum obtained for infinite compression. This design rule is discussed in this section. First, the general case is considered. Subsequently, as an example, it is applied to an FM receiver that contains a soft-limiter.

General Expression for the Critical Amplitude Compression Level

The reduction of the first-order FM receiver output noise power established by amplitude compression is schematically visualized in figure 6.24. This figure depicts the output noise power described by expression (6.47), normalized to the output noise power observed in case of infinite compression ($C_{n,1}(A) = 0$), as function of the inverse first-order compression factor $C_{n,1}(A)$.

In the case of infinite compression, all amplitude noise is eliminated and the output noise power reaches the minimum possible level, determined by the quadratic frequency noise.

In the absence of any compression, i.e. when $G(R) = R^2$ and $C_{n,1}(A) = 2$, the amplitude noise and frequency noise are passed in equal proportions. In practice, this is the maximum amplitude noise level encountered in FM demodulators. Larger values of $C_{n,1}(A)$, corresponding to $G(R) = R^a$ with $a > 2$, do

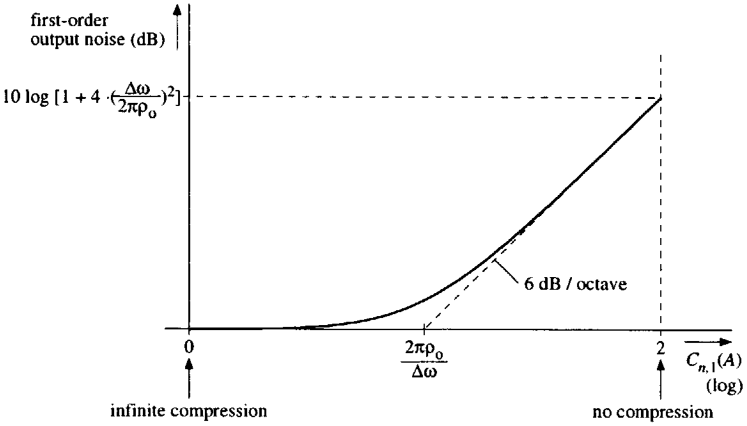


Figure 6.24: Reduction of the first-order output noise power as function of the applied level of amplitude compression.

not make sense since click noise is already completely suppressed for $G(R) = R^2$; a further increase of the exponent a results only in a disproportionate increase in the continuous noise. The deterioration of the output SNR introduced by amplitude noise, in comparison to the maximum possible SNR attained with infinite compression is illustrated by the following example. Consider the case of a rectangular IF and baseband filter with, according to table 6.1, $\rho_0 = W/(2\pi\sqrt{3})$, and the RMS frequency deviation ratio $\Delta\omega/W = 5$. In that case, it follows from (6.47) that the noise level in the absence of compression ($C_{n,1}(A) = 2$) exceeds the level obtained with infinite compression ($C_{n,1}(A) = 0$) by 25 dB.

According to (6.47), the minimum attainable noise level is approached within 3 dB, and further significant reduction by incrementing the amplitude compression level becomes impossible once $C_{n,1}(A)$ has been decreased below the 'critical' value

$$C_{n,1}(A) = C_{\text{cr},1} \overset{\text{def}}{=} \frac{2\pi\rho_0}{\Delta\omega}. \tag{6.48}$$

This level of compression corresponds to the intersection of the asymptotes in figure 6.24, where the amplitude noise power and frequency noise power are equal. The total output noise is thus only 3 dB higher than the theoretical minimum. Thus, a value of $C_{n,1}(A)$ that is much smaller than $C_{\text{cr},1}$, i.e. a compression level that is much larger than the 'critical level', makes no sense.

Example: Critical Compression Level of a Soft-Limiter

As an illustration of the previously discussed theory, we determine the critical compression level, $1/C_{\text{cr},1}(A)$ for an amplitude compressor implemented by a soft-limiter.

Consider an FM demodulator that is preceded by a soft-limiter which processes the input wave subjected to demodulation and the reference wave, such that the compressor transfer $G(R) = G_{\mathrm{sl}}^2(R)$. The transfer $G_{\mathrm{sl}}(R)$ was determined in Section 6.1.4. The inverse first-order compression factor corresponding to $G(R)$ is given by (6.12).

For a rectangular IF and baseband filter, with $\rho_0 = W/(2\pi\sqrt{3})$, and a deviation ration $\Delta\omega/W = 5$, the critical level of compression is about

$$\frac{1}{C_{\mathrm{cr},1}} \geq \sqrt{3}\frac{\Delta\omega}{W} \approx 8.7. \tag{6.49}$$

By substitution of (6.12) for $C_{\mathrm{cr},1}$, and numerical (or graphical) solution, it follows from this expression that this level of compression requires a limiter overdrive $1/x = A/K$ of at least $A/K \approx 2.5$. For larger overdrives, the optimum output SNR is approached within 3 dB.

6.4 Second-Order Continuous Noise

This section determines the contribution of the second-order noise to the demodulator and receiver output signal. As explained in Section 5.4.5, this noise represents the "modulation" of the compressor small signal transfer $C_{n,1}(A)$.

Section 6.4.1 derives a time-domain expression for the second-order output noise. Section 6.4.2 and Section 6.4.3 determine its autocorrelation and spectral density respectively. Section 6.4.4 determines the receiver output noise power, while Section 6.4.5 considers its implications on amplitude compressor design.

6.4.1 Time Domain Second-Order Noise

A time domain expression for the second-order noise is obtained from the second-order terms of the four-dimensional Taylor series of the demodulator output signal, given by (6.15). The second-order terms of that expression, $y_{\mathrm{dem},2}(t)$, are given by:

$$\begin{aligned} y_{\mathrm{dem},2}(t) = {} & \frac{n_{s,i}^2}{2}\frac{\partial^2 y_{\mathrm{dem}}}{\partial n_{s,i}^2} + \frac{n_{s,q}^2}{2}\frac{\partial^2 y_{\mathrm{dem}}}{\partial n_{s,q}^2} + \frac{\dot{n}_{s,i}^2}{2}\frac{\partial^2 y_{\mathrm{dem}}}{\partial \dot{n}_{s,i}^2} \\ & + \frac{\dot{n}_{s,q}^2}{2}\frac{\partial^2 y_{\mathrm{dem}}}{\partial \dot{n}_{s,q}^2} + n_{s,i}n_{s,q}\frac{\partial^2 y_{\mathrm{dem}}}{\partial n_{s,i}\partial n_{s,q}} + n_{s,i}\dot{n}_{s,i}\frac{\partial^2 y_{\mathrm{dem}}}{\partial n_{s,i}\partial \dot{n}_{s,i}} \\ & + n_{s,i}\dot{n}_{s,q}\frac{\partial^2 y_{\mathrm{dem}}}{\partial n_{s,i}\partial \dot{n}_{s,q}} + n_{s,q}\dot{n}_{s,i}\frac{\partial^2 y_{\mathrm{dem}}}{\partial n_{s,q}\partial \dot{n}_{s,i}} \\ & + n_{s,q}\dot{n}_{s,q}\frac{\partial^2 y_{\mathrm{dem}}}{\partial n_{s,q}\partial \dot{n}_{s,q}} + \dot{n}_{s,i}\dot{n}_{s,q}\frac{\partial^2 y_{\mathrm{dem}}}{\partial \dot{n}_{s,i}\partial \dot{n}_{s,q}}. \end{aligned} \tag{6.50}$$

Fortunately, five of the ten partial derivatives of $y_{\mathrm{dem}}(t)$ in this expression equal zero. Elaboration of the five nonzero derivatives finally results in the following expression:

$$\begin{aligned} y_{\mathrm{dem},2}(t) = \frac{G(A)}{A^2}\bigg\{ & C_{n,2}(A)\dot{\varphi}\frac{n_{s,i}^2}{2} + C_{n,1}(A)\dot{\varphi}\frac{n_{s,q}^2}{2} + \\ & [C_{n,1}(A) - 1]\, n_{s,i}\dot{n}_{s,q} - \dot{n}_{s,i}n_{s,q}\bigg\}, \end{aligned} \tag{6.51}$$

where $C_{n,2}(A)$ denotes the inverse second-order compression factor, defined by definition 2, given in Section 5.4.5. This factor basically represents the modulation of $C_{n,1}(A)$ by the input noise.

The first two noise terms in (6.51) are due to the second-order amplitude noise, since they contain the factor $\dot{\varphi}(t)$. This follows by investigation of the general expression for the demodulator output signal, given by 6.13: the contributions of the amplitude noise to the demodulator output noise are all multiplied by the message $\dot{\varphi}(t)$. The last two terms in (6.51) basically represent the second-order amplitude noise. Further, the last two terms in (6.51) also contain the second-order noise×noise component.

Finally, it should be noticed from expression (6.51) that the second-order noise is determined by both the inverse first-order and inverse second-order compression factors of the amplitude compressor transfer $G(A)$.

6.4.2 Autocorrelation Function of the Second-Order Noise

The autocorrelation function of the second-order noise components is determined in a way similar to the autocorrelation function of the first order noise, described in Section 6.3.2.

As discussed in Section 6.3.2, the cross-correlation between the second-order noise $y_{\mathrm{dem},2}(t)$ and the first-order noise $y_{\mathrm{dem},1}(t)$ equals zero, as a consequence of theorem 1.

On the basis of theorem 2, however, it is observed that the cross-correlation between the second-order noise and the message signal is generally nonzero, due to the fact that it consist of expectations of an *even* number of Gaussian random variables. These terms basically represent a second-order approximation of the signal suppression effect, discussed in Section 6.1.2. It was shown there that signal suppression is negligible in hard-limiters, which can be viewed as a practical 'worst-case' situation. With finite compression, signal suppression is usually even less important, and is therefore ignored in the conclusion.

According to the procedure outlined in Section 6.2.2, the autocorrelation of the second-order noise, denoted by $R_{\mathrm{dem},2}(\tau)$, can be expressed as a product of conditional expectations of $n_{s,i}$, $n_{s,q}$ and derivatives, followed by an expectation over φ and $\dot{\varphi}$. An example of such a calculation is made in Appendix B for the first-order noise components. Appendix D gives the expressions for all second-order correlation functions, required for the calculation of $R_{\text{dem},2}(\tau)$. By application of the appropriate weighting factors to these correlation functions, which follow directly from (6.51), the final result becomes

$$R_{\text{dem},2}(\tau) = \frac{G^2(A)}{A^4}\Big\{ \\ (\mu-\lambda)^2\left[R_{nn}^2(\tau)+\sigma_n^4\right]R_{\dot\varphi\dot\varphi}(\tau) \\ +(\mu-3\lambda+2)^2R_{nn}^2(\tau)\mathrm{E}\left[\dot\varphi_1\dot\varphi_2\cos 2\left(\varphi_1-\varphi_2\right)\right] \\ +2(\mu-3\lambda+2)(1-\lambda)R_{nn}(\tau)R_{\dot n n}(\tau)\mathrm{E}\left[(\dot\varphi_1+\dot\varphi_2)\sin 2\left(\varphi_1-\varphi_2\right)\right] \\ +2(1-\lambda)^2\left[R_{nn}(\tau)R_{\dot n\dot n}(\tau)-R_{\dot n n}^2(\tau)\right]\mathrm{E}\left[\cos 2\left(\varphi_1-\varphi_2\right)\right] \\ +2\lambda^2\left[R_{nn}(\tau)R_{\dot n\dot n}(\tau)+R_{\dot n n}^2(\tau)\right]\Big\}, \quad (6.52)$$

where, for convenience, the factors μ and λ are defined as

$$\mu \overset{\text{def}}{=} \tfrac{1}{2}C_{n,2}(A), \quad \lambda \overset{\text{def}}{=} \tfrac{1}{2}C_{n,1}(A). \quad (6.53)$$

Expression (6.52) is in accordance with the results obtained in [26] for $\mu = \lambda = 0, 0.5, 2$ (the same cases as mentioned in Section 6.3.2), and the absence of modulation, i.e. $\varphi \equiv 0$. Note the similarity between this expression and (6.44) for the first-order noise. Both autocorrelations consist of the same terms for all types of amplitude compression. The compressor transfer affects only the scaling factors of these components.

6.4.3 Power Spectral Density of the Second-Order Noise

This section considers the power spectral density of the second-order demodulator output noise. First, a general expression is derived. Subsequently, a simplified expression for wideband FM is discussed.

General Expression

The spectral density of the second-order noise is obtained by Fourier transformation of $R_{\text{dem}2}(\tau)$. Although somewhat more elaborate, the calculation of this spectrum, outlined in Appendix E, proceeds along the same lines as the calculation of $S_{\text{dem},1}(\omega)$.

Expressed in terms of the probability density of twice the message signal, i.e. $2\dot\varphi(t)$, that is denoted by $p_{2\dot\varphi}(.)$, and the spectrum $S_{n^2}(\omega) = S_n(\omega) * S_n(\omega)$,

$S_{\mathrm{dem},2}(\omega)$ follows as

$$
\begin{aligned}
S_{\mathrm{dem},2}(\omega) = \frac{G^2(A)}{A^4}\Bigg\{ & (\mu-\lambda)^2\sigma_n^4 S_{\dot\varphi}(\omega) \\
& + \frac{(\mu-\lambda)^2}{2\pi}\int_{-\infty}^{\infty} S_{n^2}(\omega-y)S_{\dot\varphi}(y)\mathrm{d}y \\
& + (1-\lambda)^2\omega^2\int_{-\infty}^{\infty} S_{n^2}(\omega-y)p_{2\dot\varphi}(y)\mathrm{d}y \\
& + (1-\lambda)(\mu-\lambda)\omega\int_{-\infty}^{\infty} yS_{n^2}(\omega-y)p_{2\dot\varphi}(y)\mathrm{d}y \\
& + \frac{1}{4}(\mu-\lambda)^2\int_{-\infty}^{\infty} y^2S_{n^2}(\omega-y)p_{2\dot\varphi}(y)\mathrm{d}y \\
& + \frac{\lambda^2}{\pi}\int_{-\infty}^{\infty} y(2y-\omega)S_n(\omega-y)S_n(y)\mathrm{d}y\Bigg\}. \qquad (6.54)
\end{aligned}
$$

This spectrum has a similar appearance as the first-order spectrum $S_{\mathrm{dem},1}(\omega)$ given in (6.45). That spectrum consisted of convolutions between the noise spectrum $S_n(\omega)$ and the probability density of $\dot\varphi$. The spectrum of the second-order noise mainly consists of convolutions between the noise spectrum $S_{n^2}(\omega) = S_n(\omega) * S_n(\omega)$ and the probability density of $2\dot\varphi$.

Second-Order Spectrum for Wideband FM

For wideband FM waves, considerable simplification of (6.54) is possible by application of the knowledge that the bandwidth of the message signal $\dot\varphi(t)$ is much smaller than the FM transmission bandwidth. This knowledge was already used in Section 6.3.3 to simplify the first-order output noise spectrum.

In Appendix E it is shown that for wideband FM, (6.54) reduces to

$$
\begin{aligned}
S_{\mathrm{dem},2}(\omega) \approx \frac{G^2(A)}{A^4}\Bigg\{ & \left[2(\mu-\lambda)^2(\Delta\omega)^2 + (1-\lambda)^2\omega^2\right]S_{n^2}(\omega) \\
& + \frac{\lambda^2}{\pi}\int_{-\infty}^{\infty} y(2y-\omega)S_n(\omega-y)S_n(y)\mathrm{d}y\Bigg\}. \qquad (6.55)
\end{aligned}
$$

In this expression, the first term in (6.54) is ignored since it contributes only to the, usually negligible, signal suppression effect. Further, similar to the noise spectrum $S_n(\omega)$ in the first-order demodulator output noise, the second-order noise $S_{n^2}(\omega)$ is subjected to quadratic shaping, and white shaping proportional to the FM message power contents $(\Delta\omega)^2$. The quadratic shaped component represents the second-order frequency noise, which is shaped by the differentiation inside the FM demodulator, while the 'white' shaped represents the second-order amplitude noise, that is not subjected to differentiation. The proportionality of the latter component to the message power contents $(\Delta\omega)^2$ indicates the presence of a factor $\dot{\varphi}(t)$ in the noise terms in (6.51) that corresponds to this spectral component, which is the factor that distinguishes the contributions of the amplitude noise from those of the frequency noise and the noise×noise components. The spectral density of these noise×noise components is represented by the convolution integral in (6.55), which is denoted by $S_{n,n}(\omega)$.

Further, it is shown in Appendix E that with the aid of the central limit theorem [25], $S_{n^2}(\omega)$ and $S_{n,n}(\omega)$ may generally be approximated by

$$\begin{aligned} S_{n^2}(\omega) &\overset{\text{def}}{=} \frac{1}{2\pi}\int_{-\infty}^{\infty} S_n(y)S_n(\omega - y)\mathrm{d}y \\ &\approx \frac{\sigma_n^4}{2r\sqrt{\pi}} \exp\left[-\frac{\omega^2}{4(2\pi r)^2}\right], \end{aligned} \tag{6.56}$$

$$\begin{aligned} S_{n,n}(\omega) &\overset{\text{def}}{=} \frac{1}{2\pi}\int_{-\infty}^{\infty} y(2y-\omega)S_n(\omega - y)S_n(y)\mathrm{d}y \\ &\approx \frac{4\pi^2 r\sigma_n^4}{\sqrt{2\pi\left[\left(\frac{\rho_r}{r}\right)^2 - 1\right]}} \exp\left\{-\frac{\omega^2}{2(2\pi r)^2\left[\left(\frac{\rho_r}{r}\right)^2 - 1\right]}\right\}, \end{aligned} \tag{6.57}$$

where r denotes the *radius of gyration*, defined by (5.13), and ρ_r equals the radius of gyration of the time-derivative of the input noise, i.e.

$$\rho_r \overset{\text{def}}{=} \frac{1}{2\pi}\sqrt{\frac{\int_{-\infty}^{\infty}\omega^4 S_n(\omega)\mathrm{d}\omega}{\int_{-\infty}^{\infty}\omega^2 S_n(\omega)\mathrm{d}\omega}}. \tag{6.58}$$

Thus, the entire second-order output noise spectrum can be expressed in terms of the RMS frequency deviation $\Delta\omega$, the parameters r and ρ_r that describe the demodulator input noise spectrum, and the amplitude compressor parameters μ and λ.

6.4.4 Output Power due to Second-Order Noise

This section derives an expression for the second-order receiver output noise power observed at the output of the baseband filter in terms of the parameters of the input noise, the baseband filter and the amplitude compressor. The result is used in the discussion on amplitude compressor design in Section 6.4.5.

The second-order output noise power, denoted by $P_{\text{dem},2}$, is obtained from the definition formula

$$P_{\text{dem},2} \overset{\text{def}}{=} \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{\text{dem},2}(\omega)\left|H_b(\mathrm{j}\omega)\right|^2 \mathrm{d}\omega. \tag{6.59}$$

In order to gain insight into the dependence of the second-order noise on the compressor transfer, we assume that the simplified expression (6.55) is valid. In that case, $P_{\mathrm{dem},2}$ can be expressed as

$$\begin{aligned} P_{\mathrm{dem},2} = \frac{G^2(A)}{A^4} \left|H_b(0)\right|^2 \Big\{ \\ \left[2(\mu-\lambda)^2(\Delta\omega)^2 + (1-\lambda)^2(2\pi\rho_2)^2\right] B_{N,2} S_{n^2}(0) + \\ 2\lambda^2 B_{N,3} S_{n,n}(0) \Big\}. \end{aligned} \tag{6.60}$$

The spectral intensities $S_{n^2}(0)$ and $S_{n,n}(0)$ can be determined exactly from the input noise spectrum, but can also be approximated from (6.56) and (6.57) with an error of usually less than 1 dB. If necessary, further refinement of these approximations is possible with the aid of Hermite polynomials [25].

When (6.56) and (6.57) are used, $P_{\mathrm{dem},2}$ can be expressed as

$$\begin{aligned} \frac{G^2(A)}{4p^2} \left|H_b(0)\right|^2 \frac{B_{N,2}}{2r\sqrt{\pi}} \Big\{ \\ (2\pi\rho_2)^2 \left[(1-\lambda)^2 + 2(\mu-\lambda)^2 \left(\frac{\Delta\omega}{2\pi\rho_2}\right)^2\right] \\ + 4\lambda^2 \frac{(2\pi r)^2}{\sqrt{2\left[\left(\frac{\rho_r}{r}\right)^2 - 1\right]}} \frac{B_{N,3}}{B_{N,2}} \Big\} \end{aligned} \tag{6.61}$$

As expected, this expression shows that the second-order noise power is inversely proportional to the squared input CNR; it decreases by 20 dB per decade of p.

6.4.5 Implications on Amplitude Compressor Design

Expression (6.61) contains very useful information for the synthesis of the amplitude compressor transfer: it allows derivation of the compressor transfer that minimizes the continuous demodulator output noise in terms of the FM wave characteristics and the parameters defined in Section 6.2.4. The derivation of this compressor transfer is the subject of this section.

In order to minimize the level of second-order noise, or even the total level of continuous noise observed at the demodulator output, two conditions on the parameters μ and λ that represent the amplitude compressor transfer have to be satisfied. These conditions, that follow from expression (6.47) and expression (6.60), are discussed below. As may be expected, it will be shown that the second-order noise is minimized by demodulators that apply infinite amplitude compression and those that apply no compression at all. In both cases, modulation of the amplitude compressor transfer by the noise, an important contribution to second-order noise, is absent.

Condition I: Minimization of the Second-Order Noise

It is clear from (6.61) that any amplitude compressor that minimizes the level of second-order noise has to satisfy the condition that

$$\begin{gathered}\mu - \lambda = 0 \Leftrightarrow \\ A^2 \frac{\mathrm{d}^2 G(A)}{\mathrm{d}A^2} - A\frac{\mathrm{d}G(A)}{\mathrm{d}A} = 0,\end{gathered} \tag{6.62}$$

where the latter equation is obtained by substitution of the definition equations (5.8) and (5.17) for $C_{n,1}(A) = 2\lambda$ and $C_{n,2}(A) = 2\mu$. This second-order differential equation for $G(A)$ is an example of an Euler equation, and has the general solution [27]

$$G(A) = c_0 A^0 + c_2 A^2, \tag{6.63}$$

where c_0 and c_2 are independent constants, that allow ***optimization*** of $G(A)$ to, for example, the input CNR and the various demodulator/filter parameters. However, it should be noticed that the absolute magnitude of the transfer $G(A)$ is not of interest in amplitude compressor design since it only represents an amplification of the entire receiver output signal. Therefore, the two independent parameters c_0 and c_2 result only in one degree of freedom in amplitude compressor design: the ratio C_2/c_0.

Thus, according to (6.63), any amplitude compressor transfer that establishes a linear combination of infinite compression (A^0) and no compression at all (A^2) satisfies the condition that $\mu = \lambda$, where

$$2\mu = C_{n,2}(A) = 2\lambda = C_{n,1}(A) = 2\frac{c_2 A^2}{c_2 A^2 + c_0}. \tag{6.64}$$

The only true degree of freedom left in this expression is the ratio c_2/c_0.

The presence of a degree of freedom in (6.63) explains the fact, that both demodulators that apply infinite compression, where $c_2/c_0 = 0$, and demodulators that apply no compression at all, where $c_2/c_0 \to \infty$, minimize the level of second-order noise and therefore satisfy (6.62). A general model for a demodulator that employs the amplitude compressor transfer from (6.63), and establishes an arbitrary level of compression, is depicted in figure 6.25. Notice the difference between this type of amplitude compression and compression established by a soft-limiter. The compressor in figure 6.25 *simultaneously* applies infinite compression, proportional to c_0, and no compression, proportional to c_2. The compressor output equals the linear combination of both components. No modulation of the amplitude compressor transfer occurs in this case, resulting in a low second-order noise level. A soft-limiter, however, establishes a certain finite level of compression by *alternation* of infinite compression and no compression,

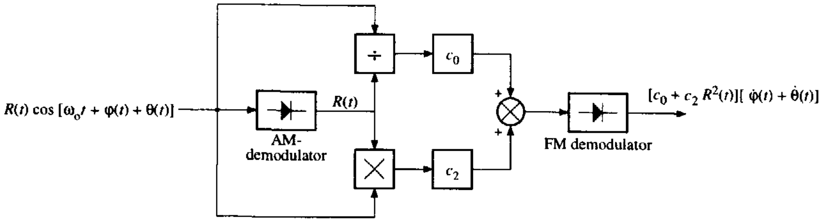

$R(t)\cos[\omega_0 t + \varphi(t) + \theta(t)]$
AM-demodulator
$R(t)$
÷
c_0
×
c_2
+
+
FM demodulator
$[c_0 + c_2 R^2(t)][\dot{\varphi}(t) + \dot{\theta}(t)]$

Figure 6.25: Demodulator configuration that minimizes the level of second-order noise.

under control of the limiter gain. Consequently, due to the frequent switching between both types of compression, considerable modulation noise has to be expected, resulting in a high second-order noise level.

Condition II: Minimization of the First-Order Noise

Once the second-order output noise has been minimized by application of the compressor transfer (6.63), the first-order output noise can be minimized by proper selection of the remaining degree of freedom: the ration c_2/c_0.

By proper design of the ratio c_2/c_0, the inverse first-order compression factor, represented by λ, can be selected such that it minimizes the total level of continuous noise $P_{\text{dem},1} + P_{\text{dem},2}$. As observed from (6.47) and (6.61), this is achieved when λ minimizes an expression of the form

$$\begin{aligned} P_{\text{dem}} &= P_{\text{dem},1} + P_{\text{dem},2} \\ &= \eta_0 + \eta_1\lambda^2 + \eta_2(1-\lambda)^2, \end{aligned} \tag{6.65}$$

where η_0, η_1, and η_2 are coefficients that follow from (6.47) and (6.60), or (6.61). The 'optimal' value of λ obtained from (6.65) equals

$$\begin{aligned} \lambda_{\text{opt}} &= \frac{\eta_2}{\eta_1 + \eta_2} \\ &= \frac{(2\pi\rho_2)^2 B_{N,2}\frac{S_{n^2}(0)}{A^4}}{\frac{2}{p}\frac{B_{N,0}}{B_{N,\text{IF}}}(\Delta\omega)^2 + 2B_{N,3}\frac{S_{n,n}(0)}{A^4} + (2\pi\rho_2)^2 B_{N,2}\frac{S_{n^2}(0)}{A^4}}, \end{aligned} \tag{6.66}$$

which corresponds to an optimal ratio c_2/c_0 equal to

$$\begin{aligned} \left.\frac{c_2}{c_0}\right|_{\text{opt}} &= \frac{\lambda_{\text{opt}}}{A^2(1-\lambda_{\text{opt}})} \\ &= \frac{(2\pi\rho_2)^2 B_{N,2}\frac{S_{n^2}(0)}{A^4}}{A^2\left[\frac{2}{p}\frac{B_{N,0}}{B_{N,\text{IF}}}(\Delta\omega)^2 + 2B_{N,3}\frac{S_{n,n}(0)}{A^4}\right]}. \end{aligned} \tag{6.67}$$

For example, when the IF filter and baseband filter are rectangular and have a bandwidth W_n and W respectively, (6.66) yields the optimal compression level obtained with the aid of (6.47) and (6.60),

$$\lambda_{\text{opt}} = \frac{4 - 3\frac{W}{W_n}}{8 - 4\frac{W}{W_n} + 4\left(\frac{W_n}{W}\right)^2 - 6\frac{W_n}{W} + 96p\left(\frac{\Delta\omega}{W}\right)^2}. \tag{6.68}$$

When e.g. $W_n = 12W$, and $\Delta\omega = 5W$, it follows that the optimal level of compression approaches infinity, i.e. $\lambda_{\text{opt}} \approx 0$, as should be expected; compression reduces the amplitude noise and therefore also the total continuous noise level.

In the presence of click noise, the optimal value of λ becomes a function of the input CNR, that may differ significantly from zero, as will be shown in Section 6.6.

6.5 Generalized Click Noise Model

In Section 5.4.3, it was shown that the click noise model provides a very compact description of the FM threshold effect, useful for engineering purposes. Further, finite compression of the demodulator input carrier amplitude, instead of the usually applied infinite compression, was proposed as a means to establish a trade-off between continuous noise and the perceptively very unpleasant click noise, in cases where operation above threshold cannot be guaranteed.

In order to judge this trade-off on its merits, a quantitative description for the amount of click noise observed at the demodulator output during finite amplitude compression is required. Since such a description is not provided by literature, this section will extend the 'Rician' click model [6] to include arbitrary types of amplitude compression.

An outline is as follows. Section 6.5.1 briefly reviews Rice's click model and shows that the required extension can be achieved by a modification of the click pulse area. Section 6.5.2 outlines the procedure followed in Sections 6.5.3 through 6.5.5 to obtain an expression for the click pulse area. The resulting model is subsequently applied to three FM demodulators with different types of amplitude compression, in sections 6.5.6 through 6.5.8.

6.5.1 Outline of the Model

As discussed in Section 5.4.3, the click model proposed in [5, 6] is a simplified description of a particular kind of noise excursions that result in a pulse in the frequency noise with significant low frequency contents. In this section, we focus on its mathematical formulation and the extension to arbitrary types of amplitude compression.

Click Noise Model for Infinite Compression

The original model, valid for infinite compression only, approximates the shape of the click pulses by a Dirac impulse, which according to [28–30] is a reasonable assumption as far as the spectral contents at baseband frequencies is concerned. Further, these pulses are considered to be uncorrelated with the continuous demodulator output noise, and with each other. The latter assumption is allowed around the threshold [31] due to the wide separation of click pulses in time. For example, in FM audio broadcasting, only a few (1-10) clicks per second are generated at the threshold CNR [6]. The mathematical consequence of these observations is that click noise can be described as a stochastic train of positive and negative impulses (corresponding to counter-clockwise and clockwise origin encirclements), distributed over time according to two independent Poisson processes t_k and t_l, with average rate N_+ and N_- respectively:

$$\dot{\theta}_{\text{click}}(t) = \sum_{k=-\infty}^{\infty} 2\pi\delta\left(t - t_k\right) - \sum_{l=-\infty}^{\infty} 2\pi\delta\left(t - t_l\right). \tag{6.69}$$

In this expression, the factor 2π represents the area of the click pulses (see Section 5.4.3). Thus in essence, the click model contains two key parameters: the click area and the average click rate. The white power spectral density of the click noise, given by (5.11), is entirely determined by these parameters.

Modifications for Finite Compression

In order to extend the model to finite amplitude compression, we should investigate in which way the click area and average click rate are affected by the level of compression.

As observed from the discussion in Section 5.4.3, the average rates N_+ and N_- are entirely determined by the amplitude compressor input signal $r(t) = s(t) + n(t)$, and are therefore independent of the applied type of compression. In fact, N_+ and N_- describe only the rate of cycles by which the noise $n(t)$ advances/delays the noisy FM wave $r(t)$ in comparison to the noise free FM wave $s(t)$, which is entirely independent of the applied type of amplitude compression. Consequently, it is legitimate to use these same rates N_+ and N_- in an extended click model.

The observation in [5, 6] that the area of a click pulse at the FM demodulator output equals 2π is entirely based on the assumption that the demodulator output signal equals the instantaneous frequency of the FM input wave $r(t)$, i.e. that infinite compression is applied. When finite compression is applied, however, this is no longer true, as observed from (6.13). In that case, the output signal, and therefore also the click pulse area observed at the demodulator output, becomes dependent on the input amplitude $R(t)$. As a consequence, in case of finite compression, the area of the click pulses is generally not equal to 2π, and, in addition, differs among the click pulses. For this case, the factor 2π in expression (6.69) should therefore be replaced by a stochastic variable ξ_k, representing the area of the pulse generated in $G\left[R(t)\right]\dot{\theta}(t)$ at the instant $t = t_k$. Further, it should be questioned whether the clicks still resemble Dirac impulses, or become widened. The Dirac impulses in (6.69) should therefore generally be replaced by pulses $p_k(t)$ of a stochastically determined shape and unity area. The click noise in case of finite compression may therefore be expressed as

$$G[R(t)]\dot{\theta}_{\text{click}}(t) = \sum_{k=-\infty}^{\infty} \xi_k p_k\left(t - t_k\right) - \sum_{l=-\infty}^{\infty} \xi_l p_l\left(t - t_l\right). \tag{6.70}$$

Spectral Density for Finite Compression

The power density spectrum corresponding to the stochastic pulse train in(6.70) generally differs from (5.11) for the infinite compression case. It is known [6, 30] that as long as the clicks are independent, the double-sided click noise spectral density equals the average total click rate $N_+ + N_-$ times the (average) power spectral density of the pulses $p_k(t)$[1].

Although the resulting spectrum is generally colored instead of white, its shape at low frequencies, i.e. in the baseband region, will still closely resemble the spectrum of a Dirac impulse train. This was also observed in [30], where the Dirac impulses in the click noise where replaced by Gaussian pulses. The differences in shape basically result in discrepancies at high frequencies, i.e. above the baseband.

Therefore, in order to model the effect of finite compression on the click noise at baseband frequencies, the click pulses can be appropriately represented as Dirac impulses with area ξ, equal to the average area of the individual pulses. The corresponding spectrum may therefore be expressed at baseband frequencies as

$$S_{\text{click}}(\omega) \approx \xi^2 \left(N_+ + N_-\right). \tag{6.71}$$

Thus, according to this expression, the click noise power density spectrum in case of finite compression is easily obtained once the average click pulse area ξ is known. Determination of this area is the subject of the subsequent sections.

6.5.2 Concept of the Procedure

This section summarizes the procedure followed in sections 6.5.3 through 6.5.5 to determine the average click pulse area for arbitrary types of compression

[1] In [6, 30], the spectrum equals the average rate times the squared modulus of a deterministic pulse spectrum. This squared-modulus spectrum is in fact equivalent to the power density of a stochastic pulse.

applied to the demodulator input FM wave.

The area ξ is an average over all clicks observed sequentially in time. Such a time average is generally hard to compute since the start and stop instants of the pulses are difficult, if not impossible, to identify from the demodulator output signal.

Therefore, the procedure discussed in this section uses the (assumed) property of ergodicity to replace the time average by an ensemble average over the input noise $n(t)$. The assumption of ergodicity is usually allowed, since nearly all physical noise processes are ergodic. Thus, by this assumption, the time average ξ is assumed to equal the area of a single click pulse, averaged over the ensemble of the input noise $n(t)$. In this way, the problem can be defined in terms of a set of time-independent stochastic variables and their joint a PDF: instead of a time-average over all click pulses, the average click pulse area is obtained from an ensemble average of a single pulse, where the average is taken over the input noise $n(t)$.

Another observation that significantly simplifies the procedure is that the click noise model, although not described by Rice [6] in this way, is essentially defined in terms of polar coordinates; the condition for the occurrence of a click is that noise *encirclements* have a radius larger than the FM phasor $\vec{s}$. For that reason, Section 6.5.3 formulates the problem in terms of polar coordinates with the origin located at the tip of the noise-free FM phasor $\vec{s}$. The noise phasor $\vec{n}$ is described by means of its (stochastic) radius R_n and phase φ_n relative to $\vec{s}$, as depicted in figure 6.26. In this figure, $\vec{v}$ denotes the velocity vector of the noise $\vec{n}$.

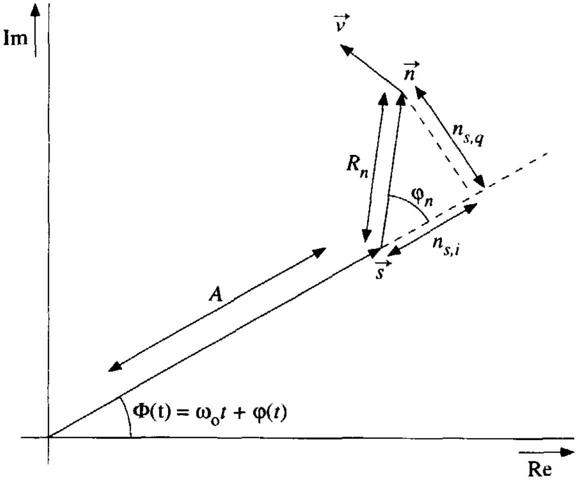


Figure 6.26: Amplitude compressor input signal $r(t) = s(t) + n(t)$ expressed in polar coordinates.

The calculation of the average click pulse area ξ, that employs the polar model of the demodulator input wave depicted in figure 6.26, is schematically outlined in figure 6.27. The procedure starts from the general expression for

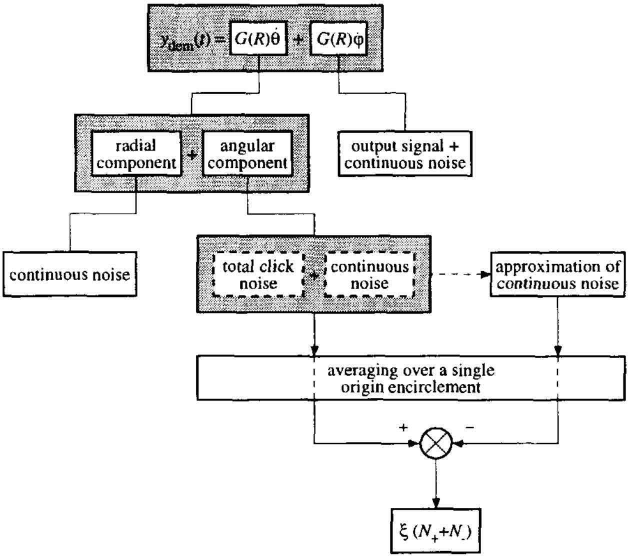


Figure 6.27: Calculation of the average click pulse area ξ.

the demodulator output signal, given by (6.13). In this expression, the term $G(R)\dot{\varphi}(t)$ can be ignored since it contains amplitude noise only. Click noise is part of the demodulator output frequency noise, and is therefore entirely included in $G(R)\dot{\theta}(t)$.

With the aid of the polar representation of the input signal, $G(R)\dot{\theta}(t)$ can be decomposed into a radial component that represents the fluctuations in the length of the noise phasor, and an angular component that contains the rotational movements of $\vec{n}$. As will be shown in Section 6.5.3, the radial component, which is proportional to the time-derivative $\dot{R}_n$ of R_n, does not contribute to the click noise since it does not result in encirclements of $\vec{n}$ around the origin. Consequently, the click noise is entirely included in the angular component, which is proportional to the time-derivative $\dot{\varphi}_n$ of φ_n.

An investigation of the click pulse structure, considered in Section 6.5.4, shows that the angular demodulator output noise component generally consists of a mixture of click noise and continuous frequency noise. As described in Section 6.5.5, the average click noise pulse area ξ is derived from the average value of this component, calculated over one encirclement of $\vec{n}$ around the origin. The contribution of the continuous noise to this average can be approximated from the first and second-order noise components derived in sections 6.3 and 6.4. Subtraction of this continuous component from the total average angular component then yields the contribution of the click noise. Further, in advance of the averaging procedure, the angular noise frequency $\dot{\varphi}_n$, that describes the click rate, is replaced by its absolute value $|\dot{\varphi}_n|$ in order to prevent a zero result for the average click noise contribution. Without this substitution, the average click noise equals $\xi\,(N_+ - N_-)$, i.e. the average area times the net click rate. However, since clockwise and counter-clockwise rotations of the noise phasor are usually equally likely, i.e. $N_+ = N_-$, this average usually equals zero. The substitution of $\dot{\varphi}_n$ with its absolute value $|\dot{\varphi}_n|$ replaces the net click rate by the total click rate $N_+ + N_-$, such that the average total click noise becomes $\xi\,(N_+ + N_-)$. The average click area is finally obtained through division by the total click rate $N_+ + N_-$, which is known already from Rice's theory, or, alternatively, can be derived from the extended theory for the case of infinite compression, where the area is known to be $\xi = 2\pi$.

6.5.3 Demodulator Output Noise in Polar Format

In order to write the part of the demodulator output noise that contains the click noise, $G[R(t)]\dot{\theta}(t)$, in polar format, the components of the input noise $n(t)$ are expressed as

$$n_{s,i}(t) = R_n(t)\cos\varphi_n(t), \tag{6.72}$$

$$n_{s,q}(t) = R_n(t)\sin\varphi_n(t). \tag{6.73}$$

Thus, the noise encircles the tip of the FM phasor $\vec{s}$ with a radius R_n. Substitution of these equations into (2.16) yields an expression for the amplitude $R(t)$ in terms of R_n and φ_n. In order to obtain an expression for $\dot{\theta}$, this frequency noise component can be expressed in terms of the polar noise components R_n, φ_n and their derivatives as

$$\frac{\mathrm{d}\theta(t)}{\mathrm{d}t} = \frac{\partial\theta}{\partial R_n}\dot{R}_n + \frac{\partial\theta}{\partial\varphi_n}\dot{\varphi}_n. \tag{6.74}$$

The first component on the right hand side (RHS) of this expression corresponds to radial movements (it is independent of $\dot{\varphi}_n$), caused by the time-dependency of $R_n(t)$, while the second component corresponds to angular movements of fixed radius R_n (it is independent of $\dot{R}_n$). Finally, the relevant component of the demodulator output noise may be expressed as

$$\begin{aligned} G\left[R(t)\right]\dot{\theta}(t) &= G\left(\sqrt{R_n^2 + 2AR_n\cos\varphi_n + A^2}\right)\left(\frac{\partial\theta}{\partial R_n}\dot{R}_n + \frac{\partial\theta}{\partial\varphi_n}\dot{\varphi}_n\right) \\ &= G\left(R_n,\varphi_n\right)\frac{\partial\theta}{\partial R_n}\dot{R}_n + G\left(R_n,\varphi_n\right)\frac{\partial\theta}{\partial\varphi_n}\dot{\varphi}_n. \end{aligned} \tag{6.75}$$

Both partial derivatives of this expression can be determined with the aid of (2.17), (6.72) and (6.73) as

$$\frac{\partial\theta}{\partial\varphi_n} = \frac{R_n^2 + AR_n\cos\varphi_n}{A^2 + 2AR_n\cos\varphi_n + R_n^2}. \tag{6.76}$$

$$\frac{\partial\theta}{\partial R_n} = \frac{AR_n\sin\varphi_n}{A^2 + 2AR_n\cos\varphi_n + R_n^2}. \tag{6.77}$$

From (6.77) it follows that the radial component of the demodulator output noise, the first term in (6.75), does not contribute to the click noise and can therefore be ignored in the conclusion. Mathematically, this follows from the observation that (6.77) is proportional to $\sin\varphi_n$, while the amplitude transfer $G(R)$ fluctuates synchronously with $\cos\varphi_n$. The phase noise φ_n is uniformly distributed, as shown in Appendix F, which means that the average contribution of the first term in (6.75) to a single encirclement of the origin equals zero.

Consequently, the click noise, and the average click area have to be determined from the angular component of the demodulator output noise, denoted by $n_{o,\mathrm{ang}}(t)$, which can be expressed as

$$\begin{aligned} n_{o,\mathrm{ang}}(t) &= G\left(R_n, \varphi_n\right)\frac{\partial\theta}{\partial\varphi_n}\dot{\varphi}_n \\ &= G\left(\sqrt{R_n^2 + 2AR_n\cos\varphi_n + A^2}\right)\frac{\partial\theta}{\partial\varphi_n}. \end{aligned} \tag{6.78}$$

6.5.4 Angular Demodulator Output Noise Components

The angular demodulator output noise $n_{o,\mathrm{ang}}(t)$ generally consists of two contributions; click noise and the angular component of the continuous demodulator output noise. In order to obtain an expression for the average click area, both components need to be separated. For that purpose, this section separately investigates the three factors that constitute the angular demodulator output noise $n_{o,\mathrm{ang}}(t)$.

Noise Frequency

The factor $\dot{\varphi}_n$ in $n_{o,\mathrm{ang}}(t)$ equals the radial frequency of the input noise $n(t)$, which represents the angular movements of the noise phasor $\vec{n}$. The other factors in the angular noise term depend only on the polar position coordinates R_n and φ_n, and serve only as a position dependent weighting factor. Since $\dot{\varphi}_n$ represents both clockwise and counter-clockwise motion, we replace it with its modulus $|\dot{\varphi}_n|$ in all subsequent calculations in order to assure that the final result contains the total click rate $N_+ + N_-$ instead of the net rate $N_+ - N_-$, which is usually zero.

Click Pulse Shape in case of Infinite Compression

The factor $\frac{\partial\theta}{\partial\varphi_n}$, given by expression (6.76), represents the shape of a single click pulse in case of infinite compression (where $G(R_n, \varphi_n)$ is a constant), as a function of the polar position coordinates R_n and φ_n. Note that the time t is not explicitly included in this expression: the click pulse is expressed as function of the stochastic variables R_n and φ_n only. The average (expected) value of this expression is considered to be equal to the time average of the click pulse area, on the basis of the assumed property of ergodicity. An investigation of the composition of such click pulses yields important information about the contribution of the continuous noise to the angular demodulator output noise, as shown below.

A remarkable property of (6.76) is observed after calculation of the expectation over the uniformly distributed phase φ_n, i.e. over one encirclement of the noise $\vec{n}$, that identifies it as the click pulse in case of infinite compression. The result is

$$\mathrm{E}\left\{\frac{\partial\theta}{\partial\varphi_n}\right\}_{\varphi_n} = \frac{1}{2}\left[1 - \mathrm{sgn}\,(A - R_n)\right]. \tag{6.79}$$

This expression clearly demonstrates that only noise encirclements with a radius larger than A, i.e. those that encircle the origin, yield a nonzero contribution to the frequency noise $\theta(t)$ during one encirclement of the input noise $\vec{n}$. This is exactly the property that distinguishes a click excursion of $\vec{n}$ from any other noise excursion. Subsequently, by taking the expectation over $|\dot{\varphi}_n|$ of this expression, the click noise contribution becomes proportional to $2\pi r$, i.e. the click area times the radius of gyration, which is a measure of the average number of encirclements of $\vec{n}$.

A deeper insight into the structure of the click pulse is obtained by plotting expression (6.76) as function of φ_n. Figure 6.28 depicts the result when $R_n = 0.9A$, which according to (6.79) corresponds to a doublet-pulse of zero net area. Figure 6.29 depicts the result when $R_n = 1.1A$, corresponding to a click pulse of area 2π. Note that both plots contain the time t as an implicit parameter only via φ_n: every value of φ_n corresponds to one time instant during the occurrence of a click pulse. Further, the relation between φ_n and t is generally nonlinear. Usually, the part of the excursion located in the left half-plane is traversed in a very short time, which means that the time-axis around $\varphi_n = \pi$ is expanded in comparison to the φ_n-axis. The part of the excursion around $\varphi_n = 0$, however, is usually traversed relatively slowly which means that the time-axis is compressed around $\varphi_n = 0$.

Concerning the composition of doublets and clicks, it can be seen from figures 6.28 and 6.29 that both the doublet and the click consist of two components:

- an impulsive component of area π;

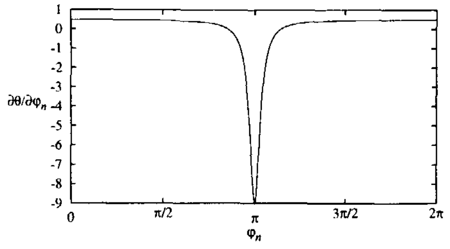


Figure 6.28: Doublet pulse as a function of φ_n, obtained from (6.76) for $R_n = 0.9A$.

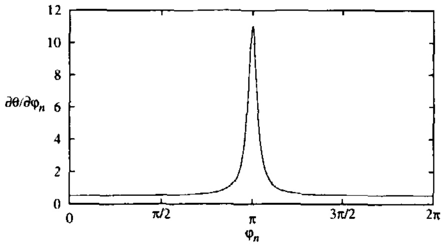


Figure 6.29: Click pulse as a function of φ_n, obtained from (6.76) for $R_n = 1.1A$.

- a phase-independent component of area π.

In the doublet pulse, both components are of *opposite polarity*, resulting in an exactly zero net area. In the click pulse, however, both components have *the same polarity*, resulting in a net area of 2π.

The phase independent component of the click is actually the average value of the continuous noise during the click. This follows from the fact that it is present for all values of the noise radius R_n, i.e. not just for $R_n > A$, and all values for φ_n. True click noise, as the impulsive component, is present only in the neighborhood of phase values $\varphi_n = \pi$. This observation has some consequences for the modeling of click noise in case of finite compression. In that case, part of the continuous component of the click is already contained in the second-order continuous noise model described in Section 6.4. This contribution obviously needs to be removed from the click noise component that is used for the calculation of ξ, in a way that is discussed in Section 6.5.5.

Amplitude Compressor Transfer

The first factor, the amplitude compressor transfer $G(R_n, \varphi_n)$, serves as weighting factor for both components of the click pulses. It (usually) partially suppresses the impulsive click component, at the cost of a usually somewhat enhanced continuous component. This is considered in depth in subsequent sections.

6.5.5 Expression for the Average Click Pulse Area

This section derives an expression for the average click pulse area, with the aid of the polar description derived in Section 6.5.3, and the investigation of the click pulse shape discussed in Section 6.5.4.

In essence, the average click pulse area is determined from the average value of the angular demodulator output noise. As observed in Section 6.5.4, the angular noise consists of an impulsive (click) component and a continuous component. The continuous component is partially already contained in the models for the continuous output noise, discussed in sections 6.3 and 6.4. Subtraction of these components finally yields the average total click noise that yields the average click pulse area.

Average Total Angular Noise

The average angular output noise equals the expected value of the angular demodulator output frequency component in (6.75), with $\dot{\varphi}_n$ replaced by $|\dot{\varphi}_n|$, in order to assure a nonzero result. Besides the total average click noise, equal to $\xi\left(N_+ + N_-\right)$, this expectation also contains a contribution of the average continuous noise, denoted by μ_c. The expectation can therefore be written as

$$\mathrm{E}\left[G\left(R_n, \varphi_n\right) \frac{\partial\theta}{\partial\varphi_n} |\dot{\varphi}_n|\right]_{R_n, \dot{R}_n, \varphi_n, \dot{\varphi}_n} = \xi\left(N_+ + N_-\right) + \lambda_c. \tag{6.80}$$

Since $G\left(R_n, \varphi_n\right)$ and $\frac{\partial\theta}{\partial\varphi_n}$ are independent of $\dot{R}_n$ and $\dot{\varphi}_n$, two of the four expectation operations in (6.80) can already be elaborated without loss of generality. This yields

$$\xi\left(N_+ + N_-\right) + \lambda_c = \mathrm{E}\left[G\left(R_n, \varphi_n\right) \frac{\partial\theta}{\partial\varphi_n} \mathrm{E}\left(|\dot{\varphi}_n| \big| R_n, \varphi_n\right)_{\dot{R}_n, \dot{\varphi}_n}\right]_{R_n, \varphi_n}. \tag{6.81}$$

The inner expectation in this expression denotes the average absolute value of the radial noise frequency as a function of the phasor-plane position coordinates R_n and φ_n. Therefore, it represents the average total click rate. The other two factors in (6.81) represent the angular frequency noise in case of infinite compression, and the amplitude compressor transfer that serves as a weighting factor.

In Appendix F is shown that the inner expectation equals

$$\mathrm{E}\left(|\dot{\varphi}_n| \big| v, \varphi_n\right)_{\dot{R}_n, \dot{\varphi}_n} = 4r\sqrt{\pi p} \exp\left[-pv^2\left(1+u^2\right)\right] + 2rpuv \exp\left(-pv^2\right) \operatorname{erf}\left(uv\sqrt{p}\right), \tag{6.82}$$

where r denotes the radius of gyration defined by (5.13), $v = R_n/A$, and $u = \dot{\varphi}/(2\pi r)$ represents the FM message signal, normalized to the average zero-crossing rate of the input noise, and p denotes the demodulator input CNR. Similar to the procedure followed in [6], a stochastic message signal can be included in the expressions by taking the expectation of (6.82) over $\dot{\varphi}$. In the absence of modulation, $u = 0$ and (6.82) reduces to a more tractable Gaussian-like function.

Note that (6.82) differs from the corresponding expression for the click rate "$H_+ (t_1)$" in [6], since in that reference an approximation is applied that obtains the click rates from an event-crossing problem (see Section 5.4.3). Expression (6.82) contains no such approximation and therefore yields the exact click rate.

Average Continuous Angular Noise

The contribution of the continuous angular noise to (6.80), denoted by μ_c, can be approximated with the aid of the expressions for the first and second-order continuous noise derived in sections 6.3 and 6.4 in the following way.

A second-order approximation for the average continuous, angular demodulator output noise corresponds to the terms in (6.42) and (6.51) that, after substitution of (6.72) and (6.73), are proportional to the noise frequency $\dot{\varphi}_n$. Only these components yield a contribution to (6.80). The only terms in (6.42) and (6.51) that satisfy this condition are those that contain the noise component $\dot{n}_{s,i}(t)$ or $\dot{n}_{s,q}(t)$. This leaves only three terms; the first order term $\dot{n}_{s,q}(t)$, and the second-order terms $n_{s,i}(t)\dot{n}_{s,q}(t)$, and $\dot{n}_{s,i}(t)n_{s,q}(t)$.

By substitution of (6.72) and (6.73), it follows that μ_c can be expressed as

$$\begin{aligned}\lambda_c &\approx G(A)\mathrm{E}\left\{\left[\frac{R_n}{A}\cos\varphi_n - \frac{R_n^2}{A^2}\cos 2\varphi_n + C_{n,1}(A)\frac{R_n^2}{A^2}\cos^2\varphi_n\right]|\dot{\varphi}_n|\right\} \\ &= \mathrm{E}\left[G(A)C_{n,1}(A)\frac{R_n^2}{2A^2}\mathrm{E}\left(|\dot{\varphi}_n||R_n\right)\right]_{R_n} .\end{aligned} \tag{6.83}$$

Thus, the continuous noise yields a nonzero contribution whenever finite compression is applied, i.e. when $C_{n,1}(A)$ is nonzero.

Note that the latter expectation in (6.83) over the noise phase φ_n can be determined freely, since the expectation (6.82) is independent of φ_n.

Total Average Click Noise

An approximate expression for the total average value of the click noise is obtained by subtraction of (6.83) from (6.81).

It should further be noted, that a click is generated only when the noise radius R_n exceeds A, i.e. when $v > 1$. Only in that case does a rotation of the noise phasor $\vec{n}$ encircle the origin. Therefore, all expectations over R_n are taken over the interval $R_n \in [A, \infty)$. In this way, we arrive at

$$\begin{aligned}\xi\left(N_+ + N_-\right) &= \int_1^\infty \mathrm{E}\left[\xi\left(N_+ + N_-\right) + \mu_c|\, v\right]_{\varphi_n} - \mathrm{E}\left(\mu_c|\, v\right)_{\varphi_n} \mathrm{d}v \\ &= \int_1^\infty \left\{\mathrm{E}\left[G\left(Av, \varphi_n\right)\frac{\partial\theta}{\partial\varphi_n}|v\right]_{\varphi_n} - G(A)C_{n,1}(A)\frac{v^2}{2}\right\}\mathrm{E}\left(|\dot{\varphi}_n||v\right)\mathrm{d}v.\end{aligned} \tag{6.84}$$

The total click rate $N_+ + N_-$ in this expression can be approximated by the existing theory given in [6], or determined exactly from

$$N_+ + N_- = \frac{1}{2\pi} \int_1^\infty \mathrm{E}\left(|\dot{\varphi}_n| | v\right) \mathrm{d}v, \tag{6.85}$$

which follows from (6.79), (6.82) and (6.84) for the case of infinite compression, i.e. $C_{n,1}(A) = 0$. The use of (6.84) is demonstrated for some relevant examples in the subsequent sections.

6.5.6 Application to Infinite Compression

As an illustration, this section applies the extended click noise model derived in sections 6.5.3 through 6.5.5 to the (trivial) case of infinite compression.

In the case of full compression, the amplitude compressor transfer $G(R)$ is a proportionality constant that we assume to equal unity. Consequently, the first-order inverse compression factor $C_{n,1}(A)$ vanishes, which means according to (6.83) that in this case the continuous demodulator output noise, as described in sections 6.3 and 6.4, does not contribute to the expected angular output noise component. Further, from (6.79) it is observed that the expected value of $G\left(R_n, \varphi_n\right) \frac{\partial \theta}{\partial \varphi_n}$ over φ_n, used in (6.84) equals unity over the entire integration interval of v, and zero elsewhere.

In order to simplify the calculations, we assume in this case that no modulation is present, i.e. that $u = 0$. In that case, expression (6.84) becomes, after substitution of (6.82),

$$\begin{aligned} \xi\left(N_+ + N_-\right) &= \int_1^\infty 4r\sqrt{\pi p} \exp\left[-pv^2\right] \mathrm{d}v \\ &= 2\pi r \left[1 - \operatorname{erf}\left(\sqrt{p}\right)\right]. \end{aligned} \tag{6.86}$$

According to the exact result for $N_+ + N_-$ given by (5.12), the obtained result equals exactly 2π times the total click rate for the case of infinite compression, which means that we obtain $\xi = 2\pi$, as expected.

6.5.7 Application to No Compression

As a second example, this section applies the generalized click model to another trivial example; the case in which no compression is applied. In this case, there is obviously no click noise, which means that the generalized model should yield a zero average click pulse area.

The "amplitude compressor" transfer in this case is proportional to the square of the input FM amplitude $R(t)$. Thus, for the angular component of the output noise we obtain from (6.75) and (6.76),

$$G\left(R_n, \varphi_n\right) \frac{\partial \theta}{\partial \varphi_n} = R_n^2 + AR_n \cos\varphi_n = A^2\left(v^2 + v\cos\varphi_n\right). \tag{6.87}$$

The expected value of this expression over φ_n equals $A^2v^2 = G(A)v^2$.

Further, from (5.8) the compression factor $C_{n,1}(A) = 2$. Thus, the contribution of the continuous noise to the expected value of (6.87), obtained from (6.83) also equals $G(A)v^2$. Thus, in this case the average value of the angular output noise is completely determined by the continuous noise. Consequently, according to (6.84), the total average click noise $\xi\,(N_+ + N_-)$ observed at the output equals zero. The average click pulse area equals $\xi \equiv 0$, as expected, since no click noise is generated.

6.5.8 Application to Finite Compression: Soft-Limiter

This section applies the generalized click model to an FM demodulator preceded by a soft-limiter of the type depicted in figure 6.13. The amount of click noise produced by an FM demodulator with this type of amplitude compression cannot be described by the Rician click model. A characteristic property of finite compression is that the click pulse area generally becomes a function of the input CNR. The effect is explained for a soft-limiter with the aid of a phasor representation. Subsequently, numerical results obtained with the extended model are discussed.

Dependence of the Click Area on the Input CNR

A characteristic property of systems that employ finite compression is that the average click area generally becomes a function of the input CNR as a result of the FM input carrier amplitude's contribution to the output signal. For a soft-limiter, this effect can be explained with the aid of figure 6.30 that depicts the motion of the noisy FM phasor $\vec{r}$ relative to the motion of the noise-free FM wave $\vec{s}$. The circle of radius K around the origin in this figure represents the linear operating region of the limiter. If $R(t) < K$, the demodulator output signal is proportional to the square of the input signal, while it quickly saturates for larger values, as described by (6.11).

At high input CNRs, the click excursions of $\vec{n}$ closely encircle the origin, as explained in Section 5.4.4. Consequently, nearly all these click paths cross through the linear operating region of the limiter, which causes a significant part of the corresponding click pulse at the demodulator output to be suppressed. At low CNRs the click excursions penetrate deeper into the left half plane (LHP), until they finally completely encircle the linear region. In the latter case, no click suppression at all occurs and the demodulator response becomes similar to the response obtained in the case of infinite compression.

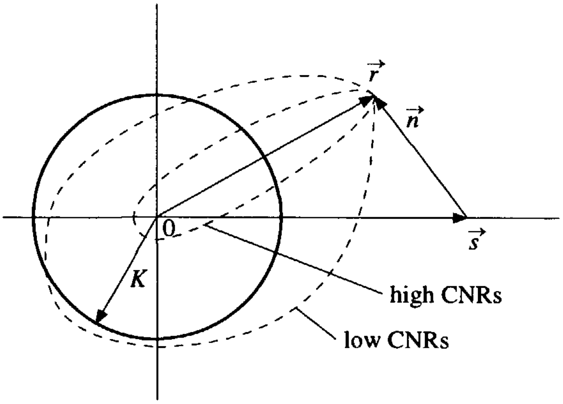


Figure 6.30: Phasor plane representation of clicks in a soft-limiting FM demodulator.

Numerical Results

Numerical results for this system were obtained for the case of no modulation, i.e. $u = 0$, by expressing the soft-limiter transfer $G(R) = G_{\mathrm{sl}}^2(R)$ from (6.11) in terms of the noise parameters v and φ_n, and application of numerical integration to (6.84).

The effect of the limiter gain, represented in the calculations by the "inverse over-drive" $x = K/A$ (see figure 6.13), which is zero for infinite compression on the shape of the click pulses is illustrated by figure 6.31. The curves in this figure were obtained by averaging over the noise radius R_n, for several values of the noise phase φ_n. It is clearly demonstrated by this figure that when the limiter gain decreases, the impulsive component of the click is significantly suppressed. This agrees with the model in figure 6.30. There, a decrease of the gain results in a larger radius of the circle around the origin, and consequently in an increased probability that the click excursion crosses through this circle. Note that it is possible that the curves in this figure become negative for certain values of φ_n; this was the cause of the zero net area of a doublet pulse.

Figure 6.32 depicts the average click area as a function of the limiter gain for input CNRs of 10, 5 and 0 dB. This plot clearly shows the reduction of the click pulse area obtained by finite compression, i.e. a finite gain. For small input CNRs, the reduction becomes less effective due to the on average larger radius of the click excursions of the noise phasor $\vec{n}$.

Finally, figure 6.33 depicts the click noise power observed at the demodulator output as function of the input CNR, normalized on the power observed in case of infinite compression. For large limiter gain values, the click noise power quickly approaches the power observed in the case of infinite compression when the input CNR decreases. This is due to the increased average length of the

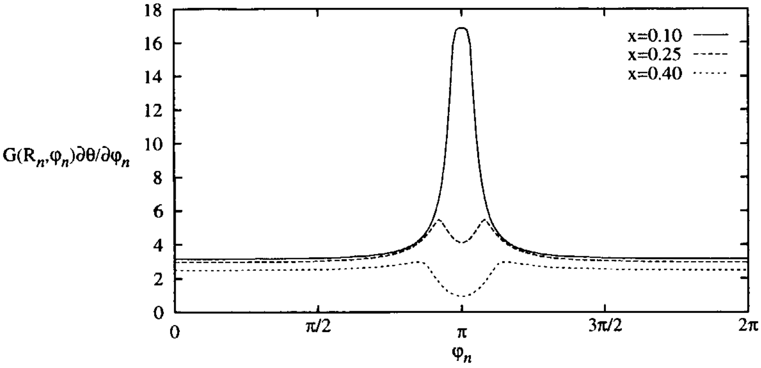


Figure 6.31: Click pulse shape as a function of the inverse limiter overdrive $x = K/A$ at an input CNR of 8.5 dB.

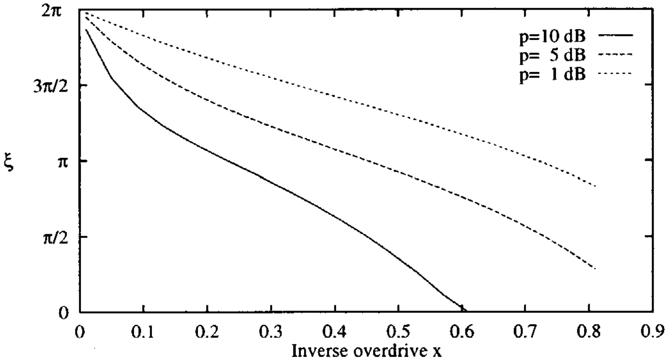


Figure 6.32: Average click pulse area as a function of the inverse limiter overdrive $x = K/A$ for several input CNRs.

noise phasor $\vec{n}$, and the correspondingly increased probability that the noise excursions completely encircle the circle of the linear region in figure 6.30. For small limiter gain values, the increase starts at a much lower input CNR due to the larger linear operating region of the limiter in that case. Below 0 dB, the curves become invalid since the inverse compression factor $C_{n,1}(A)$ no longer accurately describes the level of continuous noise. For CNRs above 0 dB, the curves of figure 6.33 closely resemble the curves given in [32–34], which correspond to the probability that the noise excursions exceed the bounds of the limiter's linear operating region in the LHP.

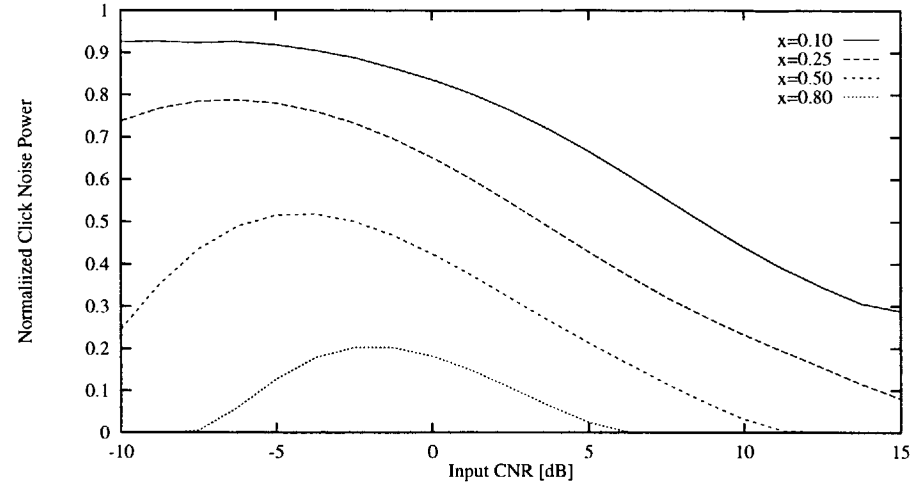


Figure 6.33: Normalized click noise power as a function of the input CNR, for various limiter gain values.

6.6 Output Signal-to-Noise Ratio

The quality of the demodulator response to a noisy FM wave is usually expressed in terms of the signal-to-noise ratio, i.e. the ratio of the signal power to the noise power observed at the demodulator output. This criterion is especially suited as a measure of the deterioration of the intelligence due to continuous noise. It is however questionable whether the SNR is also a suitable measure of the deterioration of the intelligence due to click noise. Although click noise is only present in the demodulator output signal for a very small fraction of time, it produces a tremendous amount of noise energy during this time, which completely dominates the demodulator output signal. The SNR, as a measure of the signal quality, spreads this energy over time, and thereby probably underestimates the deteriorative effect.

However, due to the lack of a more convenient measure, this section uses the receiver output SNR as a measure of the output signal quality, and describes its dependence on the transfer characteristic of the amplitude compressor and the parameters that characterize the IF filter and the baseband output filter. If necessary, perception can be included in this expression by assigning different weighting factors to the various noise components. In this section, however, such weighting factors are omitted. The results obtained in previous sections are summarized, and subsequently combined to determine the 'optimum' amplitude compressor transfer that maximizes the demodulator output SNR as a function of the input CNR.

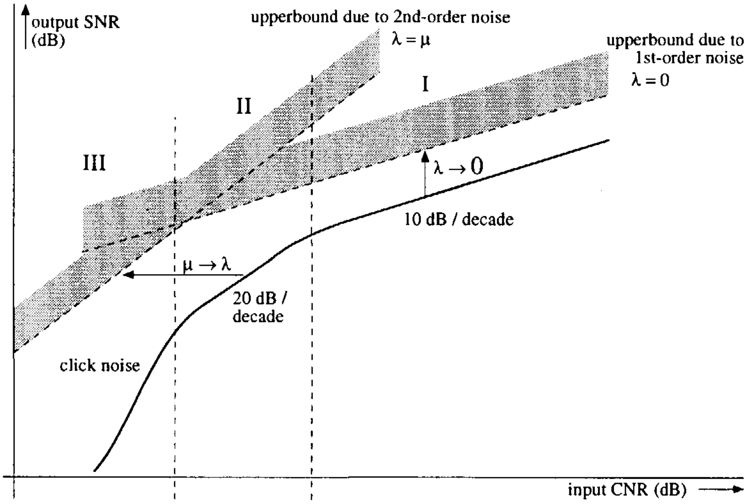


Figure 6.34: Threshold curve of FM receivers.

Section 6.6.1 considers the dependence of the threshold curve, i.e. the output SNR versus input CNR, on the amplitude compressor transfer and the filter parameters. Section 6.6.2 derives the optimum amplitude compressor.

6.6.1 Threshold Curves

As discussed previously, the demodulator output SNR is generally characterized by three types of noise: first and second-order continuous noise, and click noise. Each of these noise contributions affects a different part of the demodulator threshold curve and is determined by different demodulator parameters.

This section briefly describes the effect of each noise component on the threshold curve, and summarizes the selections of both amplitude compression factors $C_{n,1}(A)$ and $C_{n,2}(A)$, or, equivalently, λ and μ, that minimize the contributions to the demodulator output noise power. Finally, a general expression for the output SNR is discussed.

Throughout the discussion, the threshold curve sketched in figure 6.34 is used as a reference.

First-Order Noise

First-order noise determines the demodulator output SNR at high input CNRs, when the noise is small compared to the input FM wave, i.e. region I in figure 6.34. In this region, the output SNR increases by 10 dB per decade of the input CNR.

The first-order output noise power level is determined by the first-order inverse compression factor $C_{n,1}(A) = 2\lambda$. The maximum possible SNR is attained for $C_{n,1}(A) = 0$, corresponding to the case in which all amplitude noise is suppressed.

As shown in Section 6.3.4, the maximum SNR is approached within 3 dB when the established level of compression exceeds the *critical level of compression*, i.e. when

$$C_{n,1}(A) < \frac{2\pi\rho_0}{\Delta\omega}. \tag{6.88}$$

Below this compression level, i.e. for large $C_{n,1}(A)$, the output SNR decreases by 6 dB per octave of $C_{n,1}$.

Second-Order Noise

Second-order noise introduces an asymptote into the output SNR that increases by 20 dB per decade of the input CNR, region II in figure 6.34, and dominates at intermediate or low CNRs.

In receivers with finite amplitude compression, such as those equipped with a soft-limiter, second-order noise is mainly due to modulation of the first-order compression factor $C_{n,1}(A)$ by the noise, and may already become noticeable at CNRs of 10-15 dB.

It was shown in Section 6.4.4 that minimization of the second order noise requires the first and second-order inverse compression factors to be equal. Under that condition, an optimal value for the compression factors $\mu = \lambda$ can be determined, that shifts the asymptote in figure 6.34 as far to the left as possible. The condition $\mu = \lambda$ can only be satisfied by means of compressors that establish infinite compression, no compression, or linear combinations of these two, i.e.

$$G(A) = c_0 + c_2 A^2. \tag{6.89}$$

The ratio of the constants c_0 and c_2 can be used to establish the optimum values of the first-order (and second-order) inverse compression factor, as a function of the input CNR.

Click Noise

Click noise introduces a steep, exponential decay of the output SNR when the input CNR drops below the threshold, as depicted in region III of figure 6.34.

In case of infinite compression, the click pulse area equals 2π, while in the absence of compression, click noise is absent, corresponding to a zero click pulse area. As discussed in the previous sections, the area is generally dependent on the input CNR in case of finite compression. At low CNRs, it approaches 2π, while at high CNRs it approaches zero.

With the aid of (6.71) and (6.38), the amount of click noise power observed at the baseband filter output can be expressed as

$$P_{\text{clk}} = \xi^2 \left(N_+ + N_-\right) \left|H_b(0)\right|^2 B_{N,1}, \tag{6.90}$$

where the click area ξ can be obtained from (6.84). Expression (6.84) also shows that ξ is proportional to $G(A)$, i.e. the value of the compressor transfer at the 'quiescent point' $R(t) = A$. The click rates N_+ and N_- are not affected by the level of compression, as discussed previously.

Expression for the Output SNR

An expression for the receiver output SNR, observed at the output of the baseband filter, is obtained as follows.

The message signal $\dot{\varphi}(t)$ is assumed to be completely passed by the baseband output filter. According to (6.13), the signal power can therefore be expressed as

$$P_s = G^2(A) \left|H_b(0)\right|^2 (\Delta\omega)^2, \tag{6.91}$$

where $\Delta\omega$ denotes the RMS frequency deviation, in (rad/s). Note that the signal suppression effect is neglected in this expression. In case of infinite compression, this effect results in an extra factor $[1 - \exp(-p)]$ in (6.91) [11].

The first and second-order continuous noise power at the baseband filter output are given by (6.47) and (6.60) respectively, while the click noise power is given by (6.90).

The general expression for the receiver output SNR therefore becomes

$$\text{SNR} = \frac{p\frac{2B_{N,\text{IF}}}{B_{N,0}}\left(\frac{\Delta\omega}{2\pi\rho_0}\right)^2}{1 + 4\left(\frac{\Delta\omega}{2\pi\rho_0}\right)^2\lambda^2 + \Gamma_{\text{snd}} + p\frac{2B_{N,\text{IF}}}{B_{N,0}}\left[\frac{\xi}{G(A)}\right]^2\left(N_+ + N_-\right)\frac{B_{N,1}}{(2\pi\rho_0)^2}}, \tag{6.92}$$

Where the contribution of the second-order noise equals

$$\Gamma_{\text{snd}} = p\frac{2B_{N,\text{IF}}}{B_{N,0}}\left\{\left[(\mu-\lambda)^2\left(\frac{\Delta\omega}{2\pi\rho_0}\right)^2 + (1-\lambda)^2\left(\frac{\rho_2}{\rho_0}\right)^2\right] B_{N,2}\frac{S_{n^2}(0)}{A^4} + 2\lambda^2 B_{N,3}\frac{S_{n,n}(0)}{A^4\left(2\pi\rho_0\right)^2}\right\}. \tag{6.93}$$

The numerator of (6.92) equals the maximum possible SNR attained in case of infinite compression, while the denominator represents the deviation from this

SNR due to amplitude noise, second order (modulation) noise, and click noise. As discussed in the subsequent section, this expression can be used to determine the optimum amplitude compressor transfer.

6.6.2 Optimal Amplitude Compressor Transfer

This section determines the optimum amplitude compressor transfer, that maximizes the demodulator output SNR. Since the optimum transfer is a function of the input CNR, as shown below, the upper bound on the SNR derived in this section can only be attained with the aid of *adaptive amplitude compression.* As mentioned previously, perceptive aspects can be included in this optimization by assigning weighting factors to the various noise components.

The discussion in Section 6.4.4 showed that as far as minimization of second-order noise is concerned, a trade-off between (first-order) continuous noise and click noise can best be established by means of linear combinations of infinite compression and no compression. Further, as far as click noise is concerned, the advantage of such a scheme is that the (effective) click pulse area observed at the demodulator output is independent of the input CNR. The trade-off therefore remains effective at low input CNRs, as opposed to the situation in a soft-limiter where the resulting click pulse area tends towards 2π at low CNRs due to saturation of the limiter.

Therefore, this section considers the compressor transfer $G(A) = c_0 + c_2A^2$ be optimal, even in the presence of click noise. In [26] a hybrid between infinite compression and no compression, called "inverse limiting" in that reference, it was already proposed as a suitable demodulator for low CNRs. However, it remained unnoticed that this configuration is the optimum configuration, and no rules of how to control the trade-off between infinite compression and "inverse limiting" were given. In this section, the optimum trade-off, represented by ratio of the coefficients c_0 and c_2, as function of the input CNR, is determined: the absolute value of c_0 and c_2 is not of interest, as discussed previously, since it represents only an amplification applied to all components of the receiver output signal.

General Expression for the Optimal Compressor Transfer

In order to determine the optimum ratio c_2/c_0, all three types of noise need to be expressed in terms of this ratio. For the first and second-order continuous noise, this is a straightforward procedure that requires only the determination of λ,

$$\lambda = \frac{c_2A^2}{c_0 + c_2A^2}. \tag{6.94}$$

The click noise is generated only in the part of the system, depicted in figure 6.25, that applies infinite compression, while it is absent in the absence of compression. The click pulse area observed at the receiver output therefore equals $2\pi c_0$, which can also be observed from figure 6.25. The contribution of the click noise to the denominator of (6.92) then becomes proportional to

$$\begin{aligned}\left[\frac{\xi}{G(A)}\right]^2 &= 4\pi^2 \frac{c_0^2}{(c_0 + c_2 A^2)^2} \\ &= 4\pi^2 (1-\lambda)^2.\end{aligned} \tag{6.95}$$

The denominator of (6.92), denoted by $D(\lambda)$, can therefore be expressed as

$$D(\lambda) = 1 + \eta_1 \lambda^2 + \eta_2 (1-\lambda)^2, \tag{6.96}$$

where η_1 and η_2 are given by

$$\eta_1 = 4\left(\frac{\Delta\omega}{2\pi\rho_0}\right)^2 + 4p\frac{B_{N,\mathrm{IF}}}{B_{N,0}} B_{N,3} \frac{S_{n,n}(0)}{A^4 \left(2\pi\rho_0\right)^2}, \tag{6.97}$$

$$\eta_2 = p\frac{2B_{N,\mathrm{IF}}}{B_{N,0}} \left[\left(\frac{\rho_2}{\rho_0}\right)^2 B_{N,2} \frac{S_{n^2}(0)}{A^4} + 4\pi^2 \left(N_+ + N_-\right) \frac{B_{N,1}}{\left(2\pi\rho_0\right)^2}\right]. \tag{6.98}$$

The optimum value of λ is then obtained as

$$\lambda_{\mathrm{opt}} = \frac{\eta_2}{\eta_1 + \eta_2}. \tag{6.99}$$

Consequently, according to (6.94), the optimum ratio of the compressor transfer parameters equals

$$\left.\frac{c_2}{c_0}\right|_{\mathrm{opt}} = \frac{\eta_2}{A^2 \eta_1}. \tag{6.100}$$

Optimal Compression for Rectangular Filters

For example, when we consider again an FM receiver with a rectangular input filter of bandwidth W_n, and a rectangular baseband filter of bandwidth W, the optimum value of λ becomes

$$\lambda_{\mathrm{opt}} = \frac{4 - 3\frac{W}{W_n} + 96\pi p \frac{N_+ + N_-}{W} \frac{W_n}{W}}{8 - 4\frac{W}{W_n} + 4\left(\frac{W_n}{W}\right)^2 - 6\frac{W_n}{W} + 96p\left(\frac{\Delta\omega}{W}\right)^2 + 96\pi p \frac{N_+ + N_-}{W} \frac{W_n}{W}}. \tag{6.101}$$

Note the similarity between this expression and expression (6.68), that denotes the optimum value of λ in the absence of click noise. The click noise only introduces an additional term in the numerator and the denominator of (6.101).

Figure 6.35 depicts the threshold curves for infinite compression, no compression and optimal, adaptive compression, for a receiver with $W_n = 12W$, $\Delta\omega = 5W$, and sinusoidal modulation; the click rates N_+ and N_- are thus obtained from (5.15). According to this figure, the optimum compressor improves the output SNR with a few dB at low CNRs, below the demodulator threshold, by reduction of the level of compression, which also slightly 'smooths' the threshold. The corresponding optimum value of λ, and the optimum ratio c_2A^2/c_0 are depicted in figure 6.36 as functions of the input CNR. This figure shows that infinite compression, i.e. $\lambda = 0$ is optimal at high CNRs, while a compression level of $1/\lambda \approx 1.3$ is optimal around a CNR of 0 dB.

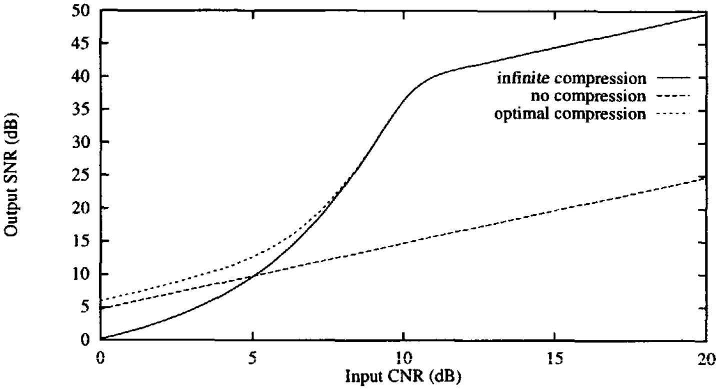


Figure 6.35: Output SNR versus input CNR for infinite compression, no compression, and optimal compression.

6.7 Verification of the Theory

The theory developed in this chapter was verified by simulations and measurements. This section discusses the results and compares the experimental data with the theoretical model.

The simulations and measurements were both performed for an FM demodulator preceded by a soft-limiter of the type depicted in figure 6.13. The demodulator was constructed such that the amplitude compressor transfer $G(R)$ in (6.13) equals the square of the soft-limiter transfer from (6.11), i.e. $G(R) = G_{\text{sl}}^2(R)$. Based on the previously developed theory, it is expected that a considerable amount of second-order noise is generated in such suboptimal demod-

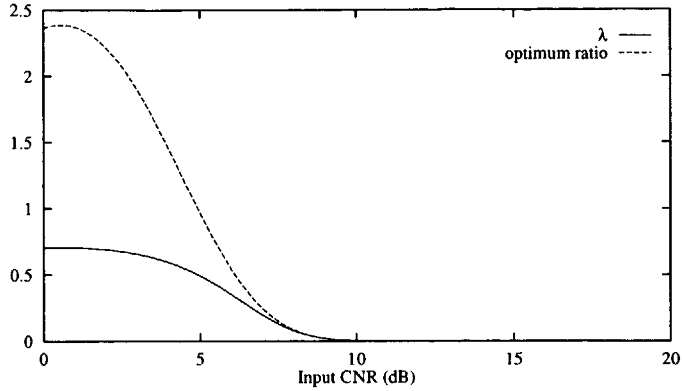


Figure 6.36: Optimum value of λ and the optimum ratio c_2A^2/c_0 as a function of the input CNR.

ulators, since, as can be derived from (6.11), the parameters μ and λ are not generally equal.

The demodulator output SNR v.s. input CNR curves were determined for various values of the parameter $x = K/A$, the ratio of the limiter's linear region (see figure 6.13) and the amplitude of the input wave.

Section 6.7.1 discusses the simulation results, while Section 6.7.2 discusses the measurements.

6.7.1 Simulation Results

The simulations were performed with the HP series IV simulation package, which allows high-level modeling of the various demodulator sub-functions. Figure 6.37 schematically depicts the setup. The input FM wave $s(t)$ is modulated by a sinusoidal wave in such a way that the frequency deviation ratio $\Delta\omega/W$ equals 5. Together with the input noise $n(t)$, this wave is filtered by an approximately rectangular filter. The bandwidth of this filter equals $5(\Delta\omega + W)$, which is considerably larger than the value obtained by Carson's formula (2.8), in order to avoid narrow-band filtering distortion. The filter at the output of the soft-limiter is identical to the input filter and retrieves the fundamental frequency

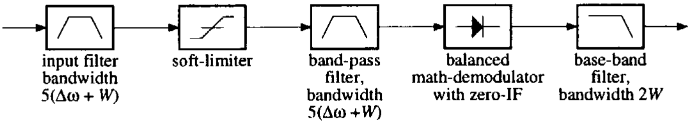

input filter bandwidth 5(Δω + W)
soft-limiter
band-pass filter, bandwidth 5(Δω +W)
balanced math-demodulator with zero-IF
base-band filter, bandwidth 2W

Figure 6.37: Simulation setup.

component of the compressed FM wave $r(t) = s(t) + n(t)$. This wave is subsequently demodulated by a balanced math-demodulator of the type discussed in Section 3.4.7. The output signal is filtered by an approximately rectangular low-pass filter with a bandwidth that equals twice the frequency of the baseband sinusoidal wave.

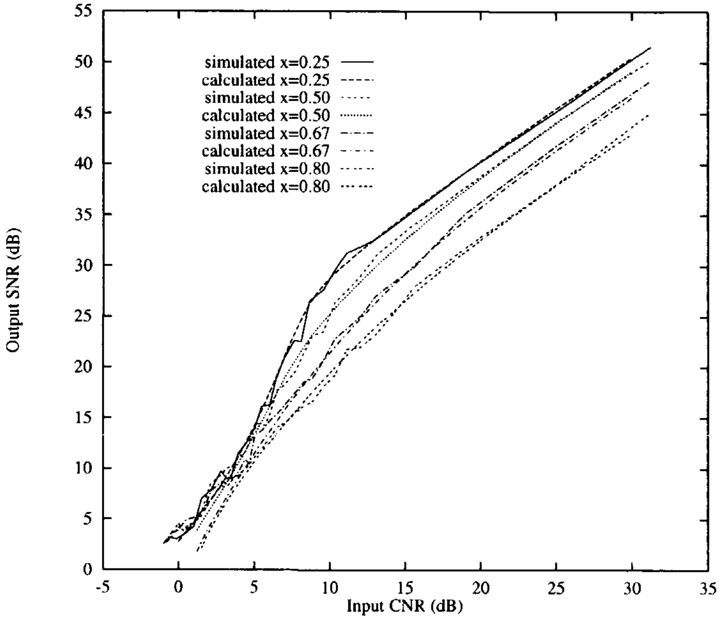


Figure 6.38: Simulated and calculated demodulator output SNR as a function of the input CNR.

After the validity of the simulator model was verified for an infinite limiter gain (hard-limiter) with the known results from literature for this case, the output SNR v.s. input CNR curves were determined for the parameter values $x = 0.25, 0.50, 0.67$ and 0.8, corresponding to limiter over-drives of $A/K = 4, 2, 1.5$ and 1.25 respectively. The results are depicted in figure 6.38, together with the theoretical curves.

In order to match the simulation results, the theoretical curves depicted in figure 6.38 required a somewhat larger value for the parameter μ than the theoretical value given by (6.53) at the "quiescent point" $R(t) = A$. The explanation for this phenomenon is found in the piece-wise linear nature of the soft-limiter transfer curve of figure 6.13, which causes a singularity in the parameter μ, as depicted in figure 6.39. This curve corresponds to the square of the amplitude transfer $G_{\rm sl}(R)$, depicted in figure 6.14.

The singularity in μ is located at $R/K = 1$, where the top of the input FM

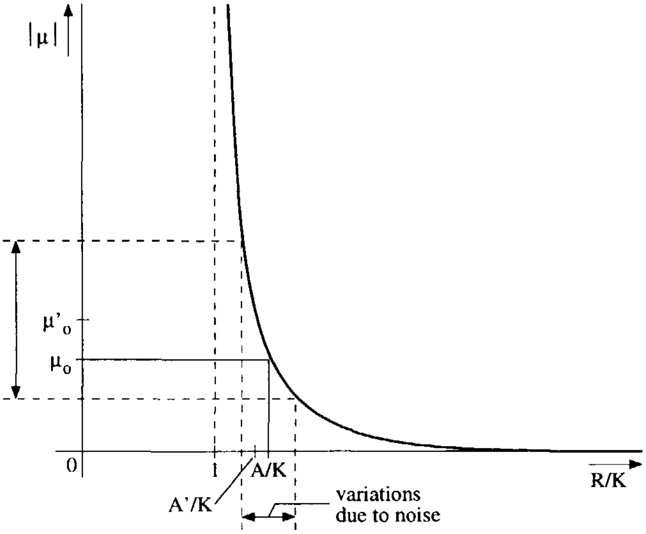


Figure 6.39: Second-order noise parameter μ for the soft-limiter transfer as a function of the normalized amplitude R/K.

wave just touches the boundaries of the soft-limiter's saturation region. The selected values for the parameter $x = K/A$, i.e. the "quiescent points" of the FM amplitude, are located quite close to the singularity in figure 6.39. The noisy limiter input amplitude $R(t)$ swings around these points, and consequently μ swings around some value μ_o (see figure 6.39). Usually, when μ is a smooth curve, the value μ_o corresponds to the quiescent point $R(t) = A$. However, in the soft-limiter case, the strong convex curvature causes μ to swing around a larger value μ'_o that corresponds to a smaller amplitude $R(t) = A'$ and a larger value of x.

With the modified values of the parameter μ, and the (unmodified) values of λ obtained from (6.53), the matching between the simulated and calculated curves is quite good.

Further, notice from figure 6.38 that the demodulator threshold curves smooth for decreasing limiter gains. The steep SNR decrease at the threshold in case of a high limiter gain becomes more gentle due to the large amount of second-order noise. Thus, the output SNR is exchanged for a smoothed threshold.

6.7.2 Measurement Results

The measurements were performed with a setup similar to the one depicted in figure 6.37.

In the measurement setup, two 8-th order Butterworth band-pass filters with a 100 kHz center frequency and 6 kHz bandwidth were used. The soft-limiter was realized by means of an emitter-degenerated differential pair. A quadrature FM demodulator, discussed in Section 3.5.8, was constructed from a 2.5μs analog delay line and an analog linear multiplier. The 100 kHz FM carrier was modulated by a 1 kHz sinusoidal wave. Due to the small bandwidth of the available filters, the frequency deviation was limited to 2 kHz, a deviation ratio $\Delta\omega/W = 2$.

With this setup, the output SNR v.s. input CNR curves were measured for the parameter values $x = 0.67, 0.85$ and 0.9. The results are depicted in figure 6.7.2

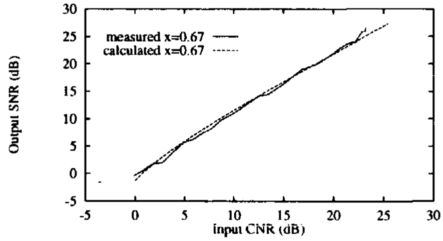


Figure 6.40: Measurement result for $x = 0.67$.

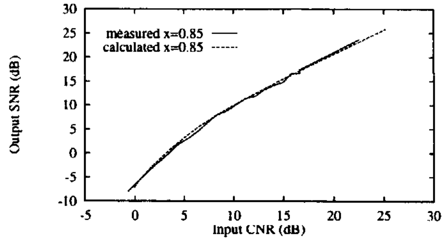


Figure 6.41: Measurement result for $x = 0.85$.

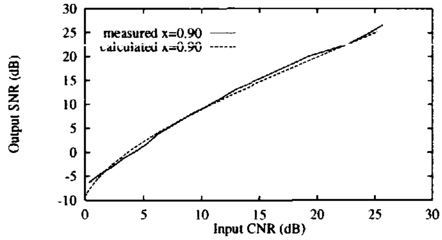


Figure 6.42: Measurement result for $x = 0.90$.

In this case, as opposed to the simulations, the calculated curves match the measurements very well for the values of the parameter μ calculated from (6.53) using, as a first approximation, the transfer $G(A) = G_{\mathrm{sl}}^2(R)$ of the soft-limiter from figure 6.13. This is explained by the fact that the transfer of the degenerated differential pair, implementing the soft-limiter, is a smooth curve instead of piece-wise linear. As a result, the characteristic for $\mu(R/K)$ is somewhat smoothed in comparison to the curve in figure 6.39 and does not show such a distinct singularity as that curve.

6.8 Conclusions

This chapter analyzed the behavior of arbitrary types of amplitude compressors that precede the FM demodulator in the presence of additive noise/interference. A newly developed mathematical model showed that an optimal amplitude compressor transfer that maximizes the demodulator output SNR can be obtained and expressed in terms of the demodulator parameters and FM wave characteristics. If necessary, perceptive aspects can be included in this optimization through the introduction of weighting factors into the various output noise power contributions. Further, it was shown that the optimum compressor transfer has to be adaptively controlled as a function of the receiver input CNR: at low input CNRs a very low level of compression should be applied, while at high CNRs a high level of compression is favorable.

Two different classes of amplitude compressors can be distinguished. The first class, which includes hard-limiters, applies compression directly to the RF carrier intensity without explicit detection of the carrier amplitude. Such compressors are characterized by the generation of the harmonics in the output signal that should not be used in the demodulation process; in order to attain an as large as possible demodulator input CNR only the fundamental harmonic should be used, which is established more or less automatically in many compressor implementations. The second class, which includes AGCs, explicitly detects the amplitude prior to the compression operation.

An outline of the general approach showed that the demodulator output noise power can generally be described by seven 'shaping' parameters that describe the characteristics of the IF filter, the demodulator and the baseband filter.

The description derived for the first-order demodulator output noise was shown to agree with results obtained in the literature for several special types of amplitude compression. Further, it was shown that the noise level observed at the baseband filter output closely approaches its lower-bound, obtained with infinite compression when the applied level of compression exceeds a critical level that can be expressed in terms of the RMS modulation and the 'shaping' parameters.

The description derived for the second-order output noise was also shown to agree with results obtained in the literature for special types of compression. Further, the minimum level of second-order noise was shown to be attained when the first-order and second-order inverse compression factors are equal. From this condition followed that the corresponding 'optimum ' amplitude compressor should realize a linear combination of infinite compression and no compression.

A newly developed, generalized click noise model was applied to an FM demodulator that establishes amplitude compression with the aid of a soft-limiter. It was shown that the area of the click pulses observed at the demodulator output is generally a function of the input CNR in case of finite compression. At low CNRs, the area approaches 2π, as in the case of infinite compression, while it approaches zero at high CNRs, as in the case of no compression.

Comparison of the theory with simulation and measurement results showed that the piece-wise linear nature of the soft-limiter transfer used by the simulator causes slight discrepancies between the theory and simulation results. The measurements matched quite well with the theory, due to the absence of piece-wise linear transfers: the transfer of practical soft-limiter circuits is relatively smooth.

References

[1] David Middleton, "The spectrum of frequency-modulated waves after reception in random noise-I", *Quarterly of Applied Mathematics*, vol. VII, no. 2, pp. 129–174, July 1949.

[2] David Middleton, "The spectrum of frequency-modulated waves after reception in random noise-II", *Quarterly of Applied Mathematics*, vol. VIII, no. 1, pp. 59–80, Apr. 1950.

[3] Nelson M. Blachman, "The demodulation of an F-M carrier and random noise by a limiter and discriminator", *Journal of Applied Physics*, vol. 20, pp. 38–47, Jan. 1949.

[4] Nelson M. Blachman, "The demodulation of a frequency-modulated carrier and random noise by a discriminator", *Journal of Applied Physics*, vol. 20, pp. 976–983, Oct. 1949.

[5] John Cohn, "A new approach to the analysis of FM threshold reception", in *Proceedings of the National Electronics Conference*, 1956, pp. 221–236.

[6] S. O. Rice, "Noise in FM receivers", in *Proceedings of the Symposium on Time Series Analysis, Brown University, 1962.* 1963, pp. 395–422, M.Rosenblatt Ed., John Wiley and Sons, New York.

[7] David Middleton, *An Introduction to Statistical Communication Theory*, McGraw-Hill Book Company, New York, 1960.

[8] Eric A. M. Klumperink, Carlo T. Klein, Bas Rüggeberg, and Ed J. M. van Tuijl, "AM suppression with low AM to PM conversion with the aid of a variable-gain amplifier", *IEEE Journal of Solid State Circuits*, vol. 31, no. 5, pp. 625–633, May 1996.

[9] Wouter A. Serdijn, *The design of Low-Voltage Low-Power Analog Integrated Circuits and their application in Hearing Instruments*, PhD thesis, Delft University of Technology, 1994.

[10] Heinrich Meyr and Gerd Ascheid, *Synchronization in Digital Communications, volume 1, Phase-Frequency-Locked Loops, and Amplitude Control*, John Wiley and Sons, New York, 1990.

[11] F. L. H. M. Stumpers, "Theory of frequency-modulation noise", *Proceedings of the IRE*, vol. 36, no. 9, pp. 1081–1092, Sept. 1948.

[12] Robert M. Gray and Thomas G. Stockham, "Dithered quantizers", *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 805–809, May 1993.

[13] Steven R. Norsworthy, "Effective dithering of sigma-delta modulators", in *Proceedings of the IEEE International Symposium on Circuits and Systems*, 1992, pp. 1304–1307.

[14] Wu Chou and Robert M. Gray, "Dithering and its effects on sigma delta and multi-stage sigma delta modulation", in *Proceedings of the IEEE International Symposium on Circuits and Systems*, 1990, pp. 368–371.

[15] Paul .C. de Jong, Gerard C.M. Meijer, and Arthur H.M. van Roermund, "A new dithering method for sigma delta modulators", *Analog Integrated Circuits and Signal Processing*, , no. 10, pp. 193–204, 1996.

[16] James C. Candy and Gabor C. Temes, Eds., *Oversampling Delta-Sigma Data Converters, Theory, Design and Simulation*, IEEE Press, New York, 1992.

[17] Nelson M. Blachman, "The Signal × Signal, Noise × Noise, and Signal × Noise output of a nonlinearity", *IEEE Transactions on Information Theory*, vol. IT-14, no. 1, pp. 21–27, Jan. 1968.

[18] James C. Springett and Marvin K. Simon, "An analysis of the phase coherent-incoherent output of the bandpass limiter", *IEEE Transactions on Communication Technology*, vol. COM-19, no. 1, pp. 42–49, Feb. 1971.

[19] Rick W. Miller and Daniel M. Hutchinson, "Bus aligned quadrature FM detector", U.S. Patent 5,596,298, January 21 1997.

[20] C. A. M. Boon, *Design of High-Performance Negative Feedback Oscillators*, PhD thesis, Delft University of Technology, 1989.

[21] David Middleton, "On theoretical signal-to-noise ratios in F-M receivers: A comparison with amplitude modulation", *Journal of Applied Physics*, vol. 20, pp. 334–351, Apr. 1949.

[22] S. O. Rice, "Mathematical analysis of random noise-I", *The Bell System Technical Journal*, vol. 23, pp. 282–332, 1944.

[23] S. O. Rice, "Mathematical analysis of random noise-II", *The Bell System Technical Journal*, vol. 24, pp. 46–156, 1945.

[24] A. Bruce Carlson, *Communication Systems, An introduction to Signals and noise in Electrical Communication*, McGraw-Hill Book Company, Singapore, 1986.

[25] Athanasios Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill Book Company, New York, 3rd edition, 1991.

[26] John H. Park, "An FM detector for low S/N", *IEEE Transactions on Communication Technology*, vol. 18, no. 2, pp. 110–118, Apr. 1970.

[27] William E. Boyce and Richard C. DiPrima, *Elementary Differential Equations and Boundary Value Problems*, John Wiley and Sons, New York, 1986.

[28] George Lindgren, "On the shape and duration of FM-clicks", *IEEE Transactions on Information Theory*, vol. 29, no. 4, pp. 536–543, July 1983.

[29] George Lindgren, "Shape and duration of clicks in modulated FM transmission", *IEEE Transactions on Information Theory*, vol. 30, no. 5, pp. 728–735, Sept. 1984.

[30] A. J. Rainal, "Power spectrum of FM clicks", *IEEE Transactions on Information Theory*, vol. 30, no. 1, pp. 122–124, Jan. 1984.

[31] Davras Yavuz and Donald T. Hess, "FM noise and clicks", *IEEE Transactions on Communication Technology*, vol. COM-17, no. 6, pp. 648–653, Dec. 1969.

[32] M.H.L. Kouwenhoven, C.J.M. Verhoeven, and A.H.M. van Roermund, "Frequency demodulator threshold minimization by application of soft-limiters", in *Proceedings of the ProRISC/IEEE Workshop on Circuits, Systems and Signal Processing*, Mierlo, November 27 - 28, 1996, pp. 195 – 200.

[33] M.H.L. Kouwenhoven, C.J.M. Verhoeven, and A.H.M. van Roermund, "A new simple design model for FM demodulators using soft-limiters for click noise suppression", in *Proceedings of the IEEE International Symposium on Circuits and Systems*, Hong Kong, June 9-12, 1997, vol. 1, pp. 265–268.

[34] M.H.L. Kouwenhoven, C.J.M. Verhoeven, and A.H.M. van Roermund, "A new design model for click noise and continuous noise in soft-limiting FM demodulators", in *Proceedings of the 13-th European Conference on Circuit Theory and Design*, Budapest, August 30 - September 3, 1997, vol. 3, pp. 1387–1392.

Chapter 7

Phase Feedback

The previous chapters were mainly concerned with the behavior of so called "conventional FM discriminators", FM demodulators that do not contain phase- or frequency feedback. Although their behavior shows many similarities with phase feedback demodulators, such as PLLs, there are at the same time some remarkable differences. For instance, although their response to external noise around the threshold does show impulse noise like the click noise phenomenon in conventional demodulators, the underlying principle is completely different. In this respect, it is known that the phase feedback or 'phase lock' mechanism is capable of suppressing part of the impulse noise and thereby achieves threshold extension.

The purpose of this chapter is to compare the main characteristics of phase feedback demodulators, especially their threshold behavior, with those of non-feedback FM demodulators. This includes an explanation of the threshold extension capabilities of phase feedback demodulators.

Section 7.1 discusses the principle of operation, in conjunction with the demodulator classification derived in Chapter 3. Section 7.2 outlines the general phase feedback demodulator model, that is used throughout the analyses. Section 7.3 considers the demodulator behavior above threshold, the counterpart of the discussion in Section 5.4.2, while Section 7.4 considers the behavior in the threshold region, the counterpart of Section 5.4.3. The consequences of the results of these analysis on phase feedback demodulator design are discussed in Section 7.5. Finally, Section 7.6 closes the chapter with conclusions.

7.1 Principles of Operation

In Chapter 3 it was shown that phase feedback demodulators can be identified as a member of the class of FM-PM conversion demodulators. This section briefly considers their principles of operation as an introduction to the investigation of their noise behavior in subsequent sections.

As far as operating principles are concerned, the main distinction between phase feedback demodulators and other, non-feedback types of FM to PM conversion demodulators is that the former implement the exact FM to PM conversion function, differentiation of the input carrier phase, whereas the latter, non-feedback demodulators or "FM discriminators" are, except for post-detection conversion demodulators, based on an approximation of that function. For example, it was shown in Section 3.5.2 that quadrature demodulators approximate differentiation of the phase by the difference of two phase values, separated in time by a fixed delay τ_d.

Section 7.1.1 derives an expression for the phase feed-back algorithm, while Section 7.1.2 briefly considers its implementation.

7.1.1 Phase Feedback Demodulation Algorithm

Mathematically, the (ideal) operation of a phase feedback demodulator may be expressed as

$$y_{\mathrm{dem}}(t) = \dot{\Phi}(t) \overset{\mathrm{def}}{=} \lim_{\tau\to 0} \frac{\Phi(t) - \Phi(t-\tau)}{\tau} = \lim_{\tau\to 0} \frac{\Delta\Phi(t,\tau)}{\tau}, \tag{7.1}$$

where $\Phi(t)$ represents the instantaneous phase of the input FM wave $s(t)$. The phasor representation of this expression is depicted in figure 7.1. In this figure,

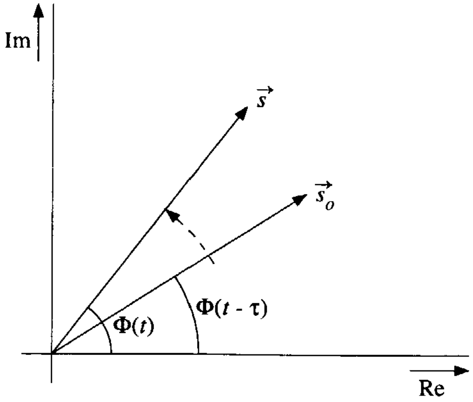


Figure 7.1: Phasor representation of phase feedback demodulators.

the phasor $\vec{s}$ corresponds to the input FM wave, with instantaneous phase $\Phi(t)$, while the phasor $\vec{s}_o$ corresponds to a reconstruction of this wave, with instantaneous phase $\Phi(t-\tau)$, generated by the demodulator. The reconstructed wave

$s_o(t)$ tracks the input FM wave $s(t)$, such that the phase error approaches zero. In that case, the control signal supplied to the input of the controlled oscillator is a reconstruction of the original FM message signal.

Although (7.1) completely describes the phase feedback demodulation algorithm, it is not directly suited for phase feedback demodulator design due to the fact that it lacks an explicit description of the demodulator architecture. Therefore we rewrite (7.1) in such a way that the feedback loop contained in this architecture is explicitly described by the expression. This is achieved by expressing the phase $\Phi(t-\tau)$ of the reconstructed FM wave $s_o(t)$ in terms of the key parameter of the phase feedback loop: the phase error $\Delta\Phi(t,\tau)$.

The required relation between the phase error $\Delta\Phi(t-\tau)$ and the reconstructed FM phase $\Phi(t-\tau)$ is obtained from (7.1) when the lim-operation is postponed to the final step of the derivation. In fact, temporary omission of this lim-operation is equivalent to a reduction of the loop gain from infinity to a finite value: a finite loop gain allows the existence of a nonzero-valued phase error $\Delta\Phi(t,\tau)$ inside the loop. In this way, the phase of the reconstructed FM wave may be expressed as:

$$\dot{\Phi}(t-\tau) = \frac{\Delta\Phi(t,\tau)}{\tau} - \frac{\mathrm{d}\Delta\Phi(t,\tau)}{\mathrm{d}t}, \tag{7.2}$$

which by integration with respect to time yields

$$\Phi(t-\tau) = \frac{1}{\tau}\int^{t} \Delta\Phi(u,\tau)\mathrm{d}u - \Delta\Phi(t,\tau). \tag{7.3}$$

When the lim-operation is applied to this expression, corresponding to a reinforcement of the loop gain towards infinity, τ approaches zero and the last term on the RHS vanishes, leaving only the integral of the phase error.

Equation (7.3) clearly demonstrates the structure of the ideal phase feedback demodulator loop. The presence of such a feedback loop is reflected by the fact that the reconstructed phase $\Phi(t-\tau)$ is a function of the phase difference $\Delta\Phi(t,\tau) = \Phi(t) - \Phi(t-\tau)$, while this phase difference is again a function of $\Phi(t-\tau)$. Further, it is observed that the loop necessarily contains a memory, represented by an integration of the phase error with respect to time. This integration is generally established by a controlled oscillator (FM modulator) in the feedback loop. Additional memories, such as those contained in a loop filter, may be present in order to satisfy various requirements but these are not essential for the demodulation operation. Finally, it is noticed that for $\tau \to 0$, the required output signal, the demodulated FM message $\dot{\varphi}(t)$, can be obtained from the integrator input. In that case, the fraction $\Delta\Phi(t,\tau)/\tau$ approaches $\omega_o + \dot{\varphi}(t)$, equal to the instantaneous frequency of the input FM wave. When the free-running frequency of the oscillator equals the carrier frequency ω_o, the carrier component is entirely realized inside the oscillator. The oscillator input signal is then the same as the demodulated message signal $\dot{\varphi}(t)$.

7.1.2 Algorithm Implementation

As observed from equation (7.3), an implementation of the phase feedback demodulation algorithm, a demodulator, consists at least of

- a controllable oscillator (to generate $s_o(t)$ with phase $\Phi(t-\tau)$);
- a phase detector (PD) (to measure the phase difference $\Delta\Phi(t,\tau)$).

In practice, the loop generally also contains

- a loop filter.

The general topology of such a system is depicted in figure 7.2.

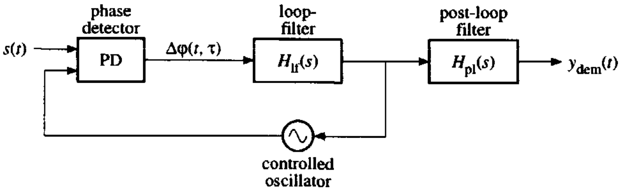


Figure 7.2: General topology of a phase feedback demodulator.

The oscillator realizes the integration in (7.3) by generating the reconstruction of the FM input wave $s_o(t)$ from the phase difference $\Delta\varphi(t,\tau)$. It adds its free-running frequency ω_o to the demodulator output signal $\dot{\varphi}(t)$, and subsequently performs an integration of the phase with respect to time. The input signal of this oscillator is the same as the demodulator output signal $y_{\text{dem}}(t)$. Further, since the oscillator is located in the feedback path of the system, its tuning range should be linear over the region of interest for the loop in order to establish a linear demodulator transfer.

The phase detector determines the phase difference $\Delta\varphi(t,\tau)$ between the input and oscillator wave. In practice, the transfer of a phase detector differs from the ideal response $\Delta\varphi(t,\tau)$ in a number of respects. In the first place, the detector output signal is not exactly linear with $\Delta\varphi(t,\tau)$, but contains a nonlinearity. This is because the detector transfer is periodic in $\Delta\varphi(t,\tau)$. As discussed in subsequent sections, this PD nonlinearity determines the threshold behavior of the demodulator. In a special type of phase feedback demodulator, so called delay locked loops (DLL), the detector nonlinearity appears to be non-periodic [1]. However, this does not result in a significantly different threshold behavior. Secondly, the phase detector output usually becomes zero for a nonzero phase difference between its inputs. In fact, this implies that such detectors behave as if a fixed, built-in phase shift is contained in one of their inputs. For example, a multiplier phase detector behaves as if a built-in phase shift of 90^0 is applied to one of the input waves since its output is zero when both (sinusoidal) inputs are in quadrature.

Although not strictly necessary for the demodulation algorithm, the demodulator usually also contains a filter $H_{\mathrm{lf}}(s)$ inside the loop, and a post-loop filter $H_{\mathrm{pl}}(s)$, as depicted in figure 7.2. These filters are included to improve the performance of the basic system, also in respect to its response to noise.

The remainder of this chapter is concerned with the noise behavior of the feedback demodulator and its dependence on the phase detector nonlinearity and the loop filter structure. It is thereby assumed that the system has reached a state, the 'lock mode', in which (7.3) holds. It should be noted however, that in practice considerable effort is required to manoeuvre the system into lock and keep it in that mode. Although omitted here, this "acquisition" behavior, and many other characteristics, are also important issues in research on such systems [2–13].

7.2 Phase Feedback Demodulator Modeling

A quantitative analysis of the phase feedback demodulator performance in the presence of noise is feasible only with the aid of the appropriate demodulator models. This section briefly discusses the generally applicable phase feedback demodulator model that is used in the various analyses described in this chapter.

Section 7.2.1 outlines the well-known low-pass equivalent model of the demodulator loop, while Section 7.2.2 considers the inclusion of the demodulator input noise into this model.

7.2.1 Low-Pass Equivalent Demodulator Model

It is well known that during phase lock the behavior of a phase feedback demodulator can be analyzed conveniently from its low-pass equivalent model, depicted in figure 7.3 [14–18]. This model, which was originally derived for a loop with a linear multiplier phase detector (PD), sinusoidal input and oscillator wave, models the oscillator by an integrator with integration constant K_o. It models

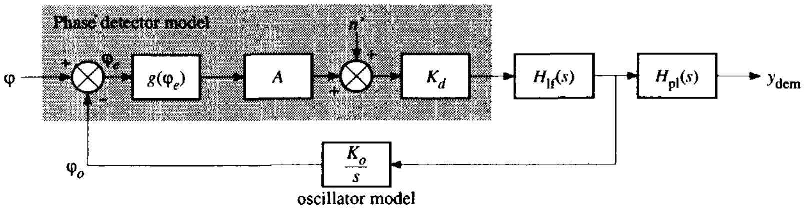


Figure 7.3: Low-pass equivalent model of a phase feedback demodulator.

the phase detector by a subtracter, a detector constant K_d, the amplitude A of the input FM wave $s(t)$, and a nonlinear transfer $g(\varphi_e)$, where

$$\varphi_e(t) \overset{\text{def}}{=} \varphi(t) - \varphi_o(t) \tag{7.4}$$

denotes the phase error inside the loop. Components in the phase detector output signal located at double the carrier frequency $2\omega_o$ are assumed to be suppressed by the loop and are therefore ignored.

For a linear multiplier-phase detector, the nonlinear transfer $g(\varphi_e)$ equals $\sin \varphi_e(t)$. In [1, 19, 20], however, is shown that the model of figure 7.3 may also be used to describe demodulators with an arbitrary phase detector, including delay-locked loops (DLL), by modification of the nonlinear transfer $g(\varphi_e)$. As shown in subsequent sections, such a system representation allows optimization of the (noise) behavior of the loop as a function of this detector transfer.

This demodulator model constitutes the basis for both the linear demodulator analysis that is valid above the threshold, and the nonlinear analysis that remains valid below the threshold. Above the threshold, some simplifications are allowed, as discussed in Section 7.3.1.

7.2.2 Noise Model

The noise $n'(t)$ that adds to the loop in figure 7.3 is an "equivalent noise source" that models the interaction between the additive input noise $n(t)$ and the oscillator output wave $s_o(t)$: it is obtained by transformation of the input noise to the phase detector output. Such a transformation is required in order to incorporate the effect of the input noise $n(t)$ on the demodulator input carrier amplitude: a noise source located at the input of the low-pass equivalent model can only account for the contribution of $n(t)$ to the phase noise of the input FM wave, while a noise source located at the PD output allows incorporation of the contribution to both the carrier phase and the carrier amplitude.

First, the properties of $n'(t)$ in a loop with a multiplier-phase detector are considered. Subsequently, the extension to arbitrary phase detectors is discussed.

Noise in a Loop with Multiplier Phase Detector

As an important example, we consider a loop with a multiplier phase detector and sinusoidal input and oscillator waves. The phase detector response to the FM input wave $s(t)$ in this loop, ignoring the double frequency terms, equals

$$\begin{aligned} s(t)s_o(t) &= -A\cos\left[\omega_o t + \varphi(t)\right] A_o \sin\left[\omega_o t + \varphi_o(t)\right] \\ &= \frac{AA_o}{2} \sin\left[\varphi(t) - \varphi_o(t)\right] \\ &= AK_d g\left(\varphi_e\right), \end{aligned} \tag{7.5}$$

where A_o denotes the amplitude of the oscillator wave. Thus, the PD constant equals $K_d = A_o/2$.

Its response to the input noise $n(t)$, again ignoring double frequency terms, is given by

$$\begin{aligned} n'(t) &= K_d^{-1} n(t) s_o(t) \\ &= -2\left[n_i(t)\cos\omega_o t - n_q(t)\sin\omega_o t\right]\sin\left[\omega_o t + \varphi_o(t)\right] \\ &= \left[-n_i(t)\sin\varphi_o(t) + n_q(t)\cos\varphi_o(t)\right]. \end{aligned} \tag{7.6}$$

The oscillator output phase $\varphi_o(t)$ is generally not equal to the input phase $\varphi(t)$, especially at low CNRs, due to the influence of the input noise $n(t)$. It follows from (7.6) that the influence of this noise on $\varphi_o(t)$ generally changes the statistics of $n'(t)$ and introduces correlation between the noise and the phase $\varphi_o(t)$ that partially depends on the input noise through the feedback loop.

However, as shown in [16, 17, 21, 22], these effects can be ignored as long as the bandwidth of the input noise $n(t)$ is large compared to the closed loop bandwidth of the demodulator since only the low-frequency range of the spectrum, located inside the loop bandwidth, is of interest. The phase $\varphi_o(t)$ introduces modulation similar to the modulation illustrated by figure 2.5 that basically affects the spectrum of $n'(t)$ around its cut-off frequency. From another point of view, the large bandwidth of $n(t)$ in comparison to the bandwidth of $\varphi_o(t)$ causes its correlation time to be much smaller than the correlation time of $\varphi_o(t)$. This in turn means that $n'(t)$ may be considered to be approximately independent of the other, 'narrow band' signals inside the loop. The statistical properties of $n'(t)$ therefore resemble those of $n_{s,i}(t)$ and $n_{s,q}(t)$, which are the low-pass equivalent noise processes of $n(t)$ (see Chapter 2).

Noise Model for Arbitrary Phase Detectors

For non-sinusoidal phase detectors, the same expressions and conclusions as derived for multiplier phase detectors hold at high CNRs. At low CNRs, however, the expressions for $n'(t)$ become slightly different [23–25]. We return to this subject in Section 7.5.1.

7.3 Response Above Threshold

As shown in Section 5.4.2, the output SNR of a conventional FM demodulator above threshold is significantly improved by the application of infinite compression to the input FM wave. The brief review in this section shows that a similar SNR improvement is achieved by the phase feedback mechanism.

Section 7.3.1 considers the demodulator modeling above threshold. Subsequently, Section 7.3.2 considers the modeling of the noise above threshold.

Finally, Section 7.3.3 derives an expression for the output SNR above threshold.

7.3.1 Linear Demodulator Model

Above threshold, when the demodulator loop is in 'lock mode', the phase error $\varphi_e(t)$ and its variance, denoted by σ_e^2, are usually very small. In that case, the nonlinear detector transfer $g(\varphi_e)$, depicted in figure 7.3, may be replaced by its first-order Taylor term, resulting in the familiar linear model of figure 7.4 [7, 16, 17]. In the absence of input noise, i.e. $n'(t) \equiv 0$, this model easily shows

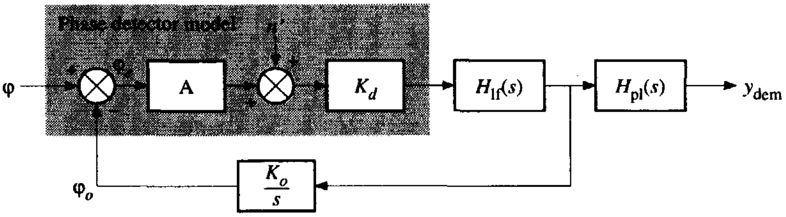


Figure 7.4: Linearized model of the demodulator during 'lock'.

that the closed loop transfer from input instantaneous frequency $\dot{\varphi}(t)$ to output signal $y_{\mathrm{dem}}(t)$ is given in the Laplace domain by

$$\frac{Y_{\mathrm{dem}}(s)}{s\Phi(s)} \overset{\mathrm{def}}{=} H_{\mathrm{dem}}(s) = \frac{AK_d H_{\mathrm{lf}}(s) H_{\mathrm{pl}}(s)}{s + AK_d K_o H_{\mathrm{lf}}(s)}. \tag{7.7}$$

When this transfer is flat over the baseband, the demodulator output signal is proportional to the message wave $\dot{\varphi}(t)$.

Besides this transfer, two other transfers are of importance in the nonlinear noise analysis described in Section 7.4. These transfers are related to the two main measures for the demodulator performance as used in the nonlinear demodulator model described in the next section: the SNR inside the demodulator closed-loop bandwidth, and the steady-state phase error.

The SNR inside the closed-loop bandwidth is related to the loop structure by means of the double-sided closed-loop noise bandwidth, denoted by W_L in (rad/s) or B_L in (Hz). This bandwidth is defined as

$$W_L = 2\pi B_L = \frac{\int_{-\infty}^{\infty} |H_{\mathrm{cl}}(\mathrm{j}\omega)|^2 \, \mathrm{d}\omega}{|H_{\mathrm{cl}}(0)|^2}, \tag{7.8}$$

where $H_{\mathrm{cl}}(s)$ denotes the "closed loop transfer" from input phase $\varphi(t)$ to oscillator output phase $\varphi_o(t)$, given by [7]

$$\frac{\Phi_o(s)}{\Phi(s)} \overset{\mathrm{def}}{=} H_{\mathrm{cl}}(s) = \frac{AK_o K_d H_{\mathrm{lf}}(s)}{1 + AK_o K_d H_{\mathrm{lf}}(s)}. \tag{7.9}$$

In essence, the steady-state phase error (SSPE), denoted by $\varphi_{e,ss}$, is the equivalent of the amplitude/phase offset encountered in non-ideal conventional FM demodulators. Such offsets were discussed in Chapter 4. The SSPE is a measure of the DC component of the PD output required to keep the oscillator at the correct frequency. It is related to the loop structure by means of the transfer $H_e(s)$, from input phase $\varphi(t)$ to loop phase error $\varphi_e(t)$ as

$$\varphi_{e,ss} \overset{\text{def}}{=} \lim_{s\downarrow 0} H_e(s)\varphi(s). \tag{7.10}$$

By inspection of the linear demodulator model of figure 7.4, it follows that

$$\frac{\Phi_e(s)}{\Phi(s)} \overset{\text{def}}{=} H_{\text{e}}(s) = \frac{s}{1 + AK_oK_dH_{\text{lf}}(s)}. \tag{7.11}$$

Thus, when the FM 'message' signal equals a constant frequency offset Ω_0, then $\Phi(s) = \Omega_o/s$ and the SSPE becomes

$$\varphi_{e,ss} = \frac{\Omega_o}{AK_oK_dH_{\text{lf}}(0)}. \tag{7.12}$$

7.3.2 Linear Noise Model

This section investigates the noise source $n'(t)$ contained in the low-pass equivalent demodulator model of figure 7.4, above the threshold. With the aid of a simplified model for this noise it is shown that phase feedback demodulators and 'conventional' FM demodulators that apply infinite amplitude compression to the input carrier amplitude behave similarly at high input CNRs.

Simplified Description for the Noise

At high input CNRs, all the different types of phase detectors respond in a similar way to noise at their input. Therefore, the behavior of phase feedback demodulators above their threshold can be analyzed using a loop with a multiplier phase detector, without loss of generality.

Reconsidering expression (7.6) for the noise in a loop with a multiplier phase detector, at high input CNRs, the oscillator phase $\varphi_o(t)$ closely resembles the message phase $\varphi(t)$ of the input FM wave $s(t)$. In that case, according to (2.14), $n'(t)$ equals

$$\begin{aligned} n'(t) &\approx -n_i(t)\sin\varphi(t) + n_q(t)\cos\varphi(t) \\ &= n_{s,q}(t). \end{aligned} \tag{7.13}$$

Thus, at high input CNRs, $n'(t)$ resembles the quadrature component of the input noise $n(t)$.The phase feedback mechanism suppresses the in-phase component of the input noise, while it passes the quadrature component.

This behavior is explained by the observation that, as far as the transfer from $n(t)$ to $n'(t)$ is concerned, the multiplier phase detector essentially behaves like a kind of synchronous detector that, according to Chapter 3, possesses phase selectivity. In case of a multiplier PD, the reference wave of this 'synchronous' detector, the oscillator output wave $s_o(t)$, is in phase quadrature with the input FM wave. Consequently, all noise and signal components in phase with this reference wave, in quadrature with the input FM wave, are passed, while those in quadrature with the reference, in-phase with the input FM wave $s(t)$, are suppressed.

Comparison with Conventional FM Demodulators

In Section 5.4.2 it was observed that a very similar behavior is achieved by amplitude compression in 'conventional' FM demodulators. In that case, infinite amplitude compression also suppresses the in-phase noise component $n_{s,i}(t)$, while it leaves the quadrature component $n_{s,q}(t)$ unaffected. Thus, above threshold, the response to external noise of a phase feedback demodulator and a conventional demodulator with infinite compression, a "limiter-discriminator", is essentially the same. *For this reason, the same SNR improvement above threshold is expected in both systems.*

7.3.3 Output SNR Above Threshold

This section derives an expression for the output SNR above threshold from the linear demodulator model of figure 7.4.

By application of the superposition principle to $\varphi(t)$ and $n'(t)$ in figure 7.4, and (7.7) for the demodulator transfer $H_{\mathrm{dem}}(\mathrm{j}\omega)$, the spectral density of the output signal $y_{\mathrm{dem}}(t)$ can be expressed as

$$\begin{aligned} S_{\mathrm{dem}}(\omega) &= |H_{\mathrm{dem}}(\mathrm{j}\omega)|^2 \left[\omega^2 S_\varphi(\omega) + \left(\frac{\omega}{A}\right)^2 S_{n'}(\omega)\right] \\ &= |H_{\mathrm{dem}}(\mathrm{j}\omega)|^2 \left[A^2 S_{\dot\varphi}(\omega) + \omega^2 S_{n'}(\omega)\right], \end{aligned} \tag{7.14}$$

where $S_{n'}(\omega) = S_n(\omega)$ denotes the density of $n'(t)$. When the closed loop transfer $H_{\mathrm{cl}}(\mathrm{j}\omega)$ is flat over the baseband region $\omega \in [-W, W]$, and the post-loop filter is assumed to be rectangular with bandwidth W, integration of (7.14) for the output SNR, in terms of the input CNR p and FM bandwidth W_n, yields

$$\mathrm{SNR}_{\mathrm{out}} = 3p\left(\frac{W_n}{W}\right)\left(\frac{\Delta\omega}{W}\right)^2, \tag{7.15}$$

identical to the maximum SNR (2.21) of a limiter-discriminator. This shows that above threshold, phase feedback and amplitude compression result in the same SNR improvement.

7.4 Response in the Threshold Region

This section investigates the demodulator behavior around the threshold, and discusses the principles of, and results obtained by, nonlinear noise analysis techniques described in literature.

The general characteristics of the behavior exhibited in the threshold region are similar to those of limiter-discriminators. In these demodulators, click noise is observed which, as discussed in Section 5.4.3, is generated whenever the noise FM input wave $r(t)$ gains or slips a cycle with respect of the noise-free wave $s(t)$. In phase feedback demodulators, impulsive noise is observed due to "cycle-slipping/skipping", which occurs when the oscillator output wave $s_o(t)$ slips/skips a cycle relative to the input wave $s(t)$. The modeling of cycle-slip noise is the subject of this section.

Section 7.4.1 discusses the nonlinear demodulator model required for the cycle-slip analysis. Section 7.4.2 discusses the conceptual model for the demodulator output noise in terms of a continuous noise and a cycle-slip noise component. Subsequently, the quantitative model for the calculation of the cycle-slip rate is outlined. For this purpose, Section 7.4.3 considers the relation between the cycle-slip rate and the probability density function (PDF) of the phase error φ_e. A formal description of this relation, the Fokker-Planck equation (F-PE), is considered in Section 7.4.4. Subsequently, Section 7.4.5 discusses the steady-state PDF of the phase error, and its relation to the demodulator loop structure. Section 7.4.6 combines the results of the previous sections, and discusses the dependence of the cycle-slip rate on the loop structure, while Section 7.4.7 compares the threshold curves of phase feedback demodulators with those of conventional demodulators.

7.4.1 Nonlinear Demodulator Model

The cycle-slip phenomenon can be explained from the periodicity of the phase detector nonlinearity $g(\varphi_e)$, and is therefore adequately described by nonlinear demodulator models only. In this section, we outline the demodulator model required by the cycle-slip noise analysis. Besides quantitative analyses, this model is also very suited for a qualitative analysis of the underlying cycle-slip mechanisms. First, the cycle-slip phenomenon is intuitively explained with the aid of the nonlinear differential equation (DE) of the phase error. Subsequently, the necessary conditions imposed on the structure of the loop in order to validate the the nonlinear analysis are considered.

Differential Equation of the Phase Error

The general mathematical description corresponding to the system in figure 7.3 due to the phase detector, is a nonlinear differential equation (DE) of the loop phase error $\varphi_e(t)$ [19, 20], given by

$$\dot{\varphi}_e = \dot{\varphi} - K_d K_o H_{\mathrm{lf}}(s)\left[Ag\left(\varphi_e\right) + n'(t)\right], \tag{7.16}$$

where $H_{\mathrm{lf}}(s)$ symbolically represents the differential equation operator of the loop filter operating on its input signal $Ag\left(\varphi_e\right) + n'(t)$.

The periodic nature of the detector nonlinearity $g\left(\varphi_e\right)$ causes stable and unstable equilibriums in (7.16), as depicted in figure 7.5 for a sinusoidal detector characteristic. Around the stable equilibriums located at the positive slopes of $g\left(\varphi_e\right)$, *negative* feedback exists within the loop that tries to restore the equilibrium whenever small noise perturbations in the phase error φ_e occur. This effect is sometimes called the loop's "*restoring force*" [19, 20].

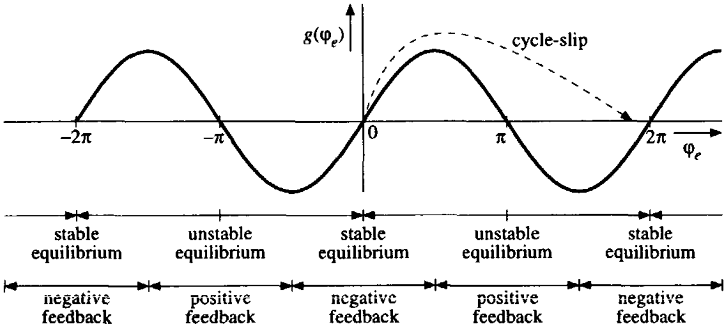


Figure 7.5: Stable and unstable equilibriums in the demodulator loop transfer.

Around the unstable equilibriums located at the negative slopes of $g\left(\varphi_e\right)$ *positive* feedback exists within the loop that drives it out of equilibrium for any perturbation in φ_e.

Consequently, when the noise is able to perturb φ_e from a stable equilibrium by an amount that is larger than π, the positive feedback rapidly increases the perturbation until φ_e arrives in the neighboring stable equilibrium, resulting in a cycle-slip.

Markov Condition

In order to validate the nonlinear analysis of the cycle-slip rate, some constraints, known as the "*Markov Conditions*" [16, 17], must be imposed on the statistics of the signals in the loop, and consequently also on the loop's structure. Here, we outline these constraints.

The analysis requires that all the signals in the loop can be described as a special type of stochastic processes, the *Markov Processes*. Markov processes are memory-less processes described by a first-order DE with a white Gaussian noise input [17]. Consequently, at every instant, their value in the near future, predicted by their time derivative, is determined only by their present value and the present value of the white noise, and not by their past. The white shape of the noise spectrum assures that this process is also memory-less; its autocorrelation function is a Dirac impulse, which means that its present value is uncorrelated with its entire past.

In a first-order loop, where $H_{\text{lf}}(s) \equiv 1$, the phase error φ_e satisfies the Markov condition whenever the input noise $n'(t)$ is wideband compared to the closed loop transfer, as observed from (7.16). In these loops, $\dot{\varphi}_e$ depends only on the (approximately) white noise $n'(t)$ and the present value of $g(\varphi_e)$.

In order to satisfy the Markov condition in higher order loops, which contain additional (dynamic) memories due to the presence of a loop filter $H_{\text{lf}}(s)$, the corresponding (m+1)-th order nonlinear DE (7.16) has to be rewritten in terms of a state-space description as a system of $(m+1)$ first-order DEs. For linear loop filters with real poles, the class of most practical interest, such a state-space description with mutually uncoupled DE's can be found.

According to [19, 20], and assuming real poles, the loop filter transfer $H_{\text{lf}}(s)$ can be written as

$$H_{\text{lf}}(s) = H_o + \sum_{k=1}^{m} \frac{H_k}{1+\tau_k s}, \tag{7.17}$$

i.e. as m parallel connected first-order filters and a direct transfer (see figure 7.6). Then the following set of $(m+1)$ first-order nonlinear DE's, equivalent

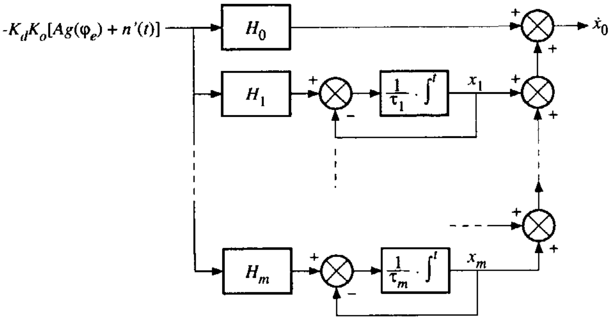

$-K_dK_o[Ag(\varphi_e)+n'(t)]$
H_0
H_1
H_m
$\frac{1}{\tau_1}\cdot\int^t$
$\frac{1}{\tau_m}\cdot\int^t$
x_1
x_m
$\dot{x}_0$
+
-

Figure 7.6: Representation of the loop filter in the analysis.

to (7.16) is then obtained:

$$\begin{aligned}
\dot{x}_0 &\overset{\text{def}}{=} \dot{\varphi}_e = \dot{\varphi}(t) - H_0 K_d K_o \left[Ag\left(\varphi_e\right) + n'(t) \right] + \sum_{k=1}^{m} x_k, \\
\dot{x}_1 &= -\frac{x_1}{\tau_1} - \frac{H_1 K_d K_o}{\tau_1} \left[Ag\left(\varphi_e\right) + n'(t) \right], \\
\vdots \quad & \quad \vdots \qquad\qquad\qquad \vdots \\
\dot{x}_m &= -\frac{x_m}{\tau_m} - \frac{H_m K_d K_o}{\tau_m} \left[Ag\left(\varphi_e\right) + n'(t) \right].
\end{aligned} \tag{7.18}$$

The analysis for filters containing complex poles or a pole in the origin (ideal integrator) is possible [20], but considerably more complex. For this reason, these cases are disregarded in the remainder of this chapter.

The state vector $\underline{x}$ of the system in (7.18) complies with the Markov conditions when the message signal $\varphi(t)$ equals a constant frequency offset Ω_o and the noise $n'(t)$ is wideband. Gaussian message signals or narrow band noise $n'(t)$ can be incorporated by writing them in terms of a state-space description with white input noise and inclusion of this description into (7.18) [16]. This rather complicated procedure will however not be considered here.

7.4.2 Nonlinear Noise Model

The output noise of a phase feedback demodulator in the threshold region can be described in similar terms to the noise at the output of non-feedback demodulators, in terms of a continuous and an impulse noise component, considered in detail in chapters 5 and 6. This section outlines the main characteristics of the phase feedback demodulator output noise. Throughout the remainder of this chapter, the term "click noise" is reserved for the impulsive noise observed at the output of 'conventional' limiter-discriminators, while the term "cycle-slip noise" is reserved for the impulsive noise in phase feedback demodulators. Further, subsequent sections do not distinguish between cycle-slips and cycle-skips.

Continuous Output Noise

The continuous noise observed at the output is due to small perturbations of the phase error $\dot{\varphi}_e(t)$ in response to the noise $n'(t)$. As discussed in Section 7.3, this noise determines the demodulator output SNR above the threshold and is conveniently calculated from the linear model.

Impulsive Cycle-Slip Noise

Similar to click noise, cycle-slip noise is conveniently modeled as a train of impulses, with the average pulse area and pulse rate as parameters. Obviously, since the underlying mechanism of cycle-slips is different from the click noise mechanism, the pulse rates are different and have to be determined by a different type of analysis. In the click noise model, the impulse approximation for the pulse shape was justified by the close and quick encirclements of the origin by the noise phasor $\vec{n}$ at CNRs around the threshold (see Section 5.4.3). In phase feedback systems, the impulse approximation is justified by the observation that the positive feedback, present in the loop during a cycle-slip, assures an extremely rapid transition between neighboring stable phase error equilibriums.

Cycle-slips may be considered as approximately independent of the continuous noise, due to their rare occurrence on the one hand, and on the other hand due to the observation that they completely overrule all other output signal components during their occurrence. Consequently, the output noise power density spectrum approximately equals the addition of the continuous noise spectral density and the density of the cycle-slip noise.

The area of the cycle-slip pulses equals exactly 2π since the stable equilibriums of $g(\varphi_e)$ are separated by $\Delta\varphi_e = 2\pi$. When the oscillator in the loop slips or skips a cycle relative to the input FM wave, its instantaneous phase changes by 2π, resulting in an impulse of area 2π in its instantaneous frequency. The demodulator output signal, observed at the oscillator input, is proportional to this frequency.

Cycle-Slip Bursts

The most important difference between the cycle-slip model and the click noise model is that, as opposed to clicks, cycle-slips cannot be considered to be mutually independent around the threshold in all circumstances.

Theoretical and experimental evidence has only been given for the independence of cycle-slips for a first-order loop [7, 16].

For higher order loops, however, dependencies between consecutive cycle slips resulting in cycle-slip bursts occur when the closed loop transfer contains poles with high mutual interaction, such as complex poles. In these systems, locking on a stable phase equilibrium requires that all other state variables x_k, each corresponding to one of the loop filter poles, also adopt (values close to) their corresponding equilibrium value, attained when $\dot{x}_k = 0$.

During a cycle-slip, the state variable $x_0 = \varphi_e$, corresponding to the oscillator pole, is 'bumped' out of its equilibrium and, in case of strong interaction, pulls the other states out of equilibrium as well. In such an unstable transition state, the system is extremely vulnerable to the occurrence of another, consecutive cycle-slip. The cycle-slip burst continues until the 'DC component' of the phase detector output has built up sufficient energy to restore all states to their equilibrium. Consequently, a burst of cycle-slips is observed whenever the magnitude of the phase detector 'DC component' is insufficient to restore all states to equilibrium within a single slip or skip. In second-order loops with a sinusoidal phase detector transfer, this is the case when the damping factor of the loop poles, ζ, satisfies the condition $\zeta < 0.9$ [7, 26].

Cycle-Slip Noise Spectral Density

In the presence of dependencies between cycle-slips, their double-sided spectral density is no longer precisely described by the equivalent of (5.11), the product of the squared noise pulse area $4\pi^2$, and the total average cycle-slip or skip rate $N_+ + N_-$. This is a consequence of the fact that such cycle-slip bursts are not Poisson distributed in time, like click noise. However, although formally incorrect, we will still use (5.11) as an approximation of the cycle-slip noise spectrum of such loops whenever necessary. This may be justified by the observation that its hardly worth while to develop a more accurate description, even if this were possible, since cycle-slip bursts are highly undesirable in virtually any demodulator intended for low CNRs. Systems with long cycle-slip bursts are therefore unsuited for such applications.

Thus, in order to calculate the power-density spectrum, we only need an expression for the average cycle-slip rate. The calculation of this rate is outlined in subsequent sections.

7.4.3 Cycle-Slips and the Phase Error Probability Density

The cycle-slip rate is strongly dependent on the PDF of the phase error process φ_ε. Therefore, knowledge of this PDF and its dependence on the structure of the demodulator loop is highly desirable in the design of phase feedback demodulators for low CNRs.

In this section, we discuss the qualitative relation between the cycle-slip rate and the phase error PDF with the aid of a charged-particle analogon. This relation constitutes the principles of the nonlinear cycle-slip rate analysis discussed in subsequent sections.

Evolution of the Phase Error PDF as a Function of Time

A qualitative analysis of the phase error PDF's evolution as a function of time is an effective means of gaining insight into the underlying probabilistic mechanisms of cycle-slips and their mathematical formulation. For this reason, we outline such an analysis here in advance of a discussion of the mathematical results.

Figure 7.7 schematically depicts the evolution of the phase error PDF in response on the stationary noise process $n'(t)$, possessing a constant variance and switched on at time $t = 0$ (see also [16, 17]). Assume that input wave $s(t)$ contains no modulation, i.e. $\varphi(t) \equiv 0$, and that the loop is locked on the

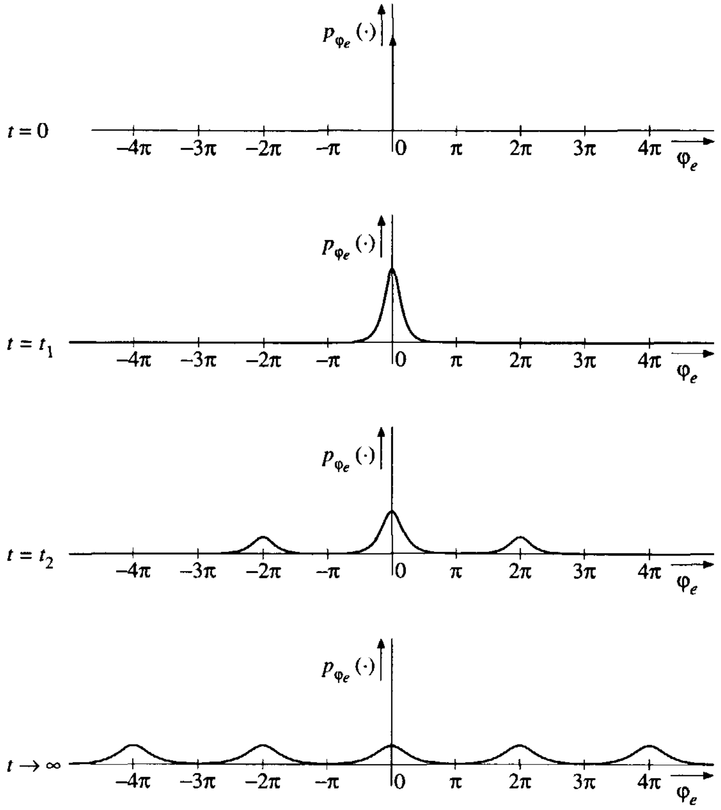


Figure 7.7: Evolution of the phase error PDF as function of time, and its relation to cycle-slips.

equilibrium $\varphi_e = 0$ prior to $t = 0$ with a zero SSPE. Further, assume a sinusoidal phase detector, as depicted in figure 7.5.

Since there is no uncertainty about the value of φ_e at $t = 0$, its initial PDF equals an impulse of unit area located at $\varphi_e = 0$,

$$p_{\varphi_e}(\varphi_e; 0) = \delta(\varphi_e). \tag{7.19}$$

As a result of the noise injection into the loop starting at $t = 0$, φ_e starts to wander around the equilibrium in an approximately Gaussian fashion. This causes the PDF to spread around φ_e, such that it becomes Gaussian, as depicted for $t = t_1$. This "diffusion" of the PDF continues as long as the restoring force of the loop is capable to keep the perturbations of φ_e smaller than π.

Eventually, $|\varphi_e|$ exceeds π, resulting in a cycle-slip. After the slip, the loop locks either on the equilibrium at $\varphi_e = -2\pi$ or at $\varphi_e = 2\pi$. From that instant on, the area underneath the PDF around the previous equilibrium position $\varphi_e = 0$ starts to shrink, while its area around $\varphi_e = 2\pi$, or $\varphi_e = -2\pi$ simultaneously starts to swell. This swelling continues until the next cycle-slip occurs, causing the loop to lock on another equilibrium, etc., as depicted for $t = t_2$.

In the steady state, reached as $t \rightarrow \infty$, all stable equilibriums have become equally likely, which means that the PDF becomes periodic while its variance becomes unbounded. As a consequence, since the PDF is of unit area, the area enclosed by it within any finite phase interval approaches zero. This solution for the steady-state PDF is rather impractical since it separately describes the statistical behavior, including the transitions between an equilibrium and its neighbors (cycle-slips), of all individual periods in the PD transfer. However, as far as the cycle-slip rate is concerned, only the total number of transitions between any neighboring equilibriums is of interest.

A better suited solution is therefore obtained by exploiting the periodicity of the PDF. Usually, the PDF is bounded to an interval $\Delta\varphi_e = 2\pi$, equal to one cycle of the PD nonlinearity. In the "periodic extension" (PE) approach [16, 17, 19, 20], this is achieved by replacement of $p_{\varphi_e}(\varphi_e)$ in the calculations by a periodic function, normalized to a period of 2π. In another approach, the "renewal process" theory [27–29], the phase process φ_e is truncated ("killed") each time it exceeds $\pm\pi$, one period of the PDF, and replaced by a new process that starts again with $\varphi_e = 0$ and the initial PDF of equation (7.19). It can be shown that both approaches are equivalent as far as cycle-slip rates are concerned. In general, the renewal process theory yields slightly more information concerning the behavior of the phase error process than the periodic extension approach.

Description of Cycle-Slips by Probability Theory

In probability theory, processes like the phase error φ_e are called "random walk" or "Brownian motion" processes [20]. For such processes, advanced theories have been developed in conjunction to other physical phenomena. In subsequent sections of this Chapter, the random walk nature of the phase error φ_e is explained with the aid of a charged-particle analogon. In this analogon, the phase error process φ_e is represented by a "probability particle" that behaves similar to an electron.

An important concept in probability theories is the so called "*probability current density*" [19, 20]. This term describes nothing more than the swell of the area enclosed by the PDF at one place in (the state-) space, and the simultaneous decrease of the area at another place, as a function of time. During such a process, probability may be considered to 'flow' from one place to the other, just like an electric current can be considered to be a flow of electric charge.

In phase feedback demodulators, cycle-slips are the cause of a "flow of probability" from an equilibrium state of the phase error process to one of its neighbors: each cycle-slip corresponds to the transition of a single 'probability particle'. Using this observation, it can be shown that the number of cycle-slips per unit time, the slip rate, corresponds to the value of the probability current density at the unstable equilibriums $\varphi_e = \pm\pi + 2k\pi$, the boundaries of one period of the phase error PDF [19, 20]. This is of course due to the fact that every passage of φ_e across these boundaries results in a cycle-slip.

With the aid of these concepts, the subsequent sections discuss the dependency of the phase error PDF in the steady-state and the cycle-slip rate on the structure of the demodulator loop.

7.4.4 Fokker-Planck Equation for the Phase Error PDF

This section discusses the meaning of the Fokker-Planck equation for the phase error PDF that gives a formal description of the relation between the PDF and the probability current density, related to the cycle-slip rate. Various probabilistic measures are defined that are required in subsequent sections for the description of the cycle-slip rate.

Interpretation of the Fokker-Planck Equation

The (steady-state) phase error PDF can be calculated with the aid of *Fokker-Planck* (F-P) techniques, which are based on the notion of probability current density discussed in the previous section.

This approach, developed in the 1930's, was first applied to phase-lock systems by Tikhonov [21, 22]. He obtained the steady state PDF of a first-order loop and a particular type of second-order loop. In the same period, others tried to obtain results for the threshold behavior in PLLs by means of various kinds of linearization techniques [14, 15, 30, 31], but none of these approaches resulted in a satisfactory solution of the problem. Viterbi [16, 17] used the F-P approach to obtain cycle-slip rates for a first order loop, considered in some more detail in [32–34], and a special type of second-order loop. Lindsey [19, 20] generalized the approach to higher order loops and arbitrary types of phase detectors. This work was subsequently elaborated by many authors [27–29, 35–41].

The *Fokker Planck* equation (F-PE) is a second-order nonlinear partial DE that basically describes the flow of probability through the state-space as a function of time, of systems described by Markov Processes. Its solution equals the PDF of these processes. In this respect, it is sometimes called the "equation of flow" [19], or a "generalized diffusion equation" [16].

The joint PDF of the state variables in (7.18) is described by an $(m + 1)$-dimensional F-PE. For the cycle-slip rate analysis, however, only the marginal

PDF of the phase error process, the projection of the joint PDF in the state-space on the plane through the φ_e-axis, is of interest. By application of the proper boundary conditions, the one-dimensional F-PE for this marginal PDF can be derived from the $(m+1)$ dimensional F-PE of the joint PDF [19]. This one-dimensional equation can be expressed as

$$\frac{\partial \mathcal{J}_0(\varphi_e, t)}{\partial \varphi_e} + \frac{\partial p_{\varphi_e}(\varphi_e; t)}{\partial t} = 0, \tag{7.20}$$

where $\mathcal{J}_0$ denotes the probability current density flowing through the state-space in the direction of φ_e. This expression is equivalent to the physical law for the conservation of electric charge in differential format [42]. It states that sources of probability currents, characterized by a decrease in time of probability, are located at those positions where the net (outgoing) probability currents are nonzero.

Components of the Probability Current Density

The meaning of $\mathcal{J}_0$ is illustrated by figure 7.8. This figure depicts the observation that $\mathcal{J}_0$ can be subdivided into a 'drift' component and a 'diffusion' component [19, 20].

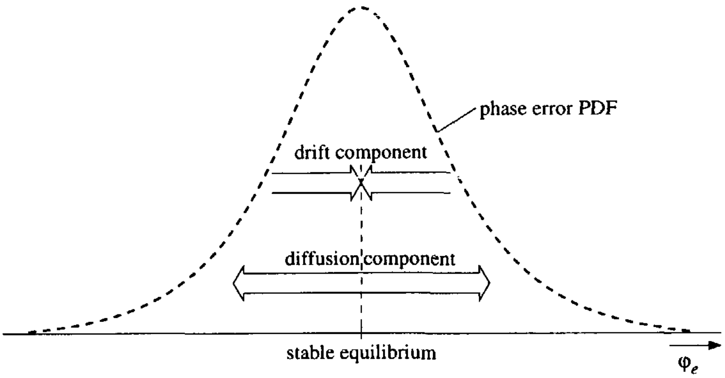


Figure 7.8: Components of the probability current density $\mathcal{J}_0$.

The drift component, denoted by $\mathcal{J}_{0,\text{drift}}$, is due to the restoring force of the loop that drives it towards its stable equilibriums. It equals the product of the probability density and some position and time-dependent velocity v,

$$\mathcal{J}_{o,\text{drift}}(\varphi_e, t) = v(\varphi_e, t)\, p_{\varphi_e}(\varphi_e; t). \tag{7.21}$$

It is directed towards the stable equilibriums since it increases the probability that φ_e attains values close to these equilibriums.

The diffusion component, denoted by $\mathcal{J}_{o,\text{diff}}$, represents the random motion of φ_e, eventually resulting in cycle-slips. This component, directed outwards the stable equilibriums due to its tendency to increase the variance of φ_e, equals a diffusion constant D times the gradient of the probability density, which in this respect may be considered to be the concentration of 'probability particles':

$$\mathcal{J}_{o,\text{diff}}\left(\varphi_e, t\right) = -D\frac{\partial p_{\varphi_e}\left(\varphi_e; t\right)}{\partial \varphi_e}. \tag{7.22}$$

The velocity v and diffusion constant D describe the influence of the 'medium' on the probability currents, and are thus determined by the state-space description (7.18) of the loop filter and the phase detector nonlinearity. Differences in behavior between phase feedback demodulators are therefore due to differences in these parameters.

Restoring Force and Potential Function

Two valuable quantities that complete the description of the phase error process are the restoring force $h_0\left(\varphi_e, t\right)$ and the potential energy function $U_0\left(\varphi_e, t\right)$ corresponding to this force. The restoring force $h_0\left(\varphi_e, t\right)$ is defined as

$$h_0\left(\varphi_e, t\right) \overset{\text{def}}{=} \frac{v\left(\varphi_e, t\right)}{D} = \beta - \alpha g\left(\varphi_e\right). \tag{7.23}$$

In the electric analogon of this expression, the velocity $v\left(\varphi_e, t\right)$ corresponds to the product of the particle mobility and the electric field strength, while the restoring force $h_0(.)$, the ratio $v\left(\varphi_e, t\right)/D$, corresponds to the ratio of the electric field strength and the thermal voltage $U_T = kT/q$. The electric field applies a force to charged particles, resulting in a drift current. The division through U_T represents the effect that the influence of this force decreases when the random kinetic energy of the particles, thermal random motion, increases.

In the stochastic phase feedback demodulator model, $h_0(.)$ represents a stochastic field that applies a force to probability particles, resulting in a drift component of the probability current density. Due to the division by the diffusion constant D, (7.23) already includes the effect, the influence of this field-force decreases when the demodulator input CNR decreases: the demodulator input noise $n(t)$ increases the kinetic energy of the probability particles through random motion, in a similar way as thermal energy increases the random motion in the electric analogon.

The second equality in (7.23) relates the phase error PDF to physically interpretable characteristics of the demodulator loop by means of the parameters α and β. As will be explained in Section 7.4.5, α represents the "effective" SNR inside the demodulator closed-loop bandwidth. Thus, this parameter represents the strength of the stochastic field force, in comparison to the random motion of the phase error. At high CNRs, large values of α, the restoring force is large, while at low CNRs, the force is weak. The PD transfer in (7.23) represents the fact that the restoring force is a function of the position in the state-space: by definition the force vanishes at the equilibriums, while it is nonzero outside these equilibriums. The minus-sign represents the fact that the force drives the particles towards the stable equilibriums. The parameter β equals the product of α and the steady-state phase error, as subsequently shown. A nonzero value of β, corresponding to a nonzero SSPE, reduces the influence of the restoring force on the probability particles. In the presence of an SSPE part of the restoring force is required to keep the local oscillator at the correct frequency, which leaves a smaller 'force' to counteract the random motion of probability particles.

The potential function $U_0(\varphi_e, t)$ is defined as the integral of the restoring force, similar to the relation between the electric field and the electric potential:

$$\begin{aligned} U_0(\varphi_e, t) &\overset{\text{def}}{=} -\int^{\varphi_e} h_0(x, t)\,\mathrm{d}x \\ &= -\beta\varphi_e + \alpha\int^{\varphi_e} g(x)\mathrm{d}x. \end{aligned} \tag{7.24}$$

The maxima and minima of this potential function correspond to the positions at which no restoring force is present, to the unstable and stable equilibriums of $g(\varphi_e)$ respectively. The instantaneous value of the phase process φ_e may be interpreted as a (probability) particle that exhibits a Brownian motion and is subject to the potential 'field' $U_0(\varphi_e, t)$.

7.4.5 Steady-State Solution for the PDF

This section discusses the steady-state phase error PDF and its relation to the structure of the demodulator loop. This PDF determines the dependence of the cycle-slip rate on the loop structure. The theoretical results discussed in this section constitute the discussion on phase feedback demodulator design in Section 7.5.

Expression for the Steady-State Solution

In the steady state, the phase error PDF becomes time-independent, which means that its time derivative in (7.20) vanishes. Consequently, according to (7.20), (7.21) and (7.22), the drift and diffusion components of $\mathcal{J}_0$ are in balance in the steady state,

$$v(\varphi_e, t)\, p_{\varphi_e}(\varphi_e; t) = D\frac{\partial p_{\varphi_e}(\varphi_e; t)}{\partial\varphi_e}. \tag{7.25}$$

An expression for the steady-state PDF can be obtained by solving (7.25). According to [19, 20], this results in

$$p_{\varphi_e}(\varphi_e) = C_0 \exp\left[-U_0(\varphi_e)\right] \int_{\varphi_e}^{\varphi_e+2\pi} \exp\left[U_0(x)\right] dx, \tag{7.26}$$

where C_0 follows from the condition that the area underneath $p_{\varphi_e}(\varphi_e)$ in the interval $\varphi_e \in [-\pi, \pi]$ equals unity, while $U_0(.)$ denotes the potential function given by (7.24).

In order to complete the solution for the PDF, the parameters α and β need to be expressed in terms of parameters of the demodulator loop and the input noise $n'(t)$. However, exact expressions for these parameters cannot be found for second or higher order loops since they contain some conditional expectations that are subject to the "chicken-egg" problem; their evaluation requires the PDF, while the PDF requires their evaluation. In order to solve this problem, one may choose between measurement of these expectations from an already implemented system, simulation, or approximation. For synthesis, the latter option is the most convenient one since it reveals the relation between the key parameters of the demodulator loop and the loop statistics.

With the aid of a Linear Mean Square Estimate (LMSE), Lindsey [19] obtains approximate expressions for α and β, which in terms of the parameters used in this chapter become

$$\alpha \approx \frac{2A}{N_o H_0 K_o K_d}\left(1 - \frac{S_G(0)}{2\sigma_G^2}\sum_{k=1}^{m}\frac{H_k}{\tau_k H_o}\right), \tag{7.27}$$

$$\beta \approx \frac{2A}{N_o H_0 K_o K_d}\left[\frac{\Omega_o}{A H_0 K_o K_d} - \overline{g}\sum_{k=1}^{m}\frac{H_k}{H_0}\left(1 + \frac{S_G(0)}{2\sigma_G^2 \tau_k}\right)\right], \tag{7.28}$$

where N_o represents the power spectral density of the low pass equivalent input noise $n_i(t)$, $n_q(t)$ (see Chapter 2). Further, $\overline{g}$ denotes the expected value, the DC component, of the PD nonlinearity $g(\varphi_e)$ in figure 7.3, σ_G^2 denotes its variance, and $S_G(0)$ its power spectral density of $G(\varphi_e) = g(\varphi_e) - \overline{g}$ at $\omega = 0$. These three parameters are yet undetermined and cannot be obtained exactly from the nonlinear model of the loop. However, fortunately, they can be approximated with the aid of the linear model of figure 7.4, as discussed subsequently.

In advance of a detailed discussion on the relation of (7.27) and (7.28) to the demodulator loop structure, it is already possible to get a grasp on the meaning of these expressions. In the first place, it is observed that α is related to the SNR inside the loop bandwidth since it contains the carrier amplitude in the numerator, and the noise power spectral density times a measure for the DC loop gain in the denominator. The parameter β is related to the steady-state phase error since it is a function of the constant frequency offset Ω_o, which is used as a model for the message modulation, and the average value $\overline{g}$ of the PD nonlinearity.

Secondly, the summations over k in both expressions represent the effect of a loop filter on the phase error PDF. Such a filter introduces additional states into the system and additional dimensions into the corresponding state-space. Consequently, in order to establish proper operation of the loop, the phase detector output signal, the restoring force, should not only drive the phase error into one of its stable equilibriums but should also drive all other state-variables to their equilibriums. In comparison to a first-order system, a higher-order loop therefore has to satisfy far more severe requirements in order to operate properly, which explains the minus-sign of the summations in (7.27) and (7.28).

Finally, it is observed that the SNR inside the loop bandwidth, and therefore also α and β, decreases when the DC loop gain increases. The loop gain affects the signal and the noise inside the loop in the same way such that the SNR measured in a *fixed* bandwidth is not affected by it. However, an increase in the loop gain also results in an increase in the closed-loop bandwidth, and therefore in an increased level of noise inside the loop bandwidth.

Relation between PDF and Loop Structure

Insight into the meaning of (7.27) and (7.28), and their relation to the structure of the demodulator loop, is gained from the low-pass equivalent demodulator loop model depicted in figure 7.9, obtained by substitution of the loop filter from figure 7.6 into the general model of figure 7.3. When this loop is degenerated

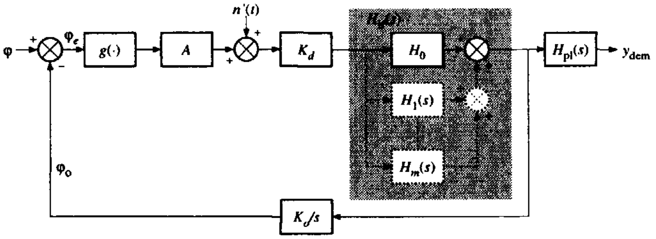


Figure 7.9: Low-pass equivalent model of the demodulator, containing the loop filter of figure 7.6.

to a first-order loop by setting $H_k = 0$ for $1 \leq k \leq m$, the first factor in (7.27) and (7.28) becomes equal to a factor that is usually called the "SNR inside the (first-order) loop bandwidth" [16, 17, 19, 20], which equals twice the CNR within the loop bandwidth, thus

$$\alpha_0 \overset{\text{def}}{=} \alpha\big|_{H_{\text{lf}}(s)=H_0} = \frac{2A}{N_o H_0 K_o K_d} = \frac{A^2}{N_o B_{L,0}} = 2\left(\frac{W_n}{W_{L,0}}\right) p, \tag{7.29}$$

where $B_{L,0}$ denotes the noise bandwidth of the degenerated loop.

Further, for the same degenerated loop, the first term inside the brackets in (7.28) equals the SSPE that would be obtained with the linear model,

$$\varphi_{e,\mathrm{ss},0} \overset{\mathrm{def}}{=} \varphi_{e,\mathrm{ss}}\big|_{H_{\mathrm{lf}}(s)=H_0} = \frac{\Omega_o}{AH_0K_oK_d}. \tag{7.30}$$

For a first-order loop, expressions (7.27) and (7.28) are no longer approximations, but become exact [16]. In that case, $\alpha = \alpha_0$ and $\beta = \alpha_0\varphi_{e,\mathrm{ss},0}$, corresponding to a restoring force given by

$$\begin{aligned} h_0\left(\varphi_e\right)\big|_{H_{\mathrm{lf}}(s)=H_0} &= \alpha_0\left[\varphi_{e,\mathrm{ss},0} - g\left(\varphi_e\right)\right] \\ &= 2\left(\frac{W_n}{W_{L,0}}\right)p\left[\varphi_{e,\mathrm{ss},0} - g\left(\varphi_e\right)\right]. \end{aligned} \tag{7.31}$$

For higher order loops, these expressions merely describe the properties of the degenerated, first-order loop, the 'intrinsic first-order' behavior, *minus* some correction term that depends on the position of the poles. This minus sign reflects the effect that the effective SNR inside the loop is smaller in higher order loops than in first-order loops.

As suggested by Viterbi [16], approximate expressions for the undetermined parameters $S_G(0)$ and σ_G^2 can be obtained from the linear model as follows. The transfer from the point in the loop where the noise is injected, to the output of the linearized detector transfer $g\left(\varphi_e\right) = \varphi_e$ equals $H_{\mathrm{cl}}(s)/A$. Consequently, $S_G(\omega)$ approximately equals

$$S_G(\omega) \approx S_{\varphi_e}(\omega) = |H_{\mathrm{cl}}(\mathrm{j}\omega)|^2\frac{N_o}{A^2}. \tag{7.32}$$

With the aid of this expression, we obtain

$$\frac{S_G(0)}{2\sigma_G^2} \approx \frac{|H_{\mathrm{cl}}(0)|^2}{\frac{1}{\pi}\int_{-\infty}^{\infty}|H_{\mathrm{cl}}(\mathrm{j}\omega)|^2\mathrm{d}\omega} = \frac{1}{2B_L}, \tag{7.33}$$

where B_L denotes the double-sided noise bandwidth of the loop in Hz. By extrapolation of Viterbi's approach, an approximate expression for $\overline{g}$ can also be obtained from the linear model. It is observed that this parameter equals the SSPE,

$$\overline{g} \approx \varphi_{e,\mathrm{ss}}. \tag{7.34}$$

Alternatively, the definition formula for $\overline{g}$, the expectation of $g\left(\varphi_e\right)$ taken over φ_e, could be solved numerically, where the phase error PDF contains $\overline{g}$ as variable.

With the aid of these approximate expressions, α and β may finally be expressed as

$$\alpha \approx 2\left(\frac{W_n}{W_{L,0}}\right) p\left(1-\sum_{k=1}^{m}\frac{H_k}{2H_0 B_L \tau_k}\right), \tag{7.35}$$

$$\begin{aligned}\beta &\approx 2\left(\frac{W_n}{W_{L,0}}\right) p\left[\varphi_{e,\mathrm{ss},0}-\varphi_{e,\mathrm{ss}}\sum_{k=1}^{m}\frac{H_k}{H_0}\left(1+\frac{1}{2B_L\tau_k}\right)\right] && (7.36)\\ &= \alpha\left[\varphi_{e,\mathrm{ss},0}-\varphi_{e,\mathrm{ss}}\sum_{k=1}^{m}\frac{H_k}{H_0}\right] && (7.37)\\ &= \alpha\varphi_{e,\mathrm{ss}}. && (7.38)\end{aligned}$$

The last expression for β is an interesting result because it is the higher order analogon of the exact expression obtained for the first order loop. It shows the previously mentioned fact that β is proportional to the SSPE.

Thus, by substitution of (7.35) and (7.38) into (7.24) and (7.26), it is concluded that the phase error PDF is determined by the CNR inside the bandwidth of the degenerated loop, the SSPE of the loop itself, the noise bandwidth B_L and the loop filter parameters H_k, τ_k. This representation of the phase error PDF will be used in the investigations of cycle-slip rates in the next section.

Example: Steady-State PDF for a Sinusoidal Nonlinearity

An indication of the correctness of the results is obtained by plots of the PDF of a sinusoidal PD for various values of the input CNR.

Figure 7.10 depicts the results for $\beta = 0$, a zero SSPE, while figure 7.11 depicts the results for $\beta/\alpha = 0.75$. The equilibrium of the phase error is located at the value of φ_e where the restoring force equals zero, where, according to (7.23), $\beta/\alpha = g(\varphi_e)$. In case of a sinusoidal PD nonlinearity, $\beta/\alpha = 0.75$ corresponds to an equilibrium, an SSPE, of $\varphi_e = \varphi_{e,\mathrm{ss}} = \arcsin(\beta/\alpha) = 0.85$ (rad). Note that this result does not follow from (7.38). In that expression, the PD nonlinearity was ignored through the application of the approximate expression (7.34). Thus, $\overline{g} \approx \varphi_{e,\mathrm{ss}}$ should actually be replaced by $\overline{g} \approx g(\varphi_{e,\mathrm{ss}})$. Figure 7.10 clearly demonstrates the spreading of the PDF, the "diffusion of probability", when the input CNR decreases.

At high CNRs, the PDF approaches a Dirac impulse centered at the steady-state phase error $\varphi_{e,\mathrm{ss}}$. The curves at figure 7.10 are therefore centered at $\varphi_e = 0$ (rad), while those in figure 7.11 are centered at $\varphi_e = 0.85$ (rad).

At low CNRs, the PDF gradually approaches the PDF of the phase of the input noise $n(t)$, a uniform density with a zero mean value [43]. Since the SSPE equals the zero-mean of the noise phase in figure 7.10, all PDF curves have their maximum value at $\varphi_e = 0$. The maximum of the PDF curves in

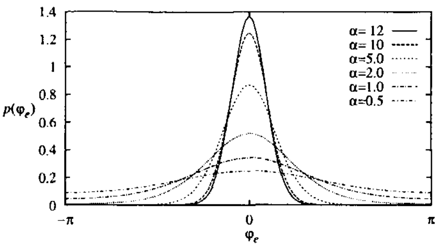


Figure 7.10: Phase error PDF for $\beta = 0$.

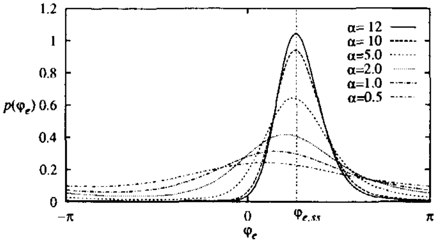


Figure 7.11: Phase error PDF for $\beta = 0.75\alpha$.

figure 7.11 gradually moves from $\varphi_e = \varphi_{e,\text{ss}}$ towards the mean of the noise phase, $\varphi_e = 0$, for decreasing CNRs.

7.4.6 Derivation of Cycle-Slip Rates from the PDF

This section uses the theory discussed earlier to obtain an expression for the average cycle-slip rate. First, it is shown that the results obtained for the steady-state phase error PDF yield useful information about the dependence of the cycle-slip rate on the structure of the demodulator loop and, as discussed in Section 7.5, may be used to optimize the loop configuration. Subsequently, the expression for the average cycle-slip rate is discussed and compared to the click rate in limiter-discriminators.

Cycle-Slip Rate and the Loop Structure

The relation between the cycle-slip rate and the structure of the demodulator loop can be analyzed with the aid of the results described in the previous section for the steady-state PDF.

In the probabilistic model for the demodulator described previously, the cycle-slip rate was represented by the probability current density that flows across the bounds at $\varphi_e = \pm\pi$, the instable equilibriums of the phase error DE. Because the restoring force equals zero here, this current density consists only of the diffusion component $\mathcal{J}_{0,\text{diff}}$.

The dependence of this diffusion current on the loop structure is described by the potential function $U_0(\varphi_e)$, which is plotted in figure 7.12 for a sinusoidal and an ideal sawtooth phase detector for various values of α. Due to the diffusion component of the probability current density, the phase error φ_e, which may in this respect be represented by a particle, performs a Brownian motion through this potential field.

A cycle-slip occurs whenever φ_e moves across one of the maxima in this potential field. In lock, φ_e is positioned at a stable equilibrium, at the bottom

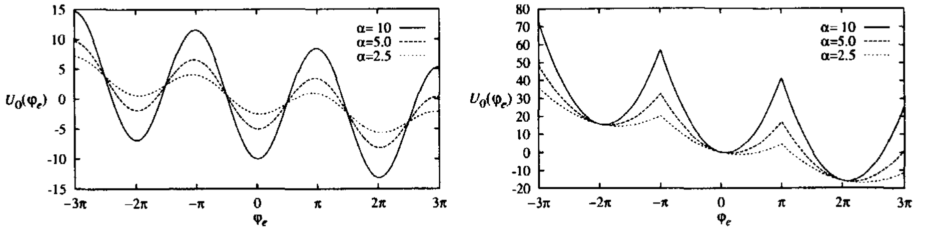


Figure 7.12: Potential functions $U_0\left(\varphi_e\right)$ of a sinusoidal phase detector (a) and an ideal sawtooth detector (b), for various values of the effective SNR α.

of one of the potential wells. Thus, in order to minimize the cycle-slip rate, the wells at the stable equilibriums should be made as deep as possible, while the potential barriers at the unstable equilibriums should be made as high as possible. Notice that the absolute level of the potential function is not of interest to the behavior of the loop: only the potential differences that exist between different values of φ_e are of interest. This is similar to the way in which an absolute potential in electronic circuits is meaningless: only the potential difference with respect to a predefined reference, e.g. a "ground" potential, contains relevant information about the circuit behavior.

The depth of the wells and height of the barriers is proportional to α and thus to the input CNR, as should be expected. At high CNRs, the wells are very deep, which means that cycle-slips are very unlikely to occur, as should be expected. Expression (7.35) implies that, since the sign of the correction term for higher order loops is negative, a first-order loop is likely to produce the smallest possible amount of cycle-slips, a fact that is known to be true in practice [7]. For higher order loops, the cycle-slip rate will be larger. Further, as observed by comparing figures 7.12(a) and 7.12(b), the shape of the PD nonlinearity has a profound influence on the wells and barriers.

The parameter β, which is proportional to the SSPE, introduces a linear slope in the potential function that increases the potential barriers at one side, but decreases them at the other side of the wells. Since the latter effect increases the total cycle-slip rate, β, and thus the SSPE, should be made as small as possible.

Further, it is observed that locking is possible only when $U_0\left(\varphi_e\right)$ has minima, i.e. when the expression

$$g\left(\varphi_e\right)=\frac{\beta}{\alpha}=\varphi_{e,\mathrm{ss}} \tag{7.39}$$

can be satisfied at a positive slope of $g\left(\varphi_e\right)$. This condition is known as the *steady-state lock-limit* [7].

Expression for the Cycle-Slip Rate

The cycle-slip rate can be calculated from the phase error PDF through the following procedure. First, it should be recalled that the net rate, the difference between the positive cycle-slip rate N_+, that increases the phase error and the negative cycle-slip rate N_- that decreases the phase error, corresponds to the magnitude of the probability current density at the unstable equilibriums, as discussed previously. At these values of the phase error φ_e the drift component of the probability current vanishes, leaving only the diffusion component described by (7.22). This diffusion component can be expressed in terms of the parameters α and β, which describe the dependence of the PDF on the loop structure, through (7.24) and (7.26).

An expression for the total cycle-slip rate $N_+ + N_-$ is obtained through combining the expression for the net cycle-slip rate $N_- - N_-$ with an expression for the ratio N_+/N_-. According to [19, 35], it follows from the physical analogon of the probability current density that the ratio of N_+ and N_- equals

$$\frac{N_+}{N_-} = \exp(2\pi\beta). \tag{7.40}$$

This expression shows that β, which is determined by the SSPE, introduces an imbalance between positive and negative cycle-slips, corresponding to unequal heights of the potential barriers at both sides of the potential wells, as discussed earlier.

With the aid $N_+ - N_-$, N_+/N_- and results obtained in [19], the total rate of slipping cycles can then be expressed as

$$N_+ + N_- = C_0 \frac{2B_{L,0}}{\alpha_0} \coth(\pi\beta) \left| \exp(-2\pi\beta) \exp\left(\alpha \int_{-\pi}^{\pi} g(x)\mathrm{d}x\right) - 1 \right|, \tag{7.41}$$

where C_0 denotes the normalization constant of the phase error PDF (7.26), $B_{L,0}$ equals the double sided noise bandwidth of the degenerated loop in (Hz), and α_0 the corresponding SNR, given by (7.29).

For a loop with sinusoidal PD, the total cycle-slip rate, denoted by $N_{\mathrm{sin,tot}}$, (7.41) can be written, by combining the results from [19] and [35], as

$$N_{\mathrm{sin,tot}} = \frac{B_{L,0}\beta \coth(\pi\beta)}{\pi\alpha_0 \left[I_0^2(\alpha) + 2\sum_{m=1}^{\infty} (-1)^m \frac{I_m^2(\alpha)}{1+\left(\frac{m}{\beta}\right)^2} \right]}. \tag{7.42}$$

Although this expression looks quite intractable, the series in the denominator converges quickly, and allows truncation after only a few terms. The Bessel-functions $\mathrm{I}_m(\alpha)$ behave approximately exponentially. Consequently, when the

SNR inside the loop, α, increases, the cycle-slip rate decays exponentially. For small β, $\beta\coth(\pi\beta) \approx \left[1 + (\pi\beta)^2/3\right]/\pi$, while for large β, $\beta\coth(\pi\beta) \to |\beta|/\pi$. This shows that $\beta = 0$, i.e. a zero-valued SSPE, yields the smallest possible cycle-slip rate.

Comparison between Cycle-Slip Rate and Click Rate

It is interesting to compare expression (7.42) with expression (5.12) for the click rate in a conventional demodulator in the absence of modulation. For this purpose, we take $\beta = 0$, resulting in the smallest possible cycle-slip rate, denoted by $N_{\text{sin,min}}$:

$$N_{\text{sin,min}} = \frac{B_{L,0}}{\pi^2 \alpha_0 I_0^2(\alpha)}. \tag{7.43}$$

This result agrees with the cycle-slip rate for the first-order loop obtained in [16] when $\alpha = \alpha_0$. For high input CNRs, an asymptotic expansion may be applied to the Bessel function in the denominator [44, 45]. If we further assume a first-order loop, the result becomes

$$N_{\text{sin,min}} \to \frac{2B_L}{\pi} \exp\left[-2p\left(\frac{W_n}{W_L}\right)\right]. \tag{7.44}$$

For high input CNRs, expression (5.12) for the click rate becomes

$$N_{\text{clk}} \to \frac{r\exp(-p)}{\sqrt{\pi p}}. \tag{7.45}$$

Although both rates increase exponentially for decreasing CNRs, the cycle-slip rate is seen to be related to the *closed loop bandwidth*, which is in the same order of magnitude as the message bandwidth, while the click rate is related to the *FM transmission bandwidth.* Consequently, the cycle-slip rate can be considerably smaller than the click rate.

These "threshold extending" capabilities of phase feedback demodulators are due to the fact that the phase feedback mechanism exploits the property of wideband FM waves that the message bandwidth is considerably smaller than the FM transmission bandwidth. A conventional demodulator does not exploit this information; except for the baseband output filter, the entire demodulator has a bandwidth that is (at least) equal to the FM transmission bandwidth.

7.4.7 Output SNR

This section combines the results of the linear and nonlinear noise analyses to produce the threshold curves for phase feedback demodulators, which, as known, relate the output SNR to the input CNR. These curves are subsequently compared to the threshold curve of a limiter-discriminator.

The calculations of the output SNR are limited to the case in which modulation is absent, where interaction between the FM message and the noise is ignored. This restriction was applied in the nonlinear analysis in order to keep the results tractable. Further, a first-order loop is assumed, which yields the highest possible SNR and lowest possible threshold. The threshold curves for phase feedback demodulators derived below therefore define an ***upper bound on the threshold extension*** that can be achieved by means of phase feedback.

If, for simplicity, a sinusoidal phase detector nonlinearity is chosen, a proper choice in practice (see Section 7.5), and the FM transmission bandwidth W_n is chosen according to Carson's formula (2.8), the following expression for the output SNR is obtained with the aid of the previously discussed results:

$$\mathrm{SNR_{ph}} = \frac{3p\left(\frac{W_n}{W}\right)\left(\frac{\Delta\omega}{W}\right)^2}{1+\frac{12\gamma p\left(\frac{W_n}{W}\right)\coth\left(\gamma p\frac{W_n}{W}\right)}{\pi\left[I_0^2\left(p\frac{W_n}{W}\right)+2\sum_{m=1}^{\infty}\frac{I_m^2\left(p\frac{W_n}{W}\right)}{1+\left(\frac{mW}{\gamma p W_n}\right)^2}\right]}}, \tag{7.46}$$

where $\gamma = \beta/\alpha$ for small values of $\varphi_{\mathrm{e,ss}}$, $\gamma \approx \varphi_{\mathrm{e,ss}}$, as obtained according to the linear demodulator model. For large SSPEs however, the nonlinearity of the PD transfer has to be included, such that $\gamma = \sin\varphi_{\mathrm{e,ss}}$. The latter equality follows from (7.39). The double-sided noise bandwidth B_L has been set to twice the single-sided message bandwidth, $B_L = 2W/(2\pi)$, by choosing a rectangular post-loop filter with a bandwidth equal to the baseband. This yields the smallest possible output noise level without attenuation of the signal by the filter.

The corresponding expression for the limiter-discriminator is given by

$$\mathrm{SNR_{LD}} = \frac{3p\left(\frac{W_n}{W}\right)\left(\frac{\Delta\omega}{W}\right)^2}{1+\sqrt{3}p\left(\frac{W_n}{W}\right)^2\left[1-\mathrm{erf}\left(\sqrt{p}\right)\right]}. \tag{7.47}$$

The most apparent difference between both expressions is that the Bessel functions in (7.46), which behave roughly exponentially, describe the threshold behavior of the phase feedback demodulator in terms of the SNR in the loop bandwidth (i.e. the baseband), while the error function in (7.47) describes the threshold in terms of the input CNR in the FM bandwidth. This important difference demonstrates the threshold extension capabilities of phase feedback systems in the reception of wideband FM waves.

Figure 7.13 depicts both expressions for a system with $\Delta\omega/W = 1$ and $W_n/W = 4$. The curves of the phase feedback demodulator have been plotted for several values of $\varphi_{e,\mathrm{ss}}$. For a negligible SSPE, the threshold is extended by almost 10 dB. However, it should be remarked that this is a somewhat too optimistic estimate. In practice, the modulation introduces a nonzero SSPE

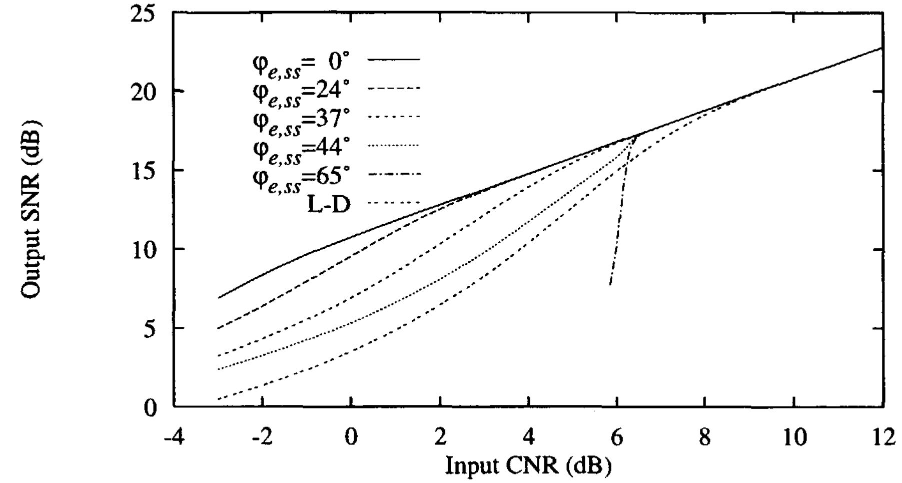


Figure 7.13: Threshold curves of a first-order phase feedback demodulator and a limiter-discriminator (L-D), for several values of $\varphi_{e,ss}$.

and, even worse, a nonzero phase error rate (frequency error). As observed from the figure, a nonzero SSPE dramatically reduces the threshold extension achieved. Modulation may even increase the threshold to a level above the limiter-discriminator threshold. When the error approaches the steady-state lock-limit at $\gamma = 1$, see equation (7.39), the threshold behavior becomes extremely "aggressive", as clearly demonstrated by the steep decay at the threshold of the curve for $\gamma = 0.9$; here, a slight decrease of the input CNR results in a dramatically reduced output SNR. In the latter case, the barriers in the corresponding potential function $U_0(\varphi_e)$ have become only slightly larger than the bottom level of the potential wells. A small amount of noise is therefore able to introduce cycle-slips and eventually to unlock the loop.

Thus, from these numerical results we conclude that in order to maintain a low cycle-slip rate, the SSPE, and probably also the dynamic phase error, should be as low as possible, as should already be expected from intuitive reasoning. This is the reason why a second-order loop appears to be a favorable choice in practice; in such loops, the SSPE can generally be reduced to a very small value by proper design of the loop filter.

7.5 Loop Design

The previous section showed that the phase detector nonlinearity and the structure of the loop filter have a profound influence on the nonlinear behavior of phase feedback demodulators in the threshold region. The implications of this analysis on phase feedback demodulator design, supplied with some additional results, are the subject of this section.

7.5.1 Phase Detector Design

The phase detector nonlinearity determines the shape of the wells and barriers in the potential function $U_0(\varphi_e)$, and thereby to a large extent the cycle-slip rate. In this section, we discuss the implications on the design of the phase detector that follow from this potential function, and consider the behavior of the detector nonlinearity $g(\varphi_e)$ in the presence of noise.

Design Implications from the Potential Function

For a minimal cycle-slip rate, the wells in the potential function should be as deep as possible, while the barriers should be as high as possible. Further, asymmetry in the barriers, as introduced by steady-state phase errors, should be prevented.

A maximum depth and height of the wells and barriers respectively is attained when the restoring force $h_o(\varphi_e)$ is as large as possible for $\varphi_e \neq 0$. As observed from (7.23), this implies that $g(\varphi_e)$ should become as large as possible for $\varphi_e \neq 0$. Further, in order to establish symmetric potential wells, $g(\varphi_e)$ should possess odd-symmetry, centered around $g(0) = 0$.

In electronic systems, supply currents and voltages define an upper bound on the magnitude of the PD output signal. With this in mind, one is lead to the conclusion that, for a given PD constant K_d, the 'optimum' phase detector nonlinearity, as far as cycle-slipping is concerned, equals [19]

$$g(\varphi_e) = \operatorname{sgn}[\sin \varphi_e], \tag{7.48}$$

realized for instance by a multiplier PD, followed by a hard-limiter, as depicted in figure 7.14. Although such a phase detector is permissible in synchronization/tracking loops of any order, including first-order, it cannot be applied in first-order phase feedback demodulator loops. In a tracking loop, the square

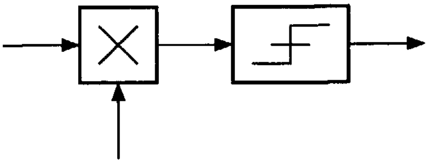


Figure 7.14: Theoretical optimal phase detector nonlinearity for minimization of the cycle-slip rate.

wave output of the phase detector is filtered by the oscillator before it becomes available at the oscillator/loop output. In demodulator loops, however, the output signal of the loop equals the oscillator input signal, and is therefore not subjected to integration (low-pass filtering) by the oscillator. In a first-order demodulator loop, the demodulator output would equal the 'jittering' squarewave at the detector output, which is a heavily distorted copy of the message signal. A post-loop filter cannot cannot eliminate all the undesired square wave components, i.e. the 'harmonics', since some of them are located inside the message bandwidth. This can only be achieved by a loop filter, present in second and higher-order loops.

Transfer Degradation in Hard-Limiter Phase Detectors

It is well known that phase detectors constructed from hard-limiters suffer from degradation in the presence of noise [7, 23, 24]. A brief explanation for the origin of this effect and its implications on phase feedback demodulator design is given below.

As discussed in Section 6.1.2, the presence of noise in the input signal of a hard-limiter results in suppression of the signal component at its output; the limiter occasionally adopts the 'wrong' output value, the one that does not correspond to the input signal, but to the noise. As shown in that section, this effect linearizes the limiter transfer from input signal component to output signal component; the higher harmonics are subject to a larger decrease than the fundamental.

In PDs constructed with hard-limiters, such as those with a triangular characteristic, the linearization of the limiter transfer causes degeneration of the PD nonlinearity, which becomes dependent on thè input CNR and approaches a sinusoid at low CNRs. This effect is illustrated by figure 7.15. At high CNRs, the harmonics of the limiter output contribute significantly to the PD transfer. However, as the the CNR decreases, the harmonics vanish, leaving only the contribution of the fundamental, similar to a multiplier PD with sinusoidal nonlinearity.

The most important consequences of this PD degradation for demodulator design are as follows.

Detector Gain As depicted in figure 7.15, not only the shape of the detector nonlinearity changes, but its gain, important for the linearized model, is also reduced. Consequently, the DC loop gain is reduced at low CNRs, which generally results in a decrease of the closed loop (noise) bandwidth. The loop gain is the central parameter of any feedback system and that constitutes the system operation. Proper operation of the system requires a sufficiently large loop gain. It should therefore be expected that reduction the loop gain in phase

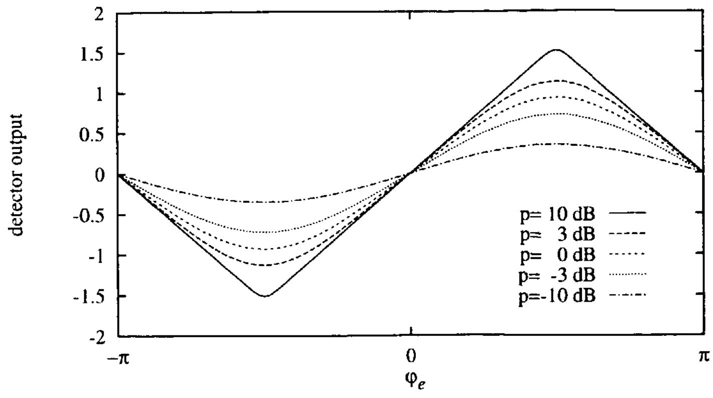


Figure 7.15: Degradation of a triangular PD nonlinearity to a sinusoid for low CNRs (calculated according to [23]).

feedback demodulators at low CNRs, due to degradation of the PD transfer, generally results in deterioration of the performance, despite a reduction of the closed-loop bandwidth. This reduction may result in the modulation limit [46] being exceeded, resulting in loss of lock.

Input Noise $n'(t)$ For high input CNRs, the expression for the noise source $n'(t)$ is the same for all types of PD nonlinearities. For low CNRs however, the expression becomes dependent on the shape of the nonlinearity. In [23] expressions for the power contained in $n'(t)$ as a function of the input CNR are calculated. Whereas the variance of $n'(t)$ at the output of a multiplier detector approaches infinity for decreasing CNRs, it saturates to the total (constant) detector output power level in detectors with hard-limiters. Although it may seem that this observation leads to the conclusion that limiter-phase detectors yield a better performance at low CNRs, it should be strongly emphasized that at the same time the DC loop gain and closed-loop bandwidth drop to zero.

Thus, in conclusion, it seems unfavorable to use hard-limiter phase detectors in phase feedback demodulators intended for low CNRs, due to the significant decrease of the DC loop gain, which will eventually cause the loop to lose lock. A multiplier phase detector, with a sinusoidal nonlinearity, perhaps *followed* by a hard-limiter, seems to be a good alternative.

7.5.2 Loop Filter Design

The design of the loop-filter in the demodulator has been one of the major subjects in phase feedback demodulator research for decades. To a large extent, this filter determines the steady-state phase error inside the loop, that may significantly increase the cycle-slip rate. The two main approaches encountered in the literature are discussed separately below.

Information Theoretical Approach

From an information theoretical point of view, the loop filter $H_{\mathrm{lf}}(s)$ is a prediction filter that supplies the controlled oscillator with information concerning the input wave's phase-value that has to be expected in the near future. In this interpretation, a phase feedback demodulator is considered as an approximate implementation of a theoretical optimum 'backward' or 'recursive' phase estimator.

Starting from this approach, numerous authors have investigated the possibilities of deriving the expressions for the optimum loop filter for various conditions, such as noisy channels and Rayleigh fading channels [47–49]. Extensive use is made of Wiener and Kalman-Bucy estimators to minimize some criterion function, usually the Minimum Mean Square Estimate (MMSE) or Maximum A Posteriori (MAP) estimate of the message. The result is often a rather complex structure consisting of several nested feedback loops and feed-forward paths. An example of such a system was discussed in Section 5.3, in conjunction with co-channel interference suppression.

A shortcoming of this approach is that it starts from the linearized demodulator model in order to keep the analysis tractable. For this reason, cycle-slips cannot be taken into account, resulting in structures that operate far from optimally in the threshold region. For instance, it is concluded that the optimal configuration for a second-order loop to minimize the MMSE requires a damping factor of $\zeta = 1/\sqrt{2}$ [48, 50]. As discussed previously, a cycle-slip analysis of such loops concludes that the damping should be $\zeta > 0.9$ in order to avoid cycle-slip bursts.

This shows that a non-adaptive configuration that behaves optimally for all CNRs does not exist; an adaptive configuration might be able to attain the optimal performance.

Fokker-Plack Approach

In the F-P approach, the information for the design of the loop filter is completely contained in the parameters α and β, for which an approximate expression is given by (7.35) and (7.38) respectively. Below, we summarize the conclusions that can be drawn from these expressions.

Noise Bandwidth It was observed in the previous section that the threshold behavior is described in terms of the input CNR that is observed within the closed loop bandwidth. For high values of this CNR, α, and β both attain large values, resulting in a narrow phase error PDF and consequently also a low cycle-slip rate. The closed-loop bandwidth should therefore not be wider than strictly necessary to track the signal properly. Requirements on the bandwidth set forward on the modulation are considered in [46, 51].

Steady-State Phase Error The SSPE is the cause of asymmetry in the phase error PDF, resulting in an increased cycle-slip rate. This error should therefore be made as small as possible, which for a given input signal (an offset frequency Ω_0) can be achieved only by maximization of the DC loop gain. As is well known, a high loop gain in a first-order loop automatically means a large noise bandwidth, which is definitely undesirable.

Order of the Loop Filter As observed from expression (7.35), the poles of the loop filter generally decrease the value of the parameter α, which for higher order loops may be viewed as the 'effective' SNR in the closed loop bandwidth. This observation is in accordance with the literature, where it is noted that the first-order loop (with zero SSPE) achieves the lowest possible cycle-slip rate [7, 16–20]. This conclusion follows from the observation that the sum of the quotients H_k/τ_k in (7.35) is always positive and therefore reduces the value of α. In fact, (7.35) states that the order of the loop filter should not be made larger than strictly necessary. Thus, higher order loops should only be used when there is a good reason to do so, e.g. in order to minimize the SSPE.

Position of the Open Loop Poles As implied by (7.35), the influence of the loop filter poles on α is small as long as

$$\frac{H_k}{2\tau_k} \ll B_L H_0, \tag{7.49}$$

where H_0 denotes the frequency-independent component (direct feed-through) of the loop filter, and B_L the double-sided noise bandwidth. This indicates that when a high loop gain is required, corresponding to a large value of H_k, the time constant τ_k should also possess a large value. The resulting loop filter therefore tends towards an 'ideal' integrator, of which the integration time constant is considerably larger than $1/(B_L H_0)$.

Direct Feed-through The direct feed-through H_0 is a means to introduce zeros into the closed loop transfer, often required to stabilize the loop. As far as the cycle-slip rates are concerned, H_0 is subject to conflicting requirements. On the one hand, it is desirable to maximize the parameter α_0, which, as opposed to α, denotes the SNR in the *degenerated*, first-order loop, since both α and β are proportional to it. To achieve this, H_0 should be as small as possible. On the other hand, a small value of H_0 means a large influence of higher order poles on α. According to (7.35), α can be maximized to obtain the optimum value of

H_0, where the dependence of α_0 and the noise bandwidth B_L should be taken into account. Lindsey's analysis [19] for a second-order loop with $H_0 = 0$, i.e. a "modified first-order loop" [7], and a zero SSPE implies that such systems are rather unfavorable. It is found that α equals twice the CNR in the noise bandwidth, just as in a first-order loop. Moreover, since the noise bandwidth and SSPE are coupled in the same (highly undesirable) way as in the first-order loop, these systems do not seem to have any advantages compared with a first-order loop.

Position of the Closed Loop Poles The pole positions of the closed loop determine whether cycle-slip bursts occur or not. According to [26], a closed-loop transfer with only real poles seems to be suitable, at least in second order systems.

Based on the foregoing discussion, a second-order loop with an ideal integrator parallel to a direct feed-through with a closed-loop damping of about one seems to be close to optimum. However, it should be noted that this is entirely based on a static analysis of the system. Insight into the behavior in the presence of modulation should be attained by a more advanced (F-PE) approach.

7.6 Conclusions

This chapter investigated the behavior of phase feedback demodulators in the presence of noise. A comparison with non-feedback demodulators was made.

It was shown that, similar to non-feedback, "conventional", demodulators the output noise of phase feedback demodulators consists of a continuous noise component and an impulsive component. The spectrum of the continuous noise component, which determines the output SNR above the threshold, is parabolically shaped by the demodulator. It was shown that this output SNR equals the maximum possible SNR, obtained in conventional demodulators by application of infinite amplitude compression.

The threshold behavior is determined by cycle-slip noise, which has many similarities with click noise in conventional receivers. Similar to click noise, cycle-slip noise consists of impulses with area 2π, that are generated whenever the local oscillator slips/skips a cycle with respect to the input wave. However, as opposed to clicks, cycle-slips may become dependent on each other, resulting in very unpleasant cycle-slip bursts. These bursts generally occur when the closed-loop transfer contains (undamped) complex poles.

The cycle-slip rate is usually significantly lower than the click rate in conventional receivers. Consequently, the threshold of phase feedback demodulators occurs at a lower CNR than the threshold of conventional demodulators. This is due to the observation that the cycle-slip rate is determined by the input CNR inside the loop bandwidth, which approximately equals the message bandwidth, while the click rate is determined by the FM transmission bandwidth.

A nonlinear analysis by means of Fokker-Planck techniques showed that the cycle-slip rate is strongly dependent on the shape of the nonlinear phase-detector transfer and the transfer of the loop filter.

The transfer of the phase detector should be anti-symmetric with respect to a zero-valued phase error. Its output signal for nonzero valued phase errors should be as large as possible in order to drive the error back to zero. At low CNRs, phase detectors containing hard-limiters at their input are generally unfavorable to multiplier phase detectors with a sinusoidal nonlinearity due to degeneration of their transfer. This degeneration reduces the detector gain, and thereby the loop bandwidth, and eventually causes loss of lock.

The loop-filter should be designed to minimize the steady-state phase error, which implies the use of an ideal integrator. This steady-state phase error significantly increases the cycle-slip rate, and may result in an extremely 'aggressive' threshold behavior, corresponding to a steep decay of the output SNR below the threshold. The message modulation, which was not included in the analysis, is believed to cause similar deterioration of the threshold. Further, a direct feed-through that introduces a zero into the loop filter transfer is advantageous for the cycle-slip rate. The order of the loop filter should not be larger than strictly necessary, while the closed loop bandwidth should not be larger than is required to accommodate the modulation.

References

[1] J.J. Spilker, "Delay-lock tracking of binary signals", *IEEE Transactions on Space Electronics and Telemetry*, vol. 9, no. 1, pp. 1–8, Mar. 1963.

[2] S.C. Gupta, "Transient analysis of a phase-locked loop optimized for a frequency ramp input", *IEEE Transactions on Space Electronics and Telemetry*, vol. 10, no. 2, pp. 79–84, June 1964.

[3] Umberto Mengali, "Acquisition behaviour of generalized tracking systems in the absence of noise", *IEEE Transactions on Communications*, vol. 21, no. 7, pp. 820–826, July 1973.

[4] Larry J. Greenstein, "Phase-locked loop pull-in frequency", *IEEE Transactions on Communications*, vol. 22, no. 8, pp. 1005–1013, Aug. 1974.

[5] B.N. Biswas, P. Banerjee, and A.K. Bhattacharya, "Phase locked loop acquisition for a noisy fm signal", *IEEE Transactions on Aerospace and Electronic Systems*, vol. 12, no. 5, pp. 545–550, Sept. 1976.

[6] M.S. Govindarajan and B.V. Rao, "On a closed form expression for phase-locked loop pull-in range", *IEEE Transactions on Communications*, vol. 24, no. 8, pp. 910–913, Aug. 1976.

[7] Floyd M. Gardner, *Phaselock Techniques*, John Wiley and Sons, New York, 2nd edition, 1979.

[8] David G. Messerschmitt, "Frequency detectors for PLL acquisition in timing and carrier recovery", *IEEE Transactions on Communications*, vol. 27, no. 9, pp. 1288–1295, Sept. 1979.

[9] Heinrich Meyr and Luitjens Popken, "Phase acquisition statistics for phase-locked loops", *IEEE Transactions on Communications*, vol. 28, no. 8, pp. 1365–1372, Aug. 1980.

[10] Ross C. Halgren, James T. Harvey, and Ian R. Peterson, "Improved acquisition in phase-locked loops with sawtooth phase detectors", *IEEE Transactions on Communications*, vol. 30, no. 10, pp. 2364–2375, Oct. 1982.

[11] Bernard S. Glance, "New phase-lock loop circuit providing very fast acquisition time", *IEEE Transactions on Microwave Theory and Techniques*, vol. 33, no. 9, pp. 747–754, Sept. 1985.

[12] Douglas N. Green, "Lock-in, tracking, and acquisition of AGC-aided phase-locked loops", *IEEE Transactions on Circuits and Systems*, vol. 32, no. 6, pp. 559–568, June 1985.

[13] Heinrich Meyr and Gerd Ascheid, *Synchronization in Digital Communications, volume 1, Phase-Frequency-Locked Loops, and Amplitude Control*, John Wiley and Sons, New York, 1990.

[14] Jean A. Develet Jr., "A threshold criterion for phase-lock demodulation", *Proceedings of the IEEE*, vol. 51, no. 2, pp. 349–356, Feb. 1963.

[15] Jean A. Develet Jr., "An analytic approximation of phase-lock receiver threshold", *IEEE Transactions on Space Electronics and Telemetry*, vol. 9, no. 1, pp. 9–12, Mar. 1963.

[16] Andrew J. Viterbi, "Phase-locked loop dynamics in the presence of noise by Fokker-Planck techniques", *Proceedings of the IEEE*, vol. 51, no. 12, pp. 1737–1753, Dec. 1963.

[17] Andrew J. Viterbi, *Principles of Coherent Communication*, McGraw-Hill, New York, 1966.

[18] F.J. Charles and W.C. Lindsey, "Some analytical and experimental phase-locked loop results for low signal-to-noise ratios", *Proceedings of the IEEE*, vol. 54, no. 9, pp. 1152–1166, Sept. 1966.

[19] W. C. Lindsey, "Nonlinear analysis of generalized tracking systems", *Proceedings of the IEEE*, vol. 57, pp. 1705–1722, Oct. 1969.

[20] William C. Lindsey, *Synchronization Systems in Communication and Control*, Prentice-Hall, Englewood Cliffs, New Jersey, 1972.

[21] V.I. Tikhonov, "The effects of noise on phase-lock oscillation operation", *Automatika i Telemakhanika*, vol. 22, no. 9, 1959.

[22] V.I. Tikhonov, "Phase-lock automatic frequency control application in the presence of noise", *Automatika i Telemakhanika*, vol. 23, no. 3, 1960.

[23] Werner Rosenkranz, "Phase-locked loops with limiter phase detectors in the presence of noise", *IEEE Transactions on Communications*, vol. 30, no. 10, pp. 2297–2304, Oct. 1982.

[24] B.N. Biswas, S.K. Ray, A.K. Bhattacharya, B.C. Sarkar, and P. Banerjee, "Phase detector response to noisy and noisy fading channels", *IEEE Transactions on Aerospace and Electronic Systems*, vol. 16, no. 2, pp. 150–158, Mar. 1980.

[25] A.H. Pouzet, "Characteristics of phase detectors in the presence of noise", in *Proceedings of the 8th Telemetering Conference*, Los Angeles, 1972, pp. 818–828.

[26] Gerd Ascheid and Heinrich Meyr, "Cycle slips in phase-locked loops: a tutorial survey", *IEEE Transactions on Communications*, vol. 30, no. 10, pp. 2228–2241, Oct. 1982.

[27] Heinrich Meyr, "Nonlinear analysis of correlative tracking systems using renewal process theory", *IEEE Transactions on Communications*, vol. 23, no. 2, pp. 192–203, Feb. 1975.

[28] William C. Lindsey and Heinrich Meyr, "Complete statistical description of the phase-error process generated by correlative tracking systems", *IEEE Transactions on Information Theory*, vol. 23, no. 2, pp. 194–202, Mar. 1977.

[29] Dietrich Ryter and Heinrich Meyr, "Theory of phase tracking systems of arbitrary order: Statistics of cycle slips and probability distribution of the state vector", *IEEE Transactions on Information Theory*, vol. 24, no. 1, pp. 1–7, Jan. 1978.

[30] R. C. Booton Jr., "The analysis of nonlinear control systems with random inputs", in *Proceedings of the Symposium on Nonlinear Circuit Analysis*, Polytechnic Institute of Brooklyn, New York, 1953, pp. 369–391.

[31] Harry L. van Trees, "Functional techniques for the analysis of the nonlinear behaviour of phase-locked loops", *Proceedings of the IEEE*, vol. 52, no. 8, pp. 894–911, Aug. 1964.

[32] Robert C. Tausworthe, "Cycle slipping in phase-locked loops", *IEEE Transactions on Communication Technology*, vol. 15, no. 3, pp. 417–421, June 1967.

[33] Robert C. Tausworthe, "Simplified formula for mean cycle-slip time of phase-locked loops with steady-state phase error", *IEEE Transactions on Communication Technology*, vol. 20, no. 3, pp. 331–337, June 1972.

[34] Donald T. Hess, "Cycle slipping in a first-order phase-locked loop", *IEEE Transactions on Communication Technology*, vol. 16, no. 2, pp. 255–260, Apr. 1968.

[35] Enrico A. Bozzoni, Giovanni Marchetti, Umberto Mengali, and Franco Russo, "An extension of Viterbi's analysis of cycle slipping in a first-order phase-locked loop", *IEEE Transactions on Aerospace and Electronic Systems*, vol. 6, no. 4, pp. 484–490, July 1970.

[36] Jack K. Holmes, "First slip times versus static phase error offset for the first- and passive second-order phase-locked loop", *IEEE Transactions on Communication Technology*, vol. 19, no. 2, pp. 234–235, Apr. 1971.

[37] James R. La Frieda and William C. Lindsey, "Transient analysis of phase-locked tracking systems in the presence of noise", *IEEE Transactions on Information Theory*, vol. 19, no. 2, pp. 155–165, Mar. 1973.

[38] William C. Lindsey and Marvin K. Simon, "Detection of digital FSK and PSK using a first-order phase-locked loop", *IEEE Transactions on Communications*, vol. 25, no. 2, pp. 200–214, Feb. 1977.

[39] Kenichi Nishiguchi and Yoshinori Uchida, "Transient analysis of the second-order phase-locked loop in the presence of noise", *IEEE Transactions on Information Theory*, vol. 26, no. 4, pp. 482–486, July 1980.

[40] Chak Ming Chie, "New results on mean time-to-first-slip for a first-order loop", *IEEE Transactions on Communications*, vol. 33, no. 9, pp. 897–903, Sept. 1985.

[41] A. L. Welti and B. Z. Bobrovsky, "Mean time to lose lock for the "Langevin"-type delay-locked loop", *IEEE Transactions on Communications*, vol. 42, no. 8, pp. 2526–2530, Aug. 1994.

[42] Marcelo Alonso and Edward J. Finn, *Fundamental University Physics volume 2, Fields and Waves*, Addison-Wesely, Reading, Massachusetts, 2nd edition, 1983.

[43] A. Bruce Carlson, *Communication Systems, An introduction to Signals and noise in Electrical Communication*, McGraw-Hill Book Company, Singapore, 1986.

[44] Milton Abramowitz and Irene A. Stegun, Eds., *Handbook of Mathematical Functions*, Dover Publications, New York, fifth edition, 1968.

[45] Wilbur B. Davenport and William L. Root, *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill Book Company, New York, 1958.

[46] Floyd M. Gardner and John F. Heck, "Angle modulation limits of a noise-free phase lock loop", *IEEE Transactions on Communications*, vol. 26, no. 8, pp. 1129–1136, Aug. 1978.

[47] Harry L. van Trees, "Analog communication over randomly-time-varying channels", *IEEE Transactions on Information Theory*, vol. 12, no. 1, pp. 51–63, Jan. 1966.

[48] Harry L. van Trees, *Detection, Estimation, and Modulation Theory-Part II, Nonlinear Modulation Theory*, John Wiley and Sons, New York, 1971.

[49] Donald L. Snyder, *The State-Variable Approach to Continuous Estimation, with applications to analog communication*, M.I.T. Press, Cambridge, 1969.

[50] R. Jaffe and E. Rechtin, "Design and performance of phase-lock circuits capable of near-optimum performance over a wide range of input signal and noise levels", *IRE Transactions on Information Theory*, vol. 1, pp. 66–76, Mar. 1955.

[51] Floyd M. Gardner and John F. Heck, "Phaselock loop cycle slipping caused by excessive angle modulation", *IEEE Transactions on Communications*, vol. 26, no. 8, pp. 1307–1309, Aug. 1978.

Chapter 8

Frequency Feedback

As opposed to phase feedback, discussed in Chapter 7, frequency feedback is not associated with a particular class of FM demodulators, but applies to any FM demodulator type. Therefore, it should be considered as an improvement technique, like amplitude compression, instead of the demodulation principle of a particular FM demodulator class. It is for this reason that frequency feedback (FFB) demodulators are not included in the classification of Chapter 3.

Originally, Chaffee [1, 2] introduced the application of negative frequency feedback to FM discriminators (without limiters) to achieve distortion reduction. The threshold extending capabilities of frequency feedback applied to FM limiter-discriminators, as qualitatively discussed in Section 5.7.1, were not recognized until the 1960's [3–5]. This discovery initiated their widespread application to such areas as reception of signals from spacecraft, satellites [6, 7], and FM broadcast reception in cars [8].

However, despite their widespread use, and the recognition of their capabilities to reduce distortion/extend the threshold, the threshold mechanism itself is still not very well understood. As opposed to 'conventional' limiter-discriminators and, to some extent, phase feedback demodulators, the modeling of this threshold mechanism is based on linear approximations that hold, at most, only down to a few dB below the threshold. This may probably be attributed to the complicated structure of such systems, containing an IF filter and an FM limiter-discriminator inside the feedback loop.

This chapter compares the threshold behavior of FFB demodulators to the response of conventional and phase feedback demodulators, by the combination and application of slight modifications of theories known from the literature.

Section 8.1 discusses a model for these demodulators that is subsequently used in the description of the demodulator response above and around the threshold, as discussed in sections 8.2 and 8.3 respectively. An example calculation of the FMFB threshold curve is discussed in Section 8.4, and compared to the threshold behavior of conventional demodulators and phase feedback demodulators in Section 8.5. Section 8.6 considers the design of FMFB systems. The conclusions are given in Section 8.7.

8.1 Frequency Feedback Demodulator Model

The various methods known in the literature for the analysis of the FFB demodulator output noise above and around the threshold all use the same low-pass equivalent demodulator model, similar to the model used to describe phase feedback demodulators (see Chapter 7).

A brief discussion of the two types of FFB demodulators that can be distinguished, Frequency Modulation Feedback Receivers (FMFB) and Dynamic Tracking Filters (DTF), is given in Section 8.1.1. Subsequently, Section 8.1.2 describes the structure of their low-pass equivalent model and its relation to the original systems.

8.1.1 Types of Frequency Feedback Demodulators

The two different types of FFB demodulators that can be distinguished are schematically depicted in figure 8.1 and 8.2. Both systems consist of an IF filter, a limiter and an FM discriminator, together enclosed by a feedback loop.

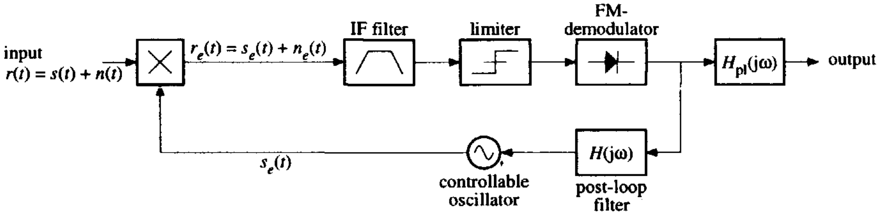


Figure 8.1: Block diagram of a Frequency Feedback Receiver (FMFB).

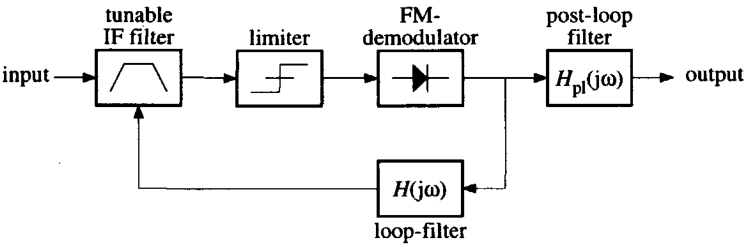


Figure 8.2: Block diagram of a Dynamic Tracking Filter (DTF).

The construction of this feedback loop however is different.

In FMFBs, sometimes called "Frequency Locked Loops" (FLL), the feedback is established by means of a controlled oscillator (FM modulator) in the feedback path and a frequency subtracter, commonly implemented by a down-conversion mixer. The instantaneous frequency of the discriminator input wave in this system equals the difference between the instantaneous frequency of the input FM wave, and the oscillator output wave, as discussed in Section 5.7.1. The filter $H(\mathrm{j}\omega)$ is a low-pass filter that eliminates high frequency (noise) components from the discriminator output signal.

In DTFs, the feedback loop is established by adaptive control of the IF filter sections with the aid of the FM discriminator output signal. This control loop is arranged such that the center-frequency of the IF filter (partially) tracks the center-frequency of the input FM wave. As discussed in Section 5.7.2, the behavior of these systems is similar to FMFBs.

The main difference between FMFBs and DTFs, as far as demodulator design is concerned, is that the DTF structure leaves more degrees of freedom to optimize its performance than an FMFB [8], and does not require a controlled oscillator. In a DTF system, the closed loop transfer can be optimized by the application of several feedback loops, whereas FMFBs cannot be optimized in this way. An exception to this made is possible by the application of multiple controlled oscillators and frequency subtracters in order to realize several feedback loops, but this solution is considered to be rather impractical in comparison with a multi-loop DTF.

8.1.2 Low-pass Equivalent Model

The FMFB and DTF can both be described by the same low-pass equivalent model, depicted in figure 8.3 [3, 7–11]. According to [7, 8, 10], identical models

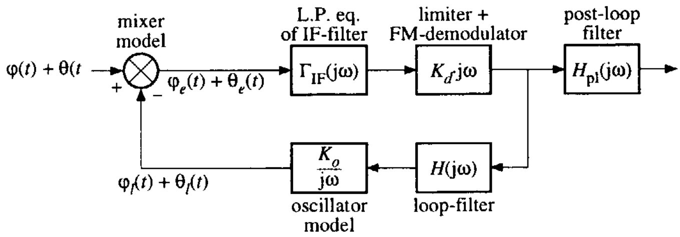


Figure 8.3: Low-pass equivalent model of FMFB and DTF systems.

for an FMFB and DTF with a single feedback loop are obtained when the FMFB IF filter, denoted by $H_{\mathrm{IF,FMFB}}(\mathrm{j}\omega)$, and the DTF IF filter, denoted by

$H_{\mathrm{IF,DTF}}(\mathrm{j}\omega)$, obey the relation

$$H_{\mathrm{IF,DTF}}(\mathrm{j}\omega) = \frac{H_{\mathrm{IF,FMFB}}(\mathrm{j}\omega)}{1 + H_{\mathrm{IF,FMFB}}(\mathrm{j}\omega)}. \tag{8.1}$$

Since most system models described in the literature are concerned with the FMFB system, the remainder of this chapter assumes that we are also dealing with an FMFB systems, although there are no fundamental differences with a single loop DTF. Multi-loop DTFs are not considered in this thesis.

Relation with the Original System

Figure 8.3 depicts the low-pass equivalent model of the FMFB system depicted in figure 8.1. The relation between the various sub-systems in both schematics is as follows. Let the center-frequency of the FM input wave $s(t)$ be represented by ω_o and the message part of its instantaneous phase by $\varphi(t)$. Further, let the center-frequency of the oscillatoe output wave $s_l(t)$ be represented by $\omega_l$and the message part of its phase by $\varphi_l(t)$.

The model implicitly assumes that only spectral components centered around the difference frequency $\omega_e = \omega_o - \omega_l$ are passed by the IF filter in the loop. All other components, corresponding to other linear combinations of ω_o and ω_l, are assumed to be suppressed. Therefore the center-frequency and message phase of the wave $s_e(t)$ at the IF filter input inside the loop may be represented by $\omega_e = \omega_o - \omega_l$ and $\varphi_e(t) = \varphi_o(t) - \varphi_l(t)$ respectively, omitting any other components.

From the discussion in Section 5.2 it follows that as long as negligible distortion is introduced, the IF filter linearly filters the message phase $\varphi_e(t)$ with its low-pass equivalent transfer $\Gamma_{\mathrm{IF}}(\mathrm{j}\omega)$, as depicted in figure 8.3. When distortion is introduced, $\Gamma_{\mathrm{IF}}(\mathrm{j}\omega)$ is supplied with additional terms that are a function of the spectrum/frequency deviation of the phase/frequency modulation $\varphi_e(t)$. The the case where $\varphi_e(t)$ is represented by Gaussian noise is considered in [12] (see also Section 5.2), while the distortion in case of single tone modulation in considered in [11].

Further, the operation of the ideal assumed FM limiter-discriminator is represented by a differentiator with gain K_d and the controlled oscillator/FM modulator by an integrator with gain K_o. Only discriminators preceded by a hard-limiter (infinite compression) are considered in this chapter. The down conversion mixer is represented by a subtracter. Obviously, the filters $H(\mathrm{j}\omega)$ and $H_{\mathrm{pl}}(\mathrm{j}\omega)$ appear unchanged in the low-pass equivalent model.

Loop Transfers

Three transfers inside this loop are of interest in subsequent sections of this chapter. These are the closed loop transfer from input phase $\Phi(\omega)$ to the oscillator output phase $\Phi_l(\omega)$ given by

$$\frac{\Phi_l(\omega)}{\Phi(\omega)} \stackrel{\text{def}}{=} H_{\text{cl}}(\text{j}\omega) = \frac{K_o K_d \Gamma_{\text{IF}}(\text{j}\omega) H(\text{j}\omega)}{1 + K_o K_d \Gamma_{\text{IF}}(\text{j}\omega) H(\text{j}\omega)}, \tag{8.2}$$

the transfer from $\Phi(\omega)$ to the phase of the wave at the mixer output, $\Phi_e(\omega)$, given by

$$\frac{\Phi_e(\omega)}{\Phi(\omega)} \stackrel{\text{def}}{=} H_e(\text{j}\omega) = \frac{1}{1 + K_o K_d \Gamma_{\text{IF}}(\text{j}\omega) H(\text{j}\omega)}, \tag{8.3}$$

and the demodulator transfer from input frequency $\text{j}\omega\Phi(\omega)$ to the demodulator output $Y_{\text{dem}}(\omega)$, given by

$$\frac{Y_{\text{dem}}(\omega)}{\text{j}\omega\Phi(\omega)} \stackrel{\text{def}}{=} H_{\text{dem}}(\text{j}\omega) = \frac{K_d \Gamma_{\text{IF}}(\text{j}\omega) H_{\text{pl}}(\text{j}\omega)}{1 + K_o K_d \Gamma_{\text{IF}}(\text{j}\omega) H(\text{j}\omega)}. \tag{8.4}$$

This model, with the optional inclusion of the distortion produced by the IF filter, is used in the determination of the demodulator output noise considered in subsequent sections.

8.2 Response Above Threshold

This section considers the output noise of an FMFB receiver above its threshold. It is shown that although the output SNR above threshold is identical to the SNR obtained with a comparable conventional discriminator, the threshold may be considerably extended. An upper bound on the attainable threshold extension is derived that follows from the above threshold model.

Section 8.2.1 shows that above threshold the same output SNR is obtained as with a limiter-discriminator. Section 8.2.2 determines the upper bound on the threshold extension that follows from the above threshold demodulator model.

8.2.1 Output SNR above Threshold

This section determines the output SNR of FMFB receivers above threshold and shows that frequency feedback does not affect the above-threshold SNR.

The low-pass equivalent model of figure 8.3, required for this investigation, applies some (justifiable) approximations to the mixer output wave $r_e(t) = s_e(t) + n_e(t)$ (see figure 8.1). In order to clarify the discussion, these approximations need to be considered first. Subsequently, the output SNR is considered.

Approximations Applied by the Low-pass Equivalent Model

The approximations applied by the low-pass equivalent model can be explained with the aid of an expression for the mixer output signal $r_e(t)$, obtained from figure 8.1. This expression is obtained as follows.

Without loss of generality, a unity mixer conversion gain may be assumed. This value is established in the models by assigning the oscillator output wave $s_l(t)$ an amplitude equal to $\sqrt{2}$. In case of other values the conversion gain may be included in the IF filter transfer or the demodulator constant K_d.

It is quite obvious that a suitable expression for $r_e(t)$ is most easily obtained by writing the noisy FM input wave $s(t) + n(t)$ in polar coordinates, as the amplitude and phase modulated wave $r(t)$ from (2.15). The mixer in figure 8.1 multiplies this wave with the oscillator output $s_l(t)$. It should be noticed that since $s_l(t)$ is the FM wave that corresponds to the (noisy) baseband FM discriminator output signal, its instantaneous frequency generally contains a message component $\varphi_l(t)$ and a phase noise component $\theta_l(t)$. With this in mind, the wave $r_e(t)$, the addition of the noise free FM wave $s_e(t)$ and the noise $n_e(t)$, may be expressed by omission of all components that are not centered around $\omega_e = \omega_o - \omega_l$, as

$$\begin{aligned} r_e(t) &= r(t)s_l(t) \\ &= \sqrt{2}R(t)\cos\left[\omega_o t + \varphi(t) + \theta(t)\right]\cos\left[\omega_l t + \varphi_l(t) + \theta_l(t)\right] \qquad (8.5) \\ &= R(t)\cos\left[\omega_e t + \varphi_e(t) + \theta_e(t)\right]. \end{aligned}$$

In this expression, the signal component of the phase equals $\varphi_e(t) = \varphi(t) - \varphi_l(t)$ and the phase noise component equals $\theta_e(t) = \theta(t) - \theta_l(t)$. Above threshold, the low-pass equivalent model applies the following approximations to (8.5).

In the first place, because it is assumed that the FM limiter-discriminator in figure 8.1 operates above its threshold, which is true for high input CNRs, it is assumed that no click noise is generated and all noise processes inside the system behave Gaussian.

Secondly, it is assumed that AM-to-PM conversion of the amplitude noise contained in $R(t)$, introduced by the IF filter, is negligible. At the current level of our considerations, this is usually permissible since such conversions are mostly due to dynamic nonlinearities in the (non-ideal) IF filter. Therefore, the amplitude noise in $R(t)$ can be ignored, since it is eventually, after IF filtering, eliminated by the hard-limiter.

Thirdly, the phase noise $\theta_e(t)$ is assumed to be small compared to the signal phase $\varphi_e(t)$, which is true for high input CNRs. In that case, the bandwidth of $r_e(t)$ is determined by the message modulation $\varphi_e(t)$ only. The IF filter bandwidth is dimensioned such that negligible distortion is introduced into the limiter-discriminator input by narrow-band filtering. At low CNRs, however, such distortion will eventually become apparent due to an increased level of the phase noise $\theta_e(t)$. The consequences of this effect are considered in Section 8.3.

Calculation of the Output SNR

It becomes clear from figure 8.3 that as long as the receiver operates above threshold, the input message phase $\varphi(t)$ and the phase noise $\theta(t)$ are treated in the same way. This already indicates that the same output SNR has to be expected as in the case of a conventional limiter-discriminator.

Calculation of the SNR yields the same conclusion. From figure 8.1 it is observed that when the demodulator transfer $H_{\text{dem}}(\mathrm{j}\omega)$ (8.4) is flat for baseband frequencies, and the post-loop filter $H_{\text{pl}}(\mathrm{j}\omega)$ is rectangular with bandwidth W, the output SNR becomes equal to the SNR of the instantaneous frequency $\dot{\varphi}(t) + \dot{\theta}(t) \approx \dot{\varphi}(t) + \dot{n}_{s,q}(t)/A$ in a bandwidth W. This SNR is given by (2.21). Consequently, frequency feedback does not affect the output SNR above the threshold, but only the distortion (dynamic range) and the position of the threshold.

8.2.2 Upper Bound on the Threshold Extension

Whereas the demodulator output noise above the threshold is entirely determined by the quadrature noise component $n_{s,q}(t)$, which is the dominant contribution to the phase noise of the input FM wave, the output noise around and below the threshold is also determined by the in-phase component $n_{s,i}(t)$, which dominates the amplitude noise. The latter component is responsible for the generation of click noise at the output of the limiter-discriminator inside the loop, and therefore initiates the FMFB threshold.

This section determines an upper bound on the threshold extension attainable with frequency feedback that follows from the description of the demodulator above the threshold. We first discuss the "frequency compression" mechanism that constitutes the extension, and subsequently the resulting upper bound. In this respect, it is shown that the FMFB responds differently to the quadrature noise (phase noise) $n_{s,q}(t)$ and the in-phase noise (amplitude noise) $n_{s,i}(t)$: $n_{s,q}(t)$ is suppressed, while $n_{s,i}(t)$ is not compressed. Finally, comparison of the FMFB threshold curve with related limiter-discriminator threshold curves explains the origin of the upper bound on the threshold extension.

Frequency Compression

Above the threshold, the frequency compression mechanism inside the loop reduces the wideband FM input wave $s(t)$ to the narrow-band wave $s_e(t)$, while it leaves the bandwidth of the wideband input noise $n(t)$, which is converted to $n_e(t)$, essentially unchanged. The reason for this behavior is that since at high

CNRs the message phase $\varphi_l(t)$ is much larger than the phase noise $\theta_l(t)$, the FM wave $s(t)$ is strongly correlated with $s_l(t)$, while the noise $n(t)$ is hardly correlated with it.

Therefore, by application of narrow-band filtering to $s_e(t) + n_e(t)$, the CNR at the input of the limiter-discriminator inside the loop can become significantly larger than the receiver input CNR, without introduction of excessive distortion. In such a system, the threshold CNR of the discriminator inside the loop corresponds to a (significantly) smaller *receiver* input CNR, which means that the threshold of the closed-loop system is "extended".

The reduced bandwidth of the FM wave $s_e(t)$ can be explained from the low-pass equivalent demodulator model as follows. The key to the bandwidth compression mechanism is the transfer $H_e(\mathrm{j}\omega)$, given by (8.3). This transfer relates the instantaneous phase of $s(t)$ to the instantaneous phase of $s_e(t)$. If, for simplicity, it is assumed that $H_e(\mathrm{j}\omega)$ is flat over the frequency range of interest, the angle modulation in $s(t)$, i.e. $\varphi(t)$, is compressed as

$$\varphi_e(t) = \frac{\varphi(t)}{1 + K_d K_o \Gamma_{\mathrm{IF}}(0) H(0)} = \frac{\varphi(t)}{1 + F_o}, \tag{8.6}$$

where F_o denotes the DC loop gain. As a result, the RMS frequency deviation of $s_e(t)$ is also reduced by $1 + F_o$. According to Carson's bandwidth formula (2.8), the bandwidth of $s_e(t)$ is also reduced in proportion. The smallest possible bandwidth, attained when the loop gain approaches infinity, equals twice the message bandwidth, i.e. $2W$.

The absence of *bandwidth* compression in $n_e(t)$ can be explained with the aid of expression (8.5) for $r_e(t) = s_e(t) + n_e(t)$. Above the threshold, $\theta(t)$ approximately equals $n_{s,q}(t)/A$, and, due to the frequency compression, $\theta_e(t)$ approximately equals $n_{s,q}(t)/\left[(1 + F_o)A\right]$. Thus, whereas frequency compression reduces the *bandwidth* of the FM wave $s(t)$, it reduces only the *magnitude* of the quadrature noise $n_{s,q}(t)$. A first order approximation of $r_e(t)$, valid at high CNRs, may then be written as

$$\begin{aligned} r_e(t) &= s_e(t) + n_e(t) \\ &\approx A\cos\left[\omega_e t + \varphi_e(t)\right] \\ &+ n_{s,i}(t)\cos\left[\omega_e t + \varphi_e(t)\right] - \frac{n_{s,q}(t)}{1 + F_o}\sin\left[\omega_e t + \varphi_e(t)\right]. \end{aligned} \tag{8.7}$$

Expression (8.7) shows that the loop compresses the quadrature noise, which represents the phase noise above threshold, while the in-phase noise, which represents the amplitude noise in $R(t)$, is not affected. Thus, as stated in the introduction of this section, the FMFB response to the in-phase noise differs considerably from its response to the quadrature noise. The latter effect is due to the absence of an amplitude feedback loop in the system; the oscillator modulates only the frequency of $s_l(t)$, and not the amplitude. Furthermore, the amplitude noise is suppressed by the limiter. The (double-sided) bandwidth of $n_{s,i}(t)$ and $n_{s,q}(t)$ roughly equals W_n, the bandwidth of the input FM wave $s(t)$. Consequently, the bandwidth of $n_e(t)$ will be in the order of W_n too.

Threshold Extension

In order to realize the largest possible threshold extension, the IF filter should be dimensioned such that the the narrow-band FM wave $s_e(t)$ just fits within its pass band, as illustrated by figure 8.4. Thus, by virtue of this filter, the discriminator input noise power is reduced, while the input FM carrier power remains unaffected.

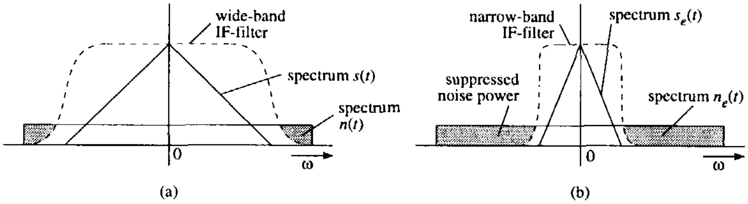


Figure 8.4: Improvement of the limiter-discriminator input CNR by narrow band filtering. (a) spectrum of the receiver input signal $s(t) + n(t)$, (b) spectrum of the compressed IF filter input signal $s_e(t) + n_e(t)$.

When the (double-sided) noise bandwidth of the IF filter inside the loop is denoted by $W_{L,\mathrm{IF}}$ in (rad/s), the FMFB receiver *maximum threshold extension* that can be achieved equals the factor

$$\Gamma_{\mathrm{id}} \overset{\mathrm{def}}{=} \frac{1}{W_{L,\mathrm{IF}} S_n(0)} \int_{-\infty}^{\infty} S_n(\omega) \mathrm{d}\omega, \tag{8.8}$$

equal to the ratio of the noise bandwidth of $n(t)$, and the IF filter noise bandwidth $W_{L,\mathrm{IF}}$. The CNR at the input of the FM discriminator inside the loop is therefore at most Γ_{id} times as large as the receiver input CNR. The threshold of the discriminator, which is determined by the in-phase noise at the IF filter output [11], is therefore reduced at most by the same factor.

Comparison with Limiter-Discriminators

It is illustrative to compare the threshold curve of the FMFB receiver with the corresponding curves of two limiter-discriminators, as depicted in figure 8.5 [11]. Curve (a) corresponds to the wideband limiter-discriminator (LD), required for demodulation of $s(t) + n(t)$ without feedback. Above threshold, its output SNR is identical to the FMFB output SNR, given by curve (d), as shown in

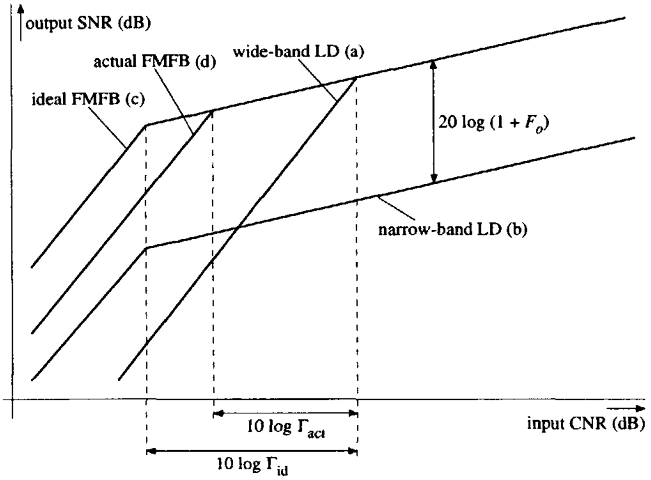


Figure 8.5: Comparison of the FMFB output SNR with two limiter discriminators. (a) wideband limiter-discriminator, (b) narrow-band limiter-discriminator (c) ideal FMFB curve (ignoring feedback noise), (d) actual FMFB curve.

Section 8.2.1. The threshold CNR of the wideband LD, however, is higher than the threshold CNR of the FMFB receiver.

Curve (b) corresponds to the narrow-band limiter-discriminator, required for demodulation of the compressed wave $s_e(t) + n_e(t)$ without feedback. Above threshold, its output SNR is roughly $20 \log [1 + F_o]$ dB smaller than the SNR of the FMFB receiver since the output signal power is proportional to the squared frequency deviation $(\Delta\omega)^2$, while the output noise power in the baseband is independent of $\Delta\omega$. The threshold of this discriminator equals the lower bound on the FMFB receiver threshold, given by curve (c), attained when the feedback noise is negligible. This is due to the fact that the level of the in-phase noise component at the discriminator input that determines the threshold is the same in both systems.

Curve (d) shows the actual behavior of the FMFB receiver. As a result of feedback noise, its threshold is generally located at a larger CNR than the threshold of an 'ideal' FMFB, given by curve (c). The actual threshold gain that is achieved by frequency feedback is represented by the distance Γ_{act} in figure 8.5, the difference between the threshold in (a) and (d). The previously discussed upper bound on the threshold extension Γ_{id}, equals the difference between the thresholds in (a) and (b), or, equivalently, (a) and (c).

8.3 Response in the Threshold Region

The threshold mechanism in FMFB receivers is not yet modeled as elegantly as the threshold mechanism in limiter-discriminators and, to some extent, phase feedback demodulators. Since the early 1960's, a vast amount of literature on the subject, see [3, 7, 9–11, 13–23], has resulted in a variety of sometimes almost fantastic theories. None of them, however, seems to put all the missing parts in place. Most theories suffer from a lack of generality, even the most extended ones. They are somehow based on experimental data, without proof of its validity in practically relevant cases.

By most relevant publications, the threshold effect in FMFB receivers is attributed to one, or a combination of the following three mechanisms that each result in a nonlinear response to noise at low CNRs:

- the threshold of the limiter-discriminator inside the loop;
- feedback of phase noise (feedback noise);
- suppression of the FM wave by the IF filter.

An important difference between the various theories is that some of them consider these mechanisms to be completely independent, while others consider these mechanisms to be mutually coupled. For example, Enloe [3] attributes the FMFB threshold to the limiter-discriminator threshold and the feedback noise, which he considers to be completely independent mechanisms. Bax [11], however, attributes the FMFB threshold to a combination of all three mechanisms and considers them to be mutually coupled. In subsequent sections, Bax' theory is adopted as the basis for the discussion on the FMFB threshold behavior.

This section investigates the threshold mechanisms in FMFB receivers by the combination and modification of several of the most important theories known so far. Subsequently, the FMFB threshold response is compared to the limiter-discriminator threshold response.

Sections 8.3.1 through 8.3.3 investigate the relation between the three previously mentioned mechanisms and the FMFB threshold response, and comment on the validity of the theories that propose that they are the dominant FMFB threshold mechanism. Subsequently, sections 8.3.4 through 8.3.6 combine these mechanism to form an FMFB threshold model.

8.3.1 Limiter-Discriminator Threshold

There is no in among any of the FMFB threshold theories that the limiter-discriminator threshold is the main cause of the FMFB threshold. This threshold mechanism, that is already present in the absence of feedback, is inherent to application of infinite compression to the discriminator input wave.

Frequency feedback cannot eliminate the limiter-discriminator threshold, but at most achieve a shift of the threshold to a lower CNR by compression of the FM bandwidth. Depending on the bandwidth of the IF filter, the threshold can at most be extended by a factor Γ_{id} (curve (c) in figure 8.5). In practice, however, a smaller extension is achieved due to the presence of other nonlinear effects in the loop.

Frutiger [9], attributes the FMFB threshold entirely to the limiter-discriminator threshold. He applies Rice's theory [24] for the limiter-discriminator to distinguish between continuous noise and click noise. The small continuous noise is supposed to be compressed by the feedback in the loop, as described in Section 8.2. Click noise, however, is believed to cause a temporary interruption of the feedback and is consequently not compressed. The FMFB threshold occurs when both noise components are of comparable strength. The threshold level obtained in this way is indeed larger than the ideal FMFB threshold of curve (c) in figure 8.5, and smaller than the wideband LD threshold of curve (a). However, although it is clear that click noise will eventually break the loop, there doesn't seem to be any reason why it would break at every click pulse. Experiments [11] indicate that clicks *are* fed-back by the loop, which contradicts Frutiger's view. Therefore, we will not adhere to his theory.

Roberts [7, 10], also entirely attributes the FMFB threshold to the limiter-discriminator threshold and applies Rice's theory to distinguish between a continuous and a click noise component. In contrast to Frutiger, however, he assumes that click noise is compressed by the feedback in the loop. Besides some (negligible) distortion introduced by the IF filter, he disregards any other nonlinear effects. Consequently, although not explicitly stated that way, his theory arrives at the ideal FMFB threshold of curve (c), which is too optimistic. This optimism is also observed from his experimental results, depicted in figure 10 of [7].

8.3.2 Feedback Noise

It seems probable that the difference between the FMFB threshold (curve (d) in figure 8.5) and the threshold of the "open-loop" discriminator (curve (b) in figure 8.5) finds its origin in the feedback mechanism of the loop. Above threshold, this same feedback mechanism is responsible for the difference in output SNR between the narrow-band LD of curve (b), and the ideal/actual FMFB threshold curves (c) and (d) in figure 8.5.

In an early paper, Enloe [3] adopts the view that besides the limiter-discriminator threshold, feedback noise also determines the FMFB threshold. Both these mechanisms are believed to introduce two separate thresholds; a limiter-discriminator threshold and a "feedback threshold".

The rather fantastic explanation given for the second threshold mechanism is generally considered to be wrong [7, 10, 11, 18]. Enloe states that due to the input mixer second order noise products between the phase noise in the oscillator and the input noise $n(t)$ become dominant at low CNRs and finally break the feedback loop at the "feedback threshold". This threshold would occur when the RMS phase deviation of the oscillator, as a result of the noise, exceeds the "magic number" 1/3.11 (rad). It has been pointed out that his conclusions are based on a series expansion that is definitely invalid at low CNRs [11, 18]: the feedback noise is too weak to introduce the FMFB threshold independently from other mechanisms. In the remainder of this chapter, we will therefore not adhere to this theory.

Bax [11] notices the influence of feedback noise, but he disagrees with Enloe's feedback threshold. In his theory, the interaction between the feedback noise and the narrow-band IF filtering increases the limiter-discriminator threshold, as discussed in the next section. Although slightly modified, Bax' theory is used as the basis for the FMFB threshold model discussed in subsequent sections.

8.3.3 Carrier Suppression by the IF Filter

The influence of nonlinear effects introduced by narrow-band IF filtering are ignored, or considered to be negligible by the vast majority of the FM threshold theories. Usually, the distortion introduced into the message by IF filtering is included in the model. However, since the feedback loop reduces the distortion, its influence on the FMFB threshold is usually negligible.

Besides this distortion, Bax's theory [11] recognizes another effect introduced by the IF filter: suppression of the FM carrier wave. This effect is considered to occur as a result of the phase noise $\theta_l(t)$ present in the oscillator output wave $s_l(t)$. As observed experimentally [5, 11], the level of this noise increases significantly at low CNRs, partially due to click noise. As a result of the mixing operation, the RMS phase/frequency deviation of $s_e(t) = s(t)s_l(t)$ at the IF filter input, and according to Carson's formula also its bandwidth, is increased. This eventually results in suppression of the FM wave, as explained below.

Figure 8.6 illustrates the suppression mechanism. The quasi-stationary approach, i.e.; the "moving finger" model, is used to model the spectral density of the FM wave $s_e(t)$ at the IF filter input.

As a result of the message modulation $\dot{\varphi}_e(t)$, $s_e(t)$ moves along the pass band of the IF filter. At low CNRs, the intensity of this modulation is increased by the frequency noise $\dot{\theta}_l(t)$, fed back from the discriminator output by the oscillator. As a result of this noise, the modulation contained in $s_e(t)$ is increased such that it occasionally drives the wave out of the IF filter pass band (see figure 8.6). During such an event, the FM wave is suppressed considerably, while the wideband noise $n_e(t)$ remains essentially unaffected. Consequently, the limiter-discriminator input CNR drops significantly, and may cause operation below its threshold.

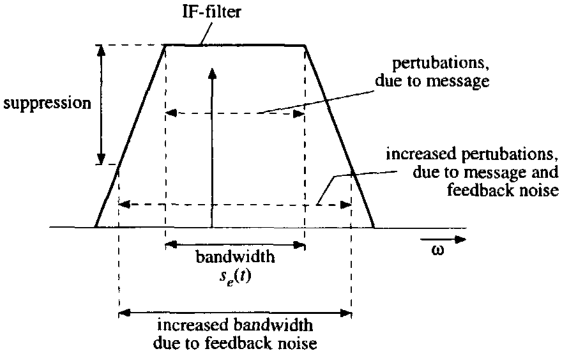

IF-filter
suppression
pertubations, due to message
increased pertubations, due to message and feedback noise
ω
bandwidth
$s_e(t)$
increased bandwidth due to feedback noise

Figure 8.6: Suppression of the FM wave by the IF filter, as a result of phase noise in the oscillator output signal.

Based on experiments, Bax decides that the FMFB threshold occurs when the discriminator is driven below its threshold for 5 – 8% of the time, which corresponds to a certain relation between the RMS frequency deviation of the oscillator output wave and the input CNR.

Although many of the details of this theory, especially the 5 – 8% conclusion are not explicitly adopted in this chapter, its principles are elaborated in subsequent sections.

8.3.4 Model of the Limiter-Discriminator Input Noise

This section starts the derivation of a model for the FMFB receiver threshold. Based on the theory developed in [11] that attributed the threshold to a combination of the limiter-discriminator threshold and carrier suppression by the IF filter, the model combines the previous threshold mechanisms in order to arrive at a model that does not contain experimentally-determined parameters.

The input CNR of the limiter-discriminator inside the loop is a key parameter in the threshold model, since it determines whether the discriminator operates above or below its threshold. As a first step in the determination of this CNR, this section derives a description for the noise observed at the limiter-discriminator input in terms of the receiver input noise $n(t)$. Subsequently, Section 8.3.5 derives an expression for the FM wave observed at this point inside the loop.

The calculation of the limiter-discriminator input noise consists of two steps. First, an expression for the mixer output noise is determined in order to obtain the limiter-discriminator input CNR, denoted by p_d. Subsequently, application of IF filtering is considered.

Mixer Output Noise

The mixer response to the input noise $n(t)$, denoted by $n'_e(t)$, is readily obtained as

$$\begin{aligned} n'_e(t) &= n(t)s_l(t) \\ &= n_{s,i}(t)\cos\left[\omega_e t + \varphi_e(t) - \theta_l(t)\right] \\ &\quad - n_{s,q}(t)\sin\left[\omega_e t + \varphi_e(t) - \theta_l(t)\right]. \end{aligned} \tag{8.9}$$

This product may be viewed as the mixer output noise component, although it should be noted that the corresponding "signal component" $s'_e(t) = s(t)s_l(t)$ also contains noise by virtue of the phase noise $\theta_l(t)$. The latter, however, is the response of the loop to the noise "source" $n(t)$ that after shifting through the mixer becomes $n'_e(t)$.

Note that (8.9) is written in terms of the input noise component $n_{s,i}(t)$ that is in-phase with $s'_e(t)$, and $n_{s,q}(t)$ that is in quadrature with $s'_e(t)$.

IF Filter Output Noise

The noise at the limiter-discriminator input, denoted by $n_d(t)$, is the IF filter response to the mixer output noise $n'_e(t)$. A difficulty in the determination of $n_d(t)$ is the presence of the phase noise $\theta_l(t)$ that introduces correlation between the amplitude and phase of $n'_e(t)$.

Fortunately, since the bandwidth of the phase noise $\theta_l(t)$ determined by the loop is much smaller than the bandwidth of the input noise $n(t)$ (and the bandwidth of $n_{s,i}(t)$, $n_{s,q}(t)$), its influence on the spectrum of $n_e(t)$ is essentially negligible, at least as far as the fraction located within the pass band of the narrow-band IF filter is concerned. A similar observation was used in Chapter 6 to disregard the modulation of the noise by the message signal. In Chapter 7, the same observation allowed the modulation introduced into the noise $n'(t)$ at the phase detector output by the controllable oscillator to be ignored.

Therefore, in virtually all cases of practical interest, the spectrum of $n_e(t)$ is essentially equal to the spectrum of $n(t)$ inside the pass band of the IF filter. For this reason, the noise $n_d(t)$ at the IF filter output is essentially independent of the FM carrier at the discriminator input. Its low-pass equivalent spectrum may thus be expressed as

$$S_{n,d}(\omega) = \left|\Gamma_{\mathrm{IF}}(\omega)\right|^2 S_n(\omega), \tag{8.10}$$

the filtered input noise spectrum.

8.3.5 Model of the Limiter-Discriminator Input Signal

This section derives an expression for the FM wave observed at the limiter-discriminator input, denoted by $s_d(t)$. As discussed in Section 8.3.3, this FM wave is subject to suppression by the IF filter.

The determination of the FM wave consists of two steps. First, the power density spectrum of the mixer output signal $s'_e(t)$ is derived. Subsequently, the effect of IF filtering on this wave is analyzed.

Spectrum of the Mixer Output FM Wave

The mixer output signal component $s'_e(t)$, which may be expressed as

$$s'_e(t) = s(t)s_l(t) = A\cos\left[\omega_e t + \varphi_e(t) - \theta_l(t)\right], \tag{8.11}$$

is modulated by the FM message $\dot{\varphi}_e(t)$ and the frequency noise $\dot{\theta}_l(t)$ that is fed back from the discriminator output. At High CNRs, $\dot{\varphi}_e(t)$ dominates the modulation and determines the bandwidth of the (compressed) wave. At low CNRs however, the frequency noise $\dot{\theta}_l(t)$ increases significantly and thereby increases the RMS frequency deviation and bandwidth of $s'_e(t)$.

According to the quasi-stationary approach discussed in Section 2.2.2, the low-pass equivalent power density spectrum of $s'_e(t)$, denoted by $S_{s,e}(\omega)$, may be approximated by the PDF of the frequency modulation $\dot{\varphi}_e(t) - \dot{\theta}_l(t)$ times 2π the power contents of the FM wave, as described by expression (2.7). When $p_{\dot{\varphi}_e}(.)$ denotes the PDF of $\dot{\varphi}_e(t)$, and $p_{\dot{\theta}_l}(.)$ denotes the PDF of $\dot{\theta}_l(t)$, the PDF of $\dot{\varphi}_e(t) - \dot{\theta}_l(t)$ equals their convolution [25]. Therefore, the spectrum of the FM wave $s_e(t)$, denoted by $S_{s,e}(\omega)$, may be expressed as

$$S_{s,e}(\omega) \approx \pi A^2 \int_{-\infty}^{\infty} p_{\dot{\varphi}_e}(x) p_{\dot{\theta}_l}(\omega - x)\mathrm{d}x. \tag{8.12}$$

The message signal $\dot{\varphi}_e(t)$ simply equals the response of the system in the absence of noise, as described by the transfer $H_e(\mathrm{j}\omega)$ from (8.3). For sinusoidal modulation and Gaussian modulation, its PDF is easily obtained. For those signals, only the variance of $\dot{\varphi}_e(t)$ differs from the variance of $\dot{\varphi}(t)$: all other characteristics of their PDFs are identical. For other types of signals, the shape of the PDF generally changes when subjected to linear filtering. The determination of the PDF may then become quite difficult [26] unless the transfer $H_e(\mathrm{j}\omega)$ is essentially constant over the entire frequency range occupied by the spectrum of the input message $\dot{\varphi}(t)$. In that case, the PDF of both waves differs only by some scaling factors.

The PDF of $\dot{\theta}_l(t)$ is difficult to calculate at low CNRs due to the presence of click noise. Rice [27] determined a rather complicated expression for the PDF of the *discriminator* output frequency noise $\dot{\theta}_d$ for the unmodulated carrier-case ($\varphi_e(t) \equiv 0$). At high CNRs however, the PDF of $\dot{\theta}_d(t)$ is Gaussian, which is still approximately true around the discriminator threshold. Therefore, since Gaussian random variables remain Gaussian after application of linear filtering, the PDF of $\dot{\theta}_l(t)$ can be roughly approximated by a Gaussian density function with a variance $\overline{\dot{\theta}_l^2}$. The variance contained in this Gaussian density function follows from the demodulator model of figure 8.3 as

$$\overline{\dot{\theta}_l^{\,2}} \approx \frac{1}{2\pi}\int_{-\infty}^{\infty}\left|\frac{K_oK_dH(\mathrm{j}\omega)}{1+K_oK_d\Gamma_{\mathrm{IF}}(\mathrm{j}\omega)H(\mathrm{j}\omega)}\right|^2 S_{\dot{\theta}_d}(\omega)\mathrm{d}\omega. \tag{8.13}$$

where $S_{\dot{\theta}_d}(\omega)$ equals the spectrum of the FM discriminator output noise, consisting of continuous noise and click noise.

Narrow Band IF Filtering

The IF filter affects the FM wave $s_e'(t)$ in essentially two different ways.

In the first place, it introduces distortion into the modulation $\varphi_e(t)$. This distortion may be determined from the "open-loop" discriminator, supplied with $s_e'(t)$ and subsequently transferred to the output by means of the transfer of the closed-loop [11]. The feedback mechanism reduces this distortion, which means that its effect on the threshold is usually negligible [7, 10, 11].

Secondly, IF filtering suppresses the part of the power density spectrum $S_{s,e}(\omega)$ that is located outside the filter pass band. Obviously, this reduces the power content of the FM wave $s_d(t)$, as illustrated by figure 8.7. The shaded area represents the fraction of the carrier power that is suppressed by the IF filter. The area enclosed by the spectrum of $s_e'(t)$ and the IF filter transfer represents the power contents of the limiter-discriminator input FM wave $s_d(t)$, denoted by $P_{s,d}$. Obviously, when the variance of the frequency noise $\dot{\theta}_l$ increases, the spectrum of $s_e'(t)$ widens, resulting in a smaller power contents of $s_d(t)$. This power contents may be expressed as

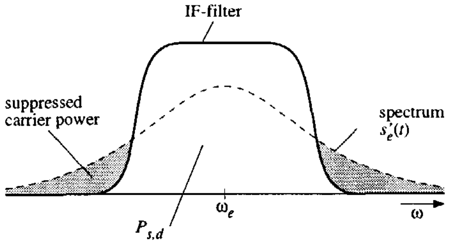


Figure 8.7: Reduction of the discriminator input carrier power at low CNRs, due to IF filtering.

$$P_{s,d} = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left|\Gamma_{\mathrm{IF}}(\omega)\right|^2 S_{s,e}(\omega)\mathrm{d}\omega. \tag{8.14}$$

Thus, the *inverse carrier-suppression factor*, denoted by Γ_c, equal to the ratio of the carrier power at the IF filter output and the IF filter input, may be expressed as

$$\Gamma_c \overset{\text{def}}{=} \frac{2P_{s,d}}{A^2} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left|\Gamma_{\text{IF}}(\omega)\right|^2 p_{\dot{\varphi}_e}(x) p_{\dot{\theta}_l}(\omega - x)\mathrm{d}x\mathrm{d}\omega. \tag{8.15}$$

Bax [11] uses a rather crude approximation for the suppression of the FM wave that underestimates the actual suppression. He basically approximates the signal suppression as a displacement of the FM wave from the center of the IF filter equal to the RMS value of $\dot{\theta}_l$. He obtains this RMS value from the linear above threshold model, ignoring click noise. In the presence of modulation, his approximation corresponds to an expression for Γ_c, equal to

$$\Gamma_{c,\text{Bax}} = \int_{-\infty}^{\infty} \left|\Gamma_{\text{IF}}\left(\dot{\varphi}_e + \dot{\theta}_{l,\text{rms}}\right)\right|^2 p_{\dot{\varphi}_e}\left(\dot{\varphi}_e\right)\mathrm{d}\dot{\varphi}_e. \tag{8.16}$$

Thus, this expression approximates the PDF of $\dot{\theta}_l(t)$, which represents the spectral density of the oscillator output wave $s_l(t)$ by a Dirac impulse located at $\omega = \dot{\theta}_{l,\text{rms}}$. A similar approximation was used to discount the modulation in the (wideband) input noise due to the message signal. However, since the bandwidth of $s_e'(t)$ and $s_l(t)$ are of the same order of magnitude, in this case this approximation is likely to introduce considerable errors.

8.3.6 FMFB Threshold Model

This section combines the models for the discriminator input noise and input signal, derived in section 8.3.4 8.3.5 respectively, to construct a model for the FMFB threshold behavior. Similar to the theories in [7, 10], this model describes the limiter-discriminator threshold by the click model.

An outline is as follows. First, an overview of the model is given. Subsequently, its components are discussed.

Overview of the Threshold Model

A low-pass equivalent FMFB model that is suitable for the calculation of the threshold is depicted in figure 8.8. This model differs from figure 8.3 in several ways. In the first place, the input noise $n(t)$ is no longer represented by the phase noise $\theta(t)$ at the input of the model, but by continuous phase noise $\theta_{d,\text{cnt}}(t)$, inserted at the IF filter output, click noise $\dot{\theta}_{d,\text{clk}}(t)$, inserted at the limiter-discriminator output (before the detector constant), and distortion $\varphi_{e,\text{dist}}(t)$, inserted at the IF filter output.

Further, both noise sources depend on the limiter-discriminator input CNR p_d instead of the receiver input CNR p. For input noise $n(t)$ with a flat spectrum, this CNR equals

$$p_d = \Gamma_c \Gamma_{\text{id}} p, \tag{8.17}$$

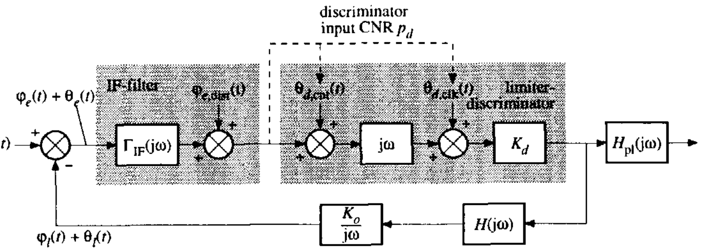

discriminator
input CNR p_d
$\varphi_e(t) + \theta_e(t)$
IF-filter
$\varphi_{e,\mathrm{dist}}(t)$
$\theta_{d,\mathrm{cnt}}(t)$
$\theta_{d,\mathrm{clk}}(t)$
limiter-
discriminator
$\varphi(t)$
+
−
$\Gamma_{\mathrm{IF}}(j\omega)$
+
+
$j\omega$
K_d
$H_{\mathrm{pl}}(j\omega)$
$\frac{K_o}{j\omega}$
$H(j\omega)$
$\varphi_l(t) + \theta_l(t)$

Figure 8.8: FMFB model used for calculation of the threshold.

where the maximum threshold extension factor Γ_{id} from (8.8) denotes the ratio of the IF filter noise bandwidth and the noise bandwidth of $n(t)$. This expression clearly shows that the carrier suppression factor Γ_c, which is smaller than unity, reduces the actual threshold extension below the maximum possible extension, represented by Γ_{id}. Thus, referring to figure 8.5, Γ_c represents the difference in the thresholds between curve (c) and curve (d).

Unfortunately, (8.17) is an implicit expression for p_d, since Γ_c depends on p_d, by means of the frequency noise variance $\overline{\dot{\theta}_l^2}$, given by (8.13). An analytic solution for p_d, or Γ_c, as function of the receiver input CNR p cannot therefore be obtained. Instead, p_d should be obtained as function of the receiver input CNR p through one out of the following approaches:

- resort to measurement of $\overline{\dot{\theta}_l^2}$ [11];
- approximation of $\overline{\dot{\theta}_l^2}$;
- graphic or iterative solving of (8.17) for p_d.

Measurement is not suitable for design purposes since it does not reveal the relations between the various parameters involved. An approximation of $\overline{\dot{\theta}_l^2}$ is not suited either, since it has to ignore the carrier suppression ($\Gamma_c \equiv 1$) in order to obtain an explicit expression. This results in an unacceptable loss of accuracy. Therefore, in subsequent sections, the relation between Γ_c (or p_d) and p is determined through graphic or iterative solving, which is not subject to the previously mentioned drawbacks of measurement and approximation.

Continuous Phase Noise

The frequency noise $\dot{\theta}_l(t)$ equals the frequency noise of the limiter-discriminator inside the loop. Therefore, in order to determine the relation between Γ_c and the discriminator input CNR p_d, the FM discriminator output noise has to be expressed in terms of p_d. Subsequently, Γ_c can be expressed in terms of p_d and the receiver input CNR p through the variance of $\dot{\theta}_l(t)$, as described by (8.15).

The continuous phase noise $\theta_{d,\mathrm{cnt}}(t)$ resulting in the continuous discriminator output frequency noise $\dot{\theta}_{d,\mathrm{cnt}}(t)$ is expressed in terms of p_d in the following way.

The discriminator input signal $r_d(t) = s_d(t) + n_d(t)$ can, on the basis of the foregoing discussion, be written as

$$\begin{aligned} r_d(t) &= s_d(t) + n_d(t) \\ &= A\sqrt{\Gamma_c}\cos\Phi_e(t) + n_{sd,i}(t)\cos\Phi_e(t) - n_{sd,q}(t)\sin\Phi_e(t) \\ &= R_d(t)\cos\left[\Phi_e(t) + \theta_d(t)\right], \end{aligned} \tag{8.18}$$

where the carrier phase $\Phi_e(t)$ equals

$$\Phi_e(t) = \omega_e t + \varphi_e(t). \tag{8.19}$$

Note that $\theta_l(t)$ is not included in $\Phi_e(t)$ since its effect on $s_d(t)$ is already described by Γ_c. The components $n_{ds,i}(t)$ and $n_{ds,q}(t)$ are in-phase and in quadrature with $s_d(t)$. They have a similar appearance as $n_{s,i}(t)$ and $n_{s,q}(t)$.

The phase noise $\theta_d(t)$ may be decomposed into a continuous component and a component that consists of phase jumps, corresponding to click noise. The continuous component may be expressed as

$$\theta_{d,\mathrm{cnt}}(t) \approx \frac{n_{sd,q}(t)}{A\sqrt{\Gamma_c\,(p_d)}}, \tag{8.20}$$

where $A\sqrt{\Gamma_c}$ denotes the amplitude of the FM wave at the IF filter output.

The corresponding power density spectrum equals

$$S_{\theta_{d,\mathrm{cnt}}}(\omega) = \frac{\left|\Gamma_{\mathrm{IF}}(\omega)\right|^2 S_n(\omega)}{A^2\Gamma_c\,(p_d)}. \tag{8.21}$$

Click Noise

The click noise component $\dot{\theta}_{d,\mathrm{clk}}(t)$ follows from the discriminator input signal and noise, as described in [24].

Its power density spectrum is given by (5.11), where the click rates N_+ and N_- depend on the discriminator input CNR p_d, instead of on the receiver input CNR p. Further, the radius of gyration r from (5.13) should be calculated from the shape of the IF filter, while the modulation $\varphi_e(t)$ instead of $\varphi(t)$ should be taken into account.

Variance of the Oscillator Frequency Noise

In order to complete the expression for Γ_c, the variance $\overline{\dot{\theta}_l^2}$ has to be expressed in terms of the loop parameters and the limiter-discriminator output noise. From figure 8.8, the expressions (8.13), (8.21) and the click noise spectral density (5.11), it follows that $\overline{\dot{\theta}_l^2}$ can be expressed as

$$\overline{\dot{\theta}_l^2} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \left|H_{\mathrm{cl}}(\mathrm{j}\omega)\right|^2 \omega^2 \frac{S_n(\omega)}{\Gamma_c A^2}\mathrm{d}\omega + 2\pi\left(N_+ + N_-\right)\int_{-\infty}^{\infty}\left|\frac{K_o K_d H(\mathrm{j}\omega)}{1+\Gamma_{\mathrm{IF}}(\mathrm{j}\omega)H_{\mathrm{j}}\omega)}\right|^2 \mathrm{d}\omega, \qquad (8.22)$$

where the click-rates N_+ and N_- are a function of the limiter-discriminator input CNR p_d.

8.4 Example Threshold Curve Calculation

As an illustration, this section determines the threshold curve of an FMFB receiver with the aid of the approach outlined in Section 8.3.6. The resulting curve is compared to the threshold curve of a comparable limiter-discriminator and phase feedback demodulator in Section 8.5.

Section 8.4.1 outlines the characteristics of the example FMFB system used in the comparison. Subsequently, Section 8.4.2 considers the solution of the implicit expression for the discriminator input CNR. Finally, Section 8.4.3 determines an expression for the FMFB output SNR as function of the receiver input CNR p.

8.4.1 System Configuration

This section outlines the characteristics of the FMFB system used in the calculation of the threshold curve. The position of the threshold of this system, as obtained from the theory, is compared in Section 8.5.3 with the measured position given in [11].

Loop Parameters

The FMFB loop is characterized by the following set of parameters:

- second-order IF filter (tuned resonant circuit), with bandwidth W_{IF};
- first-order low-pass loop-filter with -3-dB bandwidth $0.35W_{\mathrm{IF}}$;
- an additional first-order loop filter of bandwidth $3.5W_{\mathrm{IF}}$, representing the parasitic poles in the FM discriminator and controlled oscillator;
- a rectangular post-loop filter of bandwidth W.
- loop gain equal to $F_o = K_o K_d = 4$.

For such a system, the *open-loop* transfer from input phase $\Phi(\omega)$ to oscillator output phase $\Phi_l(\omega)$ can be written as

$$F_{\text{open}}(\mathrm{j}\Omega) = \frac{4}{(1+\mathrm{j}\Omega)\left(1+\mathrm{j}\frac{\Omega}{0.70}\right)\left(1+\mathrm{j}\frac{\Omega}{7}\right)}, \tag{8.23}$$

where the normalized bandwidth Ω equals $\Omega = 2\omega/W_{\text{IF}}$.

Input Signal

The input FM wave $s(t)$ is assumed to have a frequency deviation $\Delta\omega = 5W$, and a corresponding transmission bandwidth $W_n = 12W$. The noise $n(t)$ possesses the familiar characteristics, as used throughout this thesis. Its bandwidth W_n is chosen as $W_n = (1 + F_o)W_{\text{IF}} = 5W_{\text{IF}}$.

For simplicity, we consider the case of an unmodulated carrier $s(t)$ only. In the expressions for the output SNR, the "message signal" $\dot{\varphi}(t)$ is assumed to possess a power contents $P_{\dot{\varphi}} = (\Delta\omega)^2$.

Variance of the Oscillator Frequency Noise

In order to calculate the relation between the inverse carrier suppression factor Γ_c and the discriminator input CNR p_d, the variance of the frequency noise $\dot{\theta}_l(t)$ has to be expressed in terms of the loop parameters and the discriminator output noise.

By substitution of the filter transfers of the example FMFB into (8.22), this variance can be written in terms of Γ_c and the receiver input CNR p as

$$\overline{\dot{\theta}_l^2} = \frac{2}{\pi}\left(\frac{W_{\text{IF}}}{2}\right)^2 \frac{\Gamma_{\text{cnt}}}{4\Gamma_c\Gamma_{\text{id}}p} + \pi\left(N_+ + N_-\right)W_{\text{IF}}\Gamma_{\text{clk}}, \tag{8.24}$$

where $\Gamma_{\text{cnt}} \approx 13.1$ corresponds to the first integral in (8.22), representing the contribution of the continuous demodulator output noise, and $\Gamma_{\text{clk}} \approx 18.1$ corresponds to the second integral, representing the contribution of the click noise. Both integrals are expressed in terms of the normalized frequency Ω.

Further, the maximum threshold extension factor Γ_{id} for the IF filter applied in this system, and the rectangular input noise spectrum, becomes

$$\Gamma_{\text{id}} = \frac{2}{\pi}\frac{W_n}{W_{\text{IF}}}, \tag{8.25}$$

which is about 5 dB for the given system configuration.

8.4.2 Carrier Suppression and Discriminator Input CNR

The main problem in the determination of the FMFB threshold curve is to find a solution for the limiter-discriminator input CNR p_d from the implicit expression (8.17), by substitution of (8.15) for Γ_c and (8.24) for the variance $\overline{\dot{\theta}_l^2}$. This section demonstrates two approaches to obtaining a solution; a graphical approach and iteration.

Graphical Solution

A graphical solution of (8.17) for p_d is obtained from the intersection of two curves. The first curve represents the inverse carrier suppression factor Γ_c as a function of the frequency noise variance $\overline{\dot{\theta}_l^2}$, i.e. expression (8.15), that is treated as an independent variable. The second curve represents $\overline{\dot{\theta}_l^2}$ as a function of Γ_c, i.e. equation (8.24), where now Γ_c is treated as an independent parameter.

In the closed loop, both expressions are satisfied simultaneously. The solution of (8.17) therefore corresponds to the intersection of both curves in the Γ_c-$\overline{\dot{\theta}_l^2}$ plane.

An expression for Γ_c in terms of the frequency noise variance is obtained from (8.15). The PDF of the frequency noise is approximated by a Gaussian density with a variance equal to

$$\overline{\dot{\theta}_l^2} = (W_{\mathrm{IF}}/2)^2\nu. \tag{8.26}$$

For the second-order IF filter in the system under consideration, (8.15) becomes

$$\Gamma_c(\nu) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{\exp\left(-\frac{u^2}{2}\right)}{1+\nu u^2}\mathrm{d}u, \tag{8.27}$$

where $u = \Omega/\sqrt{\nu}$. This expression describes Γ_c as the area underneath the spectrum of the noise modulated FM wave $s_e(t)$, represented by the density of the frequency noise $\dot{\theta}_l$ with unit-variance, filtered by the IF filter. In terms of the normalized frequency u, the bandwidth of the IF filter is a function of the normalized variance ν. This exactly opposes the actual situation, where the density of $\dot{\theta}_l(t)$ widens for increasing ν. A closed form expression for (8.27) seems hard to find, and is actually not required. In subsequent sections, we resort to numerical calculation of this equation.

In order to express ν in terms of Γ_c, the click rate in (8.24) has to be determined. In the absence of modulation and the given IF filter, $N_+ + N_-$ may be expressed as

$$N_+ + N_- = \frac{W_{\mathrm{IF}}}{4\pi}\sqrt{\frac{\frac{2}{\pi}\Gamma_{\mathrm{id}}}{\arctan\left(\frac{2}{\pi}\Gamma_{\mathrm{id}}\right)} - 1}\left[1 - \mathrm{erf}\left(\sqrt{\Gamma_c\Gamma_{\mathrm{id}}p}\right)\right]. \tag{8.28}$$

Figure 8.9 depicts the curve corresponding to (8.27), and curves corresponding to (8.24) for $\Gamma_{\text{cnt}} = 13.1$, $\Gamma_{\text{clk}} = 18.1$, and a receiver input CNR $p = 0, 3, 7$ and 10 dB. This figure shows that for high input CNRs, the FM carrier at the discriminator input is hardly suppressed, since $\Gamma_c \approx 1$. At low CNRs however, considerable suppression is observed. For example, at $p = 0$ dB, $\Gamma_c \approx 0.4$, corresponding to a suppression of 4 dB. Notice, that for each value of p_d on the Γ_c-p_d curve, it is necessary to plot a complete curve of ν as a function of Γ_c, for a certain value of p, followed by determination of its intersection with the Γ_c-ν curve. Although insight is gained through this procedure, it is a rather inefficient means to determine the entire Γ_c-p_d curve. Therefore, for the determination of the entire curve, we resort to iteration, as discussed subsequently.

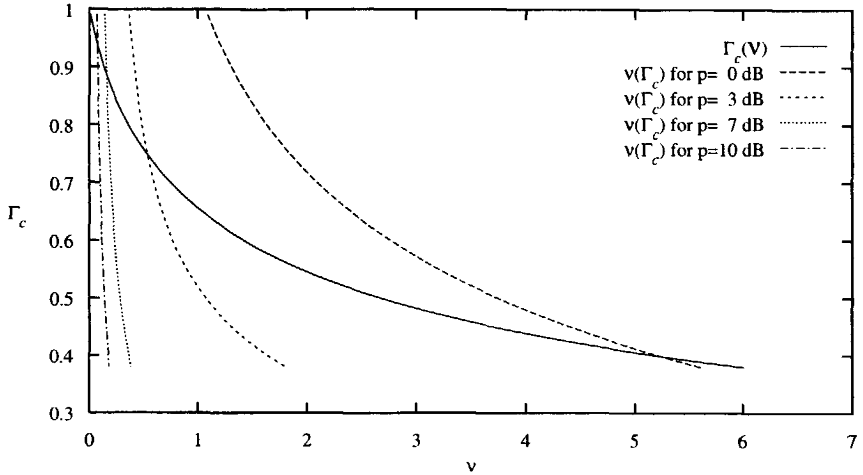


Figure 8.9: Carrier suppression factor Γ_c as a function of the normalized frequency noise variance ν for various input CNRs. The intersections correspond to the solutions for the closed loop FMFB.

Iteration

Another way of solving (8.17) is to use Picard-Iteration. For each receiver input CNR, p, the procedure starts with the assumption that no suppression occurs, i.e. $\Gamma_c = 1$. For this value, ν can be calculated. The value obtained in this way is subsequently substituted in (8.27) to obtain the new value of Γ_c, and so on.

This approach has been used to determine Γ_c as a function of the input CNR p. The iteration process was truncated when consecutive iterations for Γ_c differed by less than 0.01 (1% accuracy). The resulting curve is depicted in figure 8.10. This curve shows that the FM carrier at the discriminator input

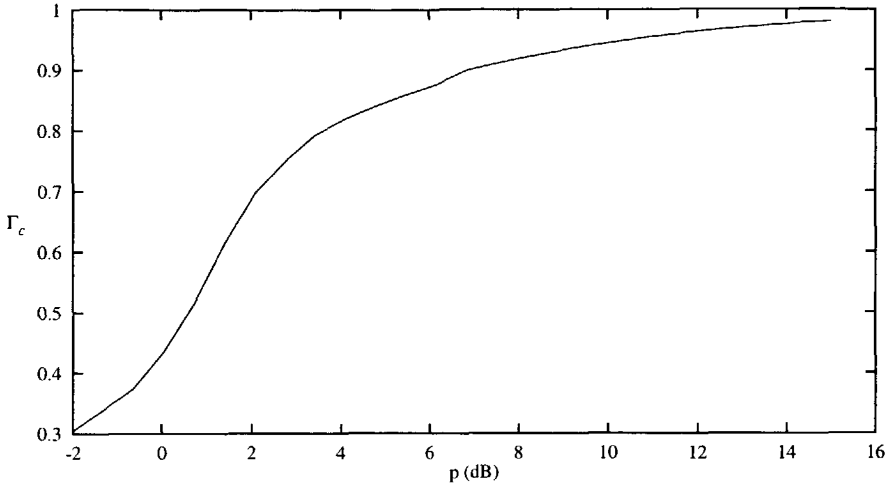


Figure 8.10: Inverse carrier suppression factor Γ_c as function of the input CNR p.

is considerably suppressed at low CNRs, which means that the accomplished threshold extension is smaller than predicted by curve (c) in figure 8.5. It should be noted that at low CNRs, the inaccuracy of the curve in this figure gradually increases as a result of the fact that $\dot{\theta}_l$ is no longer Gaussian in that region due to the click noise.

8.4.3 FMFB Receiver Output SNR

As a final step of the calculation, the output SNR of the FMFB receiver can be obtained with the aid of the curve for the inverse carrier suppression factor, depicted in figure 8.10, and the demodulator model depicted in figure 8.8. The resulting expression equals

$$\mathrm{SNR}_{\mathrm{FMFB}} = \frac{3p\Gamma_c(p)\left(\frac{W_n}{W}\right)\left(\frac{\Delta\omega}{W}\right)^2}{1+3p\frac{2}{\pi}\frac{\Gamma_c(p)}{\Gamma_{\mathrm{id}}}\left(\frac{W_n}{W}\right)^2\sqrt{\frac{\frac{\pi}{2}\Gamma_{\mathrm{id}}}{\arctan\left(\frac{\pi}{2}\Gamma_{\mathrm{id}}\right)}-1}\left[1-\mathrm{erf}\left(\sqrt{\Gamma_c(p)\Gamma_{\mathrm{id}}p}\right)\right]}. \quad (8.29)$$

The numerator of this expression equals the maximum possible output SNR, multiplied by the inverse suppression factor Γ_c. The denominator represents the contribution of the click noise.

8.5 Threshold Curve Comparison

This section compares the threshold curve of the example FMFB receiver analyzed in Section 8.4 with the curves of comparable limiter-discriminators and a comparable phase feedback demodulator in order to attain insight into the differences in the threshold behavior of these systems.

Sections 8.5.1 and 8.5.2 determine the threshold curves of comparable limiter-discriminators and a comparable phase feedback demodulator respectively. Subsequently, Section 8.5.3 compares these curves with the FMFB threshold curve through reproduction of figure 8.5 with these calculated curves.

8.5.1 Limiter-Discriminator Output SNR

The FMFB threshold curve is compared with two limiter-discriminators; a wideband discriminator, that demodulates the wideband FM wave $s(t)$, and a narrow-band discriminator, that demodulates the compressed, narrow-band FM wave $s_e(t)$.

For a fair comparison, a second-order IF filter with -3 dB bandwidth $W_{\mathrm{IF}} = W_n$ has to be placed at the input of the the wideband limiter-discriminator. The noise bandwidth of this filter equals $\frac{\pi}{2}W_n$, corresponding to $\Gamma_{\mathrm{id}} = \frac{2}{\pi}$. The equivalent bandwidth of the noise at the discriminator input however, equals only $\frac{\pi}{4}W_n$, since its spectrum extends only from $-W_n/2$ to $W_n/2$. The discriminator input CNR therefore becomes $p_d = \frac{4}{\pi}W_n$. The expression for the SNR obtained for this system, according to the theory of Chapter 5 and Chapter 6, equals

$$\mathrm{SNR}_{\mathrm{LD,w}} = \frac{3p\left(\frac{W_n}{W}\right)\left(\frac{\Delta\omega}{W}\right)^2}{1 + 3p\frac{\pi}{4}\left(\frac{W_n}{W}\right)^2\sqrt{\frac{4}{\pi}-1}\left[1 - \mathrm{erf}\left(\sqrt{\frac{4}{\pi}p}\right)\right]}. \qquad (8.30)$$

The SNR of the narrow-band limiter-discriminator is obtained from (8.30) by replacement of $\Delta\omega$ by $\Delta\omega/(1+F_o)$. Further, since its input signal is filtered by the narrow-band IF filter, the discriminator input CNR equals $p_d = \Gamma_{\mathrm{id}}p = \frac{10}{\pi}p$.

8.5.2 Phase Feedback Demodulator Output SNR

Besides the two limiter-discriminators, the FMFB threshold is also compared with a first-order phase feedback demodulator that possesses the same (closed loop) noise bandwidth as the FMFB receiver. The selectivity of both systems is therefore roughly the same. Notice that the minimum (double-sided) noise bandwidth that can be achieved by both systems equals twice the message bandwidth.

With the aid of the theory from Chapter 7, the output SNR of the phase feedback demodulator (PFB), assuming a zero-valued steady-state phase error, can be expressed as

$$\mathrm{SNR}_{\mathrm{PFB}} = \frac{3p\left(\frac{W_n}{W}\right)\left(\frac{\Delta\omega}{W}\right)^2}{1+\frac{12\left(\frac{B_L}{W}\right)^2}{\mathrm{I}_0^2\left[2p\left(\frac{W_n}{2\pi B_L}\right)\right]}}, \tag{8.31}$$

where the double-sided noise bandwidth equals

$$\begin{aligned} B_L &= \frac{W_n}{2\pi^2\Gamma_{\mathrm{id}}}\int_{-\infty}^{\infty}\left|\frac{F_{\mathrm{open}}(\mathrm{j}\Omega)}{1+F_{\mathrm{open}}(\mathrm{j}\Omega)}\frac{1+F_{\mathrm{open}}(0)}{F_{\mathrm{open}}(0)}\right|^2 \mathrm{d}\Omega \\ &\approx 1.24\frac{W_n}{2\pi^2\Gamma_{\mathrm{id}}}. \end{aligned} \tag{8.32}$$

8.5.3 Comparison

The resulting curves are depicted in figure 8.11. The curves (a) through (d) are the counterpart of the curves sketched in figure 8.5. Curve (c), of the ideal FMFB, is obtained from (8.29) by setting $\Gamma_c \equiv 1$, i.e. by ignoring the carrier suppression. Curve (e) corresponds to the phase feedback demodulator (PFB).

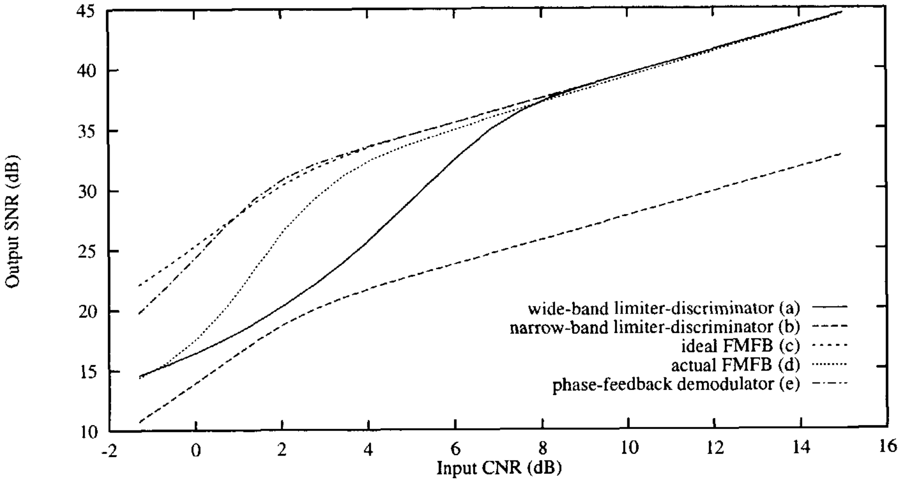


Figure 8.11: Threshold curves as a function of the FMFB receiver input CNR. (a) wideband limiter-discriminator, (b) narrow-band limiter-discriminator, (c) ideal FMFB, (d) actual FMFB, (e) PFB demodulator.

This figure shows that the carrier-suppression causes the difference between the ideal FMFB threshold, curve (c), and the actual FMFB threshold, curve (d). The upper bound on the threshold extension equals $\Gamma_{\mathrm{id}} = 10/\pi$, which corresponds to 5 dB. The actual extension is slightly smaller since $\Gamma_c < 1$. The threshold of the narrow-band discriminator, curve (b), and the threshold of ideal FMFB, curve (c), are identical, which agrees with our expectations; the minimum possible FMFB threshold equals the threshold of the narrow-band discriminator. Further, notice that as a result of the carrier suppression, the threshold of the FMFB is considerably steeper than the threshold of the limiter discriminator. This more "aggressive" threshold behavior is also observed experimentally [11], and can be heuristically viewed as the 'penalty' that is paid for the threshold extension.

It is interesting to note that the threshold of the phase feedback demodulator, curve (e), roughly coincides with the threshold of the narrow-band limiter-discriminator, curve (b), and curve (c) of the ideal FMFB. This can be explained by the observation that the phase feedback loop does not contain an IF filter, which means that the carrier suppression effect is absent. Below the threshold, the PFB curve is much steeper than the limiter-discriminator curve (b), and the ideal FMFB curve, which implies that replacement of the narrow-band limiter-discriminator in the FMFB by the PFB demodulator, as proposed in [28], is likely to result in an even worse threshold behavior, while the position of the threshold basically remains unchanged.

Some confidence in the accuracy of the actual FMFB threshold curve (d) in figure 8.11 is gained from the observation that the position of the 1 dB threshold corresponds quite accurately to the 1 dB threshold[1] measured by Bax [11] for exactly the same system. Converted into the parameters of this chapter, he measured the threshold at $p = 4.2$ dB. This is about the point where curve (d) starts to deviate significantly from the above-threshold asymptote.

8.6 Frequency Feedback Demodulator Design

In this section, we outline the implications of the previously discussed theory on the design of FMFB receivers. Although explicit rules for the optimization of FMFB receivers can hardly be given, some important remarks can be made.

The design of FMFB receivers comprises the design of the IF filter, the feedback filter $H(\mathrm{j}\omega)$ and the loop gain $K_o K_d$. These design topics are briefly considered in sections 8.6.1 through 8.6.3.

8.6.1 Design of the IF Filter

The IF filter is, after the FM demodulator, the most important subsystem in the FMFB receiver. This filter establishes the trade-off between threshold extension and reduction of the distortion in the FMFB.

[1] The 1 dB threshold equals the input CNR where the output SNR deviates by 1 dB from the above threshold asymptote described by the numerator of (8.30).

For maximum threshold extension, the IF filter bandwidth should be chosen to be as small as possible. The minimum permissible bandwidth is determined by the specification of the maximum tolerable distortion; the smaller the bandwidth, the larger the narrow-band filtering distortion. The distortion in the FMFB output signal can be calculated from the narrow-band "open-loop" discriminator, and subsequently transfered to the FMFB output by means of the loop-transfer [11].

For minimum distortion, the IF filter bandwidth should be chosen to be as large as possible. In this case, an upper bound on the bandwidth may stem from the minimum required selectivity of the system in order to suppress carriers at adjacent channels, or the maximum permissible threshold level.

Further, the most important constraint on the system is that its stability has to be guaranteed in all circumstances. In practice, this will put an upper limit on the order of the IF filter, and thereby on the selectivity of this filter.

8.6.2 Design of the Feedback Filter

The feedback filter should be dimensioned such that the loop gain of the system is essentially constant over the entire bandwidth of the message signal [11]. By this choice, all message frequencies are compressed by the same amount, which roughly minimizes the required IF bandwidth.

Further, the feedback filter should minimize the contribution of the discriminator output frequency noise to the oscillator output frequency noise. This means that it should minimize the factors Γ_{cnt} and Γ_{clk} defined in Section 8.4.1. Since the transfer corresponding to both factors is different, this minimization generally requires a trade-off between both noise contributions. The main difference between both transfers is that the poles of the IF filter $\Gamma_{\mathrm{IF}}(\mathrm{j}\omega)$ appear as zeros in the transfer of the click noise, while these zeros do not appear in the transfer for the continuous noise.

Finally, the feedback filter should be designed to guarantee the stability of the system. This may be accomplished by, for example, the introduction of some (high frequency) zeros into its transfer that appear as phantom zeros [29] in the demodulator transfer.

8.6.3 Design of the Loop gain

The most important function of the loop gain is to establish the required compression of the input FM wave's frequency deviation and the associated bandwidth compression. For maximum threshold extension, as large a loop gain as possible is required.

An upper limit to the loop gain is set by the stability requirement and, in general, also by the requirement for an as low level as possible of oscillator output frequency noise.

Roberts [7, 10] indicates the existence of a rather shallow optimum loop gain value that minimizes the variance of the oscillator output frequency noise. Although such an optimum might indeed exists at least for certain system configurations, it is questionable if his analysis is correct since it does not account for the carrier suppression by the IF filter.

8.7 Conclusions

This chapter investigated the threshold behavior of frequency feedback demodulators and made a comparison with conventional demodulators and phase feedback demodulators.

It was shown that, similar to phase feedback, frequency feedback does not improve the above-threshold output SNR of limiter-discriminators, i.e. conventional demodulators with infinite compression, but only shifts the threshold to a lower input CNR (threshold extension). An upper bound on the attainable threshold extension is defined by the threshold of a narrow-band limiter-discriminator with a bandwidth equal to the bandwidth of the IF filter inside the loop.

The FMFB threshold is due to a combination of three nonlinear effects:

- the threshold of the discriminator inside the loop;
- feedback of noise from the discriminator output;
- suppression of the FM wave at the discriminator input by the IF filter.

The feedback noise alone is too weak to cause the FMFB threshold independently. However, in combination with the suppression in the IF filter, it does cause the threshold of the limiter-discriminator inside the loop. Thus, it is concluded that all three nonlinear noise effects are mutually coupled: together they cause the FMFB threshold.

Due to the third effect, the realized threshold extension is smaller than the extension predicted by the upper bound determined by the narrow-band discriminator. This effect is due to an increase of the frequency deviation of the FM wave at the IF filter input caused by the feedback noise. A threshold model was derived by the modification and combination of some known theories that does not depend on experimentally determined parameters.

Comparison of the FMFB threshold curve with the threshold curve of limiter-discriminators showed that the threshold extension is achieved at the expense of a steeper decay of the SNR below the threshold, i.e. more "aggressive" threshold behavior. Further, it was shown that a comparable phase feedback demodulator realizes a larger threshold extension than the FMFB demodulator, which is explained by the observation that carrier suppression does not occur in phase feedback demodulators. Its threshold roughly coincides with the upper bound on the FMFB threshold extension.

Explicit design rules for the optimization of the FMFB receiver performance can hardly be given. However, it is clear that the IF filter plays a dominant role in the trade-off between the threshold extension and the reduction of distortion, the other alternative. A proper design of the loop-filter may minimize the carrier suppression effect. The stability of the closed-loop system is the main constraint on the design of the filters inside the loop, which should also account for parasitic poles introduced by the limiter-discriminator and the oscillator inside the loop.

References

[1] J.G. Chaffee, U.S. Patent 2,075,503, March 30, 1937.

[2] J.G. Chaffee, "The application of negative feedback to frequency modulation systems", *Proceedings of the IRE*, vol. 27, no. 5, pp. 317–331, May 1939.

[3] L.H. Enloe, "Decreasing the threshold in FM by frequency feedback", *Proceedings of the IRE*, vol. 50, no. 1, pp. 18–30, Jan. 1962.

[4] C.L. Ruthroff and W.F. Bodtmann, "Design and performance of a broadband FM demodulator with frequency compressive feedback", *Proceedings of the IRE*, vol. 50, no. 12, pp. 2436–2445, Dec. 1962.

[5] A.J. Giger and J.G. Chaffee, "The FM demodulator with negative feedback", *The Bell System Technical Journal*, vol. 42, pp. 1109–1135, July 1963.

[6] F. Lefrak, H. Moore, A. Newton, and L. Ozolins, "The frequency-modulation feedback system for the lunar-orbiter demodulator", *RCA Review*, vol. 27, no. 12, pp. 563–576, Dec. 1966.

[7] J.H. Roberts, "Frequency-feedback receiver as a low-threshold demodulator in FM FDM satellite systems", *Proceedings of the IEE*, vol. 115, no. 11, pp. 1607–1618, Nov. 1968.

[8] W. Bijker and W.G. Kasperkovitz, "A top-down design methodology applied to a fully integrated adaptive FM IF system with improved selectivity", in *Proceedings of the European Solid-State Circuits Conference*, Sevilla, Sept. 1993, pp. 53–56.

[9] P. Frutiger, "Noise in FM receivers with negative frequency feedback, part I & II", *Proceedings of the IEEE*, vol. 54, no. 11, pp. 1506–1520, Nov. 1966.

[10] J.H. Roberts, "Dynamic tracking filter as a low-threshold demodulator in FM FDM satellite systems", *Proceedings of the IEE*, vol. 115, no. 11, pp. 1597–1606, Nov. 1968.

[11] F.G.M. Bax, *Analysis of the FM Receiver with Frequency Feedback*, PhD thesis, Catholic University of Nijmegen, Nijmegen, The Netherlands, Oct. 1970.

[12] Edward Bedrosian and Stephen O. Rice, "Distortion and crosstalk of linearly filtered, angle-modulated signals", *Proceedings of the IEEE*, vol. 56, no. 1, pp. 2–13, Jan. 1968.

[13] Elie J. Baghdady, "The theory of FM demodulation with frequency-compressive feedback", *IRE Transactions on Communications Systems*, vol. 10, pp. 226–245, Sept. 1962.

[14] R.E. Heitzman, "A study of the threshold power requirements of FMFB receivers", *IEEE Transactions on Space Electronics and Telemetry*, vol. 8, pp. 249–256, Dec. 1962.

[15] R.M. Gagliardi, "Transmitter power reduction with frequency tracking FM receivers", *IEEE Transactions on Space Electronics and Telemetry*, vol. 9, no. 1, pp. 18–25, Mar. 1963.

[16] Jean A. Develet Jr., "Statistical design and performance of high-sensitivity frequency-feedback receivers", *IEEE Transactions on Military Electronics*, vol. 7, pp. 281–284, Oct. 1963.

[17] B.R. Davis, "Factors affecting the threshold of feedback FM detectors", *IEEE Transactions on Space Electronics and Telemetry*, vol. 10, pp. 90–94, Sept. 1964.

[18] Elie J. Baghdady and L.H. Enloe, "Decreasing the threshold in FM by frequency feedback", *Proceedings of the IEEE*, vol. 52, no. 9, pp. 1039–1044, Sept. 1964.

[19] F. Kühne, "Optimaldimensionierung von Frequenzkopplungsempfängern", *Nachrichten Technische Zeitschrift*, vol. 19, no. 6, pp. 347–351, 1966.

[20] Friedrich Kühne, "Gegenkopplungsdemodulation von frequenzmodulierten Signalen I", *Archiv Elektrische Übertragung*, vol. 21, no. 7, pp. 383–390, 1967.

[21] Friedrich Kühne, "Gegenkopplungsdemodulation von frequenzmodulierten Signalen II", *Archiv Elektrische Übertragung*, vol. 21, no. 7, pp. 507–518, 1967.

[22] J.H. Roberts, "Effects of modulation on the threshold", *Systems Technology*, vol. 2, pp. 38–44, Sept. 1967.

[23] Frank A. Cassara and Donald T. Hess, "FM threshold performance of the frequency demodulator with feedback", *IEEE Transactions on Aerospace and Electronic Systems*, vol. 8, no. 5, pp. 596–601, Sept. 1972.

[24] S. O. Rice, "Noise in FM receivers", in *Proceedings of the Symposium on Time Series Analysis, Brown University, 1962*, M. Rosenblatt, Ed. pp. 395–422, John Wiley and Sons, New York, 1963.

[25] Athanasios Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill Book Company, New York, 3rd edition, 1991.

[26] Wilbur B. Davenport and William L. Root, *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill Book Company, New York, 1958.

[27] S. O. Rice, "Statistical properties of a sine wave plus random noise", *The Bell System Technical Journal*, vol. 27, pp. 109–157, 1948.

[28] J. Frankle, "Threshold performance of analog FM demodulators", *RCA Review*, vol. 27, no. 12, pp. 521–562, Dec. 1966.

[29] E.H. Nordholt, *Design of High-Performance Negative Feedback Amplifiers*, Elsevier, Amsterdam, 1983.

Chapter 9

Conclusions

A classification of all possible FM demodulation principles is an indispensable aid in a structured approach towards FM demodulator design. It shows that direct FM demodulation by detection of the FM wave's instantaneous frequency is impossible since this frequency is not associated to the energy contained in the wave. Therefore, FM demodulation is necessarily established in an indirect way, i.e. by conversion from FM to AM, or FM to PM, and subsequent AM or PM demodulation. Two sub-classes of FM demodulators based on FM-AM conversion, four sub-classes based on FM-PM conversion combined with direct PM demodulation, and three sub-classes based on FM-PM conversion combined with indirect PM demodulation, i.e. PM-AM conversion followed by AM demodulation, exist. All FM demodulators that were encountered in the literature fit into this classification. Sub-classes of FM demodulators based on FM-AM conversion followed by indirect AM demodulation do not exist.

High performance can be attained only with FM demodulators that *prevent* the generation of a carrier-induced offset in their response to the instantaneous frequency of the input wave. This offset reduces the demodulator dynamic range; it reduces the maximum allowed signal swings, and increases the output noise level. Generation of this component can be prevented only with FM demodulators based on FM-AM conversion and subsequent AM projection detection, FM demodulators that establish FM-PM conversion with the aid of a fixed time delay, and FM demodulators that establish FM-PM conversion with the aid of phase feedback. In all other types, this offset cannot be prevented, but only eliminated afterwards. Further, the distortion can be minimized by proper design of the frequency transfer of the FM-AM and FM-PM converter.

Considerable improvement of the demodulator performance can be achieved by proper design of the FM receiver architecture that embeds the demodulator. However, since the signal processing included in this architecture is somehow based on assumptions regarding the characteristics of the received FM wave, performance degradation instead of improvement generally occurs as soon as the assumptions become invalid, e.g. due to high noise and disturbance levels. Pre-demodulation processing allows extraction of the required FM wave from the received frequency band by separation in frequency (filtering), phase (suppression of co-channel interference), and amplitude (elimination of amplitude noise). Post-demodulation processing allows the reduction of continuous demodulator output noise and impulsive output noise, called click noise. In practice, however, click detection and subsequent elimination is not effective. Finally, feedback and adaptive processing are an effective means for reducing the demodulator output noise, as long as the feedback is not disrupted by noise and disturbances.

The output noise of all types of FM demodulators consists of a continuous component and an impulsive component. The origin and behavior of the continuous component is similar for all types. The origin and behavior of the impulsive component, which is responsible for the FM threshold, is quite different.

A trade-off between continuous noise and click noise can be established by the application of finite compression to the input carrier amplitude, instead of the usual infinite compression. The contribution of amplitude noise to the demodulator output increases the level of continuous noise but reduces the level of click noise, through reduction of the average click pulse area. In order to prevent the generation of high levels of second-order modulation noise, it is favorable to establish the trade-off by a linear combination of infinite compression, and no compression, which yields a signal-to-noise ratio improvement of a few dB below the threshold.

The impulsive noise in phase feedback demodulators is due to cycle-slipping. The continuous output noise is identical to the noise obtained with a 'conventional' demodulator that applies infinite amplitude compression. A nonlinear analysis shows that considerable minimization of the cycle-slip noise is possible by proper design of the loop filter and the phase detector characteristic. A large detector gain, combined a loop bandwidth that is not larger than strictly necessary, seems profitable. Detectors with limiters at their inputs should be avoided. Loop filters with real poles, especially an 'ideal' integrator combined with a direct feed-through, seems suitable. Complex closed-loop poles should be avoided in order to avoid cycle-slip bursts. The steady-state phase error should be minimized. Generally, considerable threshold extension can be achieved.

The threshold of frequency feedback demodulators and dynamic tracking filters is due to the threshold of the discriminator inside the loop, in combination with suppression FM carrier suppression by the IF filter. The latter is due to an increase of the FM frequency deviation by feedback noise. The threshold cannot be extended below the threshold of the discriminator inside the loop. A phase feedback demodulator was observed to realize a slightly larger extension since it is not subject to carrier suppression.

Appendix A

Fourier Coefficients of the Limiter Output Signal Component

This appendix explains the derivation of expression (6.6) for the Fourier coefficients of the limiter output signal component as a function of the input CNR p. A similar derivation is given in [1], but it contains some annoying errors.

The derivation starts from expression (6.5) for the limiter output signal component $s_o(t)$. By substitution of $s(t) = A\cos\Phi(t)$, and application of the definition formula for the Fourier coefficients a_k, we obtain

$$\begin{aligned}
a_k &\overset{\text{def}}{=} \frac{\varepsilon_k}{2\pi}\int_{-\pi}^{\pi} s_{lo}\left[\Phi(t)\right]\cos k\Phi(t)\mathrm{d}\Phi(t) \\
&= \frac{\varepsilon_k}{2\pi}\int_{-\pi}^{\pi} \operatorname{erf}\left(\sqrt{p}\cos\Phi\right)\cos k\Phi \mathrm{d}\Phi \\
&= \frac{\varepsilon_k}{2k\pi}\int_{-\pi}^{\pi} \operatorname{erf}\left(\sqrt{p}\cos\Phi\right)\mathrm{d}\sin k\Phi \\
&= \frac{\varepsilon_k}{2k\pi}\operatorname{erf}\left(\sqrt{p}\cos\Phi\right)\sin k\Phi\Big|_{-\pi}^{\pi} - \frac{\varepsilon_k}{2k\pi}\int_{-\pi}^{\pi}\sin k\Phi \mathrm{d}\left[\operatorname{erf}\left(\sqrt{p}\cos\Phi\right)\right].
\end{aligned} \tag{A.1}$$

The first term in (A.1) equals zero, since $\sin(\pm k\pi) = 0$. The second term can be expanded with the aid of the definition formula for the erf function,

$$\operatorname{erf}(x) \overset{\text{def}}{=} \frac{2}{\sqrt{\pi}}\int_0^x \exp(-u^2)\mathrm{d}u, \tag{A.2}$$

as

$$\begin{aligned} a_k &= -\frac{\varepsilon_k}{2k\pi}\int_{-\pi}^{\pi}\sin k\Phi \mathrm{d}\left[\operatorname{erf}\left(\sqrt{p}\cos\Phi\right)\right] \\ &= \frac{\varepsilon_k}{2k\pi}\int_{-\pi}^{\pi}\sin k\Phi\frac{2}{\sqrt{\pi}}\exp\left(-p\cos^2\Phi\right)\sqrt{p}\sin\Phi \mathrm{d}\Phi \\ &= \frac{\varepsilon_k}{2k\pi}\sqrt{\frac{p}{\pi}}\int_{-\pi}^{\pi}\exp\left(-p\cos^2\Phi\right)\left[\cos(k-1)\Phi-\cos(k+1)\Phi\right]\mathrm{d}\Phi. \end{aligned} \tag{A.3}$$

The squared cosine in the argument of the exp(...) function can be expressed in terms of a cosine wave with double angle using the trigonometric identity $\cos^2\Phi = \frac{1}{2}+\frac{1}{2}\cos 2\Phi$. The result is

$$\begin{aligned} a_k = \frac{\varepsilon_k}{2k\pi}\sqrt{\frac{p}{\pi}}\exp\left(-\frac{p}{2}\right) \\ \int_{-\pi}^{\pi}\exp\left(-\frac{p}{2}\cos 2\Phi\right)\left[\cos(k-1)\Phi-\cos(k+1)\Phi\right]\mathrm{d}\Phi. \end{aligned} \tag{A.4}$$

The exponent inside this integral is expanded with the aid of the Jacobi-Anger formula [2, 3],

$$\exp(z\cos\varphi) = \sum_{m=0}^{\infty}\varepsilon_m \mathrm{I}_m(z)\cos m\varphi, \tag{A.5}$$

where $\mathrm{I}_m(z)$ denotes the modified Bessel function of the first kind and order m. Substitution into (A.4) gives

$$\begin{aligned} a_k = \sum_{m=0}^{\infty}\frac{\varepsilon_k\varepsilon_m}{2k\pi}\sqrt{\frac{p}{\pi}}\exp\left(-\frac{p}{2}\right)\mathrm{I}_m\left(-\frac{p}{2}\right) \\ \int_{-\pi}^{\pi}\cos(k-1)\Phi\cos 2m\Phi-\cos(k+1)\Phi\cos 2m\Phi \mathrm{d}\Phi. \end{aligned} \tag{A.6}$$

The integral in this expression can be evaluated by straightforward calculus. We obtain

$$\begin{aligned} \int_{-\pi}^{\pi}\cos(k-1)\Phi\cos 2m\Phi-\cos(k+1)\Phi\cos 2m\Phi \mathrm{d}\Phi = \\ \frac{2\pi}{\varepsilon_m}\left(\delta_{m,\frac{k-1}{2}}-\delta_{m,\frac{k+1}{2}}\right), \end{aligned} \tag{A.7}$$

where $\delta_{m,k}$ is the Kronecker delta, which equals 1 for $m=k$ and zero for $m\neq k$.

Substitution of this equation into (A.6) results in

$$
\begin{aligned}
a_k &= \sum_{m=0}^{\infty} \frac{\varepsilon_k}{k} \sqrt{\frac{p}{\pi}} \exp\left(-\frac{p}{2}\right) \mathrm{I}_m\left(-\frac{p}{2}\right) \left(\delta_{m,\frac{k-1}{2}} - \delta_{m,\frac{k+1}{2}}\right) \\
&= \sum_{m=0}^{\infty} \frac{\varepsilon_k(-1)^m}{k} \sqrt{\frac{p}{\pi}} \exp\left(-\frac{p}{2}\right) \mathrm{I}_m\left(\frac{p}{2}\right) \left(\delta_{m,\frac{k-1}{2}} - \delta_{m,\frac{k+1}{2}}\right) \\
&= \begin{cases} (-1)^{\frac{k-1}{2}} \frac{2}{k} \sqrt{\frac{p}{\pi}} \exp\left(-\frac{p}{2}\right) \left[\mathrm{I}_{\frac{k-1}{2}}\left(\frac{p}{2}\right) + \mathrm{I}_{\frac{k-1}{2}}\left(\frac{p}{2}\right)\right], & k \text{ odd} \\ 0, & k \text{ even} \end{cases} .
\end{aligned}
\tag{A.8}
$$

Substitution of $2k + 1$ for k finally yields expression (6.6) for the nonzero odd Fourier coefficients.

References

[1] William C. Lindsey, *Synchronization Systems in Communication and Control*, Prentice-Hall, Englewood Cliffs, New Jersey, 1972.

[2] David Middleton, *An Introduction to Statistical Communication Theory*, McGraw-Hill Book Company, New York, 1960.

[3] Milton Abramowitz and Irene A. Stegun, Eds., *Handbook of Mathematical Functions*, Dover Publications, New York, fifth edition, 1968.

Appendix B

Autocorrelation Function of the First-Order Noise

This appendix considers the derivation of the autocorrelation function of the first-order demodulator output noise, $R_{\mathrm{dem},1}(\tau)$, given by expression (6.44). This function is a linear combination of the autocorrelation functions and cross-correlation function of the amplitude noise contribution $\dot{\varphi}(t)n_{s,q}(t)$ and the frequency noise $\dot{n}_{s,q}(t)$.

We first derive these correlation functions, according to the procedure outlined in Section 6.2.2, and subsequently combine them to obtain the autocorrelation $R_{\mathrm{dem},1}(\tau)$.

B.1 Autocorrelation of the Amplitude Noise

The calculation of the autocorrelation function $R_{\dot{\varphi}n_{s,i}}(\tau)$ of the amplitude noise starts with the determination of the expectation over the noise processes $n_i(t)$ and $n_q(t)$. By substitution of (2.13), we obtain

$$\begin{aligned} R_{(\dot{\varphi}n_{s,i})(\dot{\varphi}n_{s,i})}(\tau) \overset{\mathrm{def}}{=} \mathrm{E}\left(\dot{\varphi}_1\dot{\varphi}_2 n_{s,i,1} n_{s,i,2} | \varphi_{1,2}, \dot{\varphi}_{1,2}\right)_{n_{i,1,2},n_{q,1,2}} = \\ \mathrm{E}\big[\dot{\varphi}_1\dot{\varphi}_2\left(n_{i,1}\cos\varphi_1 + n_{q,1}\sin\varphi_1\right) \\ \left(n_{i,2}\cos\varphi_2 + n_{q,2}\sin\varphi_2\right)|\varphi_{1,2}, \dot{\varphi}_{1,2}\big]_{n_{i,1,2},n_{q,1,2}}, \end{aligned} \tag{B.1}$$

where the subscripts "1" and "2" represent the instants $t = t_1$ and $t = t_2$ respectively. Since $n_i(t)$ and $n_q(t)$ are independent stochastic processes, this expression simplifies to

$$\begin{aligned}&\mathrm{E}\left(\dot\varphi_1\dot\varphi_2 n_{s,i,1}n_{s,i,2}|\varphi_{1,2},\dot\varphi_{1,2}\right)_{n_{i,1,2},n_{q,1,2}} = \\ &\quad \mathrm{E}\big[\dot\varphi_1\dot\varphi_2\big(n_{i,1}n_{i,2}\cos\varphi_1\cos\varphi_2 \\ &\qquad\qquad + n_{q,1}n_{q,2}\sin\varphi_1\sin\varphi_2\big)|\varphi_{1,2},\dot\varphi_{1,2}\big]_{n_{i,1,2},n_{q,1,2}}.\end{aligned} \tag{B.2}$$

Further, $n_i(t)$ and $n_q(t)$ have the same autocorrelation function $R_n(\tau)$ [1]. Therefore, (B.2) reduces to

$$\mathrm{E}\left(\dot\varphi_1\dot\varphi_2 n_{s,i,1}n_{s,i,2}|\varphi_{1,2},\dot\varphi_{1,2}\right)_{n_{i,1,2},n_{q,1,2}} = R_{nn}(\tau)\dot\varphi_1\dot\varphi_2\cos(\varphi_1-\varphi_2)\,. \tag{B.3}$$

Finally, taking the expectation over $\varphi_{1,2}$ and $\dot\varphi_{1,2}$, yields

$$R_{(\dot\varphi n_{s,i})(\dot\varphi n_{s,i})}(\tau) = R_{nn}(\tau)\mathrm{E}\left[\dot\varphi_1\dot\varphi_2\cos(\varphi_1-\varphi_2)\right]. \tag{B.4}$$

B.2 Autocorrelation of the Frequency Noise

The calculation of the autocorrelation of the frequency noise, denoted by $R_{\dot n_{s,q}}(\tau)$, starts with substitution of the time-derivative of (2.14) for $\dot n_{s,q}(t)$. This yields

$$\begin{aligned}&R_{\dot n_{s,q}\dot n_{s,q}}(\tau) \overset{\text{def}}{=} \mathrm{E}\left[\dot n_{s,q,1}\dot n_{s,q,2}|\varphi_{1,2},\dot\varphi_{1,2}\right] = \\ &\quad \mathrm{E}\big[\big(-\dot n_{i,1}\sin\varphi_1 - n_{i,1}\dot\varphi_1\cos\varphi_1 + \dot n_{q,1}\cos\varphi_1 - n_{q,1}\dot\varphi_1\sin\varphi_1\big) \\ &\quad \big(-\dot n_{i,2}\sin\varphi_2 - n_{i,2}\dot\varphi_2\cos\varphi_2 + \dot n_{q,2}\cos\varphi_2 - n_{q,2}\dot\varphi_2\sin\varphi_2\big) \\ &\qquad\qquad\qquad\qquad\qquad\qquad |\varphi_{1,2},\dot\varphi_{1,2}\big].\end{aligned} \tag{B.5}$$

Since $n_i(t)$ and $n_q(t)$ are independent, and both are uncorrelated with their derivative at the *same* instant [2, 3], this expression reduces to

$$\begin{aligned}&\mathrm{E}\left[\dot n_{s,q,1}\dot n_{s,q,2}|\varphi_{1,2},\dot\varphi_{1,2}\right] = \\ &\quad \mathrm{E}\big[\dot n_{i,1}\dot n_{i,2}\sin\varphi_1\sin\varphi_2 + \dot n_{i,1}n_{i,2}\dot\varphi_2\cos\varphi_1\cos\varphi_2 \\ &\quad + n_{i,1}\dot n_{i,2}\dot\varphi_1\cos\varphi_1\cos\varphi_2 + n_{i,1}n_{i,2}\dot\varphi_1\dot\varphi_2\cos\varphi_1\cos\varphi_2 \\ &\quad + \dot n_{q,1}\dot n_{q,2}\cos\varphi_1\cos\varphi_2 - \dot n_{q,1}n_{q,2}\dot\varphi_2\cos\varphi_1\sin\varphi_2 \\ &\quad - n_{q,1}\dot n_{q,2}\dot\varphi_1\sin\varphi_1\cos\varphi_2 + n_{q,1}n_{q,2}\dot\varphi_1\dot\varphi_2\sin\varphi_1\sin\varphi_2|\varphi_{1,2},\dot\varphi_{1,2}\big].\end{aligned} \tag{B.6}$$

When the autocorrelation of $\dot n_i(t)$ and $\dot n_q(t)$ is denoted by $R_{\dot n}(\tau)$, and the cross-correlation of $\dot n_i(t)$ and $n_i(t)$, or $\dot n_q(t)$ and $n_q(t)$ is denoted by $R_{\dot n n}(\tau)$, the result becomes

$$\begin{aligned}\mathrm{E}\left[\dot n_{s,q,1}\dot n_{s,q,2}|\varphi_{1,2},\dot\varphi_{1,2}\right] &= R_{\dot n\dot n}(\tau)\cos(\varphi_1-\varphi_2) \\ &\quad + R_{\dot n n}(\tau)\left(\dot\varphi_1+\dot\varphi_2\right)\sin(\varphi_1-\varphi_2) \\ &\quad + R_{nn}(\tau)\cos(\varphi_1-\varphi_2)\,.\end{aligned} \tag{B.7}$$

The expectation over $\varphi_{1,2}$ and $\dot\varphi_{1,2}$ finally yields

$$\begin{aligned} R_{\dot n_{s,q}\dot n_{s,q}}(\tau) = {} & R_{\dot n\dot n}(\tau)\mathrm{E}\left[\cos(\varphi_1-\varphi_2)\right] \\ & + R_{\dot n n}(\tau)\mathrm{E}\left[(\dot\varphi_1+\dot\varphi_2)\sin(\varphi_1-\varphi_2)\right] \\ & + R_{nn}(\tau)\mathrm{E}\left[\dot\varphi_1\dot\varphi_2\cos(\varphi_1-\varphi_2)\right]. \end{aligned} \tag{B.8}$$

B.3 Cross-correlation

The cross-correlation between the amplitude and frequency noise, denoted by $R_{\mathrm{cross},1}(\tau)$, is calculated as follows. By substitution of (2.13) and the derivative of (2.14), application of the independence of $n_i(t)$ and $n_q(t)$, and the knowledge that $n_i(t)$, $n_q(t)$ are uncorrelated with their derivatives at the same instant, we obtain

$$\begin{aligned} & R_{\mathrm{cross},1}(\tau) \overset{\mathrm{def}}{=} \mathrm{E}\left(\dot\varphi_1 n_{s,i,1}\dot n_{s,q,2} + \dot\varphi_2 n_{s,i,2}\dot n_{s,q,1} | \varphi_{1,2}, \dot\varphi_{1,2}\right) = \\ & \quad \mathrm{E}\big(-\dot n_{i,1}n_{i,2}\dot\varphi_2\sin\varphi_1\cos\varphi_2 - \dot n_{i,2}n_{i,1}\dot\varphi_1\sin\varphi_2\cos\varphi_1 \\ & \quad - 2n_{i,1}n_{i,2}\dot\varphi_1\dot\varphi_2\cos\varphi_1\cos\varphi_2 + \dot n_{q,1}n_{q,2}\dot\varphi_2\cos\varphi_1\sin\varphi_2 \\ & \quad + \dot n_{q,2}n_{q,1}\dot\varphi_1\cos\varphi_2\sin\varphi_1 \\ & \quad - 2n_{q,1}n_{q,2}\dot\varphi_1\dot\varphi_2\sin\varphi_1\sin\varphi_2 | \varphi_{1,2}, \dot\varphi_{1,2}\big) \\ & \quad = -2R_{nn}(\tau)\dot\varphi_1\dot\varphi_2\cos(\varphi_1-\varphi_2) + R_{\dot n n}(\tau)\left(\dot\varphi_1+\dot\varphi_2\right)\sin(\varphi_1-\varphi_2) \end{aligned} \tag{B.9}$$

The expectation over $\varphi_{1,2}$ and $\dot\varphi_{1,2}$ finally yields

$$\begin{aligned} R_{\mathrm{cross},1}(\tau) = {} & -R_{\dot n n}(\tau)\mathrm{E}\left[(\dot\varphi_1+\dot\varphi_2)\sin(\varphi_1-\varphi_2)\right] \\ & - 2R_{nn}(\tau)\mathrm{E}\left[\dot\varphi_1\dot\varphi_2\cos(\varphi_1-\varphi_2)\right]. \end{aligned} \tag{B.10}$$

B.4 Composite Autocorrelation Function

The composite autocorrelation function $R_{\mathrm{dem},1}(\tau)$ is obtained by addition of $R_{(\dot\varphi n_{s,i})(\dot\varphi n_{s,i})}(\tau)$, $R_{\dot n_{s,q}\dot n_{s,q}}(\tau)$ and $R_{\mathrm{cross},1}(\tau)$ with the appropriate weighting factors that can be obtained from expression (6.42) for the first-order component of the demodulator output signal, $y_{\mathrm{dem},1}(\tau)$.

It is observed that, ignoring the common term $G^{(}A)/A^2$, $R_{(\dot\varphi n_{s,i})(\dot\varphi n_{s,i})}(\tau)$ is weighed by $C_{n,1}^2(A)$, $R_{\dot n_s,q\dot n_{s,q}}(\tau)$ by unity, and $R_{\mathrm{cross},1}(\tau)$ by $C_{n,1}(A)$. Collecting terms then finally yields expression (6.44).

References

[1] A. Bruce Carlson, *Communication Systems, An introduction to Signals and noise in Electrical Communication*, McGraw-Hill Book Company, Singapore, 1986.

[2] Wilbur B. Davenport and William L. Root, *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill Book Company, New York, 1958.

[3] David Middleton, *An Introduction to Statistical Communication Theory*, McGraw-Hill Book Company, New York, 1960.

Appendix C

Spectrum of the First-Order Noise

This appendix considers the calculation of the first-order demodulator output noise spectrum, given by expression (6.45), and the simplified expression given by (6.46).

C.1 General Expression

The spectrum $S_{\mathrm{dem},1}(\omega)$ equals the Fourier transform of the autocorrelation function $R_{\mathrm{dem},1}(\tau)$. According to (6.44), this function consists of three terms that each are the product of two correlation functions. One correlation function represents the input noise $n(t)$, while the other one represents the modulation of this noise by the message. The output noise spectrum therefore consists of three terms equal to the convolution of the spectra that correspond to these correlation functions.

The spectra corresponding to $R_n(\tau)$, $R_{\dot{n}n}(\tau)$ and $R_{\dot{n}}(\tau)$ that are contained in $R_{\mathrm{dem},1}(t)$ can be expressed in terms of the spectrum of $n_i(t)$ and $n_q(t)$, $S_n(\omega)$, as [1]

$$R_{nn}(\tau) \overset{\mathcal{F}}{\longleftrightarrow} S_n(\omega), \tag{C.1}$$

$$R_{\dot{n}n}(\tau) = \frac{\mathrm{d}R_{nn}(\tau)}{\mathrm{d}\tau} \overset{\mathcal{F}}{\longleftrightarrow} \mathrm{j}\omega S_n(\omega), \tag{C.2}$$

$$R_{\dot{n}\dot{n}}(\tau) = -\frac{\mathrm{d}^2 R_{nn}(\tau)}{\mathrm{d}\tau^2} \overset{\mathcal{F}}{\longleftrightarrow} \omega^2 S_n(\omega). \tag{C.3}$$

The spectra corresponding to the expectations of $\varphi(t)$ and $\dot{\varphi}(t)$ can be obtained with the aid of the quasi-stationary approximation. Since $\mathrm{E}\left[\cos\left(\varphi_1 - \varphi_2\right)\right]$

equals the autocorrelation function of the zero-IF FM wave $\cos\varphi(t)$, its power density spectrum may be approximated by the probability density $p_{\dot{\varphi}}(.)$ of $\dot{\varphi}(t)$. This yields the following expressions for the spectra that correspond to the expectations in (6.44)

$$\mathrm{E}\left[\cos\left(\varphi_1-\varphi_2\right)\right] \overset{\mathcal{F}}{\longleftrightarrow} 2\pi p_{\dot{\varphi}}(\omega), \tag{C.4}$$

$$\begin{aligned}&\mathrm{E}\left[\left(\dot{\varphi}_1+\dot{\varphi}_2\right)\sin\left(\varphi_1-\varphi_2\right)\right] = \\ &\qquad -2\frac{\mathrm{d}}{\mathrm{d}\tau}\mathrm{E}\left[\cos\left(\varphi_1-\varphi_2\right)\right] \overset{\mathcal{F}}{\longleftrightarrow} -4\pi\mathrm{j}\omega p_{\dot{\varphi}}(\omega),\end{aligned} \tag{C.5}$$

$$\mathrm{E}\left[\dot{\varphi}_1\dot{\varphi}_2\cos\left(\varphi_1-\varphi_2\right)\right] = -\frac{\mathrm{d}^2}{\mathrm{d}\tau^2}\mathrm{E}\left[\cos\left(\varphi_1-\varphi_2\right)\right] \overset{\mathcal{F}}{\longleftrightarrow} 2\pi\omega^2 p_{\dot{\varphi}}(\omega). \tag{C.6}$$

The spectra corresponding to the three terms in (6.44), i.e. the frequency-domain convolutions of (C.1) and (C.6), (C.2) and (C.5), (C.3) and (C.4), can be expressed as

$$\begin{aligned}&R_{nn}(\tau)\mathrm{E}\left[\dot{\varphi}_1\dot{\varphi}_2\cos\left(\varphi_1-\varphi_2\right)\right] \overset{\mathcal{F}}{\longleftrightarrow} \\ &\qquad \int_{-\infty}^{\infty} y^2 S_n(\omega-y)p_{\dot{\varphi}}(y)\mathrm{d}y,\end{aligned} \tag{C.7}$$

$$\begin{aligned}&R_{\dot{n}n}(\tau)\mathrm{E}\left[\dot{\varphi}_1+\dot{\varphi}_2\sin\left(\varphi_1-\varphi_2\right)\right] \overset{\mathcal{F}}{\longleftrightarrow} \\ &\qquad \int_{-\infty}^{\infty} 2y(\omega-y)S_n(\omega-y)p_{\dot{\varphi}}(y)\mathrm{d}y,\end{aligned} \tag{C.8}$$

$$\begin{aligned}&R_{\dot{n}\dot{n}}(\tau)\mathrm{E}\left[\cos\left(\varphi_1-\varphi_2\right)\right] \overset{\mathcal{F}}{\longleftrightarrow} \\ &\qquad \int_{-\infty}^{\infty} (\omega-y)^2 S_n(\omega-y)p_{\dot{\varphi}}(y)\mathrm{d}y.\end{aligned} \tag{C.9}$$

Collecting terms proportional to ω^2, ω and those independent of ω, and application of the appropriate weighting factors from (6.44) finally yields (6.45).

C.2 Approximation for Wideband FM

For wideband FM waves, a simplified expression for the noise spectrum $S_{\mathrm{dem},1}(\omega)$ can be obtained by application of the fact that the FM transmission bandwidth is much larger than the bandwidth of the message signal $\dot{\varphi}(t)$.

The message signal slightly modulates the noise spectrum, described by the expectations in $R_{\mathrm{dem},1}(\tau)$. As discussed in Chapter 2, this modulation may be thought to result in slight perturbations of the center-frequency of the output noise spectrum, in the rhythm of the message. Since the noise spectrum is much wider than the message spectrum, these perturbations mainly affect the noise spectrum outside the baseband.

The part of the noise spectrum located inside the baseband can therefore be approximated by assuming that the bandwidth of the message approaches zero. In that case, the integrals in (6.45) can be approximated as

$$\int_{-\infty}^{\infty} S_n(\omega - y)p_{\dot{\varphi}}(y)\mathrm{d}y \approx S_n(\omega)\int_{-\infty}^{\infty} p_{\dot{\varphi}}(y)\mathrm{d}y, \tag{C.10}$$

$$\int_{-\infty}^{\infty} yS_n(\omega - y)p_{\dot{\varphi}}(y)\mathrm{d}y \approx S_n(\omega)\int_{-\infty}^{\infty} yp_{\dot{\varphi}}(y)\mathrm{d}y, \tag{C.11}$$

$$\int_{-\infty}^{\infty} y^2S_n(\omega - y)p_{\dot{\varphi}}(y)\mathrm{d}y \approx S_n(\omega)\int_{-\infty}^{\infty} y^2p_{\dot{\varphi}}(y)\mathrm{d}y. \tag{C.12}$$

The integral in (C.10) equals unity, by definition, since $p_{\dot{\varphi}}(.)$ is a probability density. The integral in (C.11) usually equals zero, since the message $\dot{\varphi}(t)$ is not allowed to have an offset component in FM transmission. The integral in (C.12) equals the power contained in the message signal, equal to $(\Delta\omega)^2$.

Substitution of these approximations into (6.45) yields the simplified expression (6.46).

References

[1] A. Bruce Carlson, *Communication Systems, An introduction to Signals and noise in Electrical Communication*, Singapore, 1986.

Appendix D

Correlation Functions of the Second-Order Noise

This appendix lists the equations that express the ten correlation functions required for the computation of $R_{\mathrm{dem2}}(\tau)$ in terms of the correlation functions of the input noise $n(t)$, and the message phase $\varphi(t)$.

Although the procedure is somewhat elaborate, all expressions are obtained in the same way as the three first-order correlation functions calculated in Appendix B. The weighting factor of each correlation function follows from the corresponding terms in expression (6.51) for $y_{\mathrm{dem,2}}(t)$.

The complete list of correlation functions is as follows:

$$\begin{aligned}
&\mathrm{E}\left[n_{s,i}^2\left(t_1\right) n_{s,i}^2\left(t_2\right) \dot{\varphi}\left(t_1\right) \dot{\varphi}\left(t_2\right)\right] = \\
&\qquad \sigma_n^4 R_{\dot{\varphi}\dot{\varphi}}(\tau) + 2R_{nn}^2(\tau)\mathrm{E}\left[\dot{\varphi}_1\dot{\varphi}_2\cos^2\left(\varphi_1-\varphi_2\right)\right],
\end{aligned} \tag{D.1}$$

$$\begin{aligned}
&\mathrm{E}\left[n_{s,q}^2\left(t_1\right) n_{s,q}^2\left(t_2\right) \dot{\varphi}\left(t_1\right) \dot{\varphi}\left(t_2\right)\right] = \\
&\qquad \sigma_n^4 R_{\dot{\varphi}\dot{\varphi}}(\tau) + 2R_{nn}^2(\tau)\mathrm{E}\left[\dot{\varphi}_1\dot{\varphi}_2\cos^2\left(\varphi_1-\varphi_2\right)\right],
\end{aligned} \tag{D.2}$$

$$\begin{aligned}
&\mathrm{E}\left[n_{s,i}\left(t_1\right) \dot{n}_{s,q}\left(t_1\right) n_{s,i}\left(t_2\right) \dot{n}_{s,q}\left(t_2\right)\right] = \\
&\qquad \sigma_n^4 R_{\dot{\varphi}\dot{\varphi}}(\tau) + 2R_{nn}^2(\tau)\mathrm{E}\left[\dot{\varphi}_1\dot{\varphi}_2\cos^2\left(\varphi_1-\varphi_2\right)\right] \\
&\quad +R_{nn}(\tau)R_{\dot{n}\dot{n}}(\tau)\mathrm{E}\left[\cos^2\left(\varphi_1-\varphi_2\right)\right] \\
&\quad +R_{nn}(\tau)R_{\dot{n}n}(\tau)\mathrm{E}\left[\left(\dot{\varphi}_1+\dot{\varphi}_2\right)\sin 2\left(\varphi_1-\varphi_2\right)\right] \\
&\quad +R_{\dot{n}n}^2(\tau)\mathrm{E}\left[\sin^2\left(\varphi_1-\varphi_2\right)\right],
\end{aligned} \tag{D.3}$$

$$\begin{aligned}
&\mathrm{E}\left[\dot{n}_{s,i}(t_1)\, n_{s,q}(t_1)\, \dot{n}_{s,i}(t_2)\, n_{s,q}(t_2)\right] = \\
&\quad \sigma_n^4 R_{\dot{\varphi}\dot{\varphi}}(\tau) + 2R_{nn}^2(\tau)\mathrm{E}\left[\dot{\varphi}_1\dot{\varphi}_2 \cos^2(\varphi_1 - \varphi_2)\right] \\
&\quad +R_{nn}(\tau)R_{\dot{n}\dot{n}}(\tau)\mathrm{E}\left[\cos^2(\varphi_1 - \varphi_2)\right] \\
&\quad +R_{nn}(\tau)R_{\dot{n}n}(\tau)\mathrm{E}\left[(\dot{\varphi}_1 + \dot{\varphi}_2)\sin 2(\varphi_1 - \varphi_2)\right] \\
&\quad +R_{\dot{n}n}^2(\tau)\mathrm{E}\left[\sin^2(\varphi_1 - \varphi_2)\right],
\end{aligned} \tag{D.4}$$

$$\begin{aligned}
&\mathrm{E}\left\{\dot{\varphi}(t_1)\,\dot{\varphi}(t_2)\left[n_{s,i}^2(t_1)\, n_{s,q}^2(t_2) + n_{s,i}^2(t_2)\, n_{s,q}^2(t_1)\right]\right\} = \\
&\quad 2\sigma_n^4 R_{\dot{\varphi}\dot{\varphi}}(\tau) + 4R_{nn}^2(\tau)\mathrm{E}\left[\dot{\varphi}_1\dot{\varphi}_2 \sin^2(\varphi_1 - \varphi_2)\right],
\end{aligned} \tag{D.5}$$

$$\begin{aligned}
&\mathrm{E}\left[\dot{\varphi}(t_1)\, n_{s,i}^2(t_1)\, n_{s,i}(t_2)\, \dot{n}_{s,q}(t_2) + \dot{\varphi}(t_2)\, n_{s,i}^2(t_2)\, n_{s,i}(t_1)\, \dot{n}_{s,q}(t_1)\right] = \\
&\quad -2\sigma_n^4 R_{\dot{\varphi}\dot{\varphi}}(\tau) - 4R_{nn}^2(\tau)\left[\dot{\varphi}_1\dot{\varphi}_2 \cos^2(\varphi_1 - \varphi_2)\right] \\
&\quad -R_{nn}(\tau)R_{\dot{n}n}(\tau)\mathrm{E}\left[(\dot{\varphi}_1 + \dot{\varphi}_2)\sin 2(\varphi_1 - \varphi_2)\right],
\end{aligned} \tag{D.6}$$

$$\begin{aligned}
&\mathrm{E}\left[\dot{\varphi}(t_1)\, n_{s,i}^2(t_1)\, \dot{n}_{s,i}(t_2)\, n_{s,q}(t_2) + \dot{\varphi}(t_2)\, n_{s,i}^2(t_2)\, \dot{n}_{s,i}(t_1)\, n_{s,q}(t_1)\right] = \\
&\quad 2\sigma_n^4 R_{\dot{\varphi}\dot{\varphi}}(\tau) + 4R_{nn}^2(\tau)\mathrm{E}\left[\dot{\varphi}_1\dot{\varphi}_2 \sin^2(\varphi_1 - \varphi_2)\right] \\
&\quad -R_{nn}(\tau)R_{\dot{n}n}(\tau)\mathrm{E}\left[(\dot{\varphi}_1 + \dot{\varphi}_2)\sin 2(\varphi_1 - \varphi_2)\right],
\end{aligned} \tag{D.7}$$

$$\begin{aligned}
&\mathrm{E}\left[\dot{\varphi}(t_1)\, n_{s,q}^2(t_1)\, \dot{n}_{s,q}(t_2)\, n_{s,i}(t_2) + \dot{\varphi}(t_2)\, n_{s,q}^2(t_2)\, \dot{n}_{s,q}(t_1)\, n_{s,i}(t_1)\right] = \\
&\quad -2\sigma_n^4 R_{\dot{\varphi}\dot{\varphi}}(\tau) - 4R_{nn}^2(\tau)\mathrm{E}\left[\dot{\varphi}_1\dot{\varphi}_2 \sin^2(\varphi_1 - \varphi_2)\right] \\
&\quad +R_{nn}(\tau)R_{\dot{n}n}(\tau)\mathrm{E}\left[(\dot{\varphi}_1 + \dot{\varphi}_2)\sin 2(\varphi_1 - \varphi_2)\right],
\end{aligned} \tag{D.8}$$

$$\begin{aligned}
&\mathrm{E}\left[\dot{\varphi}(t_1)\, n_{s,q}^2(t_1)\, \dot{n}_{s,i}(t_2)\, n_{s,q}(t_2) + \dot{\varphi}(t_2)\, n_{s,q}^2(t_2)\, \dot{n}_{s,i}(t_1)\, n_{s,q}(t_1)\right] = \\
&\quad 2\sigma_n^4 R_{\dot{\varphi}\dot{\varphi}}(\tau) + 4R_{nn}^2(\tau)\mathrm{E}\left[\dot{\varphi}_1\dot{\varphi}_2 \cos^2(\varphi_1 - \varphi_2)\right] \\
&\quad +R_{nn}(\tau)R_{\dot{n}n}(\tau)\mathrm{E}\left[(\dot{\varphi}_1 + \dot{\varphi}_2)\sin 2(\varphi_1 - \varphi_2)\right],
\end{aligned} \tag{D.9}$$

$$\begin{aligned}
&\mathrm{E}\left[n_{s,i}(t_1)\, \dot{n}_{s,q}(t_1)\, n_{s,q}(t_2)\, \dot{n}_{s,i}(t_2) + n_{s,i}(t_2)\, \dot{n}_{s,q}(t_2)\, n_{s,q}(t_1)\, \dot{n}_{s,i}(t_1)\right] = \\
&\quad -2\sigma_n^4 R_{\dot{\varphi}\dot{\varphi}}(\tau) - 4R_{nn}^2(\tau)\mathrm{E}\left[\dot{\varphi}_1\dot{\varphi}_2 \sin^2(\varphi_1 - \varphi_2)\right] \\
&\quad -2R_{nn}(\tau)R_{\dot{n}\dot{n}}(\tau)\mathrm{E}\left[\sin^2(\varphi_1 - \varphi_2)\right] \\
&\quad +2R_{nn}(\tau)R_{\dot{n}n}(\tau)\mathrm{E}\left[(\dot{\varphi}_1 + \dot{\varphi}_2)\sin 2(\varphi_1 - \varphi_2)\right] \\
&\quad -2R_{\dot{n}n}^2(\tau)\mathrm{E}\left[\cos^2(\varphi_1 - \varphi_2)\right].
\end{aligned} \tag{D.10}$$

Appendix E

Spectrum of the Second-Order Noise

This appendix outlines the derivation of the second-order noise spectrum $S_{\mathrm{dem},2}(\omega)$, given by (6.54). First, the general expression is considered. Subsequently, the simplifications for wideband FM are discussed. Finally, approximate expressions for the two components contained in the simplified expression are derived.

E.1 General Expression

The derivation of the second-order noise spectrum $S_{\mathrm{dem},2}(\omega)$ proceeds similar to the derivation of first-order noise spectrum $S_{\mathrm{dem},1}(\omega)$, considered in Appendix C.

Again, the spectrum equals the addition of convolutions between spectral components that represent the input noise and the message modulation respectively. The spectra corresponding to the correlation functions in (6.52) that represent the input noise can be expressed as

$$R_{nn}^2(\tau) \xleftrightarrow{\mathcal{F}} S_{n^2}(\omega) = S_n(\omega) * S_n(\omega), \tag{E.1}$$

$$R_{nn}(\tau)R_{\dot{n}n}(\tau) = \frac{1}{2}\frac{\mathrm{dR}_{\mathrm{nn}}^2(\tau)}{\mathrm{d}\tau} \xleftrightarrow{\mathcal{F}} \frac{\mathrm{j}}{2}\omega S_{n^2}(\omega), \tag{E.2}$$

$$R_{nn}(\tau)R_{\dot{n}\dot{n}}(\tau) - R_{\dot{n}n}^2(\tau) = -\frac{1}{2}\frac{\mathrm{d}^2 R_{nn}^2(\tau)}{\mathrm{d}\tau^2} \xleftrightarrow{\mathcal{F}} \frac{1}{2}\omega^2 S_{n^2}(\omega) \tag{E.3}$$

$$R_{nn}(\tau)R_{\dot{n}\dot{n}}(\tau) + R_{\dot{n}n}^2(\tau) \xleftrightarrow{\mathcal{F}} S_n(\omega) * \omega^2 S_n(\omega) - \omega S_n(\omega) * \omega S_n(\omega).) \tag{E.4}$$

Similarly, the spectra corresponding to the correlation functions that represent the modulation of the noise by the message follow with the aid of the quasi-stationary approximation as

$$\mathrm{E}\left[\cos 2\left(\varphi_1 - \varphi_2\right)\right] = 2\pi p_{2\dot{\varphi}}(\omega) \tag{E.5}$$

$$\begin{aligned}&\mathrm{E}\left[\left(\dot{\varphi}_1 + \dot{\varphi}_2\right)\sin 2\left(\varphi_1 - \varphi_2\right)\right] = \\ &\qquad -\frac{\mathrm{d}}{\mathrm{d}\tau}\mathrm{E}\left[\cos 2\left(\varphi_1 - \varphi_2\right)\right] \xleftrightarrow{\mathcal{F}} -2\pi \mathrm{j}\omega p_{2\dot{\varphi}}(\omega),\end{aligned} \tag{E.6}$$

$$\begin{aligned}&\mathrm{E}\left[\dot{\varphi}_1\dot{\varphi}_2 \cos 2\left(\varphi_1 - \varphi_2\right)\right] = \\ &\qquad -\frac{1}{4}\frac{\mathrm{d}^2}{\mathrm{d}\tau^2}\mathrm{E}\left[\cos 2\left(\varphi_1 - \varphi_2\right)\right] \xleftrightarrow{\mathcal{F}} \frac{\pi}{2}\omega^2 p_{2\dot{\varphi}}(\omega),\end{aligned} \tag{E.7}$$

where $p_{2\dot{\varphi}}(.)$ denotes the probability density of twice the message wave, $2\dot{\varphi}(t)$.

In the same way as explained in Appendix C, the convolutions between these components can be determined, and subsequently rearranged, resulting in (6.54).

E.2 Approximation for Wideband FM

For wideband FM waves, $S_{\mathrm{dem},2}(\omega)$ can be approximated in the baseband region in the same way as the first-order noise spectrum, considered in Appendix C.

The the bandwidth of the message signal is considered to be (nearly) zero, in comparison to the FM transmission bandwidth. In that case, $S_{n^2}(\omega)$ may be placed outside the integrals in (6.54), which results in

$$\begin{aligned}S_{\mathrm{dem},2}(\omega) \approx \frac{G^2(A)}{A^4}\Bigg\{&(\alpha-\beta)^2\sigma_n^4 S_{\dot{\varphi}}(\omega) \\ &+ \frac{(\alpha-\beta)^2}{2\pi}S_{n^2}(\omega)\int_{-\infty}^{\infty} S_{\dot{\varphi}}(y)\mathrm{d}y \\ &+ (1-\beta)^2\omega^2 S_{n^2}(\omega)\int_{-\infty}^{\infty} p_{2\dot{\varphi}}(y)\mathrm{d}y \\ &+ (1-\beta)(\alpha-\beta)\omega S_{n^2}(\omega)\int_{-\infty}^{\infty} y p_{2\dot{\varphi}}(y)\mathrm{d}y \\ &+ \frac{1}{4}(\alpha-\beta)^2 S_{n^2}(\omega)\int_{-\infty}^{\infty} y^2 p_{2\dot{\varphi}}(y)\mathrm{d}y \\ &+ \frac{\beta^2}{\pi}\int_{-\infty}^{\infty} y(2y-\omega)S_n(\omega-y)S_n(y)\mathrm{d}y\Bigg\}.\end{aligned} \tag{E.8}$$

The first integral in this expression equals 2π times the power contents of the message modulation, $(\Delta\omega)^2$. The second, third and fourth integral denote the area of the density, which equals unity, the expected value of $2\dot{\varphi}(t)$, which equals zero, and the the power contained in $2\dot{\varphi}(t)$, equal to $4(\Delta\omega)^2$. Substitution into (E.8) and neglecting the first term, which represents the signal suppression, finally yields the simplified spectrum given by (6.55).

E.3 Approximation with the aid of the Central Limit Theorem

This section considers the derivation of simplified expressions for the two convoluted spectral components $S_{n^2}(\omega)$ and $S_{n,n}(\omega)$, defined by (6.35), in (6.55), by the application of the central limit theorem for stochastic processes.

The Central Limit Theorem

The central limit theorem states that the probability density (PDF) of the sum of a large number of arbitrary distributed, independent random variables approaches a Gaussian density. When Σ_x denotes the sum of random variables

$$\Sigma_x = \sum_{i=0}^{n} x_i, \tag{E.9}$$

with expected value μ_x and variance σ_x^2, and n is sufficiently large, then the density of Σ_x, denoted by $p_{\Sigma_x}(u)$, approaches

$$p_{\Sigma_x}(u) \approx \frac{1}{\sigma_x\sqrt{2\pi}} \exp\left[-\frac{(x-\mu_x)^2}{2\sigma_x^2}\right]. \tag{E.10}$$

Although the theory holds only for 'large' n, values of $n = 2$ or $n = 3$ already result in close approximations when the variables x_i are of comparable magnitude.

In essence, the theorem may be regarded as a property of the convolution operation [1], because the sum of two random variables x and y, distributed according to the densities $p_x(.)$ and $p_y(.)$ respectively, equals the convolution [1]

$$p_{x+y}(u) = \int_{-\infty}^{\infty} p_x(u-v)p_y(v)\mathrm{d}v. \tag{E.11}$$

Approximation of $S_{n^2}(\omega)$

The spectral density $S_{n^2}(\omega)$ equals the convolution of the density $S_n(\omega)$ with itself. Since $S_n(\omega)$ is non-negative for all values of ω, it may be viewed, after appropriate scaling, as the probability density of a random variable n_1. The density $S_{n^2}(\omega)$ is then, according to (E.11), proportional to the PDF of the sum of two independent random variables n_1 and n_2, both distributed according to $S_n(\omega)$. This density is calculated as follows.

Scaling In order to use the central limit theorem, a density with unit area should be used. For this purpose, we define the PDF $p_n(\omega)$, that follows from the spectral density $S_n(\omega)$ as

$$p_n(\omega) \overset{\text{def}}{=} \frac{S_n(\omega)}{\int_{-\infty}^{\infty} S_n(\omega)\mathrm{d}\omega} = \frac{S_n(\omega)}{2\pi\sigma_n^2}, \tag{E.12}$$

where σ_n^2 denotes the power contained in the noise processes $n_i(t)$ and $n_q(t)$.

Expected Value The expected value of the variables n_1 and n_2, distributed according to $p_n(\omega)$, equals zero. This follows from the fact that $S_n(\omega)$ is symmetrical around $\omega = 0$. Therefore, the expected value of $n_1 + n_2$ also equals zero.

Variance The variance of n_1 and n_2 equals

$$E\left(n_1^2\right) = \int_{-\infty}^{\infty} \omega^2 p_n(\omega)\mathrm{d}\omega = (2\pi r)^2, \tag{E.13}$$

where r denotes the radius of gyration, defined by (5.13). The variance of the sum of the independent processes n_1 and n_2 therefore equals $2(2\pi r)^2$.

Density Function According to expression (E.10), the PDF of $n_1 + n_2$ can be approximated by a Gaussian density as

$$p_{n_1+n_2}(\omega) \approx \frac{1}{4\pi r\sqrt{\pi}} \exp\left[-\frac{\omega^2}{2(2\pi r)^2}\right]. \tag{E.14}$$

On the other hand, according to (E.11), this PDF equals

$$\begin{aligned} p_{n_1+n_2}(\omega) &= \int_{-\infty}^{\infty} p_n(\omega - y)p_n(y)\mathrm{d}y \\ &= \frac{1}{4\pi^2\sigma_n^4}\int_{-\infty}^{\infty} S_n(\omega - y)S_n(y)\mathrm{d}y. \end{aligned} \tag{E.15}$$

Approximate Spectral Density An approximate expression for $S_{n^2}(\omega)$ is now easily obtained from (E.14) and (E.15). By definition, $S_{n^2}(\omega)$ equals

$$S_{n^2}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_n(\omega - y)S_n(y)\mathrm{d}y. \tag{E.16}$$

Combination of this expression with (E.14) and (E.15) then yields the approximation given by (6.56).

Approximation of $S_{n,n}(\omega)$

The spectrum $S_{n,n}(\omega)$, i.e. the last integral in (6.54), can be approximated in a similar way as the spectrum $S_{n^2}(\omega)$. However, instead of expressing $S_{n,n}(\omega)$ as the convolution of two non-negative densities, we determine the approximated spectrum directly with the aid of the correlation function (E.4).

It is not difficult to show that $S_{n,n}(\omega)$ is non-negative and symmetric around $\omega = 0$. Therefore, it may be regarded, after appropriate scaling, as a PDF of the sum of a number of random variables. The PDF of these variables is determined below.

Scaling The area of $S_{n,n}(\omega)$ is obtained from (E.4) for $\tau = 0$. Thus

$$\begin{aligned}\int_{-\infty}^{\infty} S_{n,n}(\omega)\mathrm{d}\omega &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S_n(\omega - y)S_n(y)\mathrm{d}y\mathrm{d}\omega \\ &= 2\pi R_{nn}(0)R_{\dot{n}\dot{n}}(0) - R_{\dot{n}n}(0) = 2\pi\sigma_{\dot{n}}^2\sigma_n^2.\end{aligned} \tag{E.17}$$

Expected Value The expected value of the sum of random variables that corresponds to $S_{n,n}(\omega)$ equals zero since this spectrum is symmetric around $\omega = 0$.

Variance The variance of the sum of random variables is obtained from

$$\begin{aligned}\sigma_x^2 &= \frac{1}{2\pi\sigma_{\dot{n}}^2\sigma_n^2}\int_{-\infty}^{\infty}\omega^2 S_{n,n}(\omega)\mathrm{d}\omega \\ &= \frac{-1}{2\pi\sigma_{\dot{n}}^2\sigma_n^2}\frac{\partial^2}{\partial\tau^2}\left[R_{nn}(\tau)R_{\dot{n}\dot{n}}(\tau) + R_{\dot{n}n}^2(\tau)\right|\Bigg|_{\tau=0} \\ &= (2\pi r)^2\left[\left(\frac{\rho_r}{r}\right)^2 - 1\right],\end{aligned} \tag{E.18}$$

where r denotes the radius of gyration from (5.13), while ρ_r is defined as

$$\rho_r \overset{\text{def}}{=} \frac{1}{2\pi}\sqrt{\frac{\int_{-\infty}^{\infty}\omega^4 S_n(\omega)\mathrm{d}\omega}{\int_{-\infty}^{\infty}\omega^2 S_n(\omega)\mathrm{d}\omega}}. \tag{E.19}$$

Thus, ρ_r equals the radius of gyration of the time-derivative of the input noise.

Approximate Density From the foregoing discussion, the approximation for $S_{n,n}(\omega)$ is observed to be a Gaussian density of area $2\pi\sigma_n^2\sigma_{\dot{n}}^2 = 2\pi(2\pi r)^2\sigma_n^4$, and a variance given by (E.18), i.e.

$$S_{n,n}(\omega) \approx \frac{4\pi^2 r\sigma_n^2}{\sqrt{2\pi\left[\left(\frac{\rho_r}{r}\right)^2 - 1\right]}}\exp\left\{-\frac{\omega^2}{2(2\pi r)^2\left[\left(\frac{\rho_r}{r}\right)^2 - 1\right]}\right\}, \tag{E.20}$$

which equals expression (6.57).

References

[1] Athanasios Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill Book Company, New York, 3rd edition, 1991.

Appendix F

Joint Probability Density and Conditional Expectation

In this appendix we calculate the joint probability density of the four polar noise variables R_n, $\dot{R}_n$, φ_n and $\dot{\varphi}_n$, and the expected value of $|\dot{\varphi}_n|$, conditioned on the noise radius R_n and phase φ_n. These expressions are used in Section 6.5.5 to obtain an expression for the average click pulse area ξ.

F.1 Joint Probability Density

In order to obtain an expression for the joint probability density, we start from the four noise processes related to the Cartesian coordinates of the input noise, $n_{s,i}(t)$, $n_{s,q}(t)$, $\dot{n}_{s,i}(t)$ and $\dot{n}_{s,q}(t)$. As discussed in Section 6.4, these processes are Gaussian as long as all expectations are conditioned on the message signal $\dot{\varphi}$. The (optional) expectation over $\dot{\varphi}$ should therefore be performed in the final stage of the calculations.

The expression of the joint Gaussian density of the four noise processes requires evaluation of all possible (cross) correlations. The correlations that yield a nonzero result are:

$$\mathrm{E}\left(n_{s,i}^2\right) = \mathrm{E}\left(n_{s,q}^2\right) = \sigma_n^2, \tag{F.1}$$

$$\mathrm{E}\left(\dot{n}_{s,i}^2\right) = \mathrm{E}\left(\dot{n}_{s,q}^2\right) = \sigma_{\dot{n}}^2\left(1+u^2\right), \tag{F.2}$$

$$\mathrm{E}\left(n_{s,i}\dot{n}_{s,q}\right) = -\sigma_n\sigma_{\dot{n}}u, \tag{F.3}$$

$$\mathrm{E}\left(\dot{n}_{s,i}n_{s,q}\right) = \sigma_n\sigma_{\dot{n}}u, \tag{F.4}$$

where $u = \dot{\varphi}\sigma_n/\sigma_{\dot{n}} = \dot{\varphi}/(2\pi r)$. This shows that $n_{s,i}$ is correlated with $\dot{n}_{s,q}$, and $n_{s,q}$ is correlated with $\dot{n}_{s,i}$ when modulation is present.

From these expressions follows that the 2-dimensional joint Gaussian probability densities of $n_{s,i}$, $\dot{n}_{s,q}$ and $n_{s,q}$, $\dot{n}_{s,i}$ can be written as [1]

$$p\left(n_{s,i}, \dot{n}_{s,q}\right) = \frac{1}{2\pi\sigma_n\sigma_{\dot{n}}} \exp\left[-\frac{n_{s,i}^2}{2\sigma_n^2}\left(1+u^2\right) - \frac{\dot{n}_{s,q}^2}{2\sigma_{\dot{n}}^2} - \frac{n_{s,i}\dot{n}_{s,q}u}{\sigma_n\sigma_{\dot{n}}}\right], \tag{F.5}$$

$$p\left(n_{s,q}, \dot{n}_{s,i}\right) = \frac{1}{2\pi\sigma_n\sigma_{\dot{n}}} \exp\left[-\frac{n_{s,q}^2}{2\sigma_n^2}\left(1+u^2\right) - \frac{\dot{n}_{s,i}^2}{2\sigma_{\dot{n}}^2} + \frac{n_{s,q}\dot{n}_{s,i}u}{\sigma_n\sigma_{\dot{n}}}\right]. \tag{F.6}$$

These densities can be transformed into polar coordinates with the aid of (6.72), (6.73) and

$$\dot{n}_{s,i} = \dot{R}_n \cos\varphi_n - R_n\dot{\varphi}_n \sin\varphi_n, \tag{F.7}$$

$$\dot{n}_{s,q} = \dot{R}_n \sin\varphi_n + R_n\dot{\varphi}_n \cos\varphi_n. \tag{F.8}$$

Further, basic algebra shows that

$$\mathrm{d}n_{s,i}\mathrm{d}n_{s,q}\mathrm{d}\dot{n}_{s,i}\mathrm{d}\dot{n}_{s,q} = R_n^2\mathrm{d}R_n\mathrm{d}\dot{R}_n\mathrm{d}\varphi_n\mathrm{d}\dot{\varphi}_n. \tag{F.9}$$

Combination of these expressions eventually yields for the joint density in polar format:

$$p\left(R_n, \dot{R}_n, \varphi_n, \dot{\varphi}_n\right) = \\ \frac{R_n^2}{4\pi^2\sigma_n^2\sigma_{\dot{n}}^2} \exp\left[-\frac{R_n^2}{2\sigma_n^2}\left(1+u^2\right) - \frac{\dot{R}_n^2 + R_n^2\dot{\varphi}_n^2}{2\sigma_{\dot{n}}^2} - \frac{R_n^2\dot{\varphi}_n u}{\sigma_n\sigma_{\dot{n}}}\right]. \tag{F.10}$$

Since this expression does not contain the variable φ_n, this phase noise process is uniformly distributed, and independent of the other processes. Further, it is observed that the rate of change of the noise radius, $\dot{R}_n$, is independent of the other variables, while R_n and $\dot{\varphi}_n$ are correlated in the presence of modulation.

F.2 Conditional Expectation

By definition, the conditional expectation of $|\dot{\varphi}_n|$ is obtained from

$$\mathrm{E}\left(|\dot{\varphi}_n| |R_n, \varphi_n\right)_{\dot{R}_n, \dot{\varphi}_n} \overset{\mathrm{def}}{=} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\dot{\varphi}_n| \frac{p\left(R_n, \varphi_n, \dot{R}_n, \dot{\varphi}_n\right)}{p\left(\varphi_n\right)} \mathrm{d}\dot{R}_n\mathrm{d}\dot{\varphi}_n, \tag{F.11}$$

where $p\left(\varphi_n\right) = 1/(2\pi)$ denotes the marginal density of the noise phase φ_n.

Since it is observed from (F.10) that $\dot{R}_n$ is an independent zero-mean Gaussian random variable with variance $\sigma_{\dot{n}}^2$, evaluation of the inner integral in (F.11) results in

$$\mathrm{E}\left(|\dot{\varphi}_n| \middle| R_n, \varphi_n\right)_{\dot{R}_n, \dot{\varphi}_n} = \int_{-\infty}^{\infty} |\dot{\varphi}_n| \frac{R_n^2}{\sqrt{2\pi}\sigma_n^2 \sigma_{\dot{n}}} \exp\left[-\frac{R_n^2}{2\sigma_n^2}\left(1+u^2\right) - \frac{R_n^2 \dot{\varphi}_n^2}{2\sigma_{\dot{n}}^2} - \frac{R_n^2 \dot{\varphi}_n u}{\sigma_n \sigma_{\dot{n}}}\right] \mathrm{d}\dot{\varphi}_n. \tag{F.12}$$

By rewriting the argument of the exp(.) function in this expression as

$$-\frac{R_n^2}{2\sigma_n^2}\left(1+u^2\right) - \frac{R_n^2 \dot{\varphi}_n^2}{2\sigma_{\dot{n}}^2} - \frac{R_n^2 \dot{\varphi}_n u}{\sigma_n \sigma_{\dot{n}}} = -\frac{R_n^2}{2\sigma_n^2} - \frac{R_n^2}{2\sigma_n^2}\left(\frac{\dot{\varphi}_n}{2\pi r} + u\right)^2, \tag{F.13}$$

and application of the integral

$$\int_{-\infty}^{\infty} |x| \exp\left[-\frac{(x+a)^2}{2b^2}\right] \mathrm{d}x = 2b^2 \exp\left(-\frac{a^2}{2b^2}\right) + ab\sqrt{2\pi}\, \mathrm{erf}\left(\frac{a}{b\sqrt{2}}\right), \tag{F.14}$$

we obtain with $a = u$, $b = \sigma_n / R_n$ and $x = \dot{\varphi}_n/(2\pi r)$,

$$\mathrm{E}\left(|\dot{\varphi}_n| \middle| R_n, \varphi_n\right)_{\dot{R}_n, \dot{\varphi}_n} = \frac{r2\sqrt{2\pi}}{\sigma_n} \exp\left[-\frac{R_n^2}{2\sigma_n^2}\left(1+u^2\right)\right] + \frac{rR_n u}{\sigma_n^2} \exp\left(-\frac{R_n^2}{2\sigma_n^2}\right) \mathrm{erf}\left(\frac{R_n u}{\sigma_n \sqrt{2}}\right). \tag{F.15}$$

By means of the transformation $R_n = Av$, and $\mathrm{d}R_n = A\mathrm{d}v$, this expression can finally be rewritten as

$$\mathrm{E}\left(|\dot{\varphi}_n| \middle| v, \varphi_n\right)_{\dot{R}_n, \dot{\varphi}_n} = 4r\sqrt{\pi p} \exp\left[-pv^2\left(1+u^2\right)\right] + 2rpuv \exp\left(-pv^2\right) \mathrm{erf}\left(uv\sqrt{p}\right), \tag{F.16}$$

where p denotes the amplitude compressor input CNR.

References

[1] Wilbur B. Davenport and William L. Root, *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill Book Company, New York, 1958.

Summary

This thesis describes a structured approach towards the design of high-performance FM demodulators that provides insight into the principles available for the construction of these demodulators, and the various architectural measures that can be used to improve their performance. Such demodulators are applied, for example, in car radio and various types of wireless communication systems.

A brief discussion of the history of frequency modulation in Chapter 1 reveals that the existing theory on FM demodulator design lacks a unifying framework that relates all possible FM demodulation principles and the characteristics of the corresponding demodulators. Such a framework, i.e. a classification, is indispensable in a structured design approach. Further, it was observed that there is generally a large distance between the work of theoretic scientists, and practicing electronic designers in this field. It is the objective of this thesis to provide a unifying framework for FM demodulator design that bridges the gap between theoretical results and engineering practice. The thesis concentrates on the design of FM demodulators for analog FM, but the bulk of the material is also applicable to digital FM schemes.

Chapter 2 reviews the main characteristics of FM transmission and FM waves that constitute the treatises in subsequent chapters. A quasi-stationary approximation for the spectrum of FM waves is discussed, and the characteristic signal-to-noise improvement established by wideband FM through quadratic shaping of the input noise spectrum is explained.

Chapter 3 develops a classification of all possible FM demodulation principles. A brief outline of the principles of the applied design approach reveals the necessity for such a classification in demodulator design. FM demodulation through direct detection of the FM wave's instantaneous frequency appears to be impossible since this frequency is not associated to the energy of the wave. Instead, demodulation has to be established through conversion to AM, or conversion to PM, in combination with AM or PM demodulation. Further, in this respect, it appears that only direct AM demodulation is allowed, while both direct PM demodulation and indirect PM demodulation, by means of PM-AM conversion and subsequent AM demodulation, is allowed. This results in two

sub-classes of FM demodulators based on FM-AM conversion, four sub-classes based on FM-PM conversion in combination with direct PM demodulation, and three subclasses based on FM-PM conversion in combination with indirect PM demodulation. Two of the latter three subclasses are believed to be previously unknown, but at the same time have a limited practical significance. All types of demodulators that were encountered in the literature easily fit into the classification.

Chapter 4 discusses the design of the various sub-functions contained in FM demodulators. It is shown that an offset in the output signal due to the input carrier frequency deteriorates the demodulator dynamic range. It reduces the maximum allowable signal swings, and increases the output noise level. High-performance demodulation can therefore only be achieved when the generation of this offset is prevented, which is possible only with FM demodulators based on FM-AM conversion and subsequent AM projection detection, FM demodulators that establish FM-PM conversion by means of a fixed time-delay, and FM demodulators that establish FM-PM conversion by means of phase feedback. In order to avoid the offset, a zero-IF architecture, or a 'band pass' FM-AM/FM-PM converter is required. The distortion can be minimized by proper design of the converter frequency characteristic.

Chapter 5 discusses the various architectural provisions in the FM receiver architecture that can be used to improve the performance of the FM demodulator, which is embedded in the receiver. Generally, all processing performed by the receiver to improve the demodulator performance is somehow based on assumptions regarding the characteristics of the received FM wave. As soon as these assumptions become invalid, e.g. due to high noise and disturbance levels, this processing is likely to cause performance degradation instead of improvement. Pre-demodulation processing extracts the required FM wave from the received frequency band. This comprises separation in frequency, i.e. filtering, separation in phase, i.e. some kind of phase-locking in order to eliminate co-channel interference, and separation in amplitude, i.e. elimination of amplitude noise by means of limiting/amplitude compression. Post-demodulation processing reduces the level of continuous demodulator noise by base-band filtering/de-emphasis, and the level of impulse noise (click noise) by means of click detection and subsequent elimination. However, the latter type of reduction is not very effective. Frequency feedback and adaptive processing are more effective means to reduce the level of click noise, as long as the feedback/adaption is not disrupted by noise and disturbances.

The subsequent chapters consider the threshold behavior of 'conventional' demodulators, phase feedback demodulators, and frequency feedback demodulators. It is shown that in all types of demodulators, the output noise consists of a continuous noise component that behaves similarly in all three demodulators, and an impulsive component, which behaves differently and is responsible for the FM threshold.

Chapter 6 investigates application of finite amplitude compression in 'conventional' FM demodulators, instead of the usually applied infinite compression, as means to establish a trade-off between click noise and continuous noise. After an outline of the two types of amplitude compressors that can be distinguished, a newly developed model shows that finite amplitude compression increases the continuous output noise level, but at the same time reduces the click noise by reduction of the average click pulse area. A 'critical' level of compression is derived that should be exceeded in order to attain an output SNR that is at most 3 dB below the maximum possible SNR. Further, it is shown that second-order 'modulation' noise, due to 'modulation' of the compressor small-signal transfer by the noise, is minimized by a linear combination of infinite compression, and no compression. The optimum level of compression that maximizes the output signal-to-noise (SNR) ratio is derived as function of the input carrier-to-noise ratio (CNR). The theory is verified by simulations and measurements on an FM demodulator that employed a soft-limiter as (sub-optimal) amplitude compressor.

Chapter 7 considers the threshold behavior of phase feedback demodulators. Above its threshold the output SNR equals the maximum possible SNR of an FM demodulator. Further, the threshold, due to cycle-slip noise generally occurs at a considerably lower input CNR than in a conventional demodulator. A nonlinear analysis shows that the cycle-slip rate is highly dependent on the phase detector transfer and the structure of the loop filter. Generally, the steady-state phase error should be minimized, while the closed-loop transfer should not be larger than strictly necessary to accommodate the modulation. Complex poles in the closed-loop transfer should be avoided in order to avoid cycle-slip bursts. Further, an ideal integrator in combination with a direct feed-through seems close to optimum. The phase detector should not contain limiters at its input since this results in degradation of its transfer at low CNR.

Chapter 8 investigates the threshold behavior of frequency feedback receivers (FMFB) and dynamic tracking filters. Generally, the threshold of these demodulators occurs at a considerably lower CNR than the threshold of a conventional demodulator and is determined by threshold of the discriminator inside the loop and suppression of the discriminator input FM wave by the IF filter. The latter effect is due to feedback noise that increases the frequency deviation of the wave at the IF filter input in such a way that the bandwidth of this wave exceeds the bandwidth of the IF filter. This IF filter is the key element in FMFB design that determines a trade-off between threshold extension and improvement of the demodulator linearity. Further, the threshold of the discriminator inside the loop defines an upper bound on the FMFB threshold. Comparison of an example system with a comparable phase feedback demodulator showed that the phase feedback demodulator threshold is located close to the upper bound on the FMFB threshold. This is due to the fact that carrier suppression does not occur in such demodulators.

Chapter 9 presents the conclusions.

Samenvatting

Dit proefschrift beschrijft gestructureerde methode voor het ontwerp van frequentiedemodulatoren (FM) met een hoge kwaliteit en een hoge gevoeligheid. Deze methode verschaft tevens inzicht in de verschillende principes die beschikbaar zijn voor de constructie van dergelijke demodulatoren, en in de maatregelen die in de demodulator en ontvanger architectuur kunnen worden getroffen ter verbetering van de ontvangstkwaliteit. Deze systemen vinden o.a. toepassing in autoradio's en diverse andere draadloze communicatie systemen.

Een korte beschrijving van de geschiedenis van frequentie modulatie in Hoofdstuk 1 brengt aan het licht, dat een allesomvattend raamwerk, dat de verbanden tussen de verschillende typen demodulatoren aanelkaar en aan hun kwaliteit relateert, ontbreekt in de bestaande ontwerptheorie voor FM demodulatoren. Een dergelijk raamwerk, een classificatie, is onmisbaar in een gestructureerde ontwerpmethode. Verder is geconstateerd, dat het door theoretisch ingestelde wetenschappers verrichte werk vaak zodanig ver van de praktijk, uitgeoefend door elektronisch ontwerpers, ligt, dat er maar weinig gebruik van word gemaakt. Het doel van dit proefschrift is enerzijds een allesomvattend raamwerk voor het ontwerp van FM demodulatoren te ontwikkelen, en anderzijds de afstand tussen het theoretische werk en de praktijk van het elektronisch ontwerpen zo goed mogelijk te overbruggen.

Hoofstuk 2 beschrijft kort de hoofdkenmerken van FM transmissie en FM gemoduleerde signalen. Deze vormen de basis voor de beschouwingen in de daaropvolgende hoofdstukken van het proefschrift. Een quasi-stationaire benadering voor het frequentiespectrum van FM signalen, en de verbetering van de signaal-ruis verhouding, bewerkstelligd door kwadratische vervorming van het ingangsruisspectrum, worden in dit hoofdstuk behandeld.

Hoofdstuk 3 ontwikkelt een classificatie van alle mogelijke principes, beschikbaar voor het realizeren van FM demodulatie. Een kort overzicht van de principes van de gekozen aanpak maakt het grote belang van de eerdergenoemde classificatie daarin duidelijk. Het blijkt dat FM demodulatie in de directe zin, dus door detectie van de momentane frequentie van het FM signaal, onmogelijk is, omdat deze frequentie niet gerelateerd is aan de in het signaal opgeslagen energie. In plaats daarvan dient ten allen tijde een conversie naar AM of PM te worden gerealizeerd, gevolgd door AM demodulatie, danwel PM demodulatie. Verder blijkt in dit verband, dat alleen AM demodulatie in de directe zin is toegestaan, terwijl PM demodulatie zowel in directe zin, als in indirecte zin, dus door middel van PM-AM conversie en daaropvolgend AM demodulatie, mag worden toegepast. De resulterende classificatie omvat twee sub-classen gebaseerd op FM-AM conversie, vier sub-classen gebaseerd op de combinatie FM-PM conversie en directe PM demodulatie, en drie sub-classen gebaseerd op de combinatie van FM-PM conversie en indirecte PM demodulatie. Van de laatste drie sub-classen waren er twee niet eerder bekend. Deze zijn echter van weinig praktische betekenis. Alle typen FM demodulatoren die in de literatuur zijn aangetroffen kunnen eenvoudig in de ontwikkelde classificatie gerubriceerd worden.

Hoofdstuk 4 behandelt het ontwerp van de verschillende deelsystemen in FM demdulatoren. Een offset veroorzaakt door de draaggolffrequentie blijkt destructief te werken op het dynamisch bereik van de demodulator. Deze beperkt de maximale signaalslag, en verhoogt het ruisniveau. Een hoge kwaliteit kan daarom slechts bereikt worden indien het ontstaan van deze component wordt voorkomen. Dit is alleen mogelijk met demodulatoren gebaseerd op FM-AM conversie gevolgd door AM projectie-detectie, demodulatoren die FM-PM conversie bewerkstelligen met behulp van een vaste tijdvertraging, en demodulatoren die FM-PM conversie bewerkstelligen met behulp van fase-tegenkoppeling. Deze demodulatoren dienen dan tevens gebruik te maken van een directe conversie-ontvanger architectuur, of een 'banddoorlatende' FM-AM/FM-PM omzetter De distorsie dient geminimaliseerd te worden door een geschikt gekozen frequentiekarakteristiek van de omzetter.

Hoofdstuk 5 behandelt de verschillende maatregelen die in de FM ontvanger-architectuur kunnen worden getroffen ter verbetering van de ontvangstkwaliteit. In het algemeen zijn de kwaliteitsverbeterende signaalbewerkingen gebaseerd op veronderstellingen omtrend de eigenschappen van het ontvangen FM signaal. Wanneer deze veronderstellingen niet (meer) geldig zijn, bijvoorbeeld door de aanwezigheid van ruis en storingen, zullen de kwaliteitsverbeterende signaalbewerkingen met hoge waarschjnlijkheid veranderen in juist kwaliteitsverslechterende bewerkingen. Bewerkingen voorafgaand aan de eigenlijke demodulatie zijn gericht op het extraheren van het gewenste FM signaal uit de ontvangen frequentieband. Dit omvat scheiding in frequentie d.m.v. filtering, scheiding in fase d.m.v. fasevergrendeling, ter onderdrukking van co-channel interferentie, en scheiding in amplitude door middel van begrenzing/amplitude compressie. Post-demodulatie signaalbewerkingen zijn geënt op het reduceren van continue ruis in het demodulator uitgangssignaal, d.m.v. basisband filtering en de-emphasis, en reductie van impulsruis (clicks) d.m.v. click detectie en eliminatie. Dit laatste blikt echter niet effectief te kunnen worden gerealizeerd. Een veel effectievere ruis-reductie methode is het toepassen van frequentie-tegenkoppeling en/of adaptieve regeling, zolang deze niet verstoort worden door ruis en storingen.

De daaropvolgende hoofdstukken onderzoeken het gedrag van zogenaamde 'conventionele' demodulatoren, fase-tegenkoppelingsdemodulatoren en frequentie-tegenkoppelingsdemodulatoren rondom hun ruisdrempel. De uitgangsruis van al deze demodulatoren bestaat uit een continue component, die zich in alle drie de gevallen vrijwel hetzelfde gedraagt, en een impuls-component, verantwoordelijk voor de ruisdrempel, die zich verschillend gedraagt voor deze drie typen.

Hoofdstuk 6 onderzoekt de toepassing van eindige amplitude compression i.p.v. de gebruikelijke oneindige compressie, ter realizatie van een uitwisseling tussen continue ruis en impulsruis. Na een overzicht te hebben besproken van de twee onderscheidbare typen compressoren, wordt m.b.v. een nieuw ontwikkeld model aangetoond, dat eindige compressie het continue ruisniveau verhoogt, en tevens het impuls ruis-niveau verlaagt door een reductie van het gemiddelde impuls-oppervlak. Een 'kritisch' compressieniveau is afgeleid, dat dient te worden overschreden ter realizatie van een signaal-ruis verhouding (SNR) die ten hoogste 3 dB lager is dan de maximaal haalbare SNR. Verder is aangetoond, dat 'modulatieruis', veroorzaakt door modulatie van de klein-signaaloverdracht van de compressor, wordt geminimaliseerd door een lineare combinatie van oneindige compressie, en in het geheel geen compressie toe te passen. Het optimale compressieniveau, dat de SNR maximaliseert, is bepaald als functie van de ingangs CNR. De theorie is geverifiëerd door middel van simulaties en metingen aan een FM demodulator, voorafgegaan door een zachte begrenzer.

Hoofdstuk 7 onderzoekt het drempelgedrag van fase-tegenkoppelingsdemodulatoren. Boven de drempel realizeren deze demodulatoren de maximaal haalbare SNR. De drempel zelf, bepaald door 'cycle-slips', is over het algemeen gesitueerd op beduidend lagere een ingangs CNR dan bij een conventionele demodulator. Een niet-lineare analysemethode geeft aan, dat de cycle-slip frequentie in hoge mate afhankelijk is van de fasedetectoroverdracht en de structuur van het lusfilter. De statische fasefout dient geminimaliseerd te worden, terwijl de lusbandbreedte niet groter gekozen moet worden dan strict noodzakelijk is. Complexe polen in de gesloten lusoverdracht dienen vermeden te worden, ter voorkoming van cycle-slip 'bursts'. Een lusfilter bestaande uit een 'ideale' integrator en een directe overdracht lijkt de optimale configuratie dicht te benaderen. De fase-detector dient geen begrenzers aan beide ingangen te bevatten, ter voorkoming van degeneratie van de overdracht bij lage CNR's.

Hoofdstuk 8 onderzoekt het drempelgedrag van frequentietegenkoppelingsdemodulatoren (FMFB) en 'dynamic tracking filters'. De drempel van deze demodulatoren ligt ook beduidend lager dan die van conventionele demodulatoren, en wordt bepaald door de drempel van de 'discriminator' binnen de lus, en de onderdrukking van het ingangssignaal van deze discriminator door het middenfrequent filter. Dit laatste effect wordt veroorzaakt door via de tegenkoppeling teruggevoerde ruis, die de frequentiezwaai van het FM signaal aan de ingang van het filter zodanig vergroot, dat de bandbreedte ervan groter wordt dan die van het filter. Het middenfrequent filter is het centrale element in het ontwerp van FMFB ontvangers, die de uitwisseling tussen ruisdrempel en lineariteit bepaalt. De drempel van de FMFB kan niet lager zijn dan die van de discriminator binnen de lus. Vergelijking van een voorbeeld FMFB met een vergelijkbare fase-tegenkoppelingsdemodulator geeft aan, dat de laatste een drempel heeft in de buurt van de minimale drempel van de FMFB, dus in de praktijk lager dan de FMFB, doordat in deze demodulator geen onderdrukking door het middenfrequent filter plaats kan vinden.

Hoofdstuk 9 bespreekt de conclusies.

Acknowledgments

Although a Ph.D. study is basically a solitary task, it cannot be completed successfully without the support of many people. Therefore, I would like to express my gratitude to all the people that supported and encouraged me during the project in some way or another.

I am greatly indebted to my 'promotor' Prof.dr.ir. Arthur H.M. van Roermund and my 'toegevoegd promotor' Dr.ir. Chris J.M. Verhoeven for the many valuable discussions throughout the project, and the helpful comments on the manuscript of this thesis.

I would like to thank the people of Philips Research at Eindhoven, especially Dr.ing. W.G. Kasperkovitz and Ir. A. Sempel for the pleasant cooperation, and the inspiring discussions that helped me to improve the practical applicability of my research. Further, I am grateful to the Philips Company in general for its financial support and providing me the opportunity to perform my Ph.D. work in cooperation with one of the leading industrial laboratories.

Marc Joossen is acknowledged for his thorough and valuable study on the history of frequency modulation that supplied me with a wealth of information to write the introduction of this thesis.

Thanks go also to the staff of the Electronics Research Laboratory for the pleasant atmosphere. Special thanks go to Jan Nusteling, who patiently and indefatigably solved my numerous computer problems, to Rob van Beijnhem, who skillfully realized several parts of the measurement setups, to Rob Janse, who produced the vast majority of the drawings in this thesis, to Rob Otte and Jan Westra for their company, to Frank Kuijstermans for valuable discussions on limiters, and to Jan Mulder and Wouter Serdijn for pleasant cooperation and interesting discussions on noise in nonlinear circuits.

Mrs. Zaat-Jones and Mr. Simon North significantly improved the readability of this thesis by correcting my numerous linguistic errors.

I am also very grateful to my parents for their encouragement throughout my education. And last but not least, to Claudia, for her patience and understanding.

Biography

Michiel Kouwenhoven was born in Delft, the Netherlands, on the 8th of July, 1971. He started his studies in Electrical Engineering at the Delft University of Technology in 1989, and received the M.Sc. degree in 1993. Subsequently, he joined the Electronics Research Laboratory at the same university in order to prepare this Ph.D. thesis. He is currently employed as an assistant professor at this laboratory.