Jon Legarda

Feedforward Amplifiers for Wideband Communication Systems

Springer

FEEDFORWARD AMPLIFIERS FOR WIDEBAND COMMUNICATION SYSTEMS

Feedforward Amplifiers for Wideband Communication Systems

by

JON LEGARDA
CEIT and TECNUN (University of Navarra), San Sebastian, Spain

Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN-10 0-387-35137-X (HB)
ISBN-13 978-0-387-35137-7 (HB)
ISBN-10 0-387-35138-8 (e-book)
ISBN-13 978-0-387-35138-4 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

Dedication

This book is dedicated to my mom, dad, brothers and my girlfriend Oihane. Thank you all for everything.

Contents

Preface **xiii**

Acknowledgments **xvii**

1. Introduction **1**

1. Radio-Electric transmitters - Historical overview 1

1.1 Arc and spark transmitters 1

1.2 Multi-polar alternators 3

1.3 Thermoionic vacuum tubes 3

1.4 Discrete transistors 5

1.5 Integrated transistors 6

2. Digital communication systems 6

2.1 Analog signal 6

2.2 Discrete signal 7

2.3 Digital signal 7

2.4 Digital transmission 7

2.5 Electromagnetic spectrum 9

2.6 Modulation techniques 10

2.7 Signal degradation 12

2.7.1 Attenuation 12

2.7.2 Noise 13

2.7.3 Distortion 13

2.8 Basic architectures of wireless transmitters 14

2.8.1 Analog I/Q modulator vs. digital IF 15

2.8.2 Other implementations 15

3. Digital modulation 16
3.1 Applications 17
3.2 Phase Shift Keying 18
3.3 Frequency Shift Keying 19
3.4 Minimum Shift Keying 20
3.5 Quadrature Amplitude Modulation 20
3.5.1 I/Q offset modulation 22
3.5.2 Differential modulation 22
3.6 Broadband wireless access techniques 23

2. Nonlinear distortion 25
1. Harmonic distortion 27
1.1 N^{th} harmonic distortion coefficient 29
1.2 Global harmonic distortion 30
2. Intermodulation analysis 30
2.1 2 tone intermodulation 30
2.2 3 tone intermodulation 33
3. Cross modulation (X_{mod}) 34
3.1 Cross modulation (m<<1) 36
3.2 Cross modulation (m>>1) 37
4. Distortion measurement techniques 39
4.1 Harmonic distortion measurement 39
4.2 Third order distortion measurement 40
4.2.1 Two tone procedure 40
4.2.2 Third order intercept point (IP_3) 41
4.3 Second order distortion measurement 45
4.4 Cross modulation (X_{mod}) measurement 47
5. Measurements of wideband digital signals 49
5.1 Peak to average power ratio and CCDF curves 49
5.2 Modulation quality measurements 51
5.3 Code domain power 52
5.4 Adjacent Channel Leakage Ratio (ACLR) 53

3. RF power amplifiers 55
1. Classification of power amplifiers 56
2. Power amplifiers parameters 57
2.1 The efficiency rate 57
2.2 The back-off 58
2.3 Power utilization factor (PUF) 58
3. Class A 59
4. Class B 61
5. Class AB (outphasing) 63
6. Class C 65

7. Switching amplifiers 67
7.1 Class D 68
7.2 Class E 70
7.3 Class F 72
8. More operating modes 73

4. Linearization techniques 75
1. Classification of the linearization techniques 75
2. Feedback 76
2.1 RF Feedback 76
2.2 Envelope Feedback 77
2.3 Envelope and phase Feedback 78
2.4 Polar Loop 78
2.5 Cartesian Loop 79
3. Predistortion 80
3.1 RF Predistortion 81
3.1.1 Simple analog Predistortion 81
3.1.2 Compound Predistortion 83
3.2 Envelope Predistortion 83
3.3 Baseband Predistortion 84
4. Feedforward 86
5. Efficiency enhancement techniques 87
5.1 Bypassing 87
5.2 Envelope Elimination and Restoration 89
5.3 Bias Adaptation 90
5.4 LINC 90
5.5 Doherty Method 92
5.6 CALLUM 93
6. Comparison of linearization techniques 94

5. Feedforward amplifiers 97
1. Feedforward linearization technique 97
1.1 Frequency dependence 101
1.2 Amplitude and phase adjustments 105
1.3 Error amplifier distortion 107
1.4 Isolation lack 108
1.5 Main signal path loss 109
2. Feedforward in wideband communications systems 110

6. Implementation of Feedforward amplifiers 115
1. Simulation of the Feedforward architecture 116
1.1 Simulation models 116
1.2 Simulation of the error loop 117

1.3 Simulation of the distortion cancellation loop 119
1.4 The isolation effect 120
2. Selection of the error amplifier 122
3. The adjustment of the cancellation loops 124
3.1 Delay lines 124
3.1.1 Configuration 125
3.1.2 Length tolerances 125
3.1.3 Packaging 125
3.1.4 System design 126
3.2 An adjustment method for delay lines 126
4. Distortion enhanced measurement techniques 130
4.1 Spectrum analyzer mixer level optimization 131
4.1.1 Mixer level 131
4.1.2 Signal to noise ratio versus mixer level 132
4.1.3 Signal to noise ratio with external noise 134
4.1.4 Signal to distortion ratio versus mixer level 135
4.1.5 The dynamic range chart 135
4.1.6 Spectrum analyzer distortion 136
4.2 Enhanced ACLR measurements 137
4.2.1 Signal to noise ratio 137
4.2.2 Spectral regrowth 139
4.2.3 Phase noise influence 140
4.2.4 Dynamic range chart for wideband signals 141
5. Improvements of Feedforward amplifier 142
5.1.1 OIP_3 improvement 143
5.1.2 ACLR improvement 144

7. Adaptive Feedforward amplifiers 145
1. Adaptive adjusting methods for Feedforward amplifiers 145
1.1 Maximum cancellation method 146
1.1.1 Error signal minimization 146
1.1.2 Distortion cancellation 147
1.1.3 Maximum cancellation 147
1.2 Maximum output method 148
2. Distortion monitoring architectures 150
2.1 Signal correlation 150
2.2 Pilot signal detection 152
2.3 Power minimization 153
3. An output signal monitoring architecture 156
3.1 Switched RF receiver 156
3.2 Frequency down-conversion 158
3.3 Amplitude equalization 160

4. The adaptive Feedforward amplifier 162
5. Conclusions 165

References **167**

Index **175**

Preface

This work has been possible thanks to the research carried out throughout several years in the field of the linearization techniques applied to digital communication systems, particularly to those with high frequency-efficient modulation techniques.

The wireless telecommunications are more and more demanded for the Information Society. Such requirements are reflected as a great many of communication standards with specific coverage applications and, above all, with higher and higher data transmission rates.

The electromagnetic spectrum, nevertheless, is a scarce asset that can not be spread and despite the increasingly tendency to transmit in higher frequencies, the bandwidths assigned to each application are always exploited to the limit.

This context merges in the need of developing frequency-efficient modulations with widespread codification techniques that result in wideband communication systems, with strict regulations in the usable frequency bandwidths and tight restrictions in the spurious emissions over the remaining spectrum.

The radio frequency transmitters do not remain impassive to those changes, especially the power amplifiers, which efficiency and linearity directly determines the correct performance of the entire transmission system. The linearity specifications are commonly fixed by the telecommunication standards while the efficiency rates directly strikes the commercial viability of those devices.

This work tries to put into practice the Feedforward linearization technique, aimed at improving either the linearity or efficiency parameters of power amplifiers, just intended for achieving a trade-off between the

distortion specifications of the telecommunication standards and the efficiency enhancements of the transmission systems, which set, respectively, the linearity and the output level requirements of power amplifiers.

The introduction of this book tries to summarize the main characteristics of the actual context of wideband communication systems. After an historical overview, the digital communication systems and the digital modulation techniques are treated flippantly.

One the most important consequences of using high frequency-efficient modulation techniques is the nonlinear distortion that the power amplifiers introduce in the last stage of the base station transmitters. In Chapter 2 those effects are analyzed and the basic distortion measurement techniques are described.

Chapter 3 is a revision of the most important radio frequency (RF) power amplification techniques. At this point is introduced the incompatibility between the linearity and efficiency performances of any amplification device and the need of developing special circuit architectures as alternative solutions.

The linearization techniques were the solution of the scientific community to the enhancement of the amplification devices, according to both linearity and efficiency improvements. The aim of Chapter 4 lies in the compilation and description of the linearization techniques used in power amplification, emphasizing their capacities while differentiating their suitable applications.

Chapter 5 is entirely dedicated to one of those techniques, the Feedforward linearization technique, as it is the only one that is able to achieve outstanding distortion cancellation levels for wideband signals.

The challenge of this work stems from the real implementation of this linearization technique for a particular telecommunication standard (WCDMA). In Chapter 6 some useful design guidelines are proposed in order to fabricate a Feedforward amplifier prototype. Such design considerations are supported with enhanced distortion measurements carried out with a high performance spectrum analyzer.

Finally, in Chapter 7, a novel adaptive control system is introduced. The *maximum output* control method proposed in this work tries to achieve the desired trade-off between the linearity and efficiency rates according to any distortion specification fixed as an adjacent channel leakage ratio (ACLR).

This adjusting method is applied thanks to an alternative distortion monitoring system that is able to measure with enough truthfulness the output signal of any Feedforward system.

This work is enshrined in one of the present wideband communication systems, but it must be taken into account that all the recommended design guidelines are perfectly reusable in any future wideband application.

Jon Legarda

Acknowledgments

First of all I would like to express my gratitude to all those workmates that have been got involved in such ambitious project. So many years of research can not be summarised in these few lines, and the participation of each person is practically indivisible. Sincerely, thanks you all.

Obviously, I am so grateful for the support that all the colleagues of the Electronic and Communication Department of the CEIT research institute have given me throughout these five years. Their priceless company and help have given me the strength to carry on. I would like to recognize, likewise, the work of the Engineering School of TECNUN (University of Navarra) on my education thorough so many years.

In particular, I would like to make mention of those people who have helped in the translation and review of the manuscript, from its conception to completion: David Braun-Friedman and José Manuel Torres Martínez with the contribution of Oscar Pinilla Echevarría.

This work has been developed under the auspices of the SARETEK – Basque Science, Technology & Innovation Network, in collaboration with IKUSI Angel Iglesias S.A.

Finally, I would like to close these lines with a special mention of all my family members, *ama*, *aita*, *Haritz eta Eñaut* and, of course, my fellow traveller *Oihane*. I cannot find words to express my gratitude to all of you…

Thanks so much

Chapter 1

INTRODUCTION

1. RADIO-ELECTRIC TRANSMITTERS - HISTORICAL OVERVIEW

The historical evolution of the radio-electric transmitters could be separated into five different eras: the arc and spark transmitters, the multi-polar alternators, the thermoionic vacuum tubes, the discrete transistors, the integrated transistors.

All of them are summarized in the next sections [1].

1.1 Arc and spark transmitters

In the beginning of radio communications (from 1895 up to the first half of 1920), the radio-electric waves were produced by sparks, electric arcs or high frequency (HF) alternators (*Figure 1-1*).

The first spark transmitter was based on a capacitor (C), charged with the voltage of a multi-vibrator which was activated by the battery of an alternating current (AC) transformer. Once the capacitor breakdown voltage was reached, an electric spark was spontaneously generated, damped and oscillating due to the inductor effect of a LRC resonant tank. This process was repeated successively and the resultant train of waves were guided through an antenna circuit, facilitating the generation of the electromagnetic waves.

Those transmitters were relatively cheap and able to achieve moderate power levels from 500W to 5kW for low (LF) and middle frequency (MF) bandwidths.

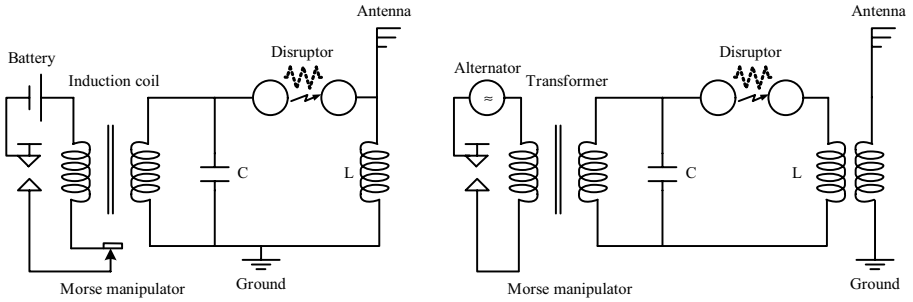


Figure 1-1. Battery and alternator spark transmitters

If an alternator was used as generator, the spark should have been created for the maximum amplitude value. In any other case, the circuit behaved as impedance, with no oscillating performance and without radiating.

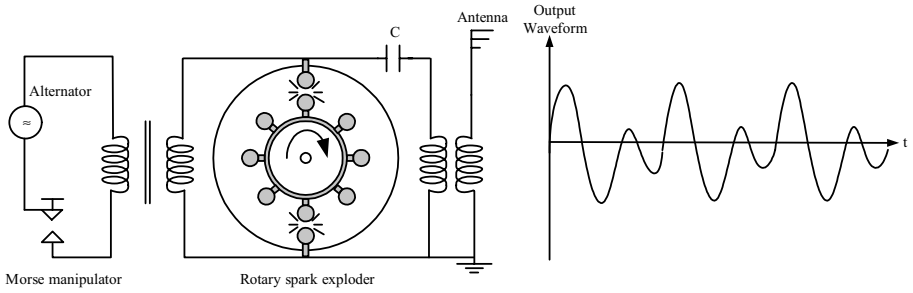


Figure 1-2. A telegraphy transmitter with a rotary exploder

The invention of the rotary spark exploder meant an outstanding improvement for these transmitters. It consisted in a revolving wheel placed between concentric electrodes and synchronized with the same frequency of the alternating voltage that loaded the capacitor. This way, the spark achieved was more regular and the wave train more sustained (*Figure 1-2*). The telegraphy transmitter used by the Titanic in 1912 was based on this technology.

Contemporary with the spark transmitters, the voltaic arc technology was the next stage in the evolution of the radio-electric transmitters. This method was based on the *negative dynamic resistance* of an electric arc discharge under specific environmental conditions (hydrogen). As a result, the wave was quite similar to a sustained wave; in fact it was the first attempt of voice transmission based on this technology with up to 1MW transmission powers.

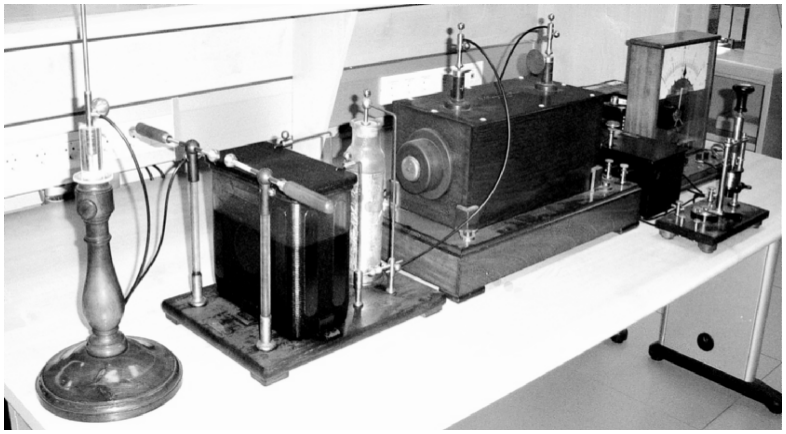


Figure 1-3. "Ducretet" spark transmitter, similar to the one used in the first wireless telegraphy transmission from the Eiffel Tower (1898). Restored in the RF laboratories of TECNUN

1.2 Multi-polar alternators

Finally, the multi-polar alternators were used in order to generate high frequencies with reasonable rotational speeds, although extremely low comparing with the actual frequencies.

The frequency range was limited to tens of kHz. Nevertheless, the power level achieved was considerable high (around 200kW) and the signal quality outstanding. Nowadays, an operative device (SAQ) is switched on once a year in Grimeton (Sweden) for amateurs delight (*Figure 1-4*).

1.3 Thermoionic vacuum tubes

The development of radio communications changed radically with the emergence of the vacuum triodes in 1909. The electrical performance of a triode is quite similar to a FET transistor, except the extra energy required during the cathode heating.

The frequency ranges were increased, the wave qualities enhanced and the transmission was made with continuous waves (CW). The radio communications had just begun its *electronic age*.

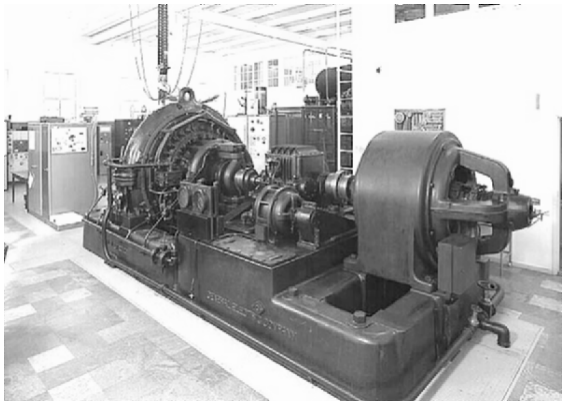


Figure 1-4. A multi-polar alternator from the Grimeton radio station (Sweden) 200kW – 17,5KHz

The RF power amplifier classification and the first impedance matching circuits come from this period. The vacuum tubes have kept their supremacy from 1920 until late 70's.

The development of the semiconductor technology has gradually displaced the vacuum techniques, with the exception of the very high power and high frequency applications. It is easy to deduce, by just looking closely at *Figure 1-5* and *Figure 1-6* that thermoionic vacuum tubes are better than semiconductor devices for those frequency and power rates.

Figure 1-5. Vacuum triodes for 150MHz and 200kW, progressive wave tube of 10GHz and 12kW and magnetron for L band with 4.800kW pulse power (Toshiba)

When the power requirements increase above 10kW, this technology becomes necessary. For example, applications like terrestrial broadcasting transmitters, satellite transponders, space communications, radar and industrial heating with pulsating waves (PW).

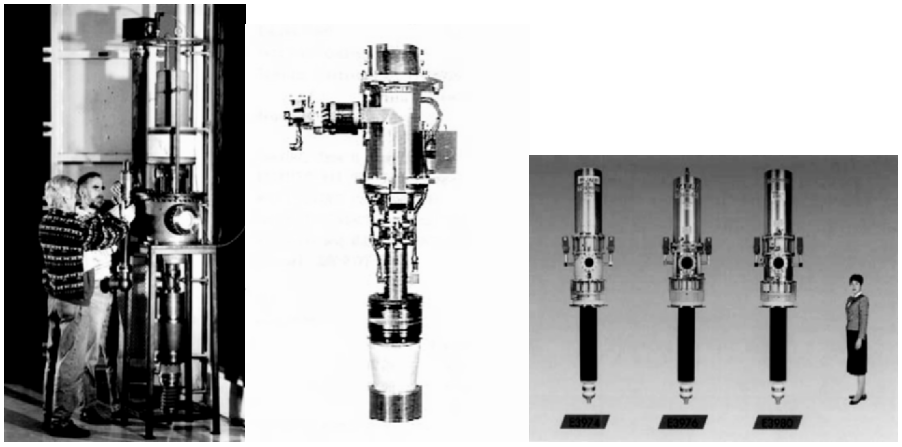


Figure 1-6. Klystrons (2.5GHz and 1.4MW), PW (10GHz and up to 100GW) and Gyrotrons

1.4 Discrete transistors

The first discrete transistors were the bipolar ones and appeared in the 60's. However, until 1970 they only were able to hardly achieve the ultra high frequencies (UHF).

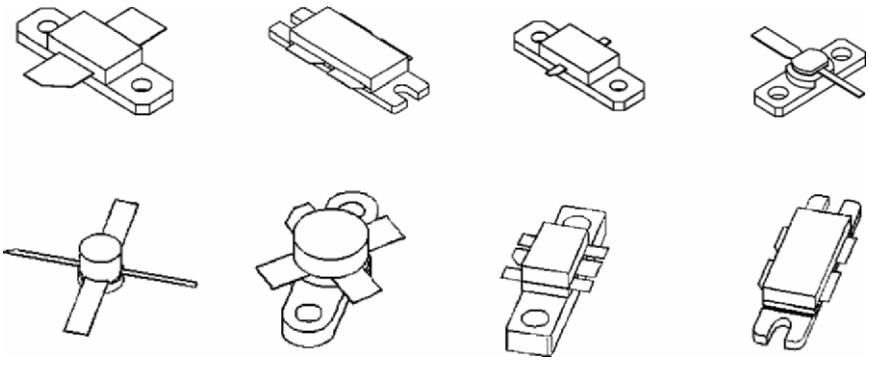


Figure 1-7. RF and microwave transistors used by Motorola (above) and Philips (below)

Throughout three decades the semiconductor amplifiers have increased both frequency and power specifications up to tens of GHz and hundreds of watts. Some pictures can be seen in the *Figure 1-8*.

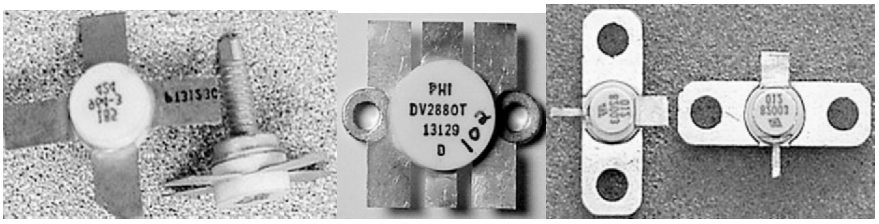


Figure 1-8. Some commercial semiconductor amplifiers

1.5 Integrated transistors

Since the beginning of 90's the new fabrication technologies based on new materials like InP, SiC, GaN,… have allowed experimental applications up to hundreds of GHz.

Those devices work with relatively low voltage and most of them are just available as integrated technology and not as discrete components. At the outside, some of them are introduced as standard functional modules, characterized as two-port circuits.

In addition, the RF engineers can also make their own designs based on the technology provided by each foundry company. Sophisticated software tools are used for circuit simulation and layout design. Once the first prototype is done, high precision test processes allow new redesigning stages until the specified requirements are fulfilled. Nevertheless, this option is sometimes rejected due to the high costs derived from the minimum fabrication units fixed by foundries and the redesigning time delay of each IC run.

2. DIGITAL COMMUNICATION SYSTEMS

According to the International Telecommunication Union (ITU), three signal types are distinguished corresponding to one of the classification used in the transmission systems.

2.1 Analog signal

An analog signal, or one of its inherent physical characteristic, continuously follows the fluctuations of the physic magnitude that represents the information to be transmitted, e.g. the output voltage of a microphone. The information is transmitted exactly and continuously.

2.2 Discrete signal

Despite the fact that the information to be transmitted changes continuously, it can be sampled in discrete values before the transmission. The transmission frequency of the discrete signal must be, at least, equal to the sample frequency in order to ensure the transmission of all the discrete values. Obviously, some information is lost during the sampling process but, it is considered acceptable when the transmitted information varies slowly in comparison with the sampling rate.

2.3 Digital signal

A digital signal itself is a discrete signal but with the information represented with predefined discrete values. The transmitted information is not continuous (discrete signal) and the transmitted values are not the real samples as they are approximated to some predefined fixed values (digital signal).

The nature of the transmitted signal gives the most important classification of the transmission systems: analog (for the first two cases) and digital (in the third case).

Thorough this section some basic concepts are summarized as far as the digital transmission systems are concerned: the signal digitalization, the electromagnetic spectrum, a signal modulation overview, the signal degradation through the transmission path and finally some basic transmission architectures.

2.4 Digital transmission

In digital systems the information, or message, to be transmitted could be discrete by itself or could be a continuous function.

In this second case, the analog function must be digitalized sampling the analog magnitude with a predefined sampling rate. In addition, a numeric value is assigned to each sampled value among a predefined number set. Each numeric value will be as similar as possible to the measured value, but this point forward the transmitted signal will never match exactly with the original message. In other words, the original analog signal is quantified and digitalized.

As the amount of the predefined numbers increase, the accuracy of the signal digitalization is better, i.e., the fidelity with the analog signal will be higher. This digitalization process is carried out by analog to digital converters (ADC).

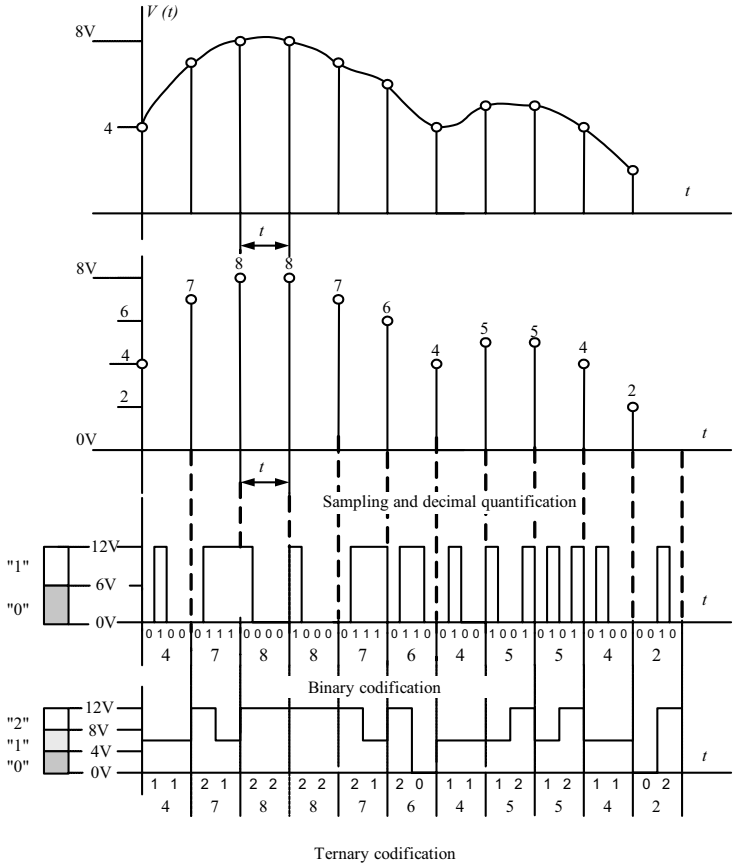


Figure 1-9. Signal digitalization before transmission

On the other hand, any numeric value can be represented in different basis or numeric systems. Usually, in digital communications, the binary system (base 2) is the most used, characterized by two symbols, “1” and “0”. Once the message is digitalized, different codification techniques can be used in order to protect the transmitted information against the interferences during the transmission.

Each codification type gives a higher or lower error protection at the expense of increasing or reducing the transmitted code. Therefore, the more robust the code is, the longer it will be.

In a binary transmission system, the “0” symbol is assigned to all the voltage values lower than a predefined value and the “1” symbol to all the voltage values above the same predefined value (6V in the example of *Figure 1-9*). This way, the transmitted information is a succession of electronic zeros and ones called pulses or bits. It does no matter the real

value of those bits; they only must be clearly identified as “0” or “1” just before the reception.

In the receptor, once the digital symbols are rebuilt and decrypted, it is possible to remake the original signal from the numeric values obtained from each symbol, correctly assembled in time and interpolated for the intermediate values. This process is done with analog to digital converters (ADC), inversely to done with the DAC.

2.5 Electromagnetic spectrum

The exceptional propagation characteristics of any of the electromagnetic energy forms turn them into a candidate for transmitting information.

The transmission channels currently most used, either guided or radiated, have an electromagnetic nature, including light and other radiations. Therefore, the messages must be converted into electromagnetic signals before the transmission.

One of the main attributes of an electromagnetic signal is its frequency. All frequencies are grouped in an electromagnetic spectrum (*Figure 1-10*), covering from subsonic frequencies up to cosmic rays. Subdivided in frequency bands, each one has unique properties for the signal transmission of a specific application.

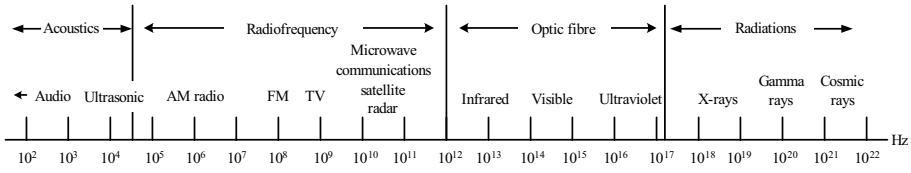


Figure 1-10. The electromagnetic spectrum

The availability of the frequency bands within the electromagnetic spectrum is a scarce asset, carefully administrated and perfectly regulated the use and assignments to any service. The radio spectrum frequency is divided in sub-bands, designated by the International Radio Consultative Committee (CCIR). Those bands are named and numbered so they can easily be assigned to specific services such as satellite, nautical communications and many others (*Table 1-1*).

Table 1-1. CCIR table for the radio-electric spectrum frequencies

Band number	Frequency ranges	Designation	λ range
2	30-300Hz	**ELF** (Extra Low Frequency).	
3	300Hz-3KHz	**VF** (Voice Frequency).	
4	3KHz-30KHz	**VLF** (Very Low Frequency).	
5	30KHz-300Hz	**LF** (Low Frequency).	Kilometric (LW)
6	300KHz-3MHz	**MF** (Medium Frequency).	Hectometric (MW)
7	3MHz-30MHz	**HF** (High Frequency).	Decametric (SW)
8	30MHz-300MHz	**VHF** (Very High Frequency).	Metric
9	300MHz-3GHz	**UHF** (Ultra High Frequency).	Decimetric
10	3GHz-30GHz	**SHF** (Super High Frequency).	Centimetric
11	30GHz-300GHz	**EHF** (Extra High Frequency).	Millimetric
12	300GHz-3THz	Infrared	
13	3THz-30THz	Infrared	
14	30THz-300THz	Infrared	
15	300THz-3PHz	Visible Light	
16	3PHz-30PHz	Ultraviolet	
17	30PHz-300PHz	X-rays	
18	300PHz-3EHz	Gamma-rays	
19	0.3 - 30EHz	Cosmic-rays	

2.6 Modulation techniques

It has been seen that the messages are first digitalized and then transmitted through the electromagnetic paths. However, there is an intermediate stage, when the digital signals are adapted to such electromagnetic path with a signal carrier that guarantees a correct and secure transmission.

This process, known as modulation, is based on the systematic modification of a single carrier according to the original message, being reversible in order to recover the information in the reception.

The modulation can be continuous or pulsed. The continuous modulation uses phase (Φ), frequency (ω) or amplitude (V) variations of a sinusoidal carrier (1.1).

$$v = V\cos(\omega t + \Phi) \tag{1.1}$$

According to which of the parameter is used, three modulation types are distinguished:

- V: amplitude modulation (AM)
- ω: frequency modulation (FM)
- ϕ: phase modulation (PM)

In *Figure 1-11* those modulations are displayed regarding to an arbitrary modulator signal (Mod).

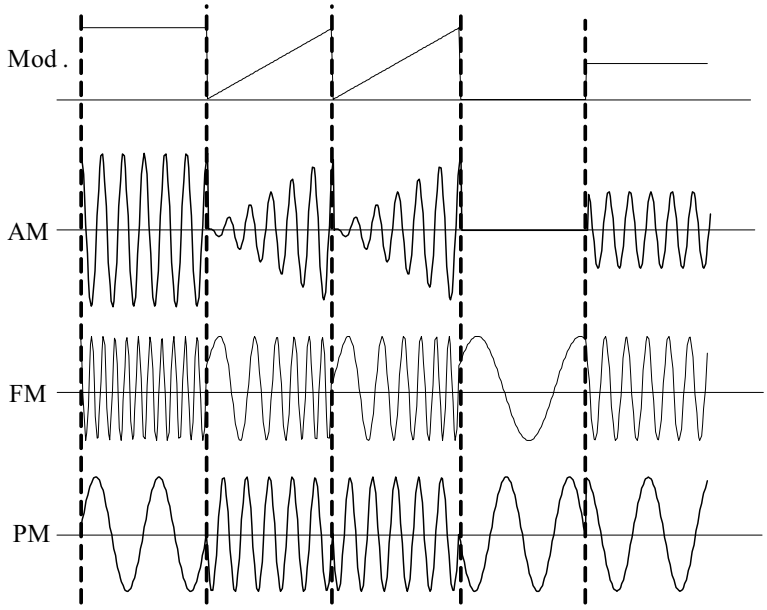


Figure 1-11. Classic continuous modulation types

The modulation is directly related to the carrier signal and the access to the transmission path. Each physical parameter is sensitive to different effects during transmission and those properties can be properly exploited.

As far as the pulse modulation is concerned pulse trains are used as carriers signals. The pulse maximum period must be lower than the sample rate, otherwise, more than one sample could occur in the same pulse gap and each pulse will have the information of more than one sample.

The previous condition leads to conclude that the carrier frequency must be, at least, higher than the sample rate. If redundant information is received at the end of the transmission path, it will not be critical for the smooth running of the transmission system.

According to which pulse parameter is changed –height, width or position– three modulation types are distinguished:

- Pulse amplitude modulation (PAM)
- Pulse width modulation (PWM)
- Pulse position modulation (PPM)

Figure 1-12 shows those modulation techniques regarding to a triangular modulation signal.

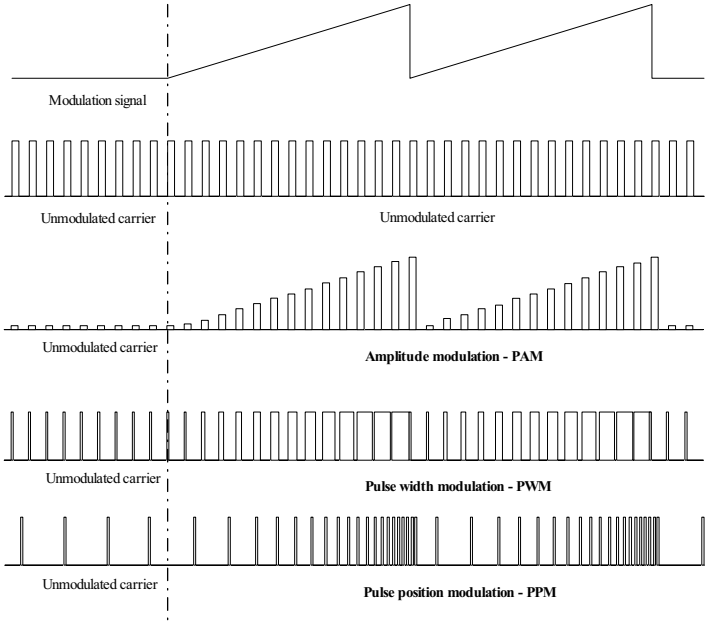


Figure 1-12. Classic pulse modulation types

2.7 Signal degradation

The signal degradation is inherent to any transmission process that manifests as attenuation, noise or distortion effects. Attenuation and noise are mostly due to the propagation path while the distortion mainly to the signal amplification.

2.7.1 Attenuation

Whatever medium access is used, there is a signal power loss attached to the signal propagation, commonly named as attenuation. This loss can be compensated with a previous amplification; however, an excessive amplification introduces some detrimental effects and increases some negative phenomena of the transmission channel.

The signal attenuation is the result of two fundamental effects. On the one hand, the interaction with the transmission path as energy dissipation, like the Joule effect in the telephonic wired transmission.

On the other hand, the energy is spread over the medium with the distance (d) between the transmitter and the receiver link. In the radiated

transmission the energy is spread through the surface of an ideal sphere with the radius equal to *d*.

2.7.2 Noise

Both the original message and the carrier signal are exposed to the undesired random addition of interference signals during the transmission, making the information reconstruction difficult. This effect is the most problematic among all the signal degradation effects, and can be "natural" or "human" by nature.

The natural noise source is a constant, random and intrinsic feature of the nature, hence it is found in all the transmissions channels. The noise can masquerade, totally or partially, the transmitted signal and it can not be totally subtracted, so it represents one of the basic challenges for the transmission techniques.

On the other hand, the external noise, also known as interference, artificial or human noise, is formed with signals that coexist in the transmission channels shared by the human activities, e.g., the radio remote transmissions, the opening and closing transitions of the telecommunication switches, radio spurious emissions, etc. The human noise is fought in its origin by the international regulations that fix the specifications of those spurious emissions [2]. If the interference origin is inaccessible, the receivers should be more selective just for the interference inside the receiver.

2.7.3 Distortion

Distortion is the signal deformation of the original message excepting the scale and delay factors suffered thorough the transmission process. In other words, the received signal is not a scaled and delayed replica of the original signal.

The distortion is produced by the transmitted signal, so if the signal disappears the distortion will disappear too, while the noise stays on.

All the distortion effects are not equally harmful. For example, the linear distortion can be totally or partially subtracted while the cross modulation does not. Nevertheless, any distortion can be avoided, or at least restricted, during the designing and dimensioning stages within acceptable ranges for a correct performance.

The Chapter 2 is entirely dedicated to the nonlinear distortion.

2.8 Basic architectures of wireless transmitters

The performance of a wireless communications system depends on the transmitter, the receiver and the air interface over which the communication takes place. This subsection shows how a digital communication transmitter works and its basic architectures are discussed [3].

Figure 1-13 shows a simplified block diagram of a digital communication transmitter that uses an I/Q modulation, commonly used in the high performance transmitters that are seen in section 3. Previously speech coding (assuming voice transmission), channel coding, and interleaving stages are carried out. Speech coding quantizes or digitalizes the analog signal and applies compression techniques to minimize the data rate and to increase the spectral efficiency. On the other hand, the channel coding and the interleaving are common techniques that provide error protection.

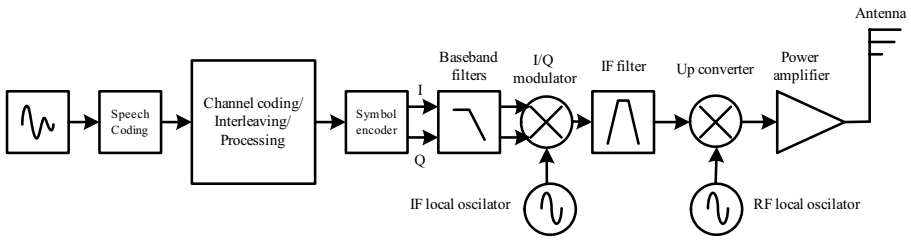


Figure 1-13. The block diagram of a digital communication transmitter

The symbol encoder translates the bit stream into the I/Q baseband signals and the symbol clock defines the frequency and the timing of the transmission of the individual symbols.

Once the I/Q baseband signals have been generated, they are filtered, smoothing the fast transitions between “0” and “1” and limiting the transmitted bandwidth. In order to improve the overall performance of the transmission system, filtering is often shared between the transmitter and the receiver, so the filters must be compatible and carefully implemented.

The filtered I/Q signals are fed into an I/Q modulator controlled by a local oscillator (LO) that can operate at either intermediate (IF) or directly at radio frequencies (RF).

Anyway, the output of the modulator is the combination of two orthogonal signals that must be up-converted to the transmission frequency.

Finally, each RF signal is often combined with other channels before being amplified by the power amplifier and transmitted by the antenna circuit.

2.8.1 Analog I/Q modulator vs. digital IF

Although digital communication transmitters can be designed using analog hardware, there is a clear trend toward the digital implementation. Therefore, the location of the digital to analog converter (DAC) varies as the digital system architectures vary.

For instance, despite baseband filters are usually implemented digitally with finite impulse response (FIR) filters, the I/Q modulator has traditionally been designed using analog hardware. Anyway in the last few years, an increasing tendency has lead to the digital implementations.

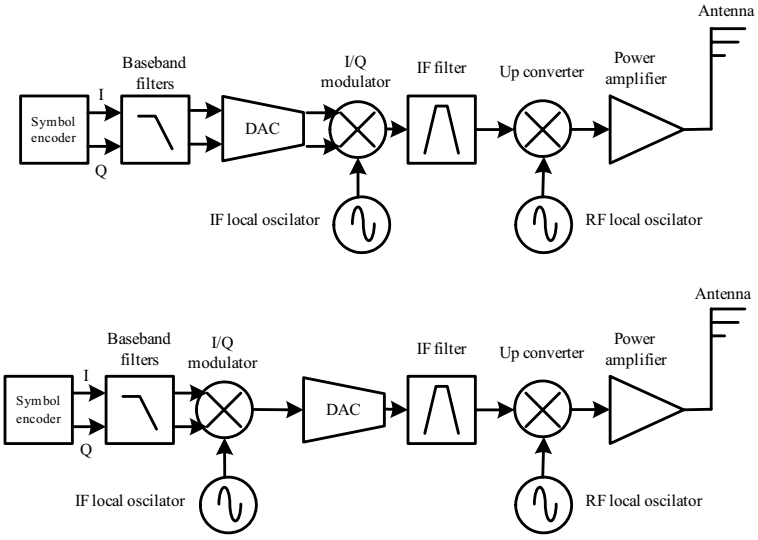


Figure 1-14. Typical analog I/Q (above) and digital IF (below) modulators

The baseband and the IF digital implementations avoid some of the typical impairments caused by the analog hardware like component aging.

2.8.2 Other implementations

In practice, there are many variations of the general block diagrams discussed above. Those variations depend mainly on the characteristics of the technology used; for example, the multiplexing technique (TDMA or CDMA) and the modulation scheme (such as OQPSK or GMSK).

For instance, GSM transmitters can be implemented using analog frequency modulators (*Figure 1-13*). Since intersymbol interference (ISI) is not as critical in GSM systems, Gaussian baseband filtering is used instead of the Nyquist filters [4].

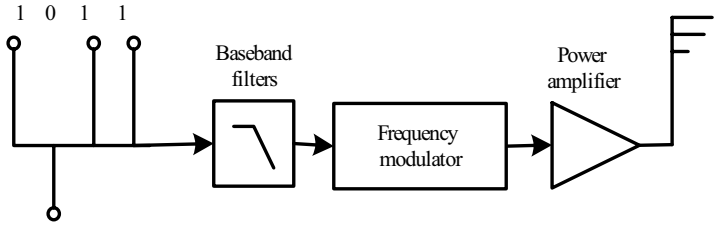


Figure 1-15. Block diagram of a GMSK transmitter with a frequency modulator

3. DIGITAL MODULATION

The amplitude (AM) and phase (PM) modulations are two continuous modulation techniques that can be modulated separately or simultaneously. Instead, the original signal can be separated into two independent components – I (Inphase) and Q (Quadrature) – that are orthogonal and do not interfere between them.

A simple way to view the amplitude and phase modulation is with the polar diagram of *Figure 1-16*. The carrier signal becomes a frequency and phase reference and the original signal is interpreted relative to it.

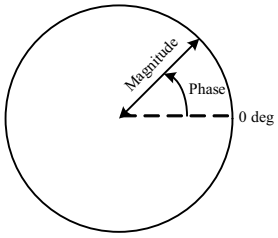


Figure 1-16. Polar diagram

Polar diagrams are the basis of many displays used in digital communications, where the signal vector is commonly described by its rectangular I/Q coordinates (*Figure 1-17*) resulting in a rectangular representation of the polar diagram. The projection of the signal vector onto the I axis is the "I" component and the projection onto the Q axis is the "Q" component.

Signals shifted by 90º, known as orthogonal or in quadrature, do not interfere between them, so they can be added up to form the output signal.

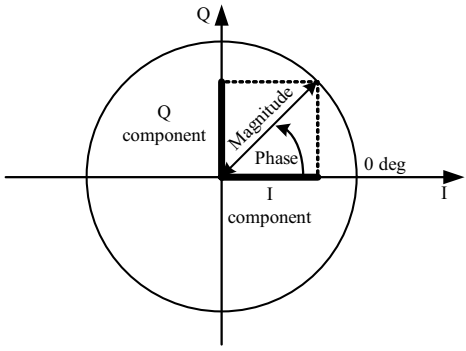


Figure 1-17. I/Q components over the polar diagram

The main advantage of the I/Q modulation is the symmetric ease of combining independent signal components into a single composite signal and later splitting it into its independent parts back.

This section covers the main digital modulation formats, their applications, the relative spectral efficiencies, and some variations of the main modulation types used in practical systems [4].

3.1 Applications

The *Table 1-2* covers some applications for different modulation techniques. Although this section is focused exclusively on wireless communications, video applications have also been included because of their similarity to other wireless communications.

In order to understand and compare the efficiencies of different modulation techniques, it is important to understand the difference between the “bit rate” and the “symbol rate” terms, because the signal bandwidth depends on the symbol rate and not on the bit rate.

The bit rate is the frequency of a bit stream. For example, a radio transmitter with an 8 bit sampler, sampling at 10kHz (voice). The bit rate, the basic bit stream rate in the radio, would be eight bits multiplied by 10.000 samples per second (80Kbit/s).

The symbol rate is the bit rate divided by the number of bits that can be transmitted with each symbol. If one bit is transmitted per symbol, then the symbol rate would be the same as the bit rate (80Kbit/s). However, if two bits are transmitted per symbol, then the symbol rate would be half of the bit rate (40Kbit/s). The more bits are sent with each symbol, the less bandwidth is used for the same transmitted data.

Table 1-2. Modulations vs. applications

Modulation	Application
MSK, GMSK	GSM, CDPD
QPSK, π/4 DQPSK	Satellite, CDMA, NADC, TETRA, PHS, PDC, LMDS, DVB-S, cable modems, TFTS
OQPSK	CDMA satellite
FSK, GFSK	DECT, RAM mobile data, AMPS, CT2, ERMES, land mobile, public safety
8, 16VSB	North American digitaly TV (ATV), broadcast, cable
8PSK	Satellite, aircraft, telemetry pilots for monitoring video systems
16QAM	Microwave digital radio, modems, DVB-C, DVB-T
32QAM	Terrestrial microwave, DVB-T
64QAM	DVB-C, modems, broadband set top boxes, MMDS
256QAM	Modems, DVB-C (Europe), Digital Video (US)

Therefore, the modulation formats that are able to use high number of symbol rates are more frequency efficient. The *Figure 1-18* shows an example. In the eight-state Phase Shift Keying (8PSK), there are eight possible states. Since $8 = 2^3$, there are three bits per symbol, so the symbol rate is one third of the bit rate.

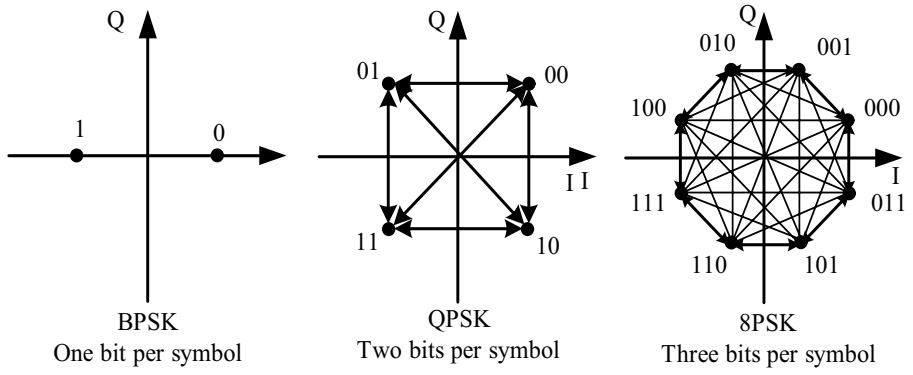


Figure 1-18. Bit rate and symbol rates

3.2 Phase Shift Keying

One of the simplest forms of digital modulation is the binary shift keying (BPSK). The phase of a constant amplitude carrier changes between 0º and

180º (*Figure 2-17*). On the I/Q diagram (*Figure 1-18*) the I component has only two different values, so the symbol rate is one bit per symbol.

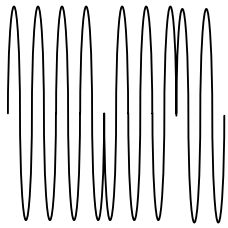


Figure 1-19. Phase shift keying (PSK)

A more common phase modulation is the quadrature phase shift keying (QPSK). It is used extensively in applications like CDMA (code division multiple access) cellular services, wireless local loops, Iridium and DVB-S (digital video broadcasting-satellite).

Two I values and two Q values are possible, so four states can be formed. Since $4 = 2^2$, the symbol rate is the half of the bit rate.

3.3 Frequency Shift Keying

Frequency modulation and phase modulation are closely related. A constant frequency shift of 1Hz means that the phase is constantly advancing at 360º per second (2π rad/sec).

Frequency shift keying (FSK) is used in many applications like wireless and paging systems, including DECT (digital enhanced cordless telephone) and CT2 (cordless telephone 2).

In FSK, the carrier frequency is changed according to the modulating signal while the amplitude remains constant. Particularly, in binary FSK (BFSK or 2FSK), “1” is represented by one frequency and “0” is represented by another frequency.

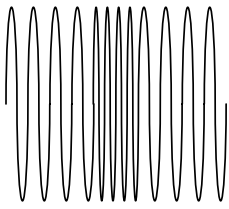


Figure 1-20. Frequency shift keying (FSK)

3.4 Minimum Shift Keying

Since a frequency shift produces an advancing or retarding phase shift, it can be detected by sampling the phase at each symbol period. Phase shifts of (2N+1) $\pi/2$ radians are easily detected with an I/Q demodulator. The polarity of the I channel at even numbered symbols and the polarity of the Q channel at odd numbered symbols conveys the transmitted data. This orthogonality between I and Q simplifies detection algorithms and reduces the power consumption in the mobile receiver.

FSK with a 90º deviation is called MSK (Minimum Shift Keying), used in the GSM (Global System for Mobile Communications) cellular standard. A phase shift of +90º represents a "1" and –90º represents a "0".

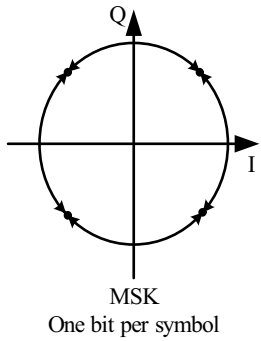


Figure 1-21. Minimum shift keying (MSK)

FSK and MSK produce constant envelope carrier signals without amplitude variations, so nonlinear power amplifiers could be used with high efficiency rates.

3.5 Quadrature Amplitude Modulation

The quadrature amplitude modulation (QAM) is used in applications like microwave digital radio, DVB-C (Digital Video Broadcasting - Cable), and modems.

In 16-state QAM (16QAM), there are four I values and four Q values, so 16 states are possible. Since $16 = 2^4$ four bits per symbol can be sent. The symbol rate is one fourth of the bit rate, so this modulation format is more spectrally efficient than any other seen before (BPSK, QPSK, 8PSK or MSK). Note that QPSK is the same as 4QAM.

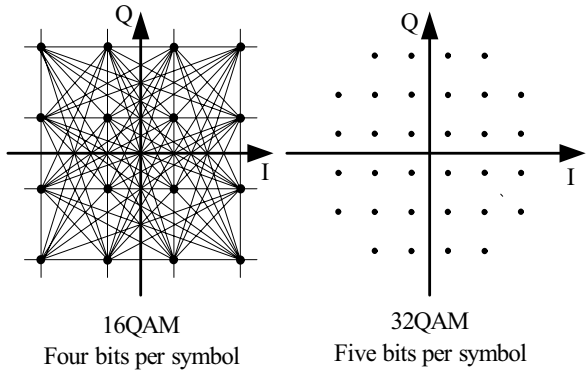


Figure 1-22. Quadrature amplitude modulation (QAM)

Currently, the implementation limits are approximately in 256QAM, though are extending up to 512 or 1024QAM.

A 256QAM system uses 16 I values and 16 Q values, giving 256 possible states. Since $256 = 2^8$, each symbol will represent eight bits. However, symbols are very close to each other and thus more subjected to errors due to noise and distortion effects. That is the reason why those modulations may have to be transmitted with high linear amplifiers, reducing the power efficiency in comparison with other modulation techniques.

In any digital modulation system, if the input signal is distorted or severely attenuated, the receiver will eventually lose the symbol lock completely and it will not be able to demodulate the signal or recover any information.

If the distortion or attenuation effects are moderate, the symbol clock will be recovered but the reception quality will be noisy. In some cases, a symbol will fall far away from its intended position and it will cross over to an adjacent position. The I/Q level detectors would misinterpret such symbol as being in the wrong location, causing undesired bit errors. Hence, QPSK is not as frequency efficient as QAM but it is more power efficient.

The modulation types outlined before form the building blocks for many transmission systems. There are three main variations on these basic building blocks that are used in communications systems: the I/Q offset modulation, the differential modulation, and the constant envelope modulation.

3.5.1 I/Q offset modulation

The first variation over a digital modulation is the offset modulation, e.g. the Offset QPSK (OQPSK) used in cellular CDMA uplink transmissions.

In QPSK, the I/Q bit streams are switched simultaneously, so the symbol clocks are correctly synchronized. In OQPSK, the I/Q bit streams are offset in their relative alignment by one bit period (one half of a symbol period), as shown in *Figure 1-23*. Since the transitions of I and Q are offset, at any given time only one of the two bit streams changes.

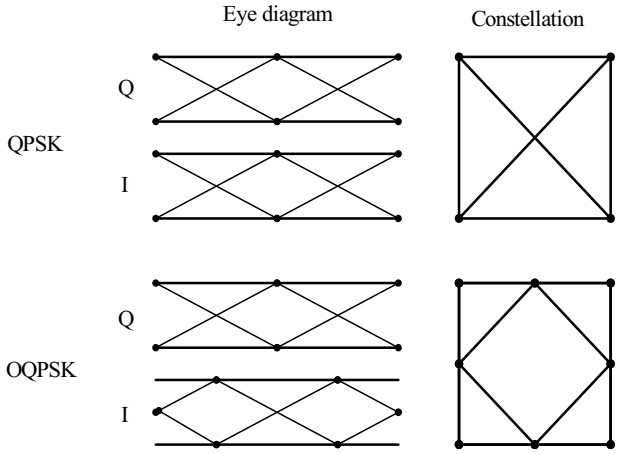


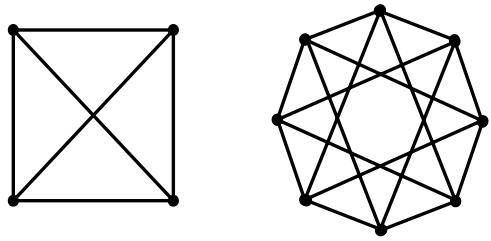
Figure 1-23. I/Q offset modulation

This way the signal trajectories are modified and the carrier amplitude will not go through or near the zero (the center of the constellation) state. The spectral efficiency is not changed but the reduction of the amplitude variations allows a more power efficient transmission.

3.5.2 Differential modulation

The second variation is the differential modulation used in QPSK (DQPSK) and 16QAM (D16QAM). Differential means that the information is not carried by the absolute state, but by the transition between states.

The $\pi/4$ DQPSK modulation format uses two QPSK constellations offset by 45º. Transitions must occur from one constellation to the other, guaranteeing that there is always a phase shift of each symbol.



QPSK
2 bits per symbol

π/4 DQPSK
2 bits per symbol

Figure 1-24. Differential modulation

Data is encoded in the magnitude and the direction of the phase shift, not in the absolute position. In π/4 DQPSK the signal trajectory does not pass through the origin, thus simplifying the transmitter design.

3.6 Broadband wireless access techniques

The electromagnetic spectrum is a coveted resource shared by many applications, thus the most communication systems use multiplexing or canalization techniques, like frequency, time, code, and geography divisions, in order to best use it.

The Frequency Division Multiple Access (FDMA) splits the available bandwidth into fixed frequency channels for each transmitter and receiver [4]. Used since 1900, it is still in use today.

Time division multiplexing separates the transmitters in time, sharing the same frequency bandwidth. The simplest type is the Time Division Duplex (TDD), which multiplexes the transmitter and receiver on the same frequency band. Modern digital radios like CT2 and DECT use TDD but multiplexing hundreds of times per second. On the other hand, the Time Division Multiple Access (TDMA) multiplexes several transmitters or receivers on the same frequency bandwidth and it is widely used in the GSM and the US NADC-TDMA systems.

CDMA is an access technique where multiple users are allowed to transmit simultaneously on the frequency bandwidths (1.23MHz). In the United States CDMA telephones a different code sequence is assigned to each terminal so that signals can be distinguished by correlating them with the overlaid sequence. This is based on codes that are shared between the base and mobile stations, 64 code channels on the downlink and unlimited on the uplink.

The last multiplexing technique is the geographical or cellular. If two transmitter/receiver pairs are far enough, they can operate on the same frequency bandwidth and they will not interfere with each other. Only few transmission systems do not use some kind of geographic multiplexing technique: clear-channel international broadcast stations, amateur stations, and some military low frequency radios broadcast around the world without any geographic boundaries.

Chapter 2

NONLINEAR DISTORTION

Nothing is linear in the nature, obviously neither the components used in the radio frequency communication systems. Sometimes, those nonlinear effects are well used in order to carry out different essential features for the design of communication devices [1].

The main problem relies on the fact that the elements that should be linear they are really not, mainly due to the actual technological limitations. The output vs. input transfer functions are not a perfect straight line (*Figure 2-1*) in all the operating range, only they can be considered partially quasilinear.

This problem is approached, when possible, by optimizing the manufacturing parameters of the nonlinear components, maximizing or minimizing the linear performance according to the each functionality, and always within specific limits. In any case, a perfect linear response will never be possible.

If nonlinear components are used for a linear purpose, like amplifiers, besides performing the amplification task, they also generate undesired effects that are going to be considered as nonlinear distortion.

The linear distortion used in some modulation techniques for amplitude or phase shifts do not generate new spectral components. However, the nonlinear distortion does.

The output signal amplitude of a nonlinear device, $y = f(x)$, can be approximated by a Taylor polynomial, in an environment of x_0 (2.1).

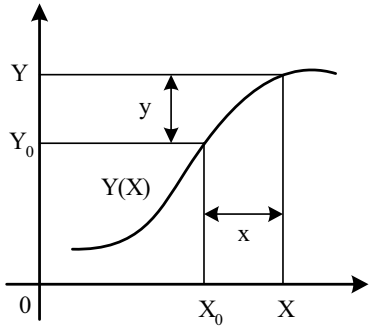


Figure 2-1. Non linear response

$$Y = f(X_0) + \left.\frac{\partial f(X)}{\partial x}\right|_{x_0} \cdot (X - X_0) + \frac{1}{2!}\left.\frac{\partial^2 f(X)}{\partial X^2}\right|_{x_0} \cdot (X - X_0)^2 + \frac{1}{3!}\left.\frac{\partial^3 f(X)}{\partial X^3}\right|_{x_0} \cdot (X - X_0)^3 + \cdots \tag{2.1}$$

The (2.1) equation can be rewritten more simply as (2.2).

$$\begin{cases} Y_0 = f(X_0) \\ y = Y - Y_0 \rightarrow y = a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + \cdots \\ x = X - X_0 \end{cases} \tag{2.2}$$

Four different distortion types are going to be studied, each one characterized with different test signals.

- A single tone for the harmonic distortion analysis.

$$x = x(t) = V_0 \cos \omega_0 t \tag{2.3}$$

- A two tone signal for the intermodulation analysis.

$$x = x(t) = V_0 (\cos \omega_1 t + \cos \omega_2 t) \tag{2.4}$$

- A triple tone signal for the third order distortions analysis in multichannel systems, also used in the triple beat distortion characterization.

$$x = x(t) = V_1 \cos \omega_1 t + V_2 \cos \omega_2 t + V_3 \cos \omega_3 t \tag{2.5}$$

- A two tone signal, one of both modulated, for the cross modulation analysis.

$$x = x(t) = V_1 \cos\omega_1 t + V_2(1 + m\cos\omega_{20} t)\cos\omega_2 t) \quad (2.6)$$

1. HARMONIC DISTORTION

Any output signal will never match exactly with the original signal when plotted over time, even the appropriated scale and delay factors are applied. The output signal will be slightly distorted, as shown in *Figure 2-2*.

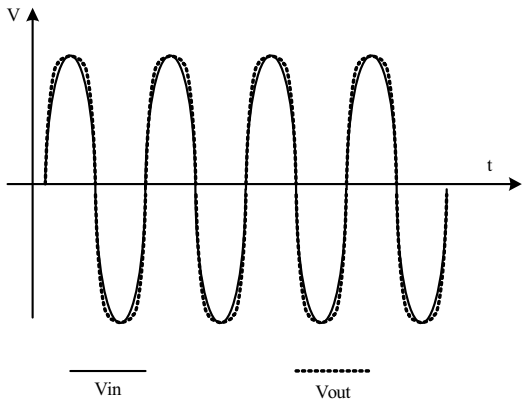


Figure 2-2. Input and output sinusoidal signals after scaling

Let's assume a pure sinusoidal tone with ω_0 frequency. The harmonic distortion can be calculated numerically with the test fixture of *Figure 2-3*

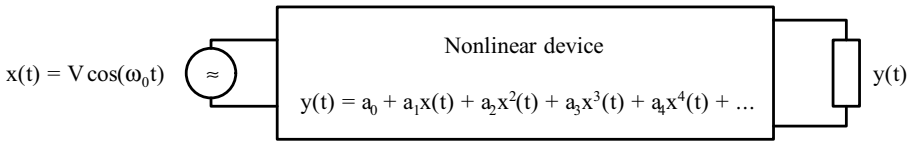


Figure 2-3. Harmonic distortion analysis

The input signal (2.7) has a constant amplitude (V) and a single frequency (ω_0)

$$x(t) = V\cos\omega_0 t \quad (2.7)$$

The transfer function of the nonlinear device will give a distorted output signal (y).

$$\begin{aligned} y(t) &= a_0 + a_1 x(t) + a_2 x^2(t) + a_3 x^3(t) + a_4 x^4(t) + \cdots \\ y(t) &= a_0 + a_1 V \cos\omega_0 t + a_2 V^2 \cos^2 \omega_0 t + a_3 V^3 \cos^3 \omega_0 t + a_4 V^4 \cos^4 \omega_0 t + \cdots \end{aligned} \tag{2.8}$$

The $\cos^n \omega t$ terms can be solved in a trigonometric way with the equations of (2.9).

$$\begin{aligned} a_0 &= a_0 \\ \cos^1 \omega t &= \cos\omega t \\ \cos^2 \omega t &= \frac{1}{2} + \frac{1}{2}\cos 2\omega t \\ \cos^3 \omega t &= \frac{3}{4}\cos\omega t + \frac{1}{4}\cos 3\omega t \\ \cos^4 \omega t &= \frac{3}{8} + \frac{1}{2}\cos 2\omega t + \frac{1}{8}\cos 4\omega t \\ \cos^5 \omega t &= \cdots \end{aligned} \tag{2.9}$$

The output signal, once those expressions have been substituted, is formed with subsequent frequency signals (nw_0) which amplitudes (V_n) are dependent on the transfer function coefficients (a_1, a_2, a_3 …).

$$y(t) = V_0 + V_1 \cos\omega_0 t + V_2 \cos 2\omega_0 t + V_3 \cos 3\omega_0 t + V_4 \cos 4\omega_0 t + \cdots \tag{2.10}$$

$$V_n \approx \frac{1}{2^{n-1}} a_n V^n \begin{cases} V_0 = a_0 + \frac{1}{2} a_2 V^2 + \frac{3}{8} a_4 V^4 + \cdots \\ V_1 = \frac{1}{1} a_1 V^1 + \frac{3}{4} a_3 V^3 + \cdots \\ V_2 = \frac{1}{2} a_2 V^2 + \frac{3}{8} a_4 V^4 + \cdots \\ V_3 = \frac{1}{4} a_3 V^3 + \cdots \\ V_4 = \frac{1}{8} a_4 V^4 + \cdots \end{cases} \tag{2.11}$$

The V_0 term is the DC value and does not transmit any information, the V_1 is the fundamental frequency, and the V_n term represents the n^{th} harmonic component.

1.1 Nth harmonic distortion coefficient

The distortion level due to any harmonic component (n^{th}) is calculated with the ratio (2.12) between its amplitude (V_n) and the amplitude of the fundamental component (V_1).

$$d_{an} = \frac{V_n}{V_1} = \frac{1}{2^{n-1}} \frac{a_n}{a_1} V^{n-1} \tag{2.12}$$

This harmonic distortion coefficient can also be represented as percentage (2.13), or in the logarithmic scale (dB) with the amplitude quadratic values (2.14).

$$d_{an}(\%) = d_{an} \times 100 \tag{2.13}$$

$$D_{an}(\text{dB}) = 10\log(d_{an}^2) = 20\log\left(d_{an}\right) \tag{2.14}$$

The output signal in a spectrum analyzer display is shown in *Figure 2-4*. The fundamental component – graphically represented as f_0 (Hz) instead of ω_0 (rad/sec) – and the harmonic components are displayed with their quadratic values in the logarithmical (dB) scale, so the distance between them will be precisely the D_{an} coefficient.

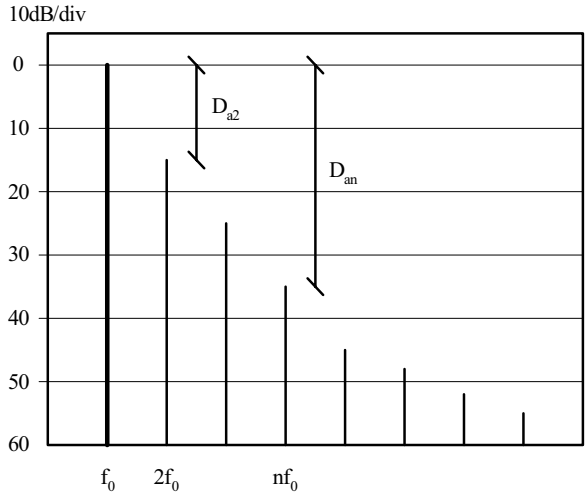


Figure 2-4. Harmonic distortion coefficients

1.2 Global harmonic distortion

In this case, the aim is to obtain a coefficient that includes all the harmonic distortion effects. It will be necessary to compare all the amplitude quadratic values with the fundamental component, because the amplitudes can not be directly added due to the fact that their correlation is null.

$$d_a^{\,2} = \frac{p_2 + p_3 + p_4 + \ldots}{p_1} = \frac{\frac{V_2^{\,2}}{2R} + \frac{V_3^{\,2}}{2R} + \frac{V_4^{\,2}}{2R} + \ldots}{\frac{V_1^{\,2}}{2R}} = \frac{V_2^{\,2} + V_3^{\,2} + V_4^{\,2} + }{V_1^{\,2}} = \sum_{n>1}\left(\frac{V_n}{V_1}\right)^2 \qquad (2.15)$$

$$d_a = \sqrt{\sum_{n>1} d_{an}^{\,2}} \qquad (2.16)$$

The global harmonic distortion can also be represented in percentage or as a logarithmical ratio.

$$D_a = 20 \log d_a \qquad (2.17)$$

The harmonic distortion, as any other nonlinear distortion, depends on the input signal amplitude, so each distortion coefficient must be associated with a specific input power level.

2. INTERMODULATION ANALYSIS

If a multi-tone signal is used as input signal instead of a single tone, the situation gets significantly complex. Several phenomena show up, besides the harmonic distortion seen before, making the signal transmission more difficult.

2.1 2 tone intermodulation

The two tone intermodulation analysis, as its name suggests, uses a two tone signal as input signal of the nonlinear device.

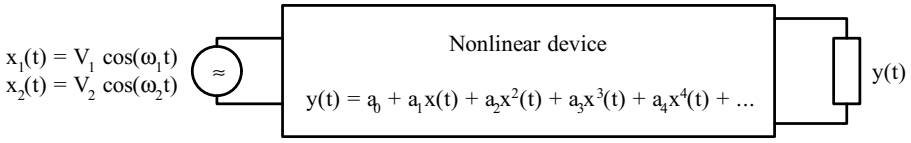


Figure 2-5. Distortion analysis with a two tone signal

Each tone has different amplitude and frequency values.

$$\begin{cases} x_1 = x_1(t) = V_1 \cos\omega_1 t \\ x_2 = x_2(t) = V_2 \cos\omega_2 t \end{cases} \rightarrow x = x_1 + x_2 \rightarrow x(t) = V_1 \cos\omega_1 t + V_2 \cos\omega_2 t \tag{2.18}$$

The transfer function of the nonlinear device is the same as in the harmonic analysis (*Figure 2-5*) and the resultant output signal (y) will have combinations of both independent tones (2.19).

$$\begin{aligned} &y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + \cdots \\ &y = a_0 + a_1 x_1 + a_1 x_2 + a_2 {x_1}^2 + a_2 {x_2}^2 + \\ &+ 2a_2 x_1 x_2 + a_3 {x_1}^3 + a_3 {x_2}^3 + 3a_3 x_1 {x_2}^2 + 3a_3 {x_1}^2 x_2 + \ldots \end{aligned} \tag{2.19}$$

First of all, all the trigonometric functions are solved (2.20).

$$\begin{aligned} &a_0 = a_0 \\ &a_1 x_1 = a_1 V_1 \cos\omega_1 t \\ &a_1 x_2 = a_1 V_2 \cos\omega_2 t \\ &a_2 {x_1}^2 = a_2 {V_1}^2 \cos^2\omega_1 t = a_2 {V_1}^2 \frac{(1+\cos 2\omega_1 t)}{2} \\ &a_2 {x_2}^2 = a_2 {V_2}^2 \cos^2\omega_2 t = a_2 {V_2}^2 \frac{(1+\cos 2\omega_2 t)}{2} \\ &2a_2 x_1 x_2 = 2a_2 V_1 V_2 \cos\omega_1 t \cos\omega_2 t = a_2 V_1 V_2 \left(\cos(\omega_1 - \omega_2)t + \cos(\omega_1 + \omega_2)t\right) \\ &a_3 {x_1}^3 = a_3 {V_1}^3 \cos^3\omega_1 t = \frac{3}{4} a_3 {V_1}^3 \cos\omega_1 t + \frac{a_3 {V_1}^3}{4} \cos 3\omega_1 t \\ &a_3 {x_2}^3 = a_3 {V_2}^3 \cos^3\omega_2 t = \frac{3}{4} a_3 {V_2}^3 \cos\omega_2 t + \frac{a_3 {V_2}^3}{4} \cos 3\omega_2 t \\ &3a_3 x_1 {x_2}^2 = \frac{3}{4} a_3 V_1 {V_2}^2 \left(2\cos\omega_1 t + \cos(2\omega_2 - \omega_1)t + \cos(2\omega_2 + \omega_1)t\right) \\ &3a_3 x_2 {x_1}^2 = \frac{3}{4} a_3 V_2 {V_1}^2 \left(2\cos\omega_2 t + \cos(2\omega_1 - \omega_2)t + \cos(2\omega_1 + \omega_2)t\right) \end{aligned} \tag{2.20}$$

Then, different terms are grouped according to the frequency components: continuous, fundamentals, harmonics, composite second order beats and intermodulation products.

- Continuous

$$y(t) = \left[a_0 + \frac{a_2(V_1^{\,2} + V_2^{\,2})}{2} + \ldots \right] + \tag{2.21}$$

- Fundamentals

$$\begin{aligned} &+ \left[a_1 V_1 + \frac{3}{4} a_3 V_1^{\,3} + \frac{6}{4} a_3 V_1 V_2^{\,2} \ldots \right] (\cos \omega_1 t) + \\ &+ \left[a_1 V_2 + \frac{3}{4} a_3 V_2^{\,3} + \frac{6}{4} V_1^{\,2} V_2 \ \ldots \right] (\cos \omega_2 t) \end{aligned} \tag{2.22}$$

- Harmonics

$$\begin{aligned} &+ \left[\frac{1}{2} a_2 V_1^{\,2} + \ldots \right] (\cos 2\omega_1 t) + \left[\frac{1}{2} a_2 V_2^{\,2} + \ldots \right] (\cos 2\omega_2 t) \\ &+ \left[\frac{1}{4} a_3 V_1^{\,3} + \ldots \right] (\cos 3\omega_1 t) + \left[\frac{1}{4} a_3 V_2^{\,3} + \ldots \right] \cos 3\omega_2 t) + \ldots \end{aligned} \tag{2.23}$$

- Composite second order beats (CSO)

$$+ [a_2 V_1 V_2 .] \cos(\omega_1 - \omega_2) t + [a_2 V_1 V_2 .] \cos(\omega_1 + \omega_2) t + \tag{2.24}$$

- Intermodulation products

$$\begin{aligned} &+ \frac{3}{4} a_3 V_1 V_2^{\,2} \cos(2\omega_2 - \omega_1) t + \frac{3}{4} a_3 V_1^{\,2} V_2 \cos(2\omega_1 - \omega_2) t + \\ &+ \frac{3}{4} a_3 V_1 V_2^{\,2} \cos(2\omega_2 + \omega_1) t + \frac{3}{4} a_3 V_1^{\,2} V_2 \cos(2\omega_1 + \omega_2) t \end{aligned} \tag{2.25}$$

The continuous, fundamental and harmonic components are expected from the harmonic distortion analysis. Moreover, the additions and subtractions of the fundamental components are the result of frequency translations, also known as composite second order beats (CSO). If this effect is not foreseen, the CSO beats could interference with the main channel. Their dependence on the second term (a_2) of the transfer function gives the name of second order distortions, and the resultant amplitude is also proportional to the mixed signal amplitudes (V_1V_2).

The rest of the components are known as intermodulation products.

Those components have a negative impact on the transmission systems. Therefore, it is recommended to keep them away from the useful frequency

bands, particularly in those wideband systems in which several frequency bands coexist.

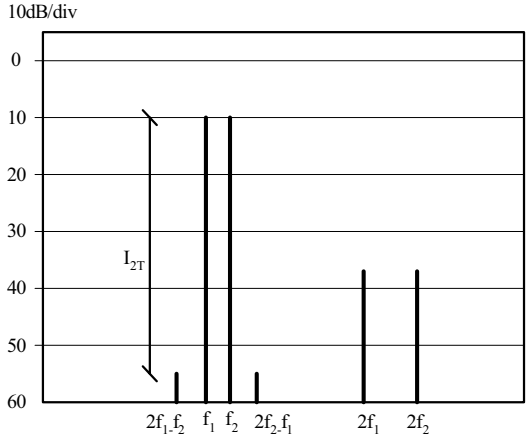


Figure 2-6. Third order intermodulation products and harmonic components

There is an especially harmful combination of intermodulation products – $(2\omega_2-\omega_1)$ and $(2\omega_1-\omega_2)$ – that fall close to the main channel (*Figure 2-6*): the third order (a_3) intermodulation products. These components are used in all the intermodulation characterizations and therefore, in any effect related to the third order distortion phenomenon.

2.2 3 tone intermodulation

In this case, three independent single carriers are used as input signals. The analysis methodology is exactly the same as in the two tone intermodulation case: solving all the trigonometric terms and grouping all the frequency components.

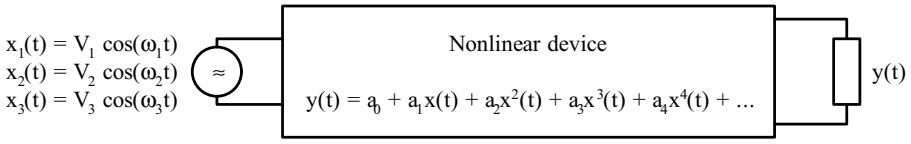


Figure 2-7. Composite triple beat (CTB) analysis

The output signal (y), besides all the previous distortion components, will have extra frequency component combinations among all the original frequencies (ω_1, ω_2 and ω_3). Now the mathematical expression of the output

signal is omitted, only it is shown the extra term due to the 3 tone intermodulation effect (2.26).

$$a_3(V_1\cos\omega_1 t + V_2\cos\omega_2 t + V_3\cos\omega_3 t)^3 \tag{2.26}$$

The 3 tone intermodulation components are collected in (2.28), once the trigonometric relations are solved and the frequency components are grouped.

$$6\cdot a_3 V_1 V_2 V_3(\cos\omega_1 t\cdot\cos\omega_2 t\cdot\cos\omega_3 t) \tag{2.27}$$

$$\left(\frac{3}{2}\cdot a_3\cdot V_1V_2V_3\right)\cos\begin{cases}(+\omega_1+\omega_2+\omega_3) & \text{a)}\\ (+\omega_1+\omega_2-\omega_3) & \text{b)}\\ (+\omega_1-\omega_2+\omega_3) & \text{c)}\\ (-\omega_1+\omega_2+\omega_3) & \text{d)}\end{cases} \tag{2.28}$$

All components depend on the third order linearity coefficient (a_3), and particularly the b, c and d ones fall close to the fundamental frequencies – provided their original frequency closeness. Alternatively, they are also used for the characterization of the third order distortion.

Note that only the four first terms (a_0, a_1, a_2, a_3) of the Taylor series have been used in the distortion analysis. Even if five or more terms were used, there will not be more than the triple cosine products, combined among them with analog sing criteria as in (2.28). This fact represents the origin of the composite triple beat (CTB) on multi carrier systems.

3. CROSS MODULATION (X_{MOD})

The cross-modulation phenomenon occurs when a modulated signal shares the same bandwidth with a single carrier.

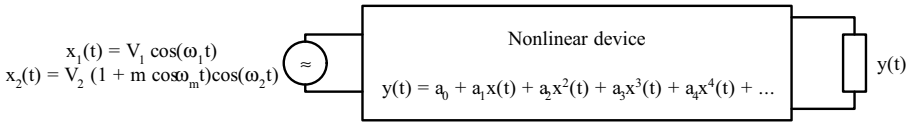


Figure 2-8. Cross-modulation analysis

Experimentally can be seen how the output single carrier is affected with a parasitic modulation, similar to the modulation of the interference signal but considerably weaker.

Two carriers with different frequencies are introduced into the nonlinear device. The first one ($x_1(\omega_1)$) is not modulated and will be considered as the interfered signal (2.29).

$$x_1 = x_1(t) = V_1 \cos\omega_1 t \tag{2.29}$$

On the other hand, the amplitude of the second carrier ($x_2(\omega_2)$) will be modulated (ω_m) which "m" as the modulation index (2.30).

$$x_2 = x_2(t) = v_2(t)\cos\omega_2 t = V_2(1 + m\cos\omega_m t)\cos\omega_2 t \tag{2.30}$$

The polynomial of the output signal (y) is again truncated in the fourth term for the same reasons as for the intermodulation case. The DC term (a_0) is also omitted because it does not have any influence in the distortion analysis. Therefore, only three terms (a_1, a_2 and a_3) are used, simplifying the transfer function of *Figure 2-8* to (2.31).

$$y = a_1 x + a_2 x^2 + a_3 x^3 \tag{2.31}$$

The analysis is similar to the intermodulation case with the exception of x_2 being a modulated signal and considering only the distortion components regarding to the interfered signal (x_1) (2.32).

$$y = V_1\left[a_1 + \frac{3}{4}a_3 V_1^{\,2} + \frac{3}{2}a_3 V_2^{\,2}\right]\cdot\cos\omega_1 t \tag{2.32}$$

Now, in (2.33), the modulation of the interference signal is introduced.

$$\begin{aligned}
&V_2(t) = V_2(1 + m\cos\omega_m t)\\
&y = V_1\left[a_1 + \frac{3}{4}a_3 V_1^{\,2} + \frac{3}{2}a_3\left(V_2(1 + m\cos\omega_m t)\right)^2\right]\cdot\cos\omega_1 t\\
&y = V_1\left[a_1 + \frac{3}{4}a_3 V_1^{\,2} + \frac{3}{2}a_3 V_2^{\,2}\cdot\left(1 + 2\cdot m\cos\omega_m t + \frac{m^2(1+\cos 2\omega_m t)}{2}\right)\right]\cdot\cos\omega_1 t\\
&y = V_1\left[a_1 + \frac{3}{4}a_3 V_1^{\,2} + \frac{3}{2}a_3 V_2^{\,2}\cdot\left(1 + \frac{m^2}{2} + 2\cdot m\cos\omega_m t + \frac{m^2\cos 2\omega_m t}{2}\right)\right]\cdot\cos\omega_1 t
\end{aligned} \tag{2.33}$$

The equation (2.33) is going to be discussed for two different cases – $m << 1$ and $m >> 1$ – in the next two sections.

3.1 Cross modulation (m<<1)

For low modulation indexes (m<<1) the m^2 factor will be insignificant against 1, so the equation (2.33) changed to (2.34).

$$\begin{aligned} y &= V_1\left[a_1 + \frac{3}{4}a_3V_1^{\,2} + \frac{3}{2}a_3V_2^{\,2}\cdot\left(1+2\cdot m\cos\omega_m t\right)\right]\cdot\cos\omega_1 t \\ y &= V_1\left[a_1 + \frac{3}{4}a_3V_1^{\,2} + \frac{3}{2}a_3V_2^{\,2} + 3a_3\cdot mV_2^{\,2}\cos\omega_m t\right]\cdot\cos\omega_1 t \end{aligned} \tag{2.34}$$

The equation (2.34) can be rewritten as (2.35) in order to convert the expression of the output signal to an amplitude modulation format, with an equivalent modulation index (m').

$$\begin{aligned} y &= \underbrace{\left(a_1 + \frac{3}{4}a_3V_1^2 + \frac{3}{2}a_3V_2^2\right)}_{K}\left[1 + \frac{3a_3\cdot V_2^2\cdot m}{\underbrace{a_1 + \frac{3}{4}a_3V_1^2 + \frac{3}{2}a_3V_2^2}_{m'}}\cos\omega_m t\right]\cdot\cos\omega_1 t \\ y &= K\left(1+m'\cos\omega_m t\right)\cos\omega_1 t \end{aligned} \tag{2.35}$$

The new modulation index depends on the interference amplitude (V_2), the interference modulation index (m), and the nonlinear transfer function (a_1, a_2 and a_3).

$$m' = \frac{3a_3\cdot V_2^2\cdot m}{\left(a_1 + \frac{3}{4}a_3V_1^2 + \frac{3}{2}a_3V_2^2\right)} \tag{2.36}$$

In a quasilinear region, theoretically, the a_3 coefficient is insignificant in comparison with a_1, so the new modulation index can be simplified to (2.37).

$$m' \approx \frac{3a_3V_2^2}{a_1}m \tag{2.37}$$

In *Figure 2-9* a cross modulation example is shown with an interfered (P1) and an interference (P2) signal. In this particular case the interference modulation index (m) is 4% (0.04). The basic AM modulation theory says that the power distance (dB_c) between any modulated carrier and its side band components is given by (2.38).

$$dB_c = 6dB - 20\log m\Big|_{m=0.04} \cong 34dB \tag{2.38}$$

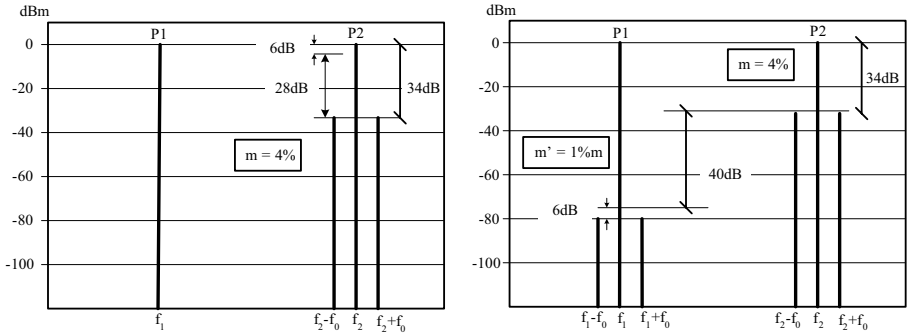


Figure 2-9. Cross modulation example (m = 0.04)

In current applications the cross modulations do not exceed the 1%. In other words, the cross-modulation index m' is around 0.01m (0.0004). In the current example, according to (2.38), this results in two cross modulated side band components $80dB_c$ from the main carrier signal.

3.2 Cross modulation (m>>1)

For high modulation indexes (m>>1) the interfered component of the output signal will be represented by (2.39).

$$
\begin{aligned}
y &= V_1\left[a_1+\frac{3}{4}a_3V_1^{\,2}+\frac{3}{2}a_3V_2^{\,2}\cdot\left(1+\frac{m^2}{2}+2\cdot m\cdot\cos\omega_m t+\frac{m^2\cos 2\omega_m t}{2}\right)\right]\cdot\cos\omega_1 t\\
y &\approx V_1\left[a_1+\frac{3}{4}a_3V_1^{\,2}+\frac{3}{2}a_3V_2^{\,2}\cdot\left(\frac{3}{2}+2\cdot m\cdot\cos\omega_m t+\frac{m^2\cos 2\omega_m t}{2}\right)\right]\cdot\cos\omega_1 t\\
y &= V_1\left[a_1+\frac{3}{4}a_3V_1^{\,2}+\frac{9}{4}a_3V_2^{\,2}\cdot+\frac{3}{2}a_3V_2^{\,2}m\cdot\left(2\cdot\cos\omega_m t+\frac{m\cdot\cos 2\omega_m t}{2}\right)\right]\cdot\cos\omega_1 t\\
y &= V_1\left(a_1+\frac{3}{4}a_3V_1^{\,2}+\frac{9}{4}a_3V_2^{\,2}\right)\cdot\left(1+\frac{3a_3V_2^{\,2}m}{a_1+\frac{3}{4}a_3V_1^{\,2}+\frac{9}{4}a_3V_2^{\,2}}\left(\cos\omega_m t+\frac{m}{4}\cos 2\omega_m t\right)\right)\cdot\cos\omega_1 t
\end{aligned}
\qquad (2.39)
$$

The equation (2.39), simplified as (2.40), shows the influence of the second order harmonics in the cross modulation distortion.

$$
\begin{aligned}
y &= K'\left(1+m'f(\omega_m)\right)\cos\omega_1 t\\
f(\omega_m) &= \cos\omega_m t+\frac{m}{4}\cos 2\omega_m t
\end{aligned}
\qquad (2.40)
$$

The new cross modulation index is slightly different to the small modulation case (m<<1), though in the quasilinear region are perfectly coincident (2.41).

$$m' = \frac{3a_3 V_2^2 m}{a_1 + \frac{3}{4}a_3 V_1^2 + \frac{9}{4}a_3 V_2^2} \rightarrow a_3 << a_1 \rightarrow m' \approx \frac{3a_3 V_2^2 m}{a_1} \tag{2.41}$$

In the example of *Figure 2-10* the modulation index (m) of the interference signal (P2) is 100% (1) and its power level is 20dB above the main signal (P1) level. The cross modulation index (m') will be again the 1% (0.01m).

Substituting into (2.40) and (2.41), the power level of the first cross modulation side band components ($f_1 \pm f_m$) will be 46dB lower than the main carrier (P1). However, the difference with the second cross modulation side band components ($f_1 \pm 2f_m$) will be 52dB.

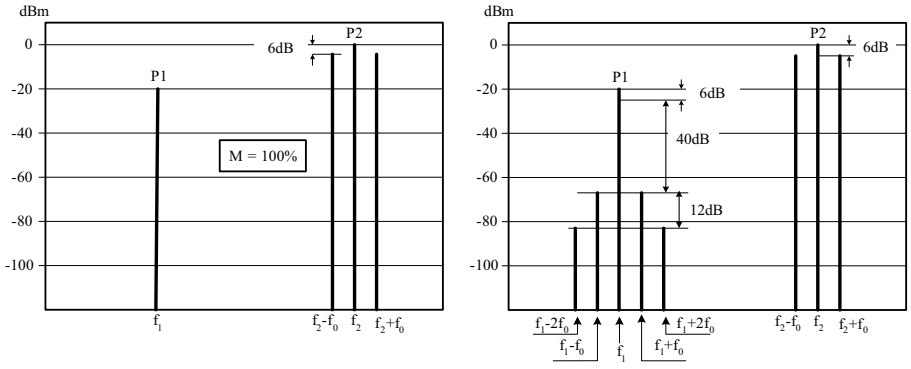


Figure 2-10. Cross modulation example (m = 1)

Therefore, as far as the cross modulation is concerned…

- It is not significantly affected by the interference modulation frequency (ω_m/f_m) and the interference modulation index (m).
- The effect of the interference signal harmonics increases according to the modulation index does.
- Improves as the linear gain (a_1) increases.
- Improves as the third order term of the nonlinear response (a_3) decreases. This is the reason why the cross modulation is considered as another third order distortion type.

Finally, the cross modulation coefficient (2.42) is usually defined as the ratio between the cross modulation index (m') and the original modulation index (m).

$$m_X = \frac{m'}{m} = \frac{3a_3 V_2^{\,2}}{a_1} \tag{2.42}$$

4. DISTORTION MEASUREMENT TECHNIQUES

The distortion phenomenon will be clearly understood with the discussion of the characterization methods used for the harmonic, intermodulation and cross modulation measurements [1].

4.1 Harmonic distortion measurement

The harmonic distortion measurement is really simple since the spectrum analyzers were developed.

The basic test fixture is shown in *Figure 2-11*. A single tone – from a RF signal generator – is used as input signal while the output signal is displayed in the spectrum analyzer.

Ideally, the input signal is noise free, but really, the signal generators introduce their own harmonic distortion. Variable filters are highly recommended at the signal generator output in order guarantee the characterization of the harmonic components introduced only by the device under test (DUT).

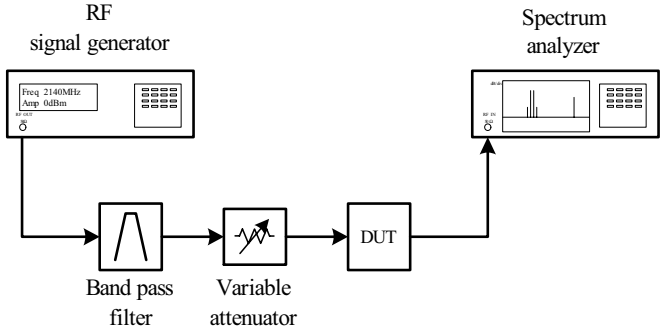


Figure 2-11. Harmonic distortion measurement test fixture

On the other hand, the harmonic components introduced by the spectrum analyzers should be also taken into account, especially with high power signals. Nowadays the spectrum analyzers include internal variable attenuators that avoid those effects.

Anyway, in order to ensure that the SA does not introduce its own harmonic distortion, its input signal is successively attenuated and the harmonic distances (D_{an}) are measured. If those distances keep constant the spectrum analyzer is not introducing any distortion. This technique is also used for the intermodulation and the cross modulation measurements.

4.2 Third order distortion measurement

4.2.1 Two tone procedure

The two tone procedure is based on measuring the intermodulation products corresponding to the ($2\omega_2$-ω_1) and ($2\omega_1$-ω_2) components that fall close to the fundamental frequencies. These intermodulation products characterize the third order distortion.

Let's take the Taylor series terms regarding to those intermodulation products (p_3).

$$\frac{3}{4}a_3\left[V_1^2V_2\left(\cos(2\omega_1+\omega_2)t+\cos(2\omega_1-\omega_2 t)\right)+V_1V_2^2\left(\cos(\omega_1+2\omega_2)t+\cos(\omega_1-2\omega_2 t)\right)\right] \quad (2.43)$$

In a two tone procedure the amplitude of the input tones must be equal (V_1=V_2=V) so the amplitude of the four intermodulation components ($|p_3|$) will be too (2.44).

$$|p_3| = \frac{3}{4}a_3V^3 \quad (2.44)$$

Remember that the input-output response of the main component is linear (2.45) while the response of the third order distortion is cubic (2.44).

$$|p_1| = a_1V \quad (2.45)$$

The ratio between the main and the third order distortion components can be expressed according to their amplitude values, as in (2.46).

$$\frac{P_1}{P_3}(\text{dB}) = 10\log\left[\frac{|V_1|}{|V_3|}\right]^2 = 20\log\left[\frac{4a_1^3}{3a_3V^2}\right] \quad (2.46)$$

But the spectrum analyzer usually displays all measurements in the logarithmical scale (dB), so the third order intermodulation distance (I_{2T}) will be too (2.47).

$$I_{2T} = \frac{P_1}{P_3}(dB) = 20\log\left[\frac{4a_1^{\ 3}}{3a_3}\right] - 2\cdot 10\log V^2 = K - 2\cdot\Delta P(dB) \qquad (2.47)$$

The main conclusion is that an input power increment (ΔP) causes a double shortage of the third order intermodulation distance (I_{2T}).

4.2.2 Third order intercept point (IP_3)

The third order distortion effect is mainly plotted in a 2D graph with output and input signals scaled in logarithmic units, like in *Figure 2-12*. The ideal input-output response of the main component (p_1) is linear so it corresponds with a $\alpha = 45^o$ ($\tan\alpha = 1$) slope in the 2D graph. The y axis at the origin refers to the gain factor of the DUT in decibels (dB).

On the other hand, the input-output response of the third order intermodulation product (p_3) is cubic so it corresponds with a $\beta = 71.57^o$ ($\tan\beta = 3$) slope in the same 2D graph.

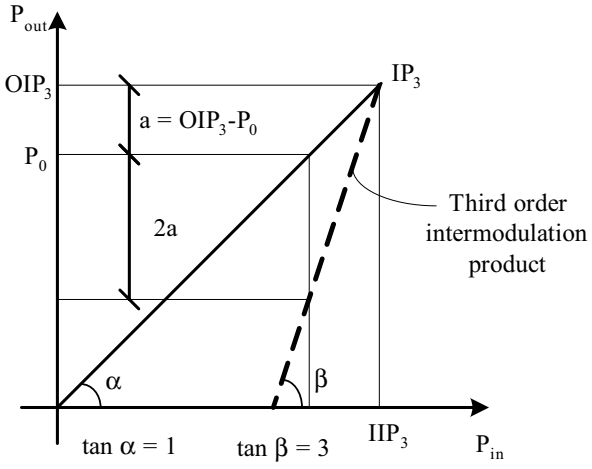


Figure 2-12. Third order intercept point (IP_3)

The distance between the main and the third order intermodulation product for a specific output signal sets the third order intermodulation distance (I_{2T} or 2a). The intercept point between both curves is known as the third order intercept point (IP_3). In amplification devices this point is referred to the output signal level as the OIP_3 avoiding the use of the gain factor. In (2.48) some equations are summarized between the OIP_3 and the I_{2T} distance, based on the *Figure 2-13*.

$$3a = I_{2T} + (OIP_3 - P_0) = 3(OIP_3 - P_0) = (OIP_3 - P_0)\tan\beta \tag{2.48}$$

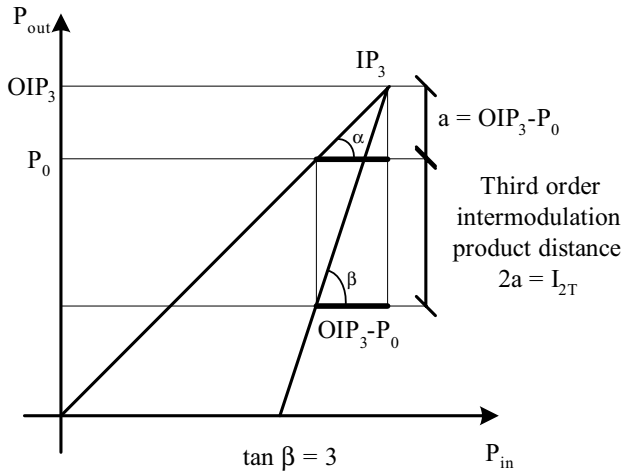


Figure 2-13. The intermodulation distance (I_{2T}) and the OIP_3

The analytical expression of the input signal amplitude for the IP_3 point is obtained from (2.49).

$$[a_1 V] = \left[\frac{3a_3 V_{IP3}^{\ 3}}{4}\right] \rightarrow V_{IP3}^{\ 2} = \frac{4a_1}{3a_3} \tag{2.49}$$

Those values are conditional on a quasilinear performance of the transfer function near the IP_3 point. However, the distortion components omitted in the mathematical analysis (a_4, a_5, a_6, etc.) start having more importance than expected as input signal increases, causing the well known saturation effect and making the IP_3 point unachievable.

The saturation effect begins once the main component of the output signal stops following the input signal with the ideal ratio. The inflexion point is known as the 1dB compression point (P_{1dB}) when the output signal is 1dB compressed from the ideal performance, as evidenced in *Figure 2-14*.

Assuming that the quasilinear condition is still acceptable at P_{1dB}, this point can be associated with the IP_3 point.

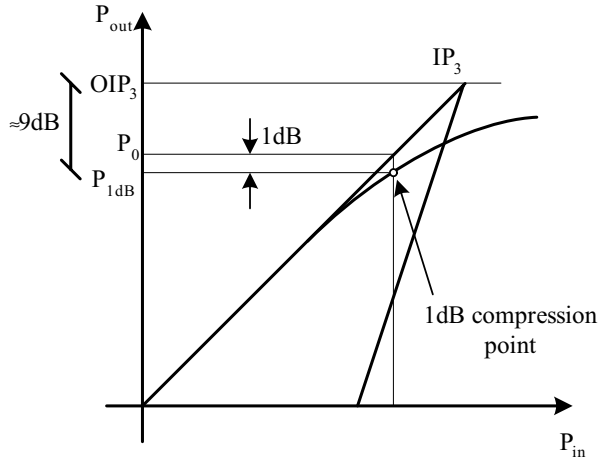


Figure 2-14. 1dB compression point and output response

The ideal and real output powers are represented in (2.50).

$$\begin{aligned} &\text{ideal} \rightarrow 20\log a_1 V \\ &\text{real} \rightarrow 20\log\left[a_1 V + \frac{3a_3 V^3}{4}\right] \end{aligned} \tag{2.50}$$

The P_{1dB} is defined when the difference between both terms is exactly 1dB.

$$20\log\left[a_1 V_{1dB} + \frac{3a_3 V_{1dB}^3}{4}\right] - 20\log\left[a_1 V_{1dB}\right] = 1\text{dB} \tag{2.51}$$

This expression can be rewritten as (2.52).

$$20\log\left[1 + \frac{3a_3}{4a_1} V_{1dB}{}^2\right] = 1\text{dB} \tag{2.52}$$

In (2.53) all terms are changed from the logarithmic to the linear scale.

$$\frac{3a_3}{4a_1} V_{1dB}{}^2 = \left(10^{\frac{1}{20}} - 1\right) \tag{2.53}$$

Now the V_{OIP3} expression obtained in (2.49) is replaced in (2.53).

$$\frac{V_{1dB}{}^2}{V_{OIP_3}{}^2} = \left(10^{\frac{1}{20}} - 1\right) \tag{2.54}$$

And finally, all terms are changed back to the logarithmic scale.

$$10\log\left[\frac{V_{1dB}{}^2}{V_{OIP_3}{}^2}\right] = 10\log\left(10^{\frac{1}{20}} - 1\right) \rightarrow P_{-1dB} - OIP_3 = -9{,}14\text{dB} \tag{2.55}$$

As a result, a thumb rule is estimated between the OIP_3 and the P_{1dB} parameters given by (2.56).

$$OIP_3[\text{dBm}] - P_{1dB}[\text{dBm}] \approx 9\text{dB} \tag{2.56}$$

The OIP_3 and the P_{1dB} are usually used as figure of merit of amplifiers in order to represent their intermodulation performance. In fact, the OIP_3 is mainly obtained from the third order intermodulation distance (I_{2T}) introduced in (2.48) and now remembered in (2.57).

$$I_{2T}[\text{dB}] = 2\left(OIP_3[\text{dBm}] - P_o[\text{dBm}]\right) \tag{2.57}$$

The measurement of the I_{2T} distance is carried out with two signal generators and a spectrum analyzer, as indicated in *Figure 2-15*. The input signal power must be properly adjusted in order to guarantee the quasilinear performance: a +1dB input signal must correspond with a +1dB in the output main signal and with a +2dB in the third order intermodulation products. The *Figure 2-16* is the display of the spectrum analyzer during the I_{2T} measurement.

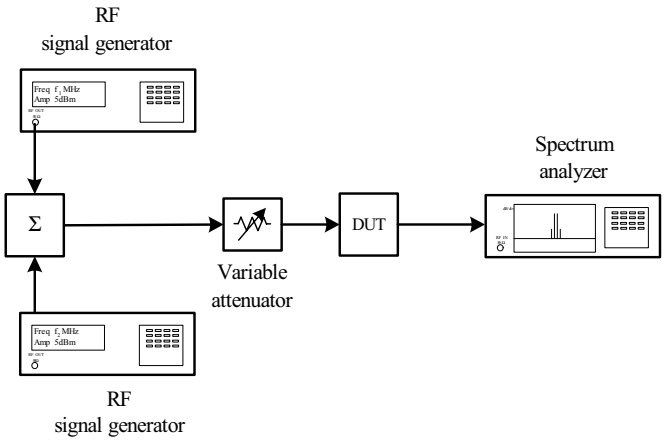


Figure 2-15. Intermodulation distance (I_{2T}) measurement test fixture

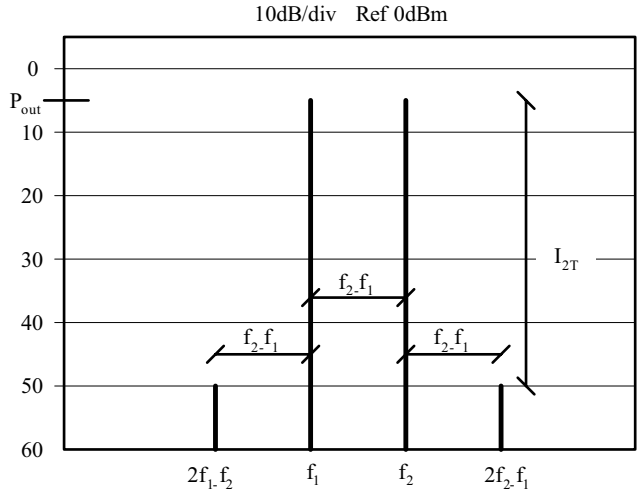


Figure 2-16. Spectrum analyzer output for I_{2T} measurement with the two tone procedure

4.3 Second order distortion measurement

Once again the two tone procedure is used for the second order distortion measurement (*Figure 2-15*).

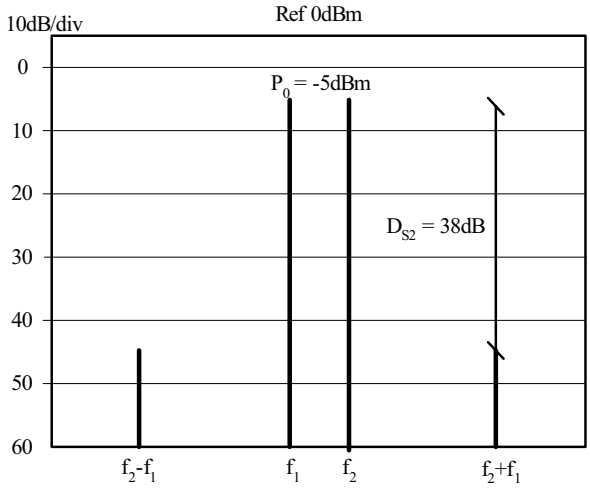


Figure 2-17. Second order distortion measurement example

As any distortion distance measurement, the second order distortion distance (D_{S2}) is also referred to the output power. Furthermore, the supposition of the quasilinear region (1dB distance decrease for 1dB input signal increase) makes possible its extrapolation to any other output level. For example, regarding the *Figure 2-17*, it is expected a 33dB distortion distance (D_{S2}) for a 0dBm output power level.

In a hypothetical never-ending quasilinear situation, the transfer function of the linear component and the second order components will coincide in a fictitious second order intercept point IP_2 (*Figure 2-18*).

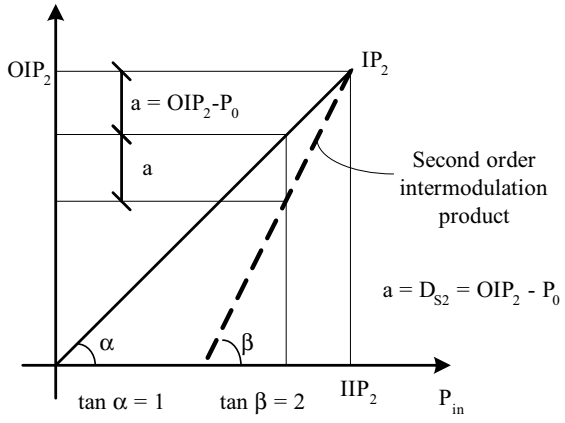


Figure 2-18. The second order intercept point (IP_2)

The OIP_2 point of *Figure 2-17* is obtained in (2.58).

$$OIP_2 = P_0 + D_{S2} = -5dBm + 38dB = 33dBm \tag{2.58}$$

This point is the most usual way to characterize the performance of the second order distortion. If the OIP_2 level is known, the expression (2.59) will give the D_{S2} distance for any output power level (N_o).

$$D_{S2}[dB] = OIP_2[dBm] - N_0[dBm] \tag{2.59}$$

4.4 Cross modulation (X_{mod}) measurement

Usually, the cross modulation (X_{mod}) is specified as the acceptable parasite modulation percentage in the interfered signal when the interfering signal is modulated at 100%.

There are several procedures to estimate the cross modulation: the modulation meters, the sound distortion meters and the most used, the spectrum analyzer (*Figure 2-19*).

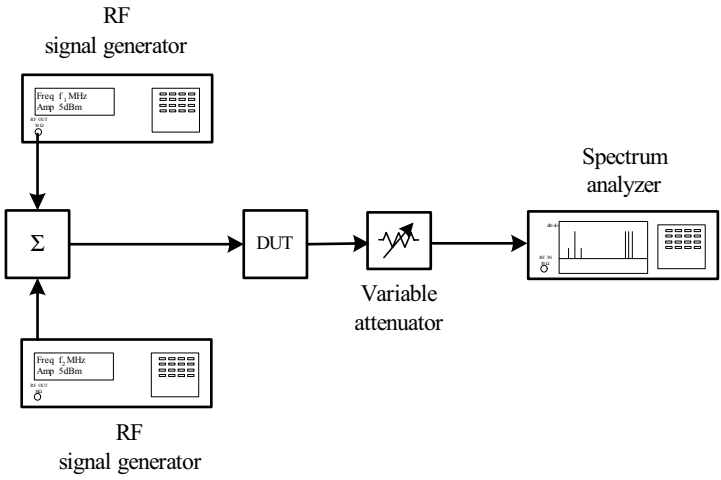


Figure 2-19. X_{mod} measurement test fixture

In a 100% modulated signal (m=1), the power of the two side bands is half of the carrier power, and as both side bands are equal, each one must have a quarter of the main signal power. In other words, the power of the side bands is always 6dB lower than the main carrier power.

On the other hand, if the modulation index is the 1% (m=0.01), the side bands power will be, moreover, 40dB lower the main carrier power (2.60).

$$\text{dB} = 20 \cdot \log(0.01) = -40\text{dB} \tag{2.60}$$

The X_{mod} measurement starts with the modulation of the interference signal (m=1) until the side bands reach their maximum value, i.e., 6dB below the main carrier (*Figure 2-20*).

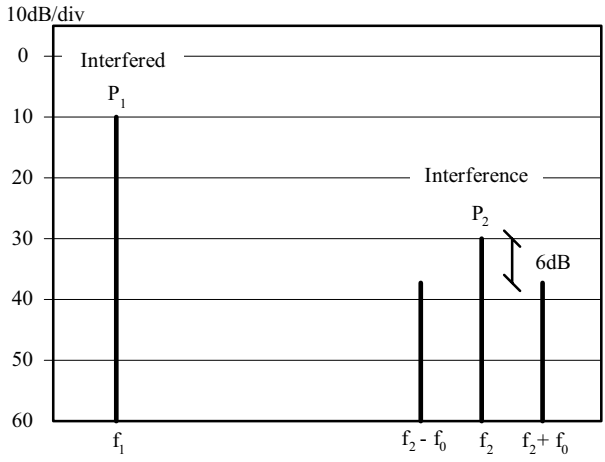


Figure 2-20. Interference signal 100% modulated (step 1)

Progressively, the interference power level (P_2) is increased until the cross modulation side bands are set 46dB below the interfered signal. At this point, it is considered a 1% trans-modulation between the interference and the interfered signals (*Figure 2-21*).

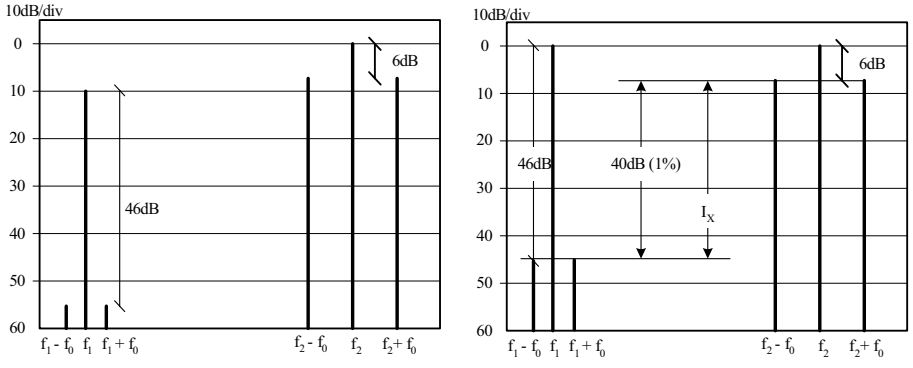


Figure 2-21. Maximum distortion for Xmod measurement

The theoretical analysis of section 3 states that the power level of the interfered signal does not have any influence over the cross modulation. This is totally true, and its level will be delimited only by the dynamic range of the spectrum analyzer. Anyway, the coincidence of the interfered and the interference signal levels is an extended recommendation.

In short, the distance between both side bands in the logarithmic scale gives the X_{mod} distance (I_X). In the *Figure 2-21* I_X is 40dB, so the trans-modulation is the 1%.

5. MEASUREMENTS OF WIDEBAND DIGITAL SIGNALS

There are several testing stages during the design of a digital transmitter and the different components are initially tested individually. When appropriate, the transmitter is fully assembled, and system tests are performed. During the design stage of the product development, verification tests are rigorous to verify the design robustness. These rigorous conformance tests verify that the design meets the system requirements thereby, ensuring interoperability with equipment from different manufacturers.

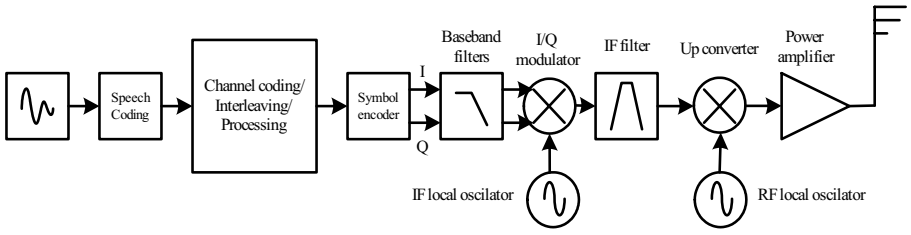


Figure 2-22. Block diagram of a digital communications transmitter

This section covers some of the most typical parameter testing: peak to average power ratio, error vector magnitude, code domain power and the adjacent leakage ratio. Most of the transmitter measurements are common to all digital communications technologies, although there are some variations in the way the measurements are performed [3].

5.1 Peak to average power ratio and CCDF curves

Peak to average power ratio and CCDF (defined below) curves are statistical measurements on the time domain. Peak to average power ratio is

the ratio of the peak envelope power to the average envelope power of a signal in a time period.

Some instruments may provide peak to average power statistics; that is, the peak envelope power is given not as an absolute peak but rather as the power level associated with a certain probability.

The power statistics of the signal can be completely characterized by performing several of these measurements and displaying the results in a graph known as the Complementary Cumulative Distribution Function (CCDF). The CCDF curve shows the probability that the power is equal to or above a certain peak to average ratio for different probabilities.

The statistics determine the headroom required in amplifiers and other components. Signals with different peak to average statistics can stress the components in a transmitter and causing different levels of distortion. CCDF measurements can be performed at different points in the transmitter to examine the statistics of the signal and the impact of the different sections on those statistics.

Peak to average power ratio and CCDF measurements are particularly important in digital communication systems because the statistics may vary. For instance, in WCDMA systems, the statistics of the signal vary depending on how many code channels and which ones are present at the same time. *Figure 2-23* shows the CCDF curves for signals with different code channel configurations with a performance spectrum analyzer (PSA) of Agilent Technologies.

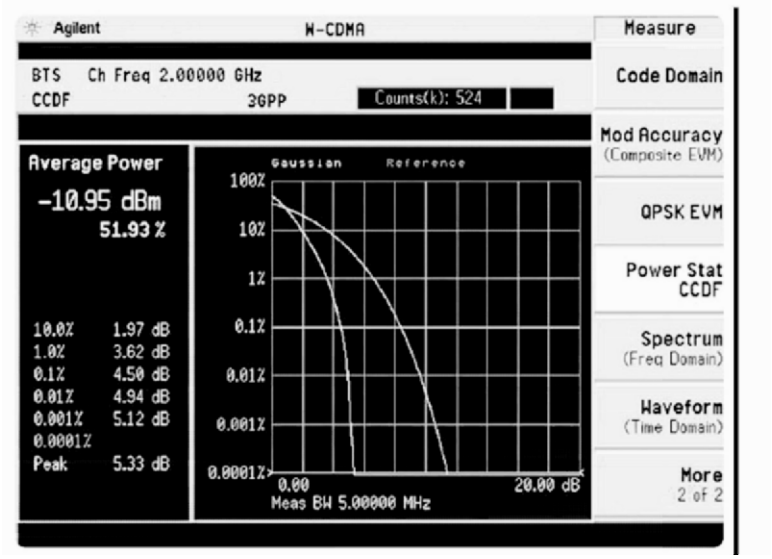


Figure 2-23. CCDF curves from a PSA (Agilent Technologies)

The more code channels transmitted, the higher the probability of reaching a given peak to average ratio will be. In systems that use constant amplitude modulation schemes, such as GSM, the peak to average ratio is relevant if the components (for example, the power amplifiers) must carry more than one carrier. There is a clear trend toward using multi-carrier power amplifiers in base station for most of the digital communications systems.

5.2 Modulation quality measurements

There are different ways to measure the quality of a digitally modulated signal. They usually involve a precise demodulation of the transmitted signal and its comparison with a mathematically generated reference signal. The definition of the actual measurement depends mainly on the modulation scheme and the standard followed. NADC and PDC, for example, use Error Vector Magnitude (EVM), while GSM uses phase and frequency errors.

The most widely used modulation quality metric in digital communications systems is the Error Vector Magnitude (EVM). The error vector is the vector difference at a given time between the ideal reference signal and the measured signal. The error vector is a complex quantity that contains a magnitude and a phase component. It is important not to confuse the magnitude of the error vector with the magnitude error, or the phase of the error vector with the phase error. A graphical description of these differences can be seen in *Figure 2-24*.

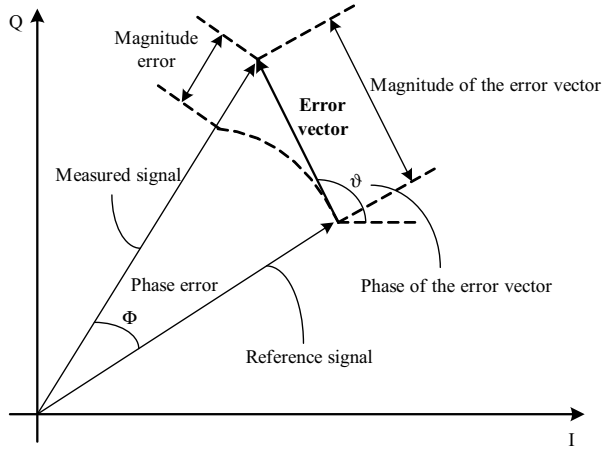


Figure 2-24. Error vector and related parameters

Error vector magnitude is the root mean square (RMS) value of the error vector over time at the symbol clock transitions. By convention, EVM is usually normalized to either the amplitude of the outermost symbol or to the square root of the average symbol power. The *Figure 2-25* shows a screen shot from the same performance spectrum analyzer (PSA) of Agilent Technologies.

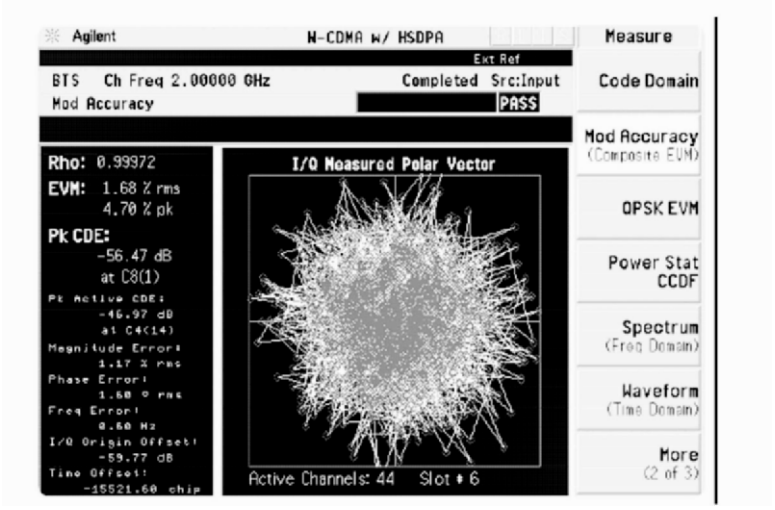


Figure 2-25. Error vector magnitude measurement with a PSA (Agilent Technologies)

5.3 Code domain power

In a code division multiple access, a multi-channel signal can be analyzed in the code domain. To analyze the composite waveform, each channel is decoded using a code correlation algorithm that determines the correlation coefficient factor for each code. Once the channels are decoded, the power in each code channel is determined.

Measuring the code domain power, shown in *Figure 2-26*, is essential for verifying that the base station is transmitting the correct power in each of the code channels. It is also important to look at the code-domain power levels of the inactive channels, which can indicate specific problems in the transmitter.

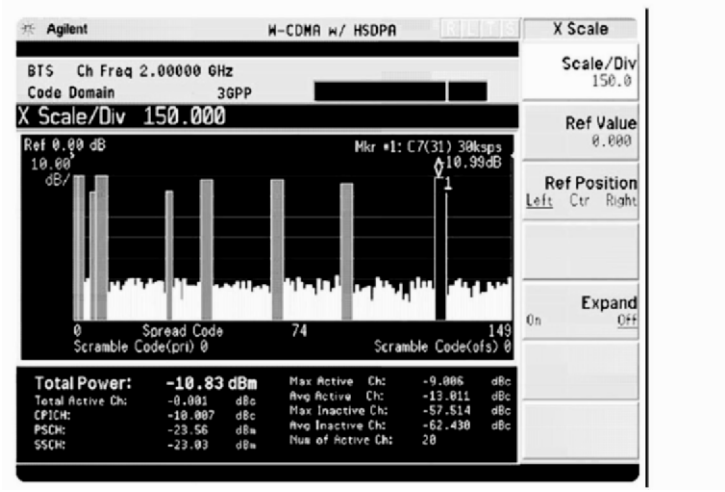


Figure 2-26. Code-domain power measurement with a PSA (Agilent Technologies)

5.4 Adjacent Channel Leakage Ratio (ACLR)

Despite of the technology used, the ACLR measurements are required to ensure that the transmitter is not interfering with adjacent and alternate channels.

The Adjacent Channel Leakage Ratio (ACLR) is usually defined as the ratio of the average power in the adjacent frequency channel to the average power in the transmitted frequency channel. For instance, in *Figure 2-27* the ACLR (dB_c) is measured for different frequency offsets.

When executing ACLR measurements, it is important to take into account the statistics of the signal transmitted. The CCDF curves can be used for this purpose. Different peak to average ratios have a different impact on the transmitter nonlinear components, such as the RF amplifier, and therefore on the ACLR as well. Higher peak to average ratios in the transmitted signal can cause more interference in the adjacent channel. Different standards have different names and definitions for the adjacent channel power (ACP) measurement. For example, for TDMA systems such as GSM, there are two main contributors to the ACP: the burst-on and burst-off transitions and the modulation itself. The GSM standard uses the Output RF Spectrum (ORFS) as ACP parameter and specifies two different measurements: ORFS due to modulation and ORFS due to switching [3].

In the case of NADC-TDMA, the ACP due to the transients and the modulation itself are also measured separately for mobile stations.

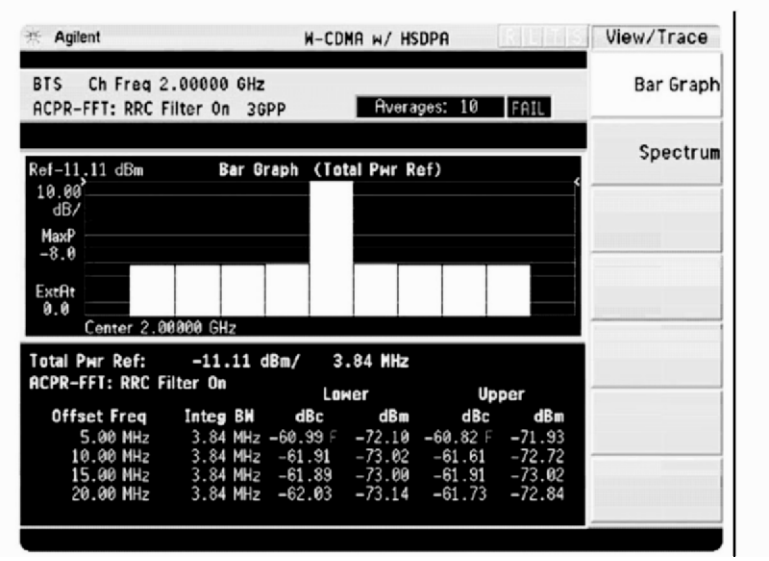


Figure 2-27. ACLR measurement with a PSA (Agilent Technologies)

Additionally, a weighting function that corresponds to the receiver base-band filter response is applied to the measurement of the base and mobile stations.

For cdmaOne systems, the ACLR is not defined in the standard but, it is often used in practice to test the specified in-band spurious emissions.

Chapter 3

RF POWER AMPLIFIERS

Despite the outstanding development achieved in the semiconductor technology in the course of the last decade, the solid state technology has not yet entirely replaced the vacuum technology in the field of power amplifiers.

Effectively, electronic tubes are still being used in high frequency applications and with high power signals and this situation will probably not change for a long period of time. However, this technological niche is not the aim of this chapter.

The brief study presented in this work focuses on the amplifiers based on semiconductor technology, much more widespread than the vacuums ones and capable of being improved by means of linearization techniques.

The wide variety of active components used as RF power amplifiers mainly depend on the frequency and power parameters defined for each application. The conventional vacuum tubes (triode type) are used for hundreds of kW and less than 1GHz. On the other hand, the vacuum valves, different from the triode type, are able to obtain hundreds of kW in the microwave range, similar to the travelling-wave tubes (TWT) used in the satellite transistors. The klystron amplifiers also belong to this frequency range; in fact, some of its variants are used in radar amplifiers. The high power tubes (magnetrons) also deserve a special mention. They work with impulsive excitations in radar transmitters and the basic ones are mostly used in the domestic microwave ovens.

As far as the semiconductor technology is concerned, silicon MOSFET devices operate in the VHF (Very High Frequency) and HF (High Frequency) bands up to hundreds of watts. However, the Indium Phosphorus

HEMT are used in the millimetric waves up to hundreds of milliwatts. Other active devices should also not be forgotten, such as bipolar (BJT), MESFET (AsGa), JFET, PHEMT…

From now on, the transistors of the sundry semiconductor technologies will be used without distinction as their basic amplification principles are valid for all of them.

1. CLASSIFICATION OF POWER AMPLIFIERS

The RF amplifiers are divided into classes or modes, identified with a capital letter that usually goes from A to F (in thermionic valves it can also be followed by a numeric sub index). In recent years, new amplification modes have been developed and identified with consecutive letters.

Traditionally, the classification of power amplifiers was related to the bias points of the transistors. Originally, this was the parameter used in the definition of the operating modes.

However, the quiescent point establishes the time (or angle) for which the amplifier is going to be "on" throughout the entire cycle of the input signal. In fact, the class A amplifiers are the only ones that complete the whole conduction cycle whereas all the others have a reduced operation angle.

Precisely, this operation angle may be the most accurate parameter for identifying, unequivocally, the amplification class type, because the bias point by itself is not able to define it unmistakably.

Really, all the amplification modes, with the exception of the class A, use some nonlinear commutation or wave reconstruction technique. Anyway, all of them are considered as amplification devices, although semantically some contradictions can be found.

As far as amplification topology is concerned, three different types could be distinguished, according to the disposition of the active elements into the amplification architecture.

Regardless of the amplification mode, the three topologies used nowadays are the unilateral, the symmetric and the complementary topologies. Some examples are shown in *Figure 3-1*.

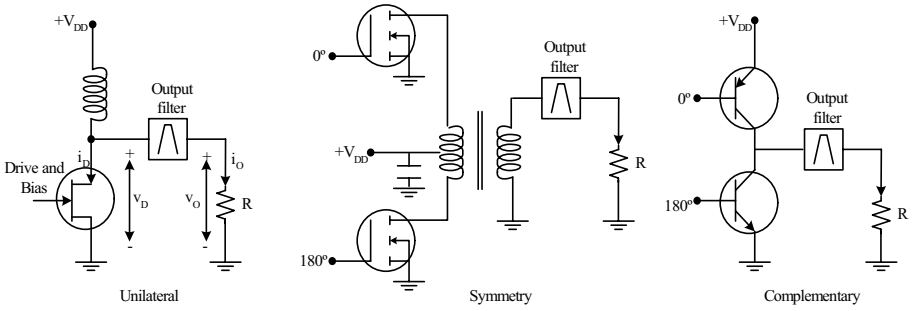


Figure 3-1. Power amplifier topologies: unilateral, symmetry and complementary

2. POWER AMPLIFIERS PARAMETERS

2.1 The efficiency rate

The efficiency, along with the linearity, is one of the most critical parameters of power amplifiers. Four efficiency types are often considered.

- Collector efficiency (η_C) is the ratio of the RF output power (P_{out}) and the DC input power (P_{DCin}).

$$\eta_C = \frac{P_{out}}{P_{DCin}} \tag{3.1}$$

- Power added efficiency (η_{PAE}). It considers the RF power increment introduced by each amplification stage. It is obtained by subtracting the RF input power (P_{RFin}) to the RF output power (P_{out}). In high gain amplifiers this parameter will practically coincide with the collector efficiency.

$$\eta_{PAE} = \frac{P_{out} - P_{RFin}}{P_{DCin}} \tag{3.2}$$

- Total efficiency (η_T). It is defined as the ratio between the RF power (P_{out}) and the input power including the RF excitation (P_{RFex}).

$$\eta_T = \frac{P_{out}}{P_{DCin} + P_{RFex}} \tag{3.3}$$

- Instantaneous efficiency. The performance of any previous efficiency rate for a given input signal in time domain is known as the instantaneous efficiency. As it has been seen in Chapter 1, some modulation techniques are based on amplitude variations. Those signals cause, inevitably, variations over the instantaneous efficiency. Therefore the mean efficiency (η_{avg}) is defined as the ratio between the output average power (P_{RFmean}) and the DC input power (P_{DCin}).

$$\eta_{avg} = \frac{P_{RFmean}}{P_{DCin}} \tag{3.4}$$

For continuous wave (CW), FM or GSM signals the amplitude envelopes are constant and their outputs are always at the maximum level. However, for other signal types their power density probability distribution function (PDF) should be considered, or more appropriate, their complementary cumulative distribution function (CCDF).

2.2 The back-off

The power amplifier back-off is defined as the headroom, in the logarithmic scale (dB), between the mean power level and the peak power level of the output signal.

The high amplitude variations used in digital communication systems oblige the RF designer to keep this power headroom when the bias point is defined. This way the amplifier guarantees the operation mode for which it has been designed even when the input signal amplitude takes the maximum value.

In some of the most popular telecommunication systems with signal widespread techniques, like the Universal Mobile Telecommunication System (UMTS), 10 to 20dB back-offs are commonly used [5].

2.3 Power utilization factor (PUF)

The amplification classes, besides the operating mode and the power efficiency, are also distinguished according to the output power, normalized to 1A output peak current and 1V output peak voltage. It is known as the power transistor utilization factor (PUF).

This merit factor is used to compare the power amplifiers output capacity, just multiplying the maximum voltage and current values by the power utilization factor.

3. CLASS A

The class A operation mode is entirely carried out between the cut-off and the saturation regions. As it is shown in *Figure 3-2*, the quiescent point (S_0) is fixed exactly in the middle of the load line (LL), between the cut-off (S_C) and the saturation points (S_S).

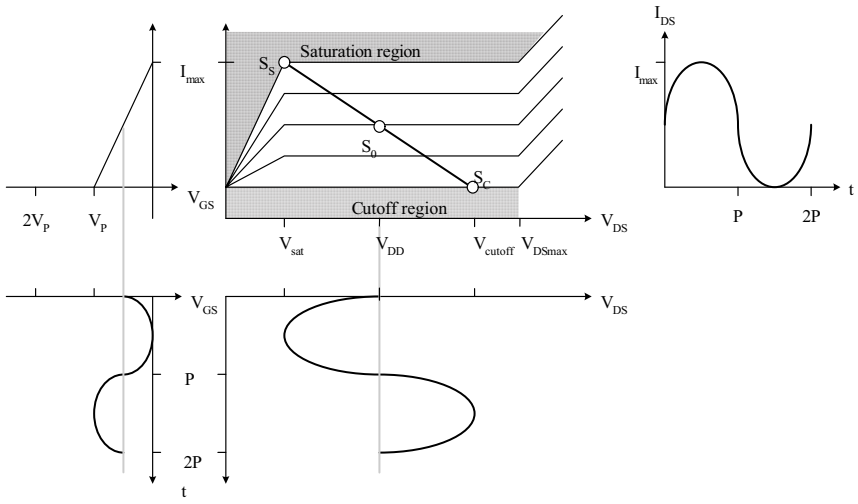


Figure 3-2. Class A load line and bias point (S_0)

Ideally, the current and voltage signals are going to be considered sinusoids. The class A amplifiers keep the transistor "on" for the entire cycle period (2P) for a given the input signal (V_{GS}). Therefore, the collector current (I_{DS}) is a constant and continuous wave like an ideal current source.

The tuning circuits used in the output collector are not inherent to their operation mode. They are used to assure an adequate harmonic suppression of the output signal and to guarantee an output impedance matching. In any case, the power analysis done in this chapter does not take into account the harmonic effects, as they have a minimum influence over the efficiency rate.

The ideal output power of any class A amplifier is shown in (3.5).

$$P_0 = \frac{v_L^2}{2R_L} \tag{3.5}$$

v_L is the voltage of the R_L load and it can not be higher than the V_{DD} supply voltage (*Figure 3-3*).

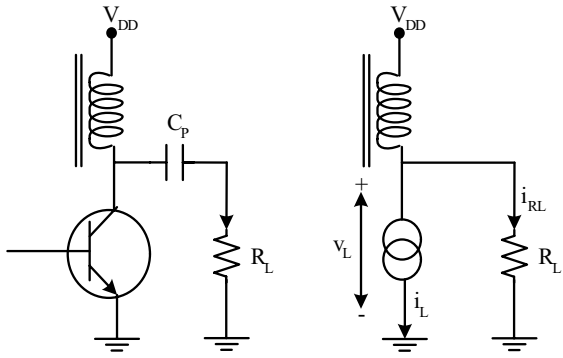


Figure 3-3. Class A amplifier and the equivalent circuit

The theoretical collector efficiency of class A amplifiers hovers around the 50%. However, the real implementations are not able to force the maximum operating level and the efficiency is reduced around the 25%, even up to the 15% depending on the linearity requirements.

On the other hand, the instantaneous efficiency, for the theoretical case, is proportional to the output power, and the average efficiency is inversely proportional to the peak to average ratio. The utilization factor is 1/8.

The gain factor of class A amplifiers is usually 3 or 4dB greater than in class B, and much more greater than in class C, due to the effective transconductance in class A is twice the class B.

As far as frequency response is concerned, the class A amplifiers can work almost to the frequency limit of the transistor device. That is the reason why they are used in microwave applications.

This operation mode is also necessary in high linear amplification. The inherently linear performance of class A amplifiers entails, unfailingly, the improvement of the nonlinear distortions according to the output power level in a monotonous way.

Some typical applications of class A amplifiers are:

- Classic linear applications, like SSB voice transmitters with third order intermodulation distance (I_{2T}) requirements of 40dB.
- TV transistors, highly sensitive to the compression of synchronism impulses.
- TV repeaters or transponders and CATV amplifiers with 60dB third order intermodulation distances (I_{2T}).

4. CLASS B

In class B configurations the transistor only amplifies one semi-cycle of the input signal, so the conduction angle is 180º and the quiescent current is zero in absence of input signal. The collector current, moreover, remains constant and it increases as the output power increases.

Since the collector current is proportional to the input signal amplitude, class B amplifiers provide considerably linear amplification. However, the on-off transitions of output signal generate important harmonic components.

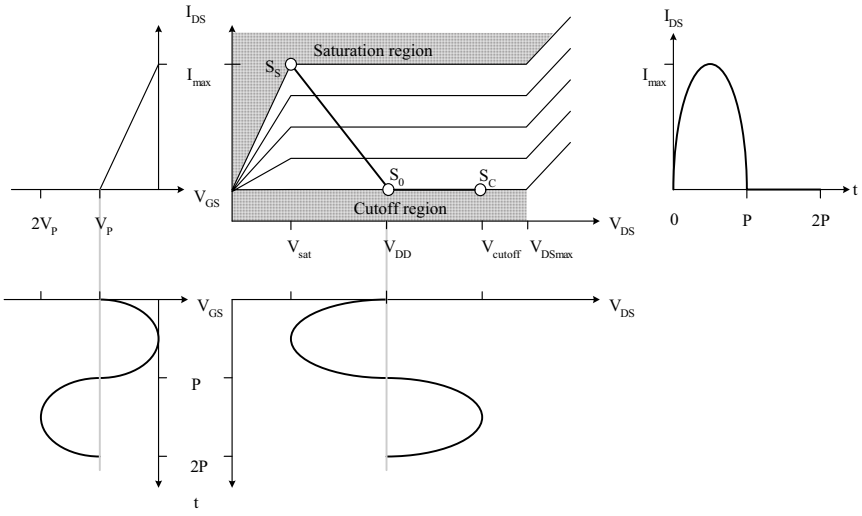


Figure 3-4. Class B load line and bias point (S_0)

If those harmonic components are out of band, which is very usual in RF amplifiers, they could be eliminated with filtering techniques (*Figure 3-5*). However, in low frequency applications with high bandwidth requirements the harmonic components fall into the band of interest and the filtering turns out unviable.

The load impedance is one of the effects which strongly affect the distortion and its nominal value for the fundamental frequency must be usually specified.

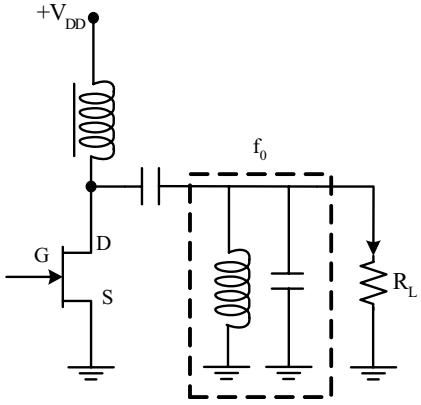


Figure 3-5. Class B load line and bias point (S_0)

Particularly, the load reactance has a high influence on the second harmonic components, often due to the transistor output capacitance. The second order harmonics of the collector current over the load reactance induces, inevitably, undesired voltage components which must be limited. If those amplitudes are, at least, the 10% of the fundamental voltage amplitude, the amplifier will be saturated to a lower power than it was expected, so the output power should have to be reduced in order to guarantee the distortion specifications.

This effect is usually avoided by adding an external capacitor that, despite the worsening caused in the gain and efficiency, it has much better effect on the distortion performance. As an experimental design rule suggests, the reactance of both the internal and the external capacitances of the second harmonic component should be 2.2 times the load resistance of the fundamental frequency component.

There are other solutions for wideband amplifiers, such as the coupled chokes, which ground the even harmonic components and keep high impedance for the fundamental frequency.

The theoretical instantaneous efficiency for the peak envelope power (PEP) of an ideal class B amplifier is 78.5%, though the real value is reduced up to the 70% in VHF applications. This efficiency reduction is due to three main reasons.

- The ohmic losses attached to the input and output matching circuits
- The distortion specification

- The quiescent collector current in bipolar transistors, really, is not zero, and usually it is around the 2% of the maximum collector current. In MOSFET transistors, this value increases up to 12%

Finally, the utilization factor is the same as in class A (0.125) although gain values are lower and, moreover, frequency dependent.

5. CLASS AB (OUTPHASING)

All the linearity requirements are not fulfilled with the class B amplification. This is the reason why new amplification architectures are developed completing the output signal cycle and improving, consequently, the distortion levels. This is possible by amplifying, with a complementary transistor, the negative semi-cycle of the input signal and coupling it, in a precise way, with the positive one. Those amplifiers are known as outphasing amplifiers.

Two configurations are mainly used: push-pull amplifiers, with two transistors connected with a transformer, and complementary configurations, with two transistors connected directly as shown in *Figure 3-6*.

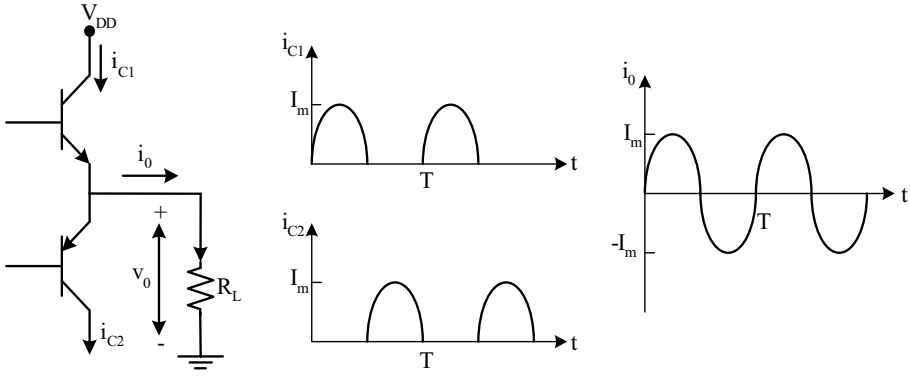


Figure 3-6. Outphasing operating principle (complementary configuration)

The outphasing principle reduces drastically the even harmonics of the output signal but introduces a crossover distortion due to the poor performance close to the cut-off point . In *Figure 3-7* this effect is displayed and its influence on the outphasing configuration output.

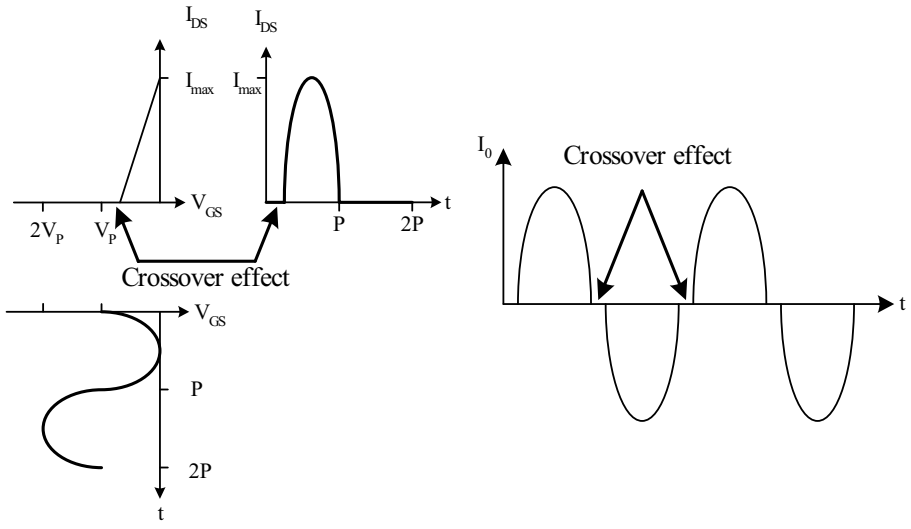


Figure 3-7. Crossover distortion in class B amplifiers

The class AB amplification mode, half-way between class A and class B, is considered as an improvement of the outphasing configuration (*Figure 3-8*). The bias point is fixed so the amplifier output becomes zero for less than one semi-cycle, and the conduction angle is slightly greater than 180º and much smaller than 360º.

Any of those situations occur when the bias point is lower than the input peak amplitude and the quiescent point is closer to the cut-off point than to the saturation point.

This amplification mode is suitable for linear amplification but without the strict requirements of class A amplifiers, for example…

- Single side band (SSB) final stage amplifiers ($I_{2T} \approx 30$dB)
- Final stages of TV transmitters with gain compression requirements
- Final stages of mobile communication base stations

In HF and VHF amplifiers the efficiency at maximum output power is around 60-65%. However, if the intermodulation specification is high enough, the mean efficiency falls up to 40%. The higher frequencies are used, the lower efficiencies are achieved.

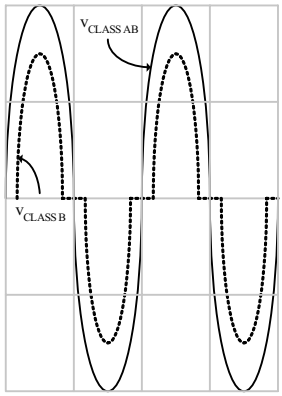


Figure 3-8. Crossover improvement with the class AB amplifiers

The gain factors of class AB amplifiers are intermediate between those achieved in class A and class B modes.

The push-pull configuration is very useful in wideband high frequency applications, but unfortunately, the complementary configuration is not suitable due to the inefficient performance of PNP transistors. The complementary solutions are limited to audio applications and to low and medium frequencies, though in integrated circuit technology up to 1GHz have been achieved.

6. CLASS C

The raison d'être of this amplification mode lies in the high power thermionic valve amplifiers. The transistors are "on" for less than a semi-cycle and inversely biased (*Figure 3-9*).

Class C transistors are deliberately brought to a nonlinear situation, so the efficiency, almost theoretically, hovers around the 100% as the conduction angle approaches to zero. Unfortunately, this technique involves an inevitable gain reduction as input power tends to infinite values. A recommended trade-off is achieved with a 150º conduction angle, resulting in an 85% efficiency rate.

The class C could be considered as a switching device that minimizes the ohmic losses with a very small conduction angle. The fundamental frequency of the impulsive periodic output signal is filtered with a tuning

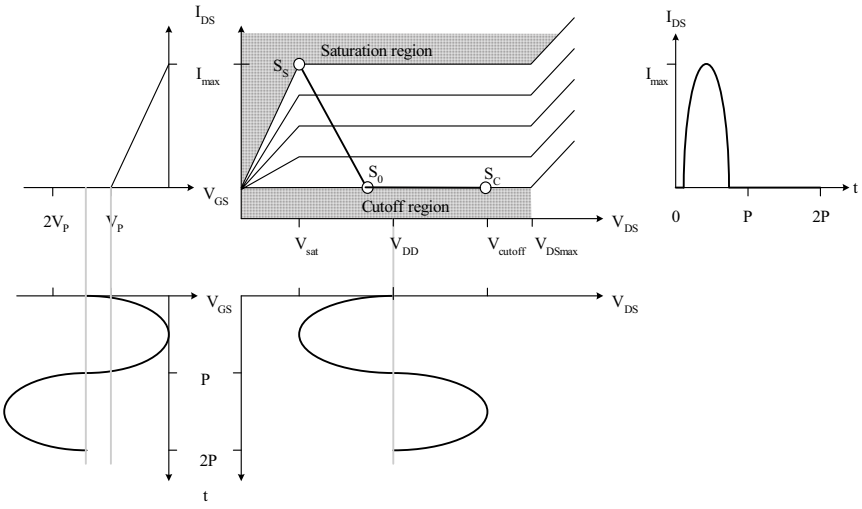


Figure 3-9. Class C load line and bias point (S_0)

circuit at the output collector. It consists of a parallel resonant circuit that grounds the harmonic components of the pulsating signal, avoiding undesired induced voltages over the output load.

If the amplifier is brought to saturation, the efficiency is stabilized and the output voltage is fixed by the supply voltage (V_{DD}). It allows a high level linear amplitude modulation since it acts as the modulator signal.

The class C is used in RF signal amplification with low linearity requirements, i.e., the mobile telephony (GSM), the mobile radiotelephony (narrowband FM) and the FM broadcasting transmitters (broadband FM). This technology is excellent in vacuum tubes but it is impracticable in solid state amplifiers, particularly with bipolar transistors. Their breakdown voltage is too low to support the inverse bias voltage of the class C mode. In some exceptional applications, an auto-biasing architecture is used base on an alternating current-direct current (AC-DC) conversion of the RF input signal (*Figure 3-10*).

The MOSFET transistors are much more suitable for this operation mode. The gate is biased to zero although the gain factor is slightly reduced (some few dB). The advantages, actually, lie in the outstanding efficiency rates.

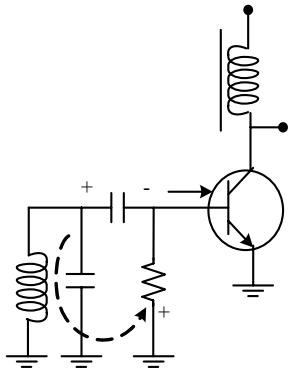


Figure 3-10. RF auto-biasing

7. SWITCHING AMPLIFIERS

The operating modes from D to S cover a wide spectrum of switching amplifiers. Their operation principle is based on the fact that the amplification is done only in the cut-off (On) and saturation (Off) regions, each one attached to one half-cycle (*Figure 3-11*).

The main goal of those techniques is the reduction of the switching time, in order to diminish the energy dissipation.

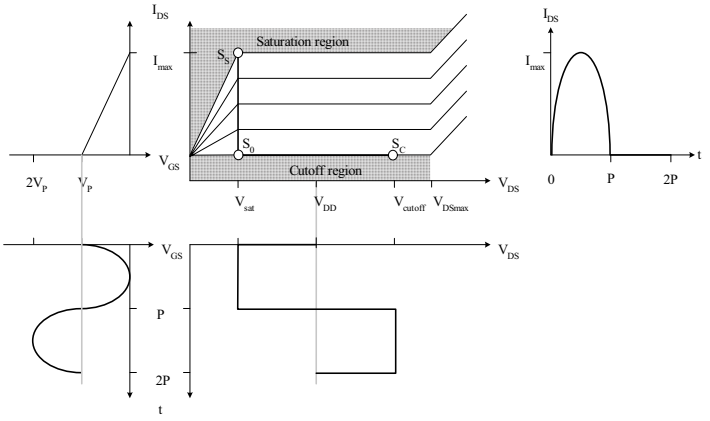


Figure 3-11. Operation principle of switching amplification

The switch rates of class D, E and F amplifiers match up with the carrier frequency, turning out as the most suitable switching amplification modes

for RF applications. Due to their inherent nonlinearity they are used in FM, CW or pulse width modulations (PWM).

As far as the semiconductor field is concerned, the field effect transistors have been used for applications up to a hundred MHz.

As regards the output power, it diminishes as frequency increases, though 10kW are achieved in medium wave and 1kW in short wave transmitters, while efficiency hovers around the 80% to 90%.

7.1 Class D

Class D amplifiers consist of two or more transistors arranged in a switching configuration and constituting an ideal bipolar switch. The theoretical output signal is a voltage or current square wave where the fundamental frequency is filtered. The *Figure 3-12* shows a class D amplifier in a complementary configuration.

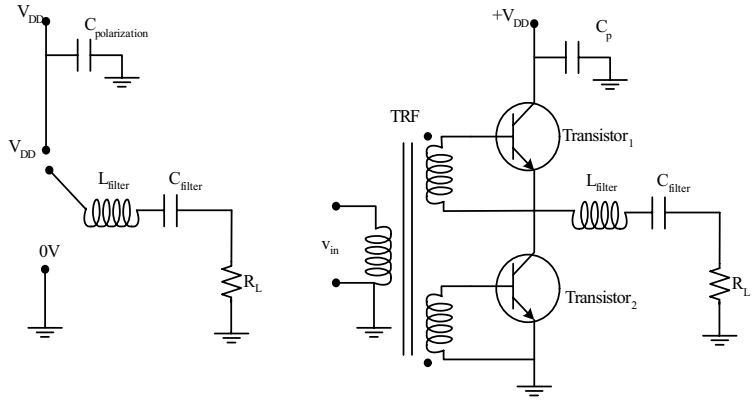


Figure 3-12. Class D amplifiers in a complementary configuration

In a final amplification stage the input waveform can be both square and sinusoidal. In the second case, it must be strong enough to saturate and cut-off the transistor in the precise time gap.

Then, a resonant circuit filters the fundamental frequency component that generates an output power (P_{OUT}) over the load resistance (R_L).

$$P_{OUT} = \left(\frac{8}{\pi^2}\right)\frac{V^2_{DD}}{R_L} \tag{3.6}$$

The utilization factor (0.159) is the highest among the power amplifiers, and if the switch rate is fast enough, the efficiency (ideally 100% in push-pull

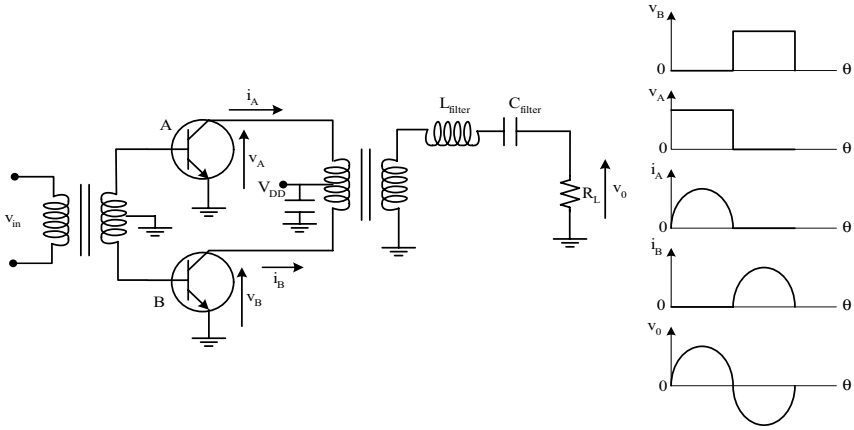


Figure 3-13. Class D amplifiers in a push-pull configuration

configurations) will not be degraded. This property is unique in class D amplifiers.

Some ohmic losses are unavoidable, attached to the saturation resistance, the switching rate and the collector capacitance. A finite switch rate obliges the transistor to remain in the active region long enough, turning into an unwanted power loss. In addition, the collector capacitance is charged and discharged once per cycle, resulting in undesired losses proportionally increased with the frequency and the DC source (V_{DD}).

It is also necessary to keep a correct phase shift between both input signals in order to guarantee a good efficiency rate.

Some designs have been done for high frequencies (HF) up to 1kW, but rarely do they exceed the VHF lower frequencies. Nevertheless, this technology has recently been tested with up to 1GHz by replacing the phase shift transformer with an equivalent integrated circuit that performs the same function.

On the other hand, special care should be taken if reactive loads are used. They may cause undesired phase shifts and unexpected inverse voltages in the transistors. The bipolar transistors are delicate when are faced with a polarity inversion and usually they are protected with diodes in anti-parallel configuration between the emitter and collector ports.

A last but not less important characteristic of these amplifiers is the parasitic capacitance of the transistor output, due to its own collector capacitance and to the output filter parasitic capacitance. If the saturation resistance is not equal to zero, whenever a transistor goes in or out of conduction, this capacitance charges and discharges through this resistance resulting in an unnecessary power consumption and heat dissipation.

7.2 Class E

The class E was developed recently in 1976, but its usefulness was not recognized until 90's, when the mobile technology development required high efficiency amplifiers.

It consists of a single transistor conceived as a switch. Theoretically, this unilateral configuration does not have any phase shift requirement between the push-pull amplifiers, as in the class D configuration. However, the potential of this amplification mode lies in the fact that the power consumption is only due to active region current.

The transistor input waveform must be perfectly squared in order to obtain the maximum efficiency and must have a 50%-50% duty cycle, falling practically to zero in the cut-off semi-cycle.

The looses attached to the class E mode are, in some cases, up to 2.3 times lower than in the class B or C modes for the same transistor, frequency and power conditions.

The basic operating principles can be explained with the *Figure 3-14*. In the cut-off state, the current of the DC choke keeps charging the C_p capacitance while, simultaneously, some current (i_L) begins to flow towards the load through the L_F-C_L resonant circuit. Once the C_p voltage is approximately three times the DC supply voltage (V_{DD}), the current, that continues flowing towards the load, starts diminishing this voltage. For the maximum efficiency, the C_p voltage must have fallen near to zero just when the transistor comes into conduction again, avoiding any power waste due to the C_p capacitor charge through the saturation resistance.

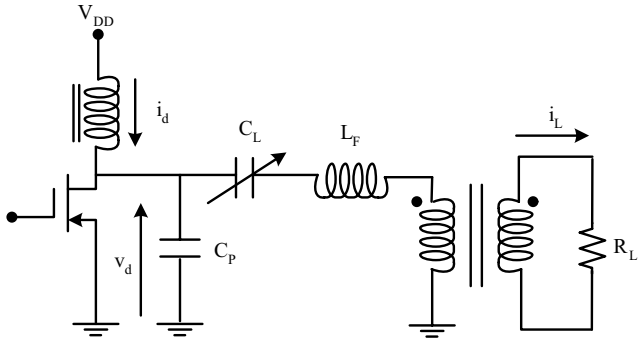


Figure 3-14. Class E basis architecture with a MOSFET transistor

The waveform of the drain voltage (v_d), shown in *Figure 3-15*, is a combination of the RF signal and the DC component of the C_p voltage.

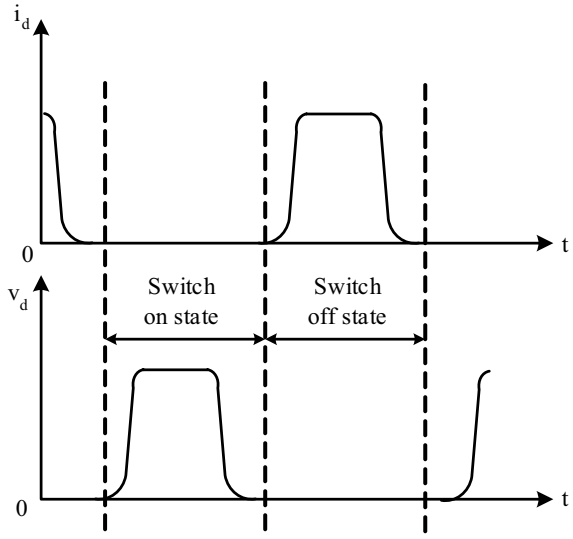


Figure 3-15. Ideal current and voltage waveforms in a class E amplifier

In an ideal class E amplifier, the drain voltage falls to zero and keeps being zero until the transistor begins to conduct again. As a consequence: the theoretical efficiency is 100%, the drain capacitance losses of the class D amplifiers are eliminated, the switching losses are reduced and the tolerance toward the component nonlinearities is good enough.

The high efficiency of class E amplifiers is the reason why they are suitable for several applications. Low cost MOSFET amplifiers, up to 1kW, have been implemented for commutation applications rather than for RF.

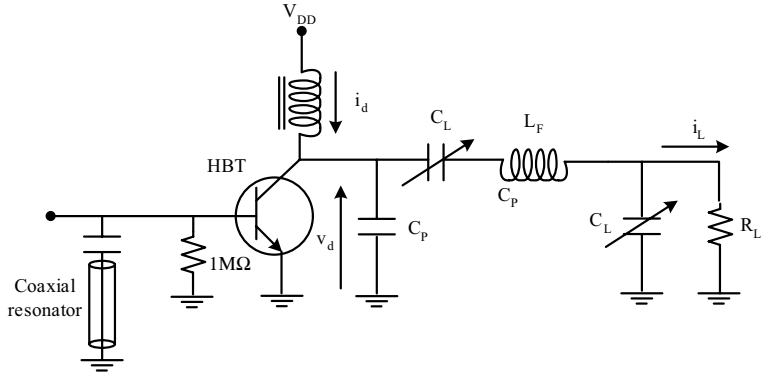


Figure 3-16. Enhanced class E amplifiers with a coaxial resonator and HBT transistor

Some improvements can be suggested in order to maximize the efficiency values, like the harmonic resonators of *Figure 3-16*. The transmission line (TEM) in the input port will perform like an open circuit for the fundamental frequency and like a shortcut for all the undesired harmonics.

7.3 Class F

Class F amplifiers were one of the first attempts to improve the efficiency of the nonlinear amplifiers. The transistors perform like a current source (saturation), with some harmonic resonators used in order to make the collector waveform square like. It is fairly similar to the enhanced class E amplifiers of *Figure 3-16*.

As it is known, the more harmonic components are used, the more squared the signal will be, and the higher efficiency rates will be achieved.

In order to improve the squared waveform, the harmonic components should take a much more specific value, so the output load should be 2 to 10 times higher for the harmonic components than for the fundamental one.

As the output filter of the class F amplifiers is much more complicated than in any other one, the output impedance will be suitable only for certain frequencies. Concentrated elements are used for low frequency applications and transmission lines for microwaves (Ka band up to 30GHz). An example is shown in *Figure 3-17*.

Some several alternatives are used among the C, E and F operating modes, but in all of them the maximum attainable efficiency rates and utilization factors depend on the harmonic components that are used (1 to 5) and on the output impedance value for each one, respectively.

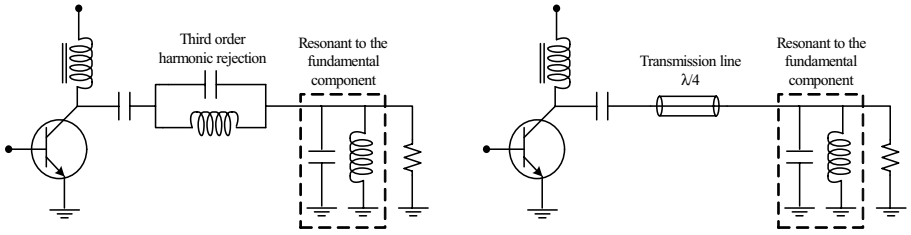


Figure 3-17. Class F output circuits: concentrate elements and transmission line

8. MORE OPERATING MODES

Many more operating modes exist than those which have been reviewed in this chapter: S; G; H... However, many of them are simple variations of the ones showed here, or they do not have specific usefulness in radio frequency applications, so they receive no more than just a mere mention.

Chapter 4

LINEARIZATION TECHNIQUES

As it has been noticed, actual and incipient communication systems require high linear amplification in their transmitters. Presently, high linear amplifiers bring with them reduced efficiency rates, so special circuit architectures have came out in recent years as alternative solutions.

The aim of this chapter lies in the compilation and description of the linearization techniques used in power amplification, emphasizing their capacities while differentiating the suitable applications for each case.

1. CLASSIFICATION OF THE LINEARIZATION TECHNIQUES

Contrary to expectations, linearization techniques are not as novel as they seem to be. In fact, some of them were proposed before the transistor was used as an amplifier device, mainly in analog transmitters of the AM broadcasting.

In recent years, the increase of the high efficiency demands have result in specific efforts to best use the electromagnetic spectrum, involving high linear amplification and, consequently, entailing efficiency reductions. This has culminated in an ongoing research on linearization techniques focused on higher frequencies, bandwidths and efficiency rates.

The classification of linearization techniques, according to one of the most generalized tendency, can be divided into two groups.

In the first group an input signal is amplify while distortion components are generated. The aim of these techniques lies in canceling such distortion modifying the input signal or directly subtracting it from the output signal. These linearization techniques are:

- Feedback
- Predistortion
- Feedforward

The linearization techniques of the second group are commonly known as efficiency enhancement techniques. Load impedance modifications, supply source variations or output signal combinations are used in order to maximize the overall transmission efficiency or likewise, the linear performance. Some of those techniques are:

- Bypassing
- Envelope Elimination and Restoration (EER)
- Envelope tracking
- Linear amplification using Nonlinear Components (LINC)
- Doherty Method
- Combined Analogue Locked Loop Universal Modulator (CALLUM)

Alternatively, the linearization can be differenced whether baseband or RF signals are used as input signal. In the first case, the linearization techniques can use any of the modulation techniques and make any frequency translations whereas in the second case they do not.

2. FEEDBACK

Under this category are grouped a set of linearization techniques based on the well known negative feedback architecture. Some of the most important are described in the following subsections.

2.1 RF Feedback

This technique is based on the standard Feedback scheme used in the low frequency amplification and control systems (*Figure 4-1*). Following this scheme, as shown in equation (4.1), the gain (G) of the amplification system is determined by the own gain of the main amplifier (A) and the feedback gain (β).

$$G = \frac{A}{1 + \beta \cdot A} \tag{4.1}$$

In the Feedback loops the distortion reduces as the gain factor does, so a trade-off must be achieved between them.

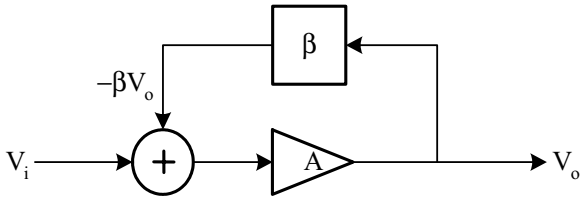


Figure 4-1. Feedback basic architecture

One of the inherent disadvantages of this technique is the stability, so it is impossible to assure a proper functionality against component aging.

Another one is the bandwidth reduction due to the feedback loop delay, reserving this technique for low frequency and narrowband applications.

These limitations initiated the emergence of new architectures based on the same concept, but comparing the low frequency signals instead of the RF ones. Those new techniques require the demodulation of the amplified signal, so the implementation is more complicated.

2.2 Envelope Feedback

This linearization technique corrects the amplitude distortion of the output signal and compensates the gain variation of the main amplifier. As shown in *Figure 4-2*, both input and output envelopes are detected and compared with a differential amplifier that simultaneously acts as an automatic gain control (AGC) over the main amplifier.

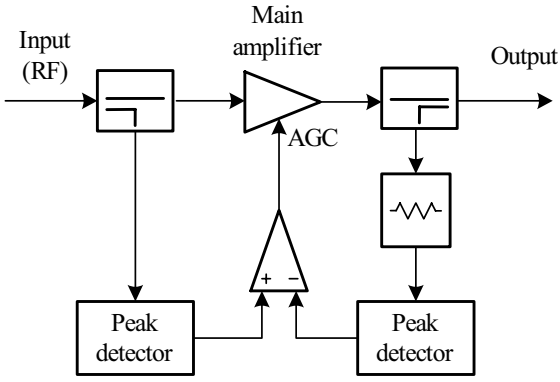


Figure 4-2. Envelope Feedback architecture

The peak detectors must do high precision measurements in order to avoid instabilities and distortion components. Hence, the dynamic range

must be high enough and the bandwidth twice as much as the amplified signal bandwidth.

This technique makes the phase error worse in an attempt to improve the output amplitude, and is usually used with linear amplifiers (class A, AB, B).

2.3 Envelope and phase Feedback

The disadvantages of the Envelope Feedback technique have led to keep on improving the basic schemes, resulting in the Envelope and Phase Feedback technique (*Figure 4-3*). As its name suggests, phase unbalances are also rectified, though precision and bandwidth requirements are still mandatory.

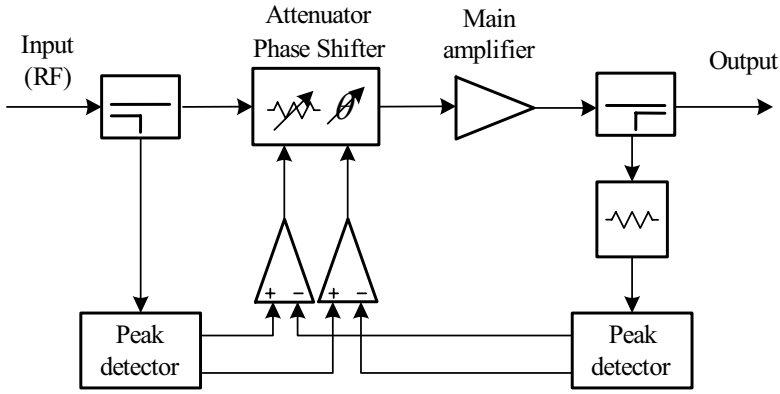


Figure 4-3. Envelope and phase Feedback architecture

2.4 Polar Loop

The Polar Loop is an alternative technique for both amplitude and phase adjustments but with intermediate frequency (IF) signals instead of the RF ones (*Figure 4-4*).

In this case, the phase and amplitude magnitudes of the input and output signals are compared. The amplitude comparison adjusts an automatic gain control (AGC) while the phase comparison acts over the voltage control oscillator (VCO).

The linearity characteristics of these Feedback amplifiers are good: 50dB distortion distance (I_{2T}) for 50% efficiency. However, the design and the component selection are critical because of their influence on the frequency

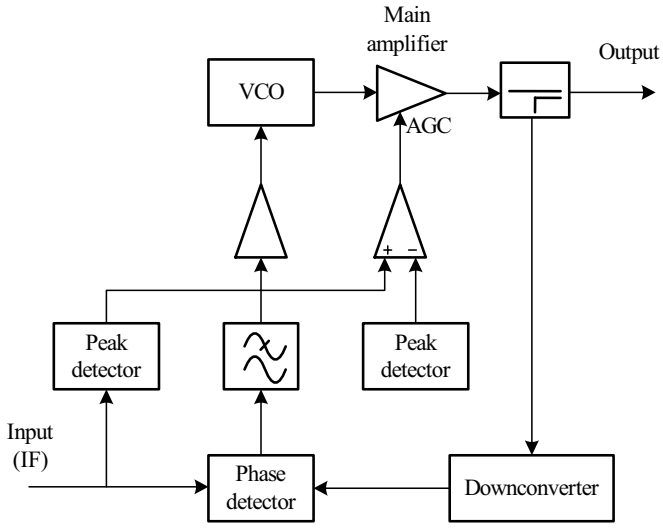


Figure 4-4. Polar Loop architecture

conversion process, the linearity performance and the unbalances of the signal comparators [6].

The bandwidths of the differential amplifiers are also critical parameters. The output phase detector bandwidth must be 5 to 10 times higher than the bandwidth of the envelope detector. In any other case, distortion components will fall far away from the phase detector range and they will be impossible to subtract.

2.5 Cartesian Loop

The Cartesian Loop is an improvement on the Polar Loop linearization technique**,** developed by Petrovic and Smith [6], but with baseband signals.

It was not an interesting linearization technique until the digital systems were developed and the I/Q baseband signals were digitally generated. Nowadays, this technique is suitable for single carrier applications and is the most promising of Feedback techniques [7-10].

The basic scheme is shown in *Figure 4-5*. In this technique there are not comparisons between the input and output amplitude and phase magnitudes, but the output signal demodulation results in a critical step that must be carefully done.

Despite the implementation difficulties, the Cartesian Loop technique is widely standardized, among other things, owing to the low production costs. The linearization improvement depends on the gain and bandwidth

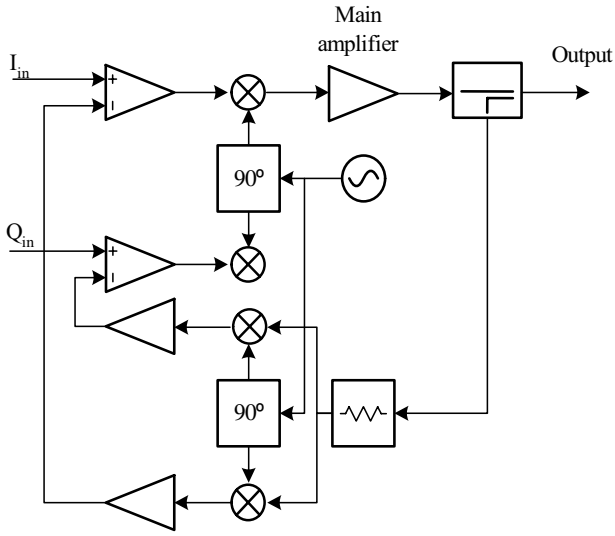


Figure 4-5. Cartesian Loop architecture

characteristics of the differential amplifiers and, above all, on the linearity of the demodulator device.

The Application Specific Integrated Circuits (ASIC) developed for Feedback amplifiers open up new possibilities in the Digital Signal Processing (DSP) field. Despite the increase of the production costs, DSP based devices allow the use of simultaneous linearization techniques, i.e., Cartesian Loop together with Predistortion techniques, improving consequently the overall linearity.

Ideally, this technique achieves 20dB to 45dB cancellation levels of the intermodulation distance (I_{2T}) with 35% to 65% efficiency rates. Really, those values get worse as the signal bandwidth increases: a 5MHz feedback loop with 500kHz gain bandwidth achieves, theoretically, a 20dB distortion cancellation. This is the reason why this technique is not suitable for wideband signals and it is usually used in TDMA or FDMA narrow band single carrier applications.

3. PREDISTORTION

The Predistortion linearization technique, as it name suggests, modifies the input signal according to the non linearity characteristics of the main amplifier transfer function. The basic architecture is shown in *Figure 4-6*.

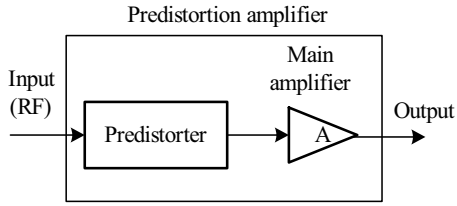


Figure 4-6. Predistortion architecture

The open loop architecture turns the Predistortion technique into an unconditionally stable system. The input signal can be predistorted in a constant way or dynamically with feedback loops. The former is, though much simpler, ineffective against environmental changes and aging effects.

In the next subsections, three particular Predistortion techniques are summarized:

- RF Predistortion
- Envelope Predistortion
- Baseband Predistortion

3.1 RF Predistortion

This technique is regularly used in high power amplifiers [11] and often together with other linearization techniques. It entails inserting one or more nonlinear devices – predistortion devices – between the input signal and the main amplifier (*Figure 4-6*). The transfer function of the distortion signal of the predistortion device must be exactly opposite to the transfer function of the distortion signal of the main amplifier. As a result, the total gain of the distortion components will be 0dB. This effect is shown in *Figure 4-7*.

This technique uses constant predistortion factors, so the main amplifier transfer function must be thoroughly characterized. Alternatively, new dynamic adjustment methods have been proposed [12] avoiding the influence of aging and environmental effects.

Simple analog predistortion devices are used for the correction of the third order intermodulation products, but if several intermodulation orders would be cancelled (5, 7…), more complex devices should be developed [13].

3.1.1 Simple analog Predistortion

The simple analog Predistortion technique avoids only the third order intermodulation products using linear components as predistortion devices. Initially, different configurations were used only with diodes and resistances,

but in the course of the years new improvements have appeared based on the field effect transistors [14].

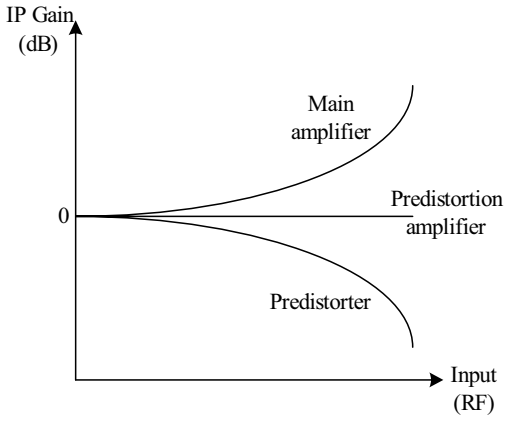


Figure 4-7. Gain factor of the intermodulation product after Predistortion

The equation of a nonlinear transfer function has been introduced in Chapter 2. Now, this formula is remembered with similar annotation but only with the odd distortion components (4.2).

$$V_{nl} = a_1 v_{in} + a_3 v_{in}^3 + a_5 v_{in}^5 + ... \tag{4.2}$$

The transfer function of the predistortion device must have only the nonlinear terms of the power amplifier transfer function (a_3, a_5 …), so the linear term (a_1) should be eliminated. One of the architectures used in the subtraction of the linear term of a nonlinear device is shown in *Figure 4-8*.

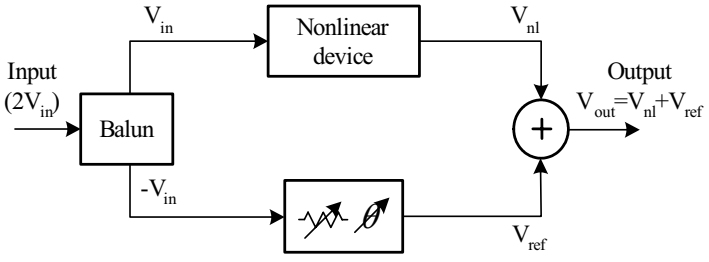


Figure 4-8. Linear component subtraction from a nonlinear device

The variable attenuator and phase shifter guarantee the equation (4.3), subtracting the linear response completely (4.4).

$$V_{ref} = -a_1 v_{in} \tag{4.3}$$

$$V_{out} = a_3 v_{in}^3 + a_5 v_{in}^5 + \ldots \tag{4.4}$$

Now the nonlinear device has changed to a "pure" nonlinear device. *Figure 4-9* shows the solution proposed by [15] that completely cancels the third order distortion generated by the main amplifier.

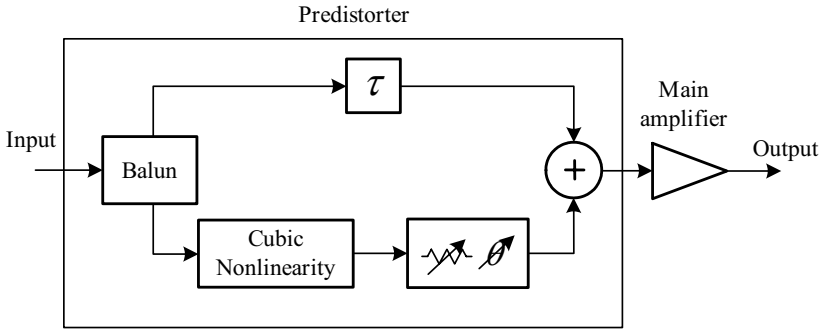


Figure 4-9. Simple analog Predistortion architecture

Despite the low production costs, this technique has three important disadvantages, well reported in [16]: the AM-PM conversion is worsened, the input power must be constant and the frequency response is narrowband.

These disadvantages have caused the irruption of the Compound Predistortion.

3.1.2 Compound Predistortion

The Compound Analog Predistortion tries to subtract all the distortion components (a_3, a_5 …) separately [17]. Usually, feedback controllers are used in the phase and amplitude adjustments for the subtraction of each distortion component.

The production costs and the implementation complexity are high, and furthermore, the frequency dependence prevents linearizing wideband signals.

3.2 Envelope Predistortion

The Envelope Predistortion technique consists of dynamically changing the input signal phase and amplitude parameters, by means of variable

attenuators and phase shifters, in order to compensate the distortion effects of the main amplifier (*Figure 4-10* and *Figure 4-11*) [16].

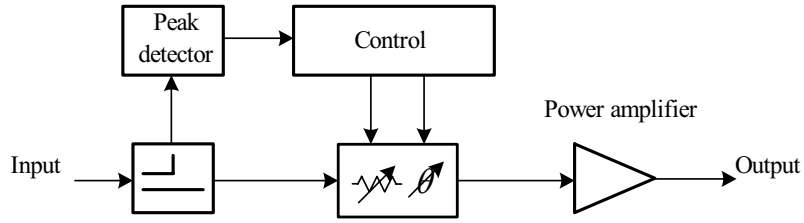


Figure 4-10. Envelope Predistortion with analog or DSP controller

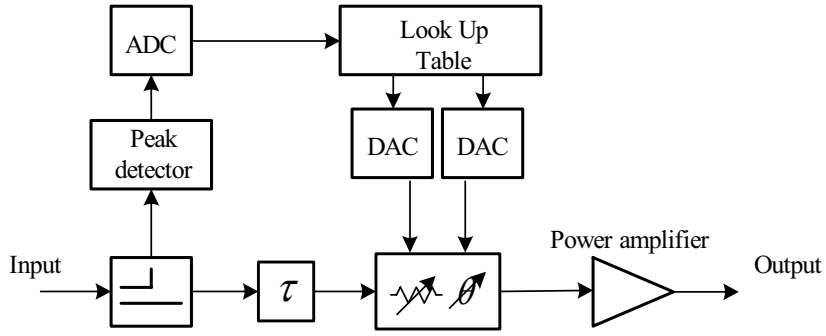


Figure 4-11. Envelope Predistortion with a LUT controller

Analog controllers, DSP and Look Up Tables (LUT) are used in this linearization technique [18]. The implementation complexity is low and the distortion is cancelled between 5 to 15dB. On the other hand, the system bandwidth is limited by the control frequency, necessarily 10 times greater than the modulator frequency [19].

The control algorithm is designed for one specific amplifier, so the aging or environmental effects could avoid the linearization system responding appropriately.

3.3 Baseband Predistortion

The Baseband Predistortion techniques, as its name suggests, use the input baseband signals instead of the RF signals. Three Baseband Predistortion techniques are the most developed: Mapping Predistortion, Complex Gain Predistortion and Analog Predistortion [20-22].

The Mapping Predistortion technique uses a large LUT in order to distort the original I/Q signals and a DSP to consider the memory effects of the main amplifier [23].

Despite the smooth running of this technique, its adaptability directly depends on the memory required for dynamic control [24, 25].

Open loop (*Figure 4-12*) and adaptive (*Figure 4-13*) architectures can be found in the state of the art, and 25dB cancellation levels are achieved in some cases [26].

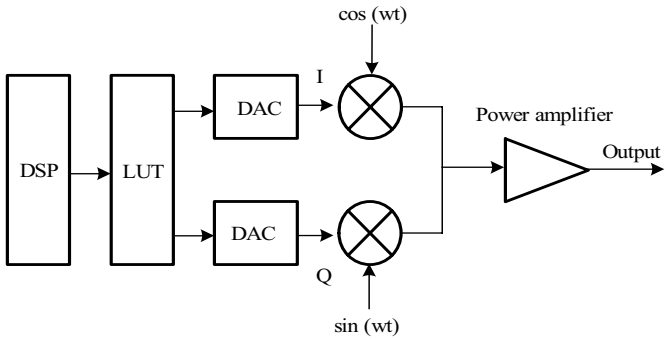


Figure 4-12. Open loop digital baseband Predistortion architecture

The Complex Gain Predistortion technique is derived from the mapping Predistortion technique and comes to minimize the amount of memory used during the control stage. In this case only the amplitude of the input signal is used instead of the I/Q signals**,** reducing the memory requirement and increasing the computational cost [21].

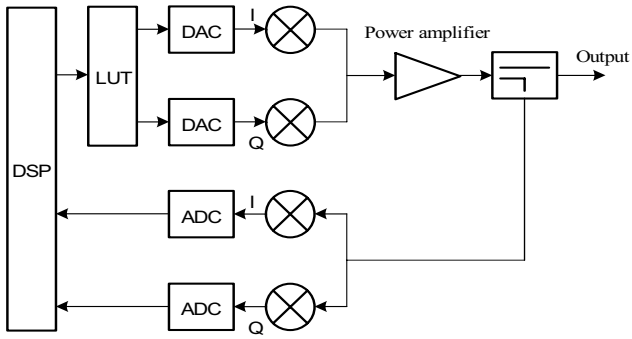


Figure 4-13. Adaptive digital baseband Predistortion architecture

On the other hand, the Analog Baseband Technique uses analog circuitry, sometimes controlled by DSP in the predistortion stage. The theoretical bandwidth of this architecture could be considered as wideband in comparison with the rest of the Predistortion techniques, sadly limited by the DSP frequencies.

The Baseband Predistortion techniques are easily implementable, low cost, and can be used in some wideband applications [27], but the I/Q signals must be available and the cancellation levels are not outstanding.

4. FEEDFORWARD

This technique was developed by H. S. Black in 1928 [28], previously to the Feedback technique, also developed by him in 1937. The Feedback technique is not unconditionally stable and the distortion cancellation is limited. The Feedforward technique however is unconditionally stable and the distortion introduced by the main amplifier can be, theoretically, completely subtracted. On the other hand, the Feedforward technique requires precise phase and gain adjustments, making the implementation really complicated. This was the reason why the Feedforward technique disappeared into obscurity in favor of the well-known Feedback technique for a long time.

With the evolution of technology, the frequency and bandwidth requirements began increasing, highlighting the Feedback limitations and recovering the Feedforward technique. In 1971 Seidel [29] linearized a TWT (Travelling-Wave Tube) amplifier, and later Bennett in 1974 [30], with the Feedforward technique. Feedforward became one of the linearization techniques most developed and implemented in audio systems, cable television and RF applications.

The basic Feedforward architecture is shown in *Figure 4-14*. The scheme consists of two cancellation loops. The first loop samples the amplified signal from the main amplifier output and extracts the distortion components, while the second loop subtracts those distortion components from the main signal [31].

This linearization technique is open loop and unconditionally stable, so theoretically suitable for any wideband application. High cancellation levels (33dB) are reported of the third order intermodulation products with the two tone procedure [32], depending on bandwidth and frequency parameters. This technique is widely discussed in Chapter 5.

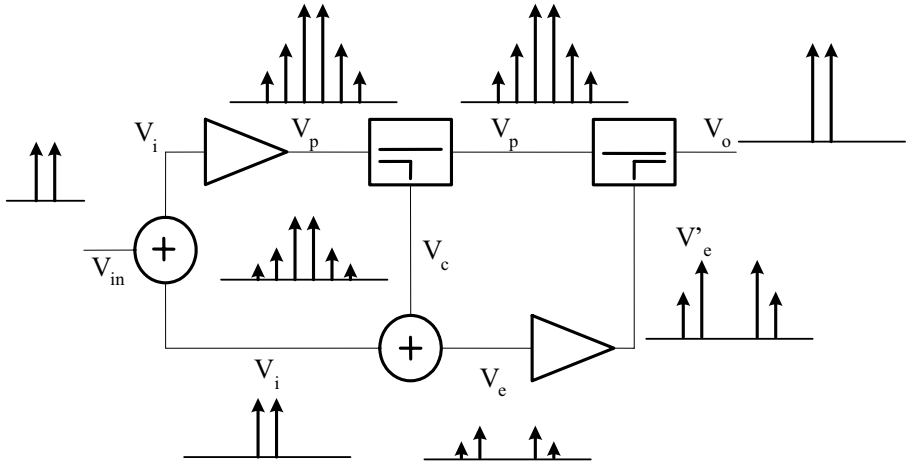


Figure 4-14. Basic Feedforward architecture

5. EFFICIENCY ENHANCEMENT TECHNIQUES

This group brings together those techniques focused on improving the efficiency of power amplifiers and maintaining their linearity performances [33]. Those techniques are:

- Bypassing
- **E**nvelope **E**limination and **R**estoration (EER)
- Envelope tracking
- **LI**near amplification using **N**onlinear **C**omponents (LINC)
- Doherty technique
- **C**ombined **A**nalogue **L**ocked **L**oop **U**niversal **M**odulator (CALLUM)

5.1 Bypassing

Normally, the headroom of power amplifiers is chosen according to the input Peak Envelope Power level (PEP). However, this point is hardly achieved most of the time, resulting in a detrimental efficiency reduction. The Bypassing technique uses independent amplifiers tailored for the normal and PEP stages (*Figure 4-15*).

When input signal level exceeds the nominal value, both amplifiers are connected between the input and the output ports. As the input power level decreases, the second amplifier and its power supply are disconnected, reducing the power consumption and maximizing the efficiency rate.

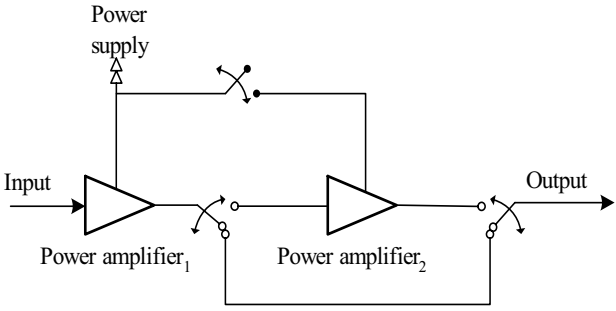


Figure 4-15. Bypassing technique switched for small signals

This technique is suitable for mobile devices, highly dependent on battery consumption. However, effects like hysteresis and commutations transients make this technique really complex.

A bypassing technique with an adaptive gate switching is shown in *Figure 4-16*.

The gate width of the upper transistor, and therefore its current and power capacity, is 20 times wider than the other one. The biasing is controlled in order to enable each transistor according to the power requirements. The electromechanical switches have been replaced by an electronic commutation, considerably reducing the switching losses to the detriment of the impedance matching.

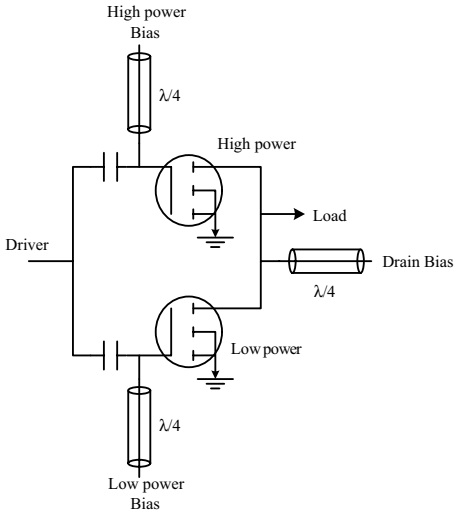


Figure 4-16. Adaptive gate switching

5.2 Envelope Elimination and Restoration

This technique, commonly known by its acronym (EER), came up as an extension of the high power AM modulators. As it has been seen in Chapter 1, in the first transmission systems the information was modulated in the envelope of the carrier signals, but as the modulation techniques improved, the information also was modulated as phase shifts, turning the EER into an useless technique.

The EER or Khan [34] technique combines a high efficiency nonlinear power amplifier with an envelope amplifier. The basic architecture consists of an envelope limiter that subtracts the amplitude modulation from the carrier signal. The resulting phase modulated signal is amplified with any of the high efficiency amplification techniques. Finally, the original modulation is restored with an RF amplifier-modulator device. The *Figure 4-17* shows the basic Khan technique applied to an RF input signal.

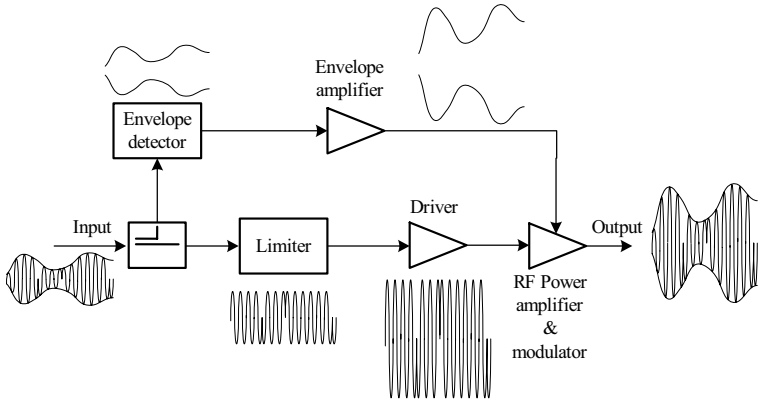


Figure 4-17. EER basic architecture with RF input signal

Recent technologies use digital signal processors (DSP) for the generation of both envelope and phase modulated signals. Unlike linear amplifiers, this technique allows a high efficient amplification in a wide power range, suitable for many signal types with different back-off requirements.

Two of the most important factors concerning the linearity performance are the envelope bandwidth and the adjustment between the envelope and the phase modulated signals. A rough design rule suggests that the envelope bandwidth must be, at least, twice the RF bandwidth, and the adjustment between the envelope and the phase signals must be higher than the 90% of the inverse bandwidth.

This technique has been widely used in single carrier or narrowband applications [35], constantly improved with new architectures [36-38].

5.3 Bias Adaptation

The Bias Adaptation technique is quite similar to the previous Khan technique. In this case the power supply of a high linear amplifier is dynamically biased according to the input envelope. The basic architecture is shown in *Figure 4-18*.

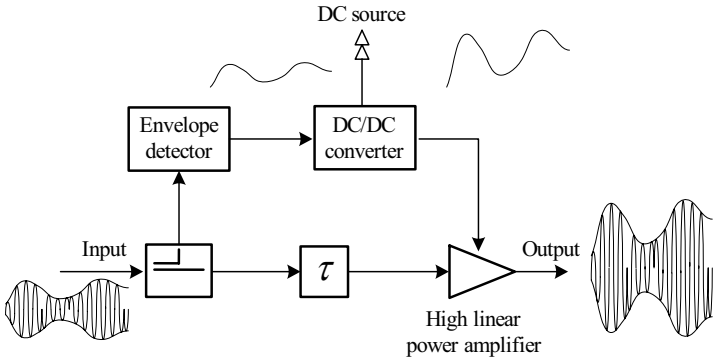


Figure 4-18. Bias adaptation technique

This technique has been used fairly often [39-41], despite the commercial amplifiers are not suitable for such biasing variations. The new implementations have use this technique in integrated circuit (IC) designs [42, 43].

The DC power consumption could be controlled with the biasing of the power supply transistor (class G), or with DC/DC converters (class H). The commutation frequency requirements of those converters are high, 10 to 20 MHz, and usually GaAs HBT even RF MOSFET transistors are used for these application-specific integrated circuits (ASIC).

5.4 LINC

In the 30's Chireix developed this linearization technique, known as Outphasing, with the purpose of improving the efficiency and cost diminishing of AM broadcasting transmitters [44]. This technique basically improves the amplitude distortion without considering the phase distortion.

Rediscovered by Cox [45, 46], this technique began to be well-known as Linear Amplification with Nonlinear Components (LINC). He modified the

original architecture so that the phase distortion could also be corrected, giving as result the scheme shown in *Figure 4-19*.

This technique is based on the equivalence (4.5) between an amplitude/phase modulated signal (x) and two phase modulated signals (S_1 and S_2).

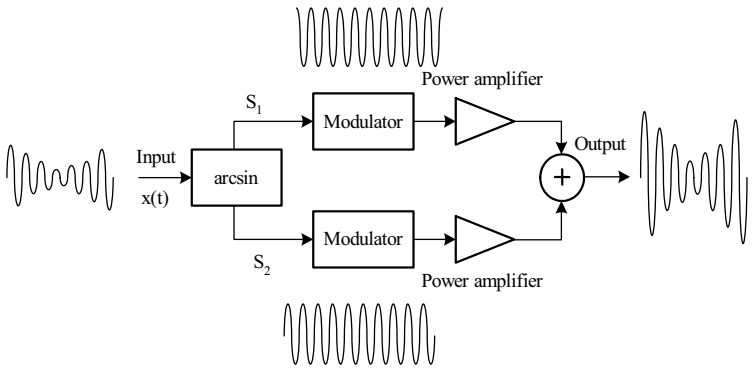


Figure 4-19. LINC technique architecture

$$
\begin{aligned}
&x(t) = E(t)\cos(\omega_c t + \Phi(t)) \\
&x(t) = \underbrace{\frac{1}{2}\cos\left(\omega_c t + \Phi(t) + \varphi(t)\right)}_{s_1} + \underbrace{\frac{1}{2}\cos\left(\omega_c t + \Phi(t) - \varphi(t)\right)}_{s_2} \\
&\varphi(t) = \cos^{-1}\left[E(t)\right]
\end{aligned}
\tag{4.5}
$$

The information to be transmitted is now phase modulated in two envelope constant signals, so high efficient amplifiers can be used without linearity restrictions. Finally, both phase-modulated signals are combined.

The main disadvantage of this technique was the generation of the S_1 and S_2 signals. Nowadays, the digital signal processors (DSP) allow their base band processing, before been modulated and up-converted in frequency [47] (see *Figure 4-20*).

The baseband digital decomposition into two envelope constant-phase modulated signals is not as efficient as in the analog case, so alternative factorization techniques have been proposed [26], as shown in *Figure 4-21*.

LINC is extremely sensitive to the signal path unbalances [48], so dynamic adjustment systems or auto-calibrating circuits are included in order to guarantee a perfect signal combining [49].

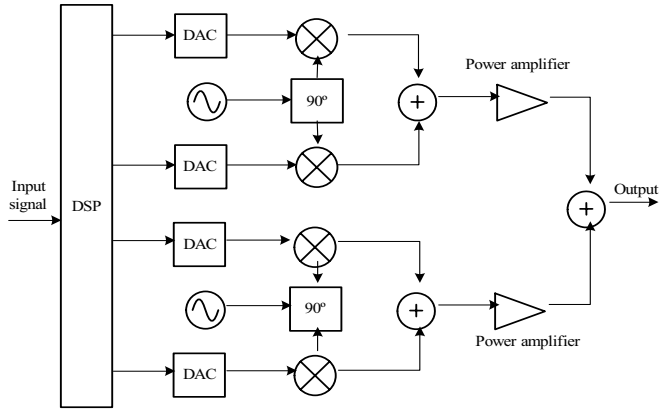


Figure 4-20. DSP based LINC architecture

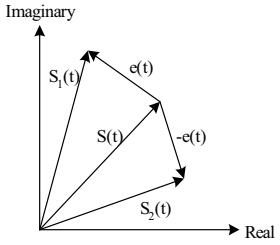


Figure 4-21. Input signal decomposition for digital LINC implementation

Two reasons make this technique an attractive research area: the good wideband performance and the combination possibility with other linearization techniques to mitigate the inherent distortion of this amplification technique. The main disadvantage of the analogical implementation is the bandwidth limitation of the input splitter.

5.5 Doherty Method

This method was developed by Bell Labs in 1936 [50]. The performance of any power amplifier is determined, among other things, by the load impedance, so this amplification technique is based on the load adjustment according to the input signal level.

The classic architecture combines two amplifiers with equal power performances, one of them (main amplifier) operating in class B and the

other one (auxiliary amplifier) in class C. The conduction angle of the auxiliary amplifier is one of the design parameters of this technique.

The main amplifier is "on" while the output signal does not exceed the half of the maximum amplitude. This point forward the auxiliary amplifier changes into conduction.

Signal splitting and combining is done with hybrid couplers, quarter wave lines or with combinations of those elements, as shown in *Figure 4-22*.

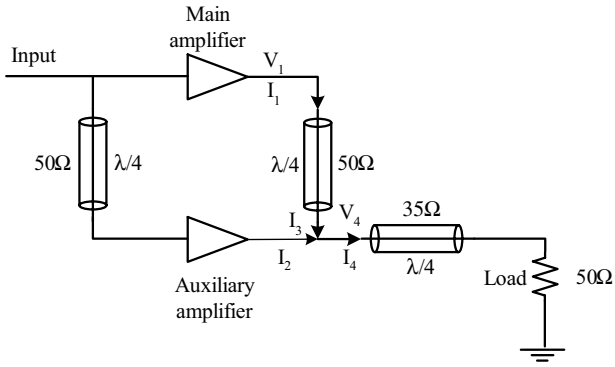


Figure 4-22. Doherty method

This technique has been improved with DSP for the power and biasing controls in order to improve the linearity performance. In the same way, more than two amplifiers could be used in order to keep a higher instantaneous efficiency for a wider dynamic range.

5.6 CALLUM

The CALLUM (Combined Analogue Locked Loop Universal Modulator) linearization technique was devised by Bateman in 1992 [51]. This technique combines two envelope constant signals generated with two feedback loops (*Figure 4-23*).

Two voltage controlled oscillators are tuned with the comparison of the input and output I/Q signals. As the information is phase modulated, nonlinear amplifiers are used with high efficiency rates. Nevertheless, the stability of the feedback loops only is guaranteed for one of the four quadrants [52, 53].

This technique has been successfully used in digital modulations [54, 55] with 60dB distortion cancellations and good stability.

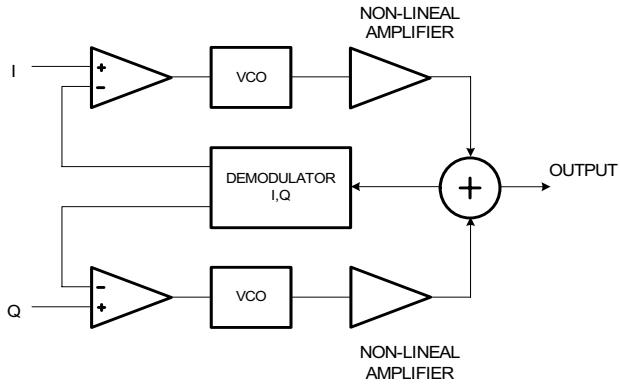


Figure 4-23. CALLUM architecture

The main disadvantage of this amplification technique lies in the unbalance of the VCO paths, which could destabilize all the linearization system.

6. COMPARISON OF LINEARIZATION TECHNIQUES

Different linearization techniques have been summarized in this chapter, each one with different advantages and disadvantages which make them suitable for specific applications. In *Table 4-1* some of the most important characteristics are compared in order to facilitate a correct selection depending on current application.

Table 4-1. Comparison of linearization techniques

Techniques		Complexity	Linearity	Bandwidth	Efficiency
Feedback	RF	Medium	Low	Low	-
	Envelope	Medium	Low	Medium	-
	Polar Loop	High	High	Low	-
	Cartesian Loop	High	High	Low	-
Predistortion	RF	Medium	Medium	Medium	-
	Envelope		Medium	Medium	-
	Baseband	High	High	Medium	-
Feedforward	Basic	High	High	High	-
Efficiency Enhancement Techniques	Bypassing	Medium	-	Low	Medium
	EER	Medium	-	Low	Medium
	Bias adaptation	Medium	-	Low	Medium
	LINC	High	-	Medium	High
	Doherty	High	-	Low	High
	CALLUM	High	-	Low	High

Those techniques can be implemented simultaneously, complementing them in order to improve any of the perform characteristics, like Feedforward with RF analogical Predistortion [56, 57], Cartesian Loop with Baseband Predistortion... and so on.

Chapter 5

FEEDFORWARD AMPLIFIERS

The Feedforward linearization technique was developed by H.S. Black in 1928 [28]. The most important characteristics of this technique in comparison with other common linearization techniques were the unconditional stability, the wideband signal linearization capability and the possibility to completely cancel the distortion components. However, the strict assembly requirements and the implementation complexity led this technique rapidly to the obsolescence.

The frequency and wideband requirements of emergent communication systems aroused the interest in the Feedforward technique. The first "new age" Feedforward implementations were initiated by Seidel in 1971 [29]. He used this technique in a TWT amplifier linearization, which was later developed by Bennett in 1974 [30]. From that time forward, as shown by [32, 58-60], the Feedforward linearization technique has become one of the linearization techniques most developed and implemented in audio systems, cable television and nowadays in RF, where every time the frequencies used are higher and the bandwidths wider.

1. FEEDFORWARD LINEARIZATION TECHNIQUE

The basic Feedforward architecture consists of two cancellation loops. The aim of the first loop (Loop 1) is to sample the distortion introduced by the main amplifier. It is commonly known as the "signal cancellation loop" or the "error loop".

On the other hand, the second cancellation loop (Loop 2) uses the distortion sample obtained from the error loop to subtract the distortion component from the amplified main signal. It is so-called the "distortion

cancellation loop". The *Figure 5-1* shows the basic Feedforward architecture with a two tone signal.

The two tone procedure is used to characterize the linearity of any active devices in the quasilinear region. The power amplifiers are usually taken to the saturation in order to maximize the efficiency rates, getting out of the ideal conditions of this procedure. Nevertheless, this method becomes useful in order to compare the linearized amplifiers, and it corresponds to the present case.

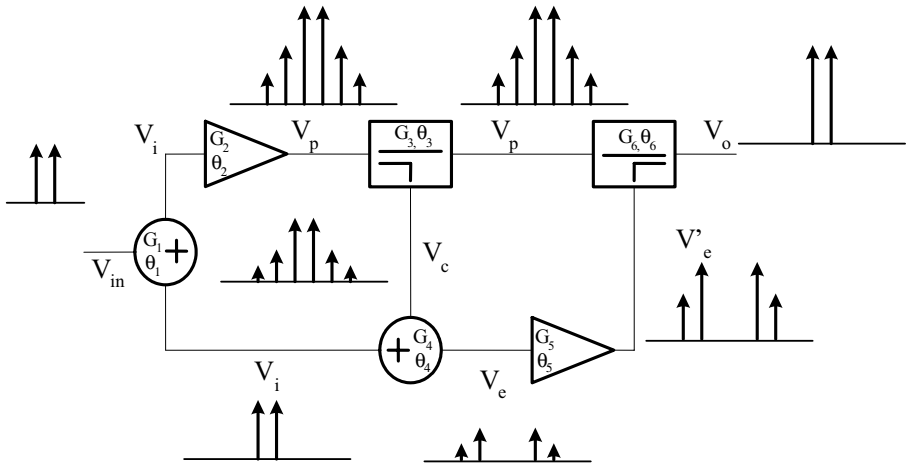


Figure 5-1. Feedforward linearization basic architecture

Two considerations have been made throughout the following analysis to allow certain simplicity in the nomenclature of the formulas.

- In most of cases, power splitters (G_1, θ_1) and power combiners (G_4, θ_4) are the components that provide the reference signal and subtract it from the main signal, respectively. In the following, the transmission gains (or losses) will be equal for both the split and the combined ports. In some other cases, hybrid and directional couplers are also used, but in any case, the conclusions of this analysis could be extrapolated to any other Feedforward scheme.
- The couplers used in the Feedforward architecture are represented only by their coupling factors (G_3 and G_6), thus depreciating the direct transmission losses. This effect will be considered separately at the end of this chapter, simplifying the final expressions but maintaining their validity.

Once it has been come to terms with those considerations, it is going to proceed to the analysis of the Feedforward performance.

The input signal (V_{in}) is divided in two by the first power splitter (G_1, θ_1). One of split signal (V_i) is amplified by the main amplifier (G_2, θ_2) and the other one (V_i) is used as a reference signal (5.1).

$$V_i = V_{in} \cdot 10^{\left(\frac{1}{20}G_1\right)} \cdot e^{j(\theta_1)} \tag{5.1}$$

The output of the main amplifiers is named as V_p (5.2).

$$V_p = V_{in} \cdot 10^{\frac{1}{20}(G_1+G_2)} \cdot e^{j(\theta_1+\theta_2)} + V_d \tag{5.2}$$

G[dB] and θ[rad] are the gain and phase transmission parameters shown in *Figure 5-1*. It is important to notice how all the distortion introduced by the main amplifier is included in the V_d term. From now to end, two signal components are going to be distinguished: the main signal component and the distortion signal component (V_d).

A directional coupler (G_3, θ_3) samples the amplified signal (V_c).

$$\begin{aligned} V_c &= V_{in} \cdot 10^{\frac{1}{20}(G_1+G_2+G_3)} \cdot e^{j(\theta_1+\theta_2+\theta_3)} \\ &+ V_d \cdot 10^{\frac{1}{20}(G_3)} \cdot e^{j(\theta_3)} \end{aligned} \tag{5.3}$$

The error loop (Loop 1) concludes with the subtraction of the reference signal (V_i) from the coupled signal (V_c) in a power combiner (G_4, θ_4). The resultant signal is so-called the error signal and will be denoted as V_e.

$$\begin{aligned} V_e &= V_{in} \cdot 10^{\frac{1}{20}(G_1+G_2+G_3+G_4)} \cdot e^{j(\theta_1+\theta_2+\theta_3+\theta_4)} \\ &+ V_{in} \cdot 10^{\frac{1}{20}(G_1+G_4)} \cdot e^{j(\theta_1+\theta_4)} \\ &+ V_d \cdot 10^{\frac{1}{20}(G_3+G_4)} \cdot e^{j(\theta_3+\theta_4)} \end{aligned} \tag{5.4}$$

The error signal has again two components: a sample of the distortion signal on the one hand, and a sample of the main signal on the other hand. The aim of the error loop is to get only the distortion component, so two phase and gain equations should be fulfilled (5.5).

$$\begin{aligned} \theta_2 + \theta_3 &= \pm\pi \\ G_3 &= -G_2 \end{aligned} \tag{5.5}$$

With those gain and phase conditions, the error signal (V_e) will have only the distortion sample (5.6).

$$V_e = V_d \cdot 10^{\frac{1}{20}(G_3+G_4)} \cdot e^{j(\theta_3+\theta_4)} \tag{5.6}$$

The distortion cancellation loop (Loop 2) tries to cancel the distortion component of the amplified signal (V_p) by subtracting this error signal (V_e) correctly scaled. A directional coupler (G_6, θ_6) is used for the distortion cancellation in order to attenuate the main signal component as less as possible.

The scaling of the error signal (V_e) is done with an auxiliary amplifier, known as the error amplifier (G_5, θ_5).

The scaled error signal will be denoted as V'_e (5.7).

$$V'_e = V_d \cdot 10^{\frac{1}{20}(G_3+G_4+G_5)} \cdot e^{j(\theta_3+\theta_4+\theta_5)} \tag{5.7}$$

The output signal (V_o) of the second directional coupler (G_6, θ_6) has one main component and two uncorrelated distortion components (5.8).

$$\begin{aligned} V_o &= V_{in} \cdot 10^{\frac{1}{20}(G_1+G_2)} \cdot e^{j(\theta_1+\theta_2)} \\ &+ V_d + V_d \cdot 10^{\frac{1}{20}(G_3+G_4+G_5+G_6)} \cdot e^{j(\theta_3+\theta_4+\theta_5+\theta_6)} \end{aligned} \tag{5.8}$$

If the distortion components must be subtracted, two gain and phase equations (5.9) should be fulfilled.

$$\begin{aligned} \theta_3 + \theta_4 + \theta_5 + \theta_6 &= \pm\pi \\ G_5 &= -G_3 - G_4 - G_6 \end{aligned} \tag{5.9}$$

In this ideal case, the output signal (V_o) will have the input signal (V_{in}) amplified by the power amplifier gain factor (G_2) and attenuated by the power splitter transmission loss (G_1), but without any distortion (V_d) (5.10).

$$V_o = V_{in} \cdot 10^{\frac{1}{20}(G_1+G_2)} \cdot e^{j(\theta_1+\theta_2)} \tag{5.10}$$

It is obvious that the amplitude and phase equations supposed in the previous analysis do not correspond with a real case. Some of the most popular non-idealities are studied in the subsequent sections: the frequency dependence, the gain and phase adjustments, the error amplifier distortion, the isolation lack and the main signal path loss.

1.1 Frequency dependence

The transmission parameters of any radiofrequency component are frequency dependent, despite in many narrowband applications the flat frequency assumption is done. Therefore, the gain and phase equations of (5.1) and (5.9) are not fulfilled for the entire band of interest, and the total signal cancellation will not be feasible.

In the following, the frequency dependence effect is going to be represented with a linear law for the amplitude and phase transmission parameters, as shown in *Figure 5-2*.

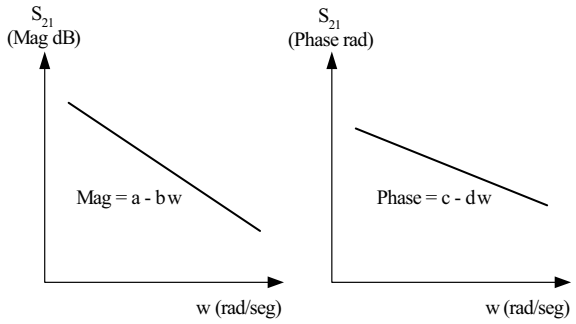


Figure 5-2. Linear laws for gain and phase frequency dependences

a and b are gain parameters, while c and d are phase parameters. All the formulas of the previous section are now rewritten with the linear laws of the frequency dependences.

First of all, the error loop signals:

$$V_i = V_{in} \cdot 10^{\frac{1}{20}(a_1 - b_1 w)} \cdot e^{j(c_1 - d_1 w)} \tag{5.11}$$

$$V_p = V_{in} \cdot 10^{\frac{1}{20}\left(\sum_{i=1}^{2} a_i - \left(\sum_{i=1}^{2} b_i\right) \cdot w\right)} \cdot e^{j\left(\sum_{i=1}^{2} c_i - \left(\sum_{i=1}^{2} d_i\right) \cdot w\right)} + V_d \tag{5.12}$$

$$\begin{aligned} V_c = {} & V_{in} \cdot 10^{\frac{1}{20}\left(\sum_{i=1}^{3} a_i - \left(\sum_{i=1}^{3} b_i\right) \cdot w\right)} \cdot e^{j\left(\sum_{i=1}^{3} c_i - \left(\sum_{i=1}^{3} d_i\right) \cdot w\right)} \\ & + V_d \cdot 10^{\frac{1}{20}(a_3 - b_3 w)} \cdot e^{j(c_3 - d_3 w)} \end{aligned} \tag{5.13}$$

$$\begin{aligned} V_e &= V_{in} \cdot 10^{\frac{1}{20}\left(\sum_{i=1}^{4} a_i - \left(\sum_{i=1}^{4} b_i\right)\cdot w\right)} \cdot e^{j\left(\sum_{i=1}^{4} c_i - \left(\sum_{i=1}^{4} d_i\right)\cdot w\right)} \\ &+ V_{in} \cdot 10^{\frac{1}{20}(a_1+a_4-(b_1+b_4)w)} \cdot e^{j(c_1+c_4-(d_1+d_4)w)} \\ &+ V_d \cdot 10^{\frac{1}{20}(a_3+a_4-(b_3+b_4)w)} \cdot e^{j(c_3+c_4-(d_3+d_4)w)} \end{aligned} \quad (5.14)$$

And the equations for the main signal cancellation:

$$\begin{aligned} c_2 + c_3 &= \pm\pi \\ a_3 &= -a_2 \end{aligned} \quad (5.15)$$

The new error signal (5.16) has an unexpected main signal component even the gain and phase equations of (5.15) are fulfilled.

$$\begin{aligned} V_e &= V_{in} \cdot 10^{\frac{1}{20}\left(a_1+a_4-\left(\sum_{i=1}^{4} b_i\right)\cdot w\right)} \cdot e^{j\left(c_1\pm\pi+c_4-\left(\sum_{i=1}^{4} d_i\right)\cdot w\right)} \\ &+ V_{in} \cdot 10^{\frac{1}{20}(a_1+a_4-(b_1+b_4)w)} \cdot e^{j(c_1+c_4-(d_1+d_4)w)} \\ &+ 10^{\frac{1}{20}(a_4-a_2-(b_3+b_4)w)} \cdot V_d \cdot e^{j(\pm\pi-c_2+c_4-(d_3+d_4)w)} \end{aligned} \quad (5.16)$$

Therefore, two additional equations must be added (5.17).

$$\begin{aligned} b_2 + b_3 &= 0 \\ d_2 + d_3 &= 0 \end{aligned} \quad (5.17)$$

The equation (5.17) establishes that the frequency response of the V_i and V_c signals must be identical. The amplitude equalization techniques are not advisable for the Feedforward amplifiers. Nevertheless, the frequency response (Δw) of the phase transfer function ($\Delta\theta$), is defined as the signal group delay (τ) and it is perfectly adjustable.

$$\tau = \frac{\Delta\theta}{\Delta w} = b_i \quad (5.18)$$

The group delay adjustments are explained in Chapter 6, but an introductory example is shown in *Figure 5-3*. $Signal_1$ and $Signal_2$ represent two UMTS downlink signals [61]. The frequency flatness is 0,2dB in both cases, but the former has a group delay difference of 46ps (1º) whereas the second one is 232ps (4º). The cancellation for the central frequency (2140MHz) is total in any case, but for the sideband frequencies this cancellation is 35dB and 29dB for the first and second cases, respectively. A

group delay error of 186ps result in a cancellation reduction of 6dB for this frequency range.

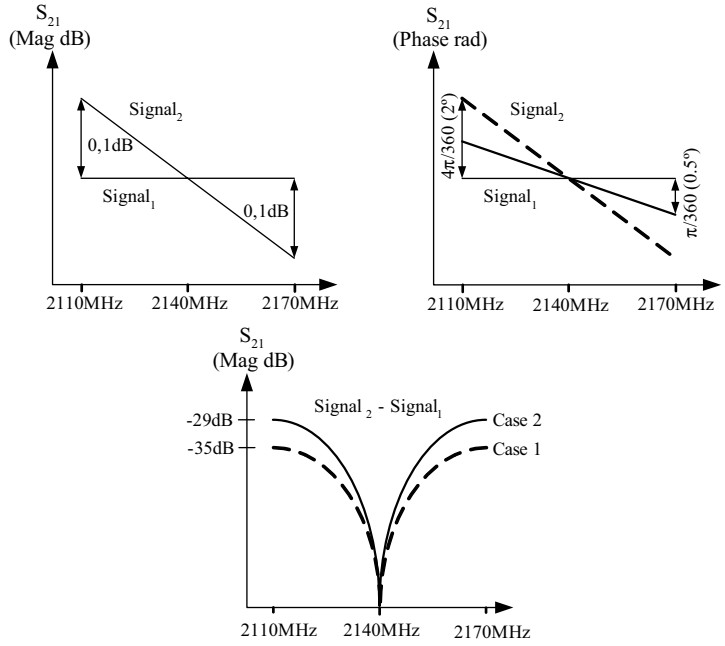


Figure 5-3. Signal cancellation (0.2dB amplitude flatness, 46ps and 232ps group delay error)

Despite supposing a perfect delay group adjustment, hardly complicated, the error signal (V_e) will still have a main signal component derived from the amplitude flatness (5.19).

$$\begin{aligned} V_e &= V_{in} \cdot \phi(b_2, b_3) \\ &+ V_d \cdot 10^{\frac{1}{20}(a_3+a_4-(b_3+b_4)\cdot w)} \cdot e^{j(c_3+c_4-(d_3+d_4)\cdot w)} \end{aligned} \tag{5.19}$$

The subtraction of the distortion component in the second loop has the same limitations (5.20).

$$\begin{aligned} V_o &= V_{in} \cdot 10^{\frac{1}{20}\left(\sum_{i=1}^{2} a_i - \left(\sum_{i=1}^{2} b_i\right)\cdot w\right)} \cdot e^{j\left(\sum_{i=1}^{2} c_i - \left(\sum_{i=1}^{2} d_i\right)\cdot w\right)} \\ &+ V_{in} \cdot \phi(b_2 + b_3) \cdot 10^{\frac{1}{20}(a_5+a_6-(b_5+b_6)\cdot w)} \cdot e^{j(c_5+c_6-(d_5+d_6)\cdot w)} \\ &+ V_d \cdot \left[1 + 10^{\frac{1}{20}\left(\sum_{i=3}^{6} a_i - \left(\sum_{i=3}^{6} b_i\right)\cdot w\right)} e^{j\left(\sum_{i=3}^{6} c_i - \left(\sum_{i=3}^{6} d_i\right)\cdot w\right)}\right] \end{aligned} \tag{5.20}$$

Consequently, even fulfilling the amplitude and phase equations of the ideal case (5.9), the distortion component will not be totally cancelled. Once again, two more equations (5.21) will be necessary in order to equalize the frequency dependence of the phase and amplitude transfer functions.

$$\begin{aligned}
&\sum_{i=3}^{6} a_i = 0 \\
&\sum_{i=3}^{6} c_i = \pm\pi \\
&\sum_{i=3}^{6} d_i = 0 \\
&\sum_{i=3}^{6} b_i = 0
\end{aligned} \tag{5.21}$$

Although group delays were perfectly coincident, the distortion component will not be totally eliminated. In addition, the output signal (V_o) will have two main signal components instead of one, due to the main signal component of the error loop (5.22).

$$\begin{aligned}
V_o = {} & V_{in} \cdot 10^{\frac{1}{20}\left(\sum_{i=1}^{2} a_i - \left(\sum_{i=1}^{2} b_i\right)\cdot w\right)} \cdot e^{j\left(\sum_{i=1}^{2} c_i - \left(\sum_{i=1}^{2} d_i\right)\cdot w\right)} \\
& + V_{in} \cdot \phi(b_2, b_3) \cdot 10^{\frac{1}{20}(a_5 + a_6 - (b_5 + b_6)\cdot w)} \cdot e^{j(c_5 + c_6 - (d_5 + d_6)\cdot w)} \begin{cases} \text{main signal} \\ \text{interference} \end{cases} \\
& + V_d \cdot \left[1 - 10^{\frac{1}{20}\left(\left(\sum_{i=3}^{6} b_i\right)\cdot w\right)}\right]
\end{aligned} \tag{5.22}$$

The expression of the output signal (5.22) can be sensibly simplified if some of the coefficients are grouped, like in (5.23).

$$V_o = V_{in} \cdot 10^{\frac{1}{20}(A_0 - B_0 \cdot w)} \cdot e^{j(C_0 - D_0 \cdot w)} + V_d \cdot \left[1 - 10^{\frac{1}{20}\left(\left(\sum_{i=3}^{6} b_i\right)\cdot w\right)}\right] \tag{5.23}$$

The Feedforward architecture is shown in the *Figure 5-4* after the delay elements (τ) have been added. Notice the main signal component of the error signal (V_e) and the residual distortion components of the output signal (V_o).

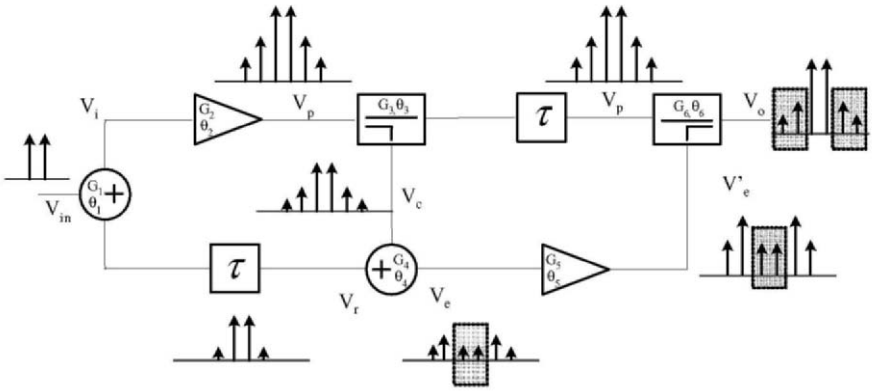


Figure 5-4. Feedforward architecture with delay elements

1.2 Amplitude and phase adjustments

The amplitude and phase equations of (5.15) and (5.21), allow the cancellation in both Feedforward loops. However, all the transmission parameters are fixed, hardly coincident with the required amplitude and phase conditions.

Variable attenuators and phase shifters are commonly used as adjustment elements. Their location on the Feedforward architecture is not standard, although any interaction with the main path should be avoided (*Figure 5-5*).

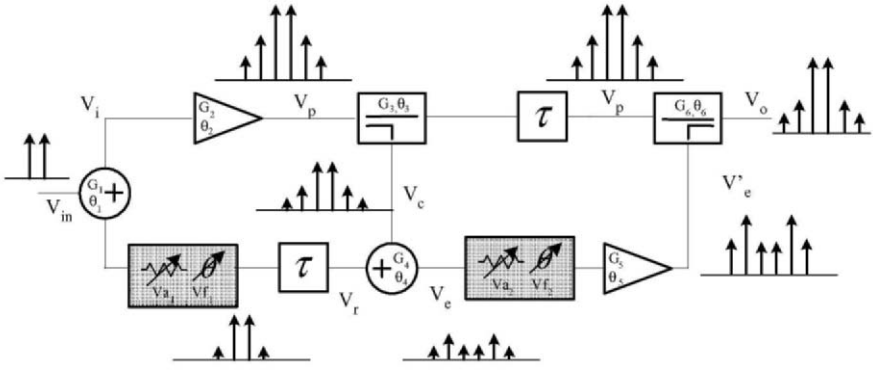


Figure 5-5. Feedforward architecture with adjustment elements

In following lines, only the error and the output signal formulas are shown. The error signal regarding the first loop adjustment parameters (Va_1, Vf_1) is rewritten in (5.24).

$$\begin{aligned} V_e &= V_{in}\cdot 10^{\frac{1}{20}\left(\sum_{i=1}^{4} a_i-\left(\sum_{i=1}^{4} b_i\right)\cdot w\right)}\cdot e^{j\left(\sum_{i=1}^{4} c_i-\left(\sum_{i=1}^{4} d_i\right)\cdot w\right)} \\ &+ V_{in}\cdot 10^{\frac{1}{20}(a_1+a_4+Va_1-(b_1+b_4)\cdot w)}\cdot e^{j(c_1+c_4+Vf_1-(d_1+d_4)\cdot w)} \\ &+ V_d\cdot 10^{\frac{1}{20}(a_3+a_4-(b_3+b_4)\cdot w)}\cdot e^{j(c_3+c_4-(d_3+d_4)\cdot w)} \end{aligned} \tag{5.24}$$

The main signal cancellation will be determined by those adjustment parameters and the first delay element, as reflected in (5.25).

$$\begin{aligned} Vf_1 &= c_2 + c_3 \pm \pi \\ Va_1 &= a_2 + a_3 \\ d_2 &+ d_3 = 0 \end{aligned} \tag{5.25}$$

The error signal will be as deduced in (5.19).

$$\begin{aligned} V_e &= V_{in}\cdot \phi(b_2, b_3) \\ &+ V_d\cdot 10^{\frac{1}{20}(a_3+a_4-(b_3+b_4)\cdot w)}\cdot e^{j(c_3+c_4-(d_3+d_4)\cdot w)} \end{aligned} \tag{5.26}$$

The output signal regarding the second loop adjustment parameters (Va_2, Vf_2) is shown in (5.27).

$$\begin{aligned} V_o &= V_{in}\cdot 10^{\frac{1}{20}(\phi(Va_2)-B_0\cdot w)}\cdot e^{j(\varphi(Vf_2)-D_0\cdot w)} \\ &+ V_d\cdot\left[1+10^{\frac{1}{10}\left(Va_2+\sum_{i=3}^{6} a_i-\left(\sum_{i=3}^{6} b_i\right)\cdot w\right)}e^{j\left(Vf_2+\sum_{i=3}^{6} c_i-\left(\sum_{i=3}^{6} d_i\right)\cdot w\right)}\right] \end{aligned} \tag{5.27}$$

Note how the main component of the output signal depends on the second loop adjusting parameters. This is an interesting effect that will be determining in the design of the control algorithm.

The new gain and phase equations for the distortion cancellation are obtained in (5.28).

$$\begin{aligned} Vf_2 &= -\sum_{i=3}^{6} c_i \pm \pi \\ Va_2 &= -\sum_{i=3}^{6} a_i \\ \sum_{i=3}^{6} d_i &= 0 \end{aligned} \tag{5.28}$$

Once again, the total subtraction of the distortion component is not possible due to the frequency dependence of the amplitude transfer function (b_3, b_4, b_5, b_6).

$$V_o = V_{in} \cdot 10^{\frac{1}{20}\left(\phi(Va_2)-B_0 \cdot w\right)} \cdot e^{j\left(\varphi(Vf_2)-D_0 \cdot w\right)} + V_d \cdot \left[1-10^{\frac{1}{20}\left(\left(\sum_{i=3}^{6} b_i\right)\cdot w\right)}\right] \tag{5.29}$$

The degrees of freedom of the adjustment parameters in the Feedforward linearization technique can be graphically seen in *Figure 5-6*.

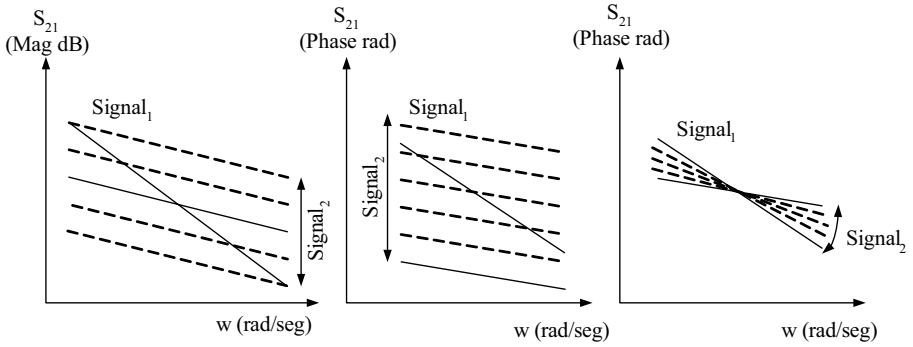


Figure 5-6. Degrees of freedom in the Feedforward linearization technique

From left to right the adjustment ranges are shown for the variable attenuators (Va_1, Va_2), the phase shifters (Vf_1, Vf_2) and the group delays (τ_1, τ_2) elements.

1.3 Error amplifier distortion

The error amplifier represents one of the highest restrictions in the Feedforward linearization technique. The gain is represented with the a_5 parameter, and it is found in expression (5.28), now rewritten as (5.30).

$$-a_5 = a_3 + a_4 + a_6 + Va_2 \tag{5.30}$$

The minus sign corroborates that a_5 must be a gain factor. It must compensate the losses of the a_3 and a_6 coupling factors, the a_4 insertion loss and the Va_2 attenuation. In real applications this gain factor hovers around 50dB. Such gain values must be achieved with two or three amplification stages, and sadly, the distortion introduce by them is not gone unnoticed.

If the error amplifier distortion ($V_{d\_erro\_amp}$) is added to the previous formulas, the output signal (V_o) will change to (5.31).

$$V_o = V_{in} \cdot 10^{\frac{1}{20}(\phi(Va_2)-B_0 \cdot w)} \cdot e^{j(\varphi(Vf_2)-D_0 \cdot w)} + V_d \cdot \left[1 - 10^{\frac{1}{20}\left(\left(\sum_{i=3}^{6} b_i\right)\cdot w\right)}\right] + V_{d\_error\_amp} \cdot 10^{\frac{1}{20}(a_6 - b_6 \cdot w)} \cdot e^{j(c_6 - d_6 \cdot w)} \quad (5.31)$$

If high linear error amplifiers are used, the Feedforward technique turns into a non efficient linearization technique.

Actually, a suitable trade-off must be taken among price, efficiency and cancellation levels, and it constitutes one of the most critical points in the design of Feedforward amplifiers.

1.4 Isolation lack

The lack of total isolation in the power splitters and the power combiners contaminates the reference signal (V_r), as shown in *Figure 5-7*. This effect changes all the formulas seen until now.

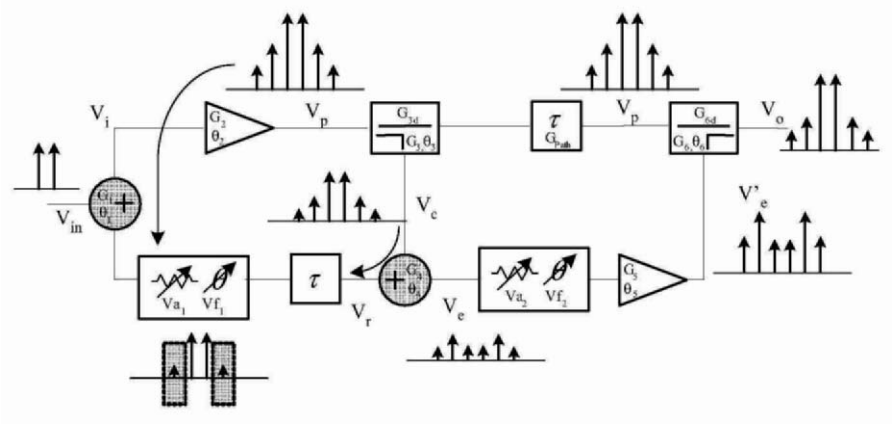


Figure 5-7. Isolation lack effect in the reference signal

The distortion components of the error signal will be affected by all the adjustment parameters and therefore the output signal will be too (5.32).

$$
\begin{aligned}
V_o &= V_{in}\cdot 10^{\frac{1}{20}\left(\phi(Va_2)-B_0\cdot w\right)}\cdot e^{j\left(\varphi(Vf_2)-D_0\cdot w\right)} \\
&+ V_{in}\cdot 10^{\frac{1}{20}\left(Va_1+Va_2+A_{isolation}-\left(B_{isolation}\right)w\right)}\cdot e^{j\left(Vf_1+Vf_2+C_{isolation}-\left(D_{isolation}\right)w\right)} \\
&+ V_d\cdot\left[1-10^{\frac{1}{20}\left(\left(\sum_{i=3}^{6}b_i\right)\cdot w\right)}\right] \\
&+ V_d\cdot 10^{\frac{1}{20}\left(Va_1+Va_2+A^*_{isolation}-\left(B^*_{isolation}\right)\cdot w\right)}\cdot e^{j\left(Vf_1+Vf_2+C^*_{isolation}-\left(D^*_{isolation}\right)\cdot w\right)} \\
&+ V_{d\_error\_amp}\cdot 10^{\frac{1}{20}\left(a_6-b_6\cdot w\right)}\cdot e^{j\left(c_6-d_6\cdot w\right)}
\end{aligned}
\tag{5.32}
$$

The most important conclusion is the influence of the four adjustment parameters over the main and the distortion components. This fact allows the design of an alternative adjustment method that will achieve a trade-off between linearity and efficiency rates.

1.5 Main signal path loss

One of the considerations in the beginning of this mathematical analysis refers to the insertion loss of the directional couplers. Now, those effects are represented with two new parameters, G_{3d} and $G_{6d,}$ added in *Figure 5-7*. The frequency dependence of these new parameters is not considered because of its irrelevance.

Finally, the transmission loss of the delay element (G_{Path}) is introduced. This parameter is also fundamental for the efficiency of the linearization technique. Note that the influence of G_{3d}, G_{6d} and G_{Path} on the main signal path loss inevitably reduces the output power of the Feedforward amplifier.

There are several techniques for the implementation of delay elements, like coaxial or microstrip transmission lines, ceramic components and some special filters. The advantages and inconveniences of each one are discussed in the next chapter.

$$
\begin{aligned}
V_s &= V_{in}\cdot 10^{\frac{1}{20}\left(G_{3d}+G_{6d}+G_{Path}+\phi(Va_2)-B_0\cdot w\right)}\cdot e^{j\left(\varphi(Vf_2)-D_0\cdot w\right)} \\
&+ V_{in}\cdot 10^{\frac{1}{20}\left(Va_1+Va_2+A_{isolation}-\left(B_{isolation}\right)w\right)}\cdot e^{j\left(Vf_1+Vf_2+C_{isolation}-\left(D_{isolation}\right)w\right)} \\
&+ V_d\cdot\left[1-10^{\frac{1}{20}\left(\left(\sum_{i=3}^{6}b_i\right)\cdot w\right)}\right] \\
&+ V_d\cdot 10^{\frac{1}{20}\left(Va_1+Va_2+A^*_{isolation}-\left(B^*_{isolation}\right)\cdot w\right)}\cdot e^{j\left(Vf_1+Vf_2+C^*_{isolation}-\left(D^*_{isolation}\right)\cdot w\right)} \\
&+ V_{d\_error\_amp}\cdot 10^{\frac{1}{20}\left(a_6-b_6\cdot w\right)}\cdot e^{j\left(c_6-d_6\cdot w\right)}
\end{aligned}
\tag{5.33}
$$

The final output signal (5.33) represents the real Feedforward performance according to all the non-idealities. Some of those effects limit the cancellation capability of the Feedforward linearization technique, but other ones suggest alternative and interesting adjustment methods, discussed in Chapter 7.

2. FEEDFORWARD IN WIDEBAND COMMUNICATIONS SYSTEMS

The two tone signal analysis of the previous section shows the operation principle of the Feedforward amplifiers and some of their most important limitations. However, the wideband modulations used in some digital communication systems introduce high amplitude variations that should have been considered.

This section checks those effects with a real wideband signal used by the UMTS communication standard [61]: a WCDMA signal settled at 2140MHz (chip rate 3.84Mcps, Peak/Avg. = 10dB @ 0.01% probability on CCDF, 3GPP Test Model 1, 64 DTCH.) .

The WCDMA signal progression through the Feedforward linearization architecture is shown in *Figure 5-8*. The equivalence with the two tone procedure is intuitive.

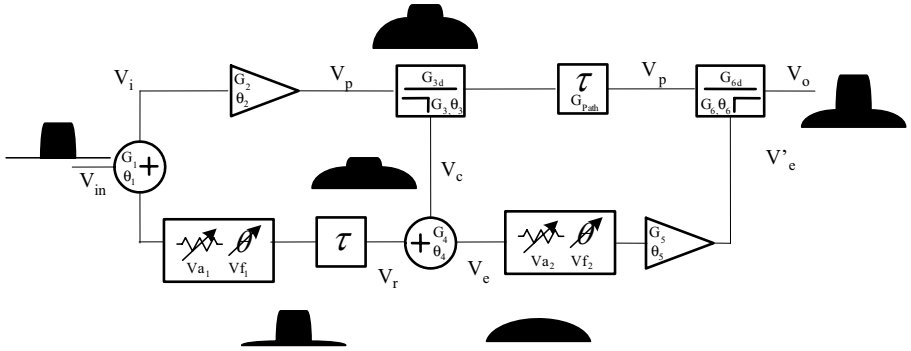


Figure 5-8. Feedforward linearization technique with wideband signal

All figures that are going to be shown are obtained from the characterization of the Feedforward prototype presented in Chapter 6. These figures represent the error (V_e) and the output (V_o) signals versus all the adjustment parameters (Va_1, Vf_1, Va_2, Vf_2). All of them verify the analytical expressions of the previous section.

First of all, the error signal level (dBm) is shown in *Figure 5-9* according to the error loop adjustment parameters (Va_1 and Vf_1). Note that if the error signal is minimized, the cancellation of the main signal is maximum.

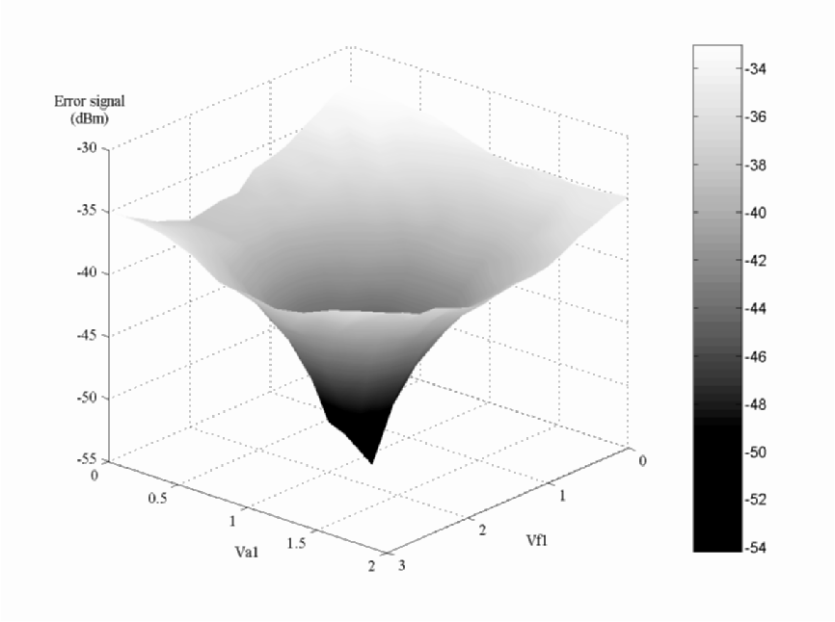


Figure 5-9. Main signal cancellation or error signal minimization vs. Va_1 and Vf_1

Ideally, the main signal component of the error signal should be totally cancelled only when the attenuation (Va_1) and phase shift (Vf_1) fulfil with the equation (5.25). However, the frequency dependence of the transmission characteristics on the one hand, and the accuracy of the delay elements on the other hand, limit this cancellation.

As far as the output signal is concerned, both the main and the distortion components are analyzed.

In *Figure 5-10* the distortion component level (dBm) is visualized vs. the signal cancellation loop adjustment parameters (Va_2, Vf_2).

The distortion signal achieves a minimum level when the second loop adjustment parameters (Va_2, Vf_2) fulfil the equation (5.28). Once again, the frequency dependence and the accuracy of the delay components limit the cancellation level.

Let's check the influence of the first loop adjustment parameters over the output distortion component (5.32). This fact is corroborated in *Figure 5-11*.

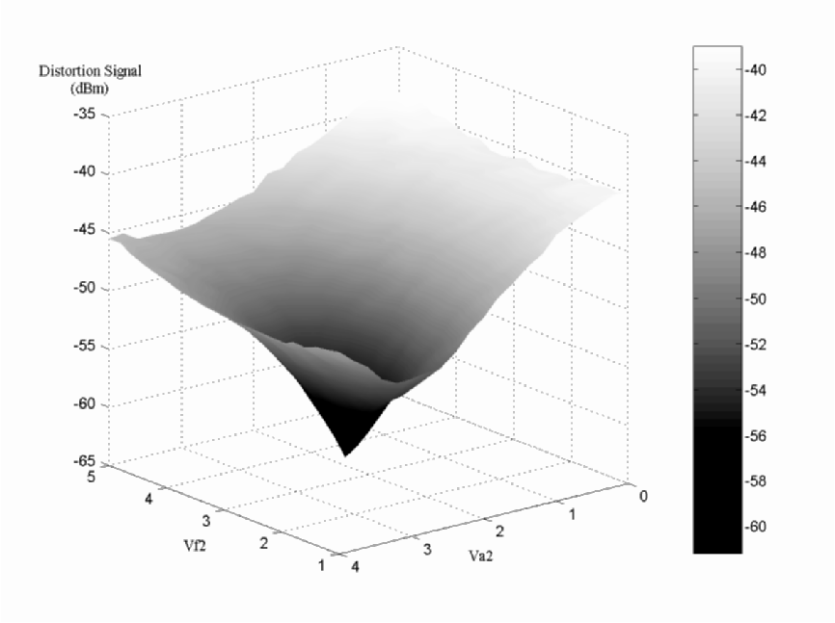


Figure 5-10. Distortion component cancellation vs. Va_2 and Vf_2

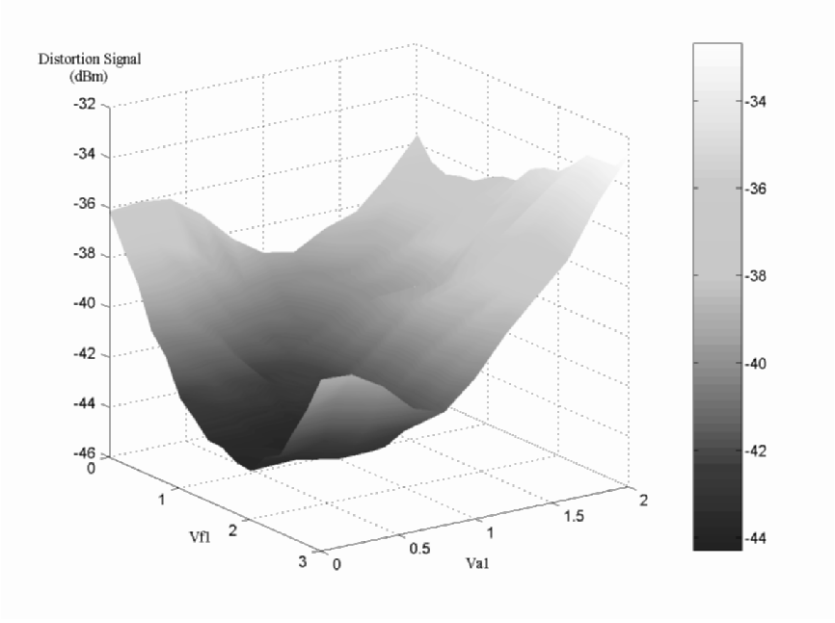


Figure 5-11. Distortion component cancellation vs. Va_1 and Vf_1

One of the main conclusions of the previous section asserts that all the adjustment parameters have some influence over the main component of the output signal. In spite of the fact that the aim of the Feedforward technique is the suppression of the power amplifier distortion, maybe the unexpected capability to improve the output level would benefit the efficiency of all the transmission system.

Therefore, it is especially interesting to verify the influence of all the adjustment parameters on the output main signal power level. Thus, in *Figure 5-12* and *Figure 5-13* the output signal power is shown vs. both loop adjustment parameters.

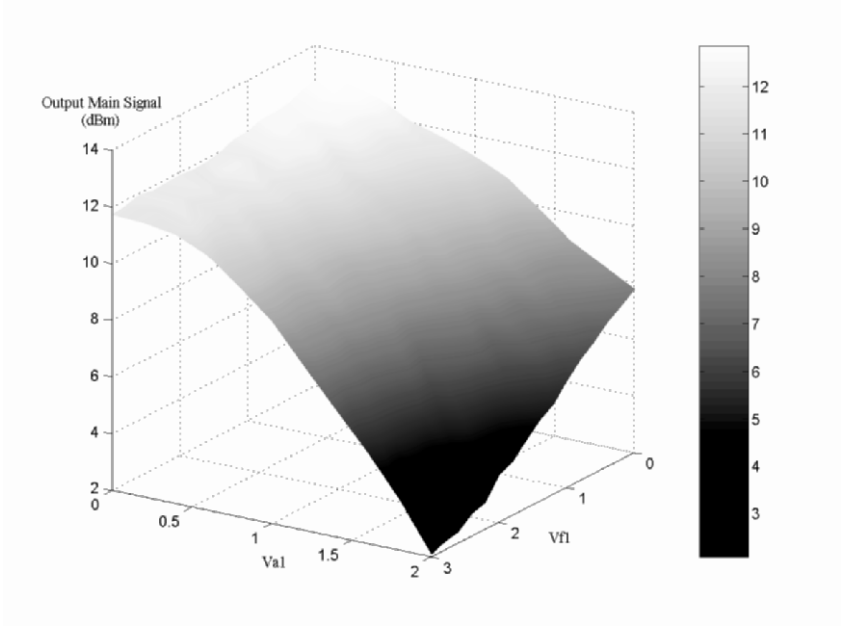


Figure 5-12. Output main signal vs. Va_1 and Vf_1

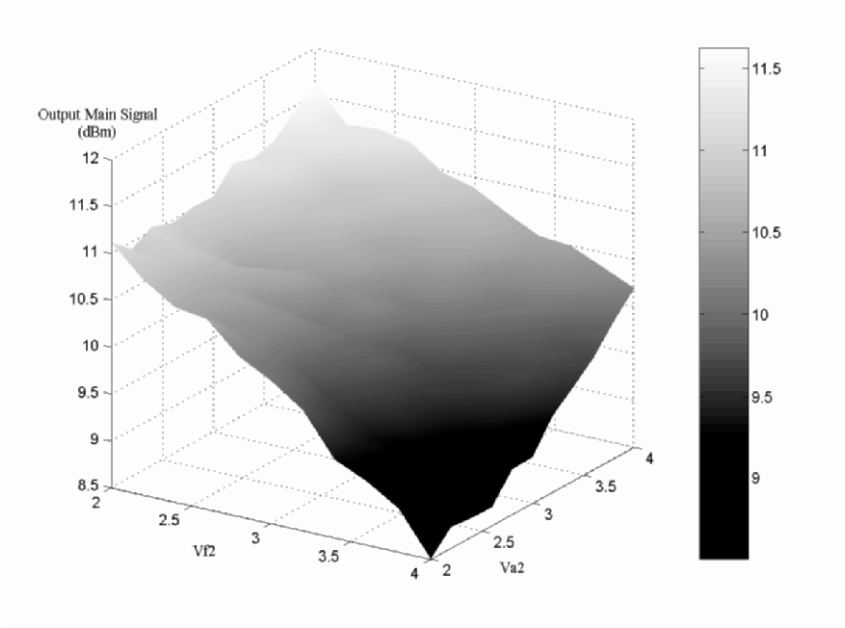


Figure 5-13. Output main signal vs. Va_2 and Vf_2

Such power level variations are not in accordance with the distortion cancellation achieved before, so an output power improvement will involve inevitably, a distortion increment and vice versa. In other words, an efficiency improvement entails a linearity worsening, as it occurs in any amplification device.

In Chapter 7 is widely discussed how this property of Feedforward amplifiers undergoes a change of direction in the actual control methods.

Chapter 6

IMPLEMENTATION OF FEEDFORWARD AMPLIFIERS

In the previous chapter a mathematical analysis has been done of the Feedforward linearization technique. The study has given the reasons that avoid a total cancellation of the distortion introduced by the power amplifier. Likewise, it has been notice the potential of the adjustment parameters to enhance the benefits of this linearization technique. This chapter is focused on suggesting a design guideline to implement Feedforward amplifiers for wideband communication systems. As introduced in Chapter 5, a UMTS downlink channel has been used as application case. The signal type used by this standard is a wideband code division multiple code (WCDMA), particularly it has been used the Test model 1 defined by the 3rd Generation Partnership Project (3GPP) [61].

Some simulations have been carried out at circuit level in order to demonstrate the conclusions obtained in the previous chapter. At the same time, a useful component selection criterion has been extracted, necessary for the prototyping stage. At this point, the error amplifiers have been specially discussed, as they constitute the bottleneck of the Feedforward amplifiers.

In addition, some design considerations are proposed for the implementation of the delay elements. This stage fixes the maximum cancellation levels that the linearization technique will achieved.

Some guidelines are introduced for the measurement of wideband signals in order not to masquerade the improvements achieved with the linearized amplifier. Those measurements have been used in the Feedforward amplifier prototype implemented with the guidelines proposed in this work.

1. SIMULATION OF THE FEEDFORWARD ARCHITECTURE

Simulation tools are highly recommended for the performance analysis during the circuit design process, even more if the models of the components used are reliable enough. In this case the simulation cases are ready to check the theoretical performance of the Feedforward technique. Particularly, the aim of this section is to corroborate the performance of the cancellation loops for a particular linearization scheme in order to select the most suitable components.

The simulations have been done with the linearization design library of the Advanced Design System (ADS) tool of Agilent Technologies.

1.1 Simulation models

The choice of the simulation model is essential, as its analogy dictates the truthfulness of the results.

The nonlinear models of the main and the error amplifiers can include memory effects or not. The memory effect models are more accurate as they consider the frequency dependence phenomena. Consequently, they are much more complex and only are used when the accuracy requirements of the power amplifier determine the success of the linearization technique (e.g. Predistortion). Several models can be found referenced [62-67].

The memory-less models are also widely used: polynomial models, AM-AM and AM-PM models.

The polynomial models are suitable only for low signal (low distortion) simulations. As the input signal level increases, distortion effects become more important and other models must be used. Some of them represent the variations of the output envelope and phase signals according to the envelope of the input signals. They are known as the AM-AM and AM-PM conversion models.

The models recommended for the Feedforward simulations are the AM-AM and AM-PM conversion models. The AM-AM is characterized with the 1dB compression point (P_{1dB}), the output saturation power (P_{SAT}) and the output third order intercept point (OIP_3). The AM-PM, on the other hand, models the output phase vs. the input power.

As far as the transmission and reflection effects are concerned, the scattering parameters are used for all the passive and active components.

The goal of the simulations should not be an exact reproduction of the real performance of the linearization system, since mathematical models are used as replicas of the real components. The aim of the simulations should be a component selection tool that helps to fence in the offer provided by the wide market of the RF components.

Therefore, the frequency dependence of the scattering parameters can be avoided as far as the amplitude flatness is concerned. It is advisable, however, to consider the group delays, as they constitute one of most important design parameters.

1.2 Simulation of the error loop

The schematic of the simulated architecture is shown in *Figure 6-1*. A harmonic balance analysis is done with a two tone signal settled at 2137.5MHz. The S-parameters and group delays of all components are achieved from their individual characterizations.

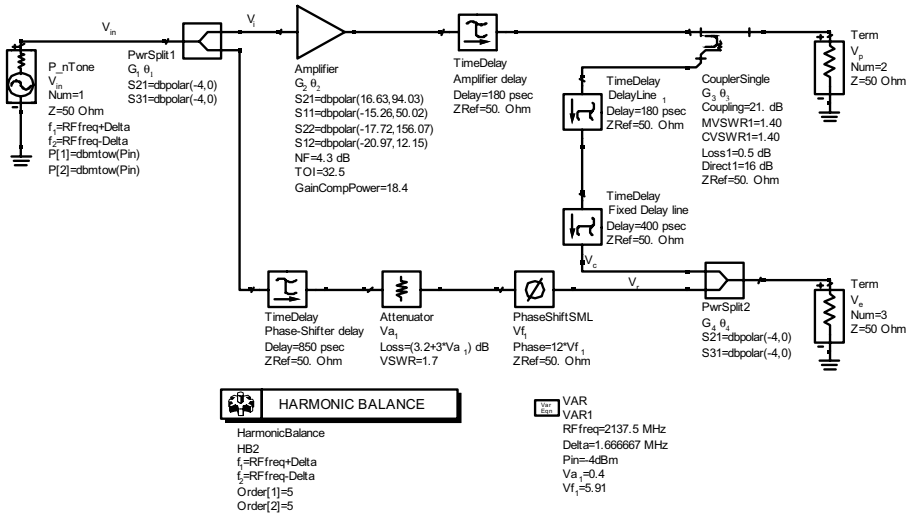


Figure 6-1. Error loop simulation

The simulation methodology is the used in the theoretical analysis. The central frequency coincides with the sixth descendent channel in the UMTS standard. Two tones are generated, 3.33MHz separated, so the third order intermodulation products are located into the adjacent channels.

$$
\begin{aligned}
f_1 &= 2137.5\text{MHz} - \frac{3.33\text{MHz}}{2} = 2135.835\text{MHz} \\
f_2 &= 2137.5\text{MHz} + \frac{3.33\text{MHz}}{2} = 2139.165\text{MHz} \\
&\left.\begin{aligned} IP_3\big|_1 &= 2f_2 - f_1 = 2142.5\text{MHz} \\ IP_3\big|_2 &= 2f_1 - f_2 = 2132.5\text{MHz} \end{aligned}\right\} IP_3 = 2137.5\text{MHz} \pm 5\text{MHz}
\end{aligned}
\tag{6.1}
$$

The simulator adjusts the Va_1, Vf_1 and $DelayLine_1$ parameters until the fundamental frequencies (f_1 and f_2) are cancelled, or at least minimized.

The *Figure 6-2* shows three significant signals of this cancellation loop: the sample of the amplified signal (V_c), the reference signal (V_r) and the subtraction between them that results in the error signal (V_e).

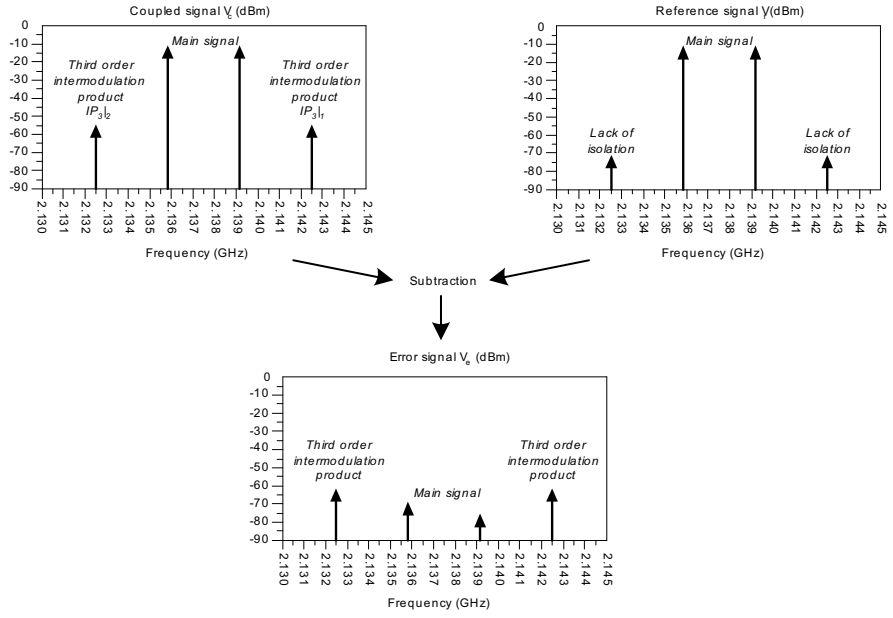


Figure 6-2. Coupled (V_c), reference (V_r) and error (V_e) signals in the first loop simulation

The reference and the error signals show two of the limitations seen in the mathematical analysis.

The lack of a total isolation in the power splitter (G_1, θ_1) and the power combiner (G_4, θ_4) is the raison d'être of the distortion component at the

reference signal (V_r). Ideally the reference signal should be a replica of the input signal, without any distortion component. This effect is analyzed later on.

1.3 Simulation of the distortion cancellation loop

The second cancellation loop tries to subtract the distortion components of the amplified signal with the distortion of the error signal. The error amplifier is considered ideal up to this point.

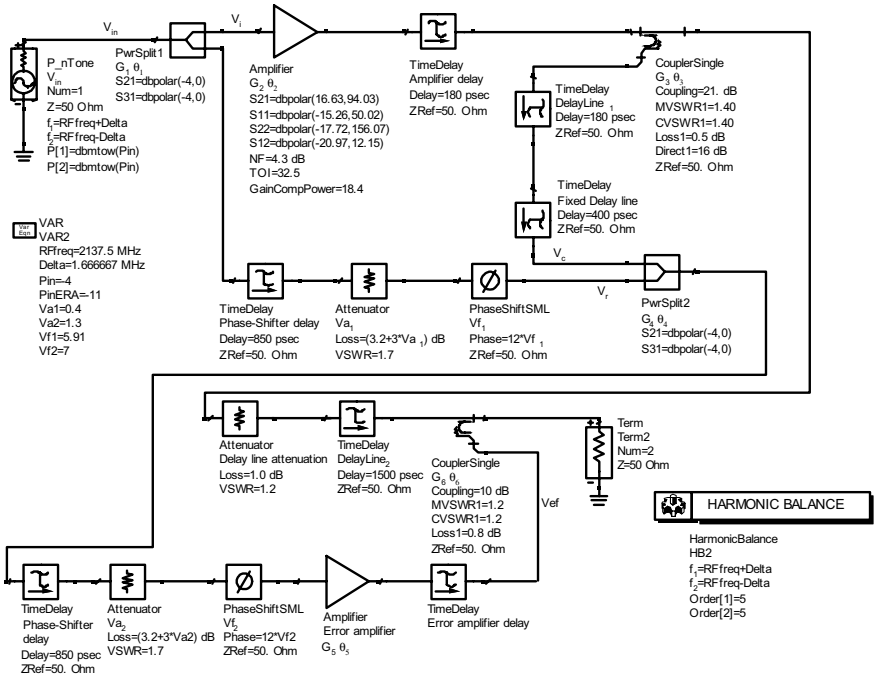


Figure 6-3. Distortion cancellation loop simulation

The adjustment method consists in variations of the Va_2, Vf_2 and $DelayLine_2$ parameters until the third order intermodulation products of the output signal are minimal or until the intermodulation distance reach the maximum value (*Figure 6-4*).

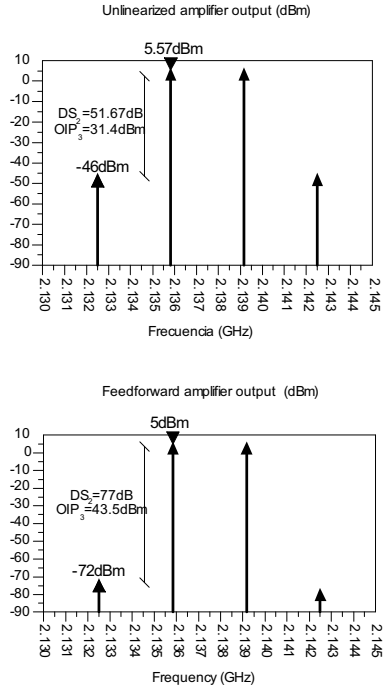


Figure 6-4. Simulated OIP_3 improvement

The simulation allows evaluating the influence of each one of the components over the linearization performance and selecting them in the offer provided by the wide market of the RF components.

1.4 The isolation effect

The lack of isolation in the power combiner (G_4, θ_4) can be simulated with a circulator placed as in *Figure 6-5*. $Direction_1$ term measures the reference signal part from the input port (V_r) and $Direction_2$ term measures the reference signal part from the isolated port (V_c).

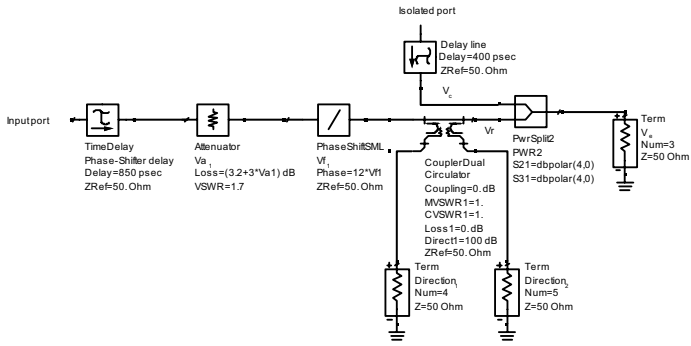


Figure 6-5. A Circulator before de power combiner: lack of isolation simulation

The simulation results shown in *Figure 6-6* prove that the coupled signal (V_c) is the main noise source of the distortion component of the reference signal.

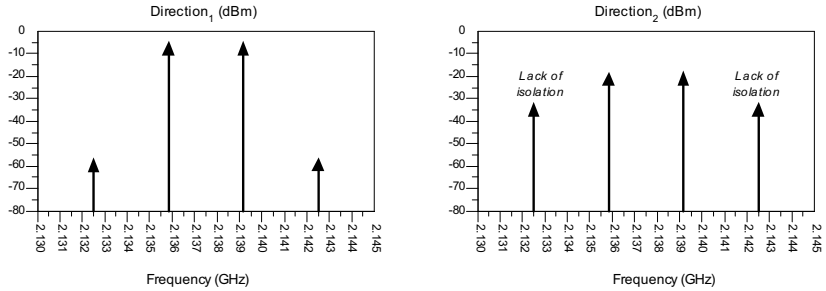


Figure 6-6. Isolation effect on reference signal

In order to avoid this effect an isolator is placed in the power combiner input, and the distortion component of the reference signal are reduced in 30dB (*Figure 6-7*).

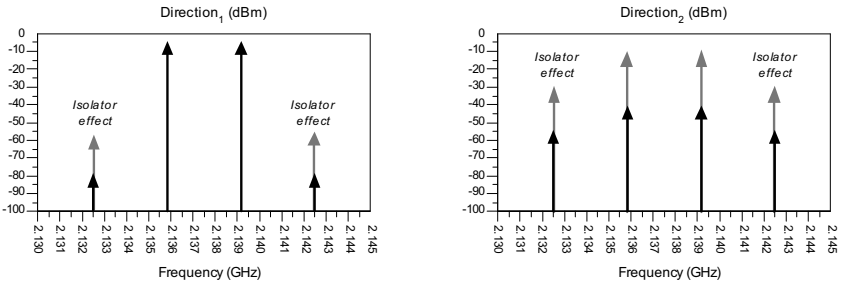


Figure 6-7. Reference signal improvement with isolator

This improvement on the isolation is rejected if the electromagnetic radiation over the air is not correctly isolated in the implementation stage. The electromagnetic compatibility (EMC) considerations must seriously applied [2].

2. SELECTION OF THE ERROR AMPLIFIER

The first consideration for the error amplifier selection is that the Feedforward technique must be designed to be competitive with the non-linearized amplifier. This means that production costs and above all the efficiency rates must be seriously taken into account.

The second consideration deals with the linearity performance of the error amplifier. The equation (5.28), now rewritten as (6.2), will help to understand such assertion.

$$-a_5 = a_3 + a_4 + a_6 + Va_2 \tag{6.2}$$

The gain of the error amplifier ($-a_5$) must compensate two coupling factors (a_3 y a_6), a power combiner loss (a_4) and the variable attenuation (Va_2). In real implementations this gain hovers around 50dB. Moreover, the distortion introduced by the error amplifier has a direct effect over the distortion component of the output signal (6.3). In other words, a trade-off is needed among gain, linearity, cost and efficiency rates.

$$\begin{aligned} V_o &= V_{in} \cdot 10^{\frac{1}{20}(\phi(Va_2)-B_0 \cdot w)} \cdot e^{j(\varphi(Vf_2)-D_0 \cdot w)} \\ &+ V_d \cdot \left[1 - 10^{\frac{1}{20}\left(\left(\sum_{i=3}^{6} b_i\right) \cdot w\right)} \right] \\ &+ V_{d\_error\_amp} \cdot 10^{\frac{1}{20}(a_6 - b_6 \cdot w)} \cdot e^{j(c_6 - d_6 \cdot w)} \end{aligned} \tag{6.3}$$

The peak power of the error amplifier (E_{pp}) depends on the maximum peak power of the main amplifier (M_{pp}), the ACLR level of the main signal (M_{ACLR}), the main signal cancellation level (M_C), the main signal path loss (G_{Path}), the second coupling factor (a_6) and an arbitrary security factor (SF) that represents the crest factor increase of the error signal [31].

$$E_{pp} = M_{pp} + 10 \cdot \log\left(3 \cdot 10^{\frac{-M_{ACLR}}{10}} + 10^{\frac{-M_C}{10}} \right) - G_{path} - G_6 + SF \tag{6.4}$$

The E_{pp} can be used as the specification of the 1dB compression point (P_{1dB}). Theoretically, the OIP_3 is approximately 9dB higher than the P_{1dB} (Chapter 2, section 4), but experimentally this factor often increases up to 12dB. As a result, the gain and linearity parameters of the error amplifier can be specified with (6.5).

$$\begin{aligned} OIP_3 &\approx E_{pp} + 12\text{dB} \\ G_5 &= G_3 + G_6 + G_4 + Va_2 - G_{path} \end{aligned} \tag{6.5}$$

A practical example (*Table 6-1*) will help to visualize the involvement of some parameters defined in (6.4) in the selection of the error amplifier.

Table 6-1. Feedforward amplifier example

Parameter type	Value
Main amplifier output power (M_p)	10dBm
Main amplifier peak power (M_{pp})	19dBm
Security factor (SF)	6dB
Loop 1 coupler (G_3)	20dB
Loop 2 coupler (G_6)	10dB
Loop 1 combiner (G_4)	4dB
Main signal path loss (G_{path})	1dB
Loop 2 variable attenuator (Va_2)	10dB

The gain parameter is fixed (45dB) by the loss factors of the directional coupler, the power combiner, the main signal path loss and the variable attenuator. However, the OIP_3 is independently affected by the ACLR level of the main amplifier and the suppression level (M_C) of the main signal component in the error loop. A graphical view is available in *Figure 6-8*.

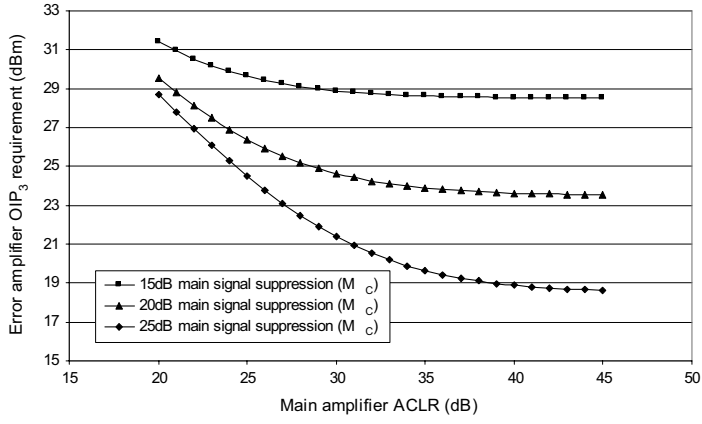


Figure 6-8. Error amplifier OIP_3 requirement vs. ACLR level for different main signal suppression levels

The main signal suppression level (M_C) is critical for high ACLR levels. This means that for small signals the cancellation of the main signal component in the error signal must be as precise as possible.

On the other hand, for high power signals (low ACLR) the M_C level will not have such influence and high linear (low efficient) error amplifiers will be needed in any case. This situation is considered as the decline of the linearization capability of the Feedforward technique, because of the error amplifier begins requiring more gain and linearity requirements than the main amplifier.

Some solutions [68] suggest the linearization of the error amplifier, though at expense of an efficiency reduction and an increase in price. A very useful thumb rule proposes that the error amplifier output level should be 10 to 15dB lower than the main amplifier's one [69].

3. THE ADJUSTMENT OF THE CANCELLATION LOOPS

The gain, loss and delay requirements of all components that constitute the Feedforward cancellation loops are obtained from the simulation stage. Some parameters are fixed by the final application, like the main amplifier gain and linearity specifications, the frequency bandwidth and the input power level. All the other ones must be conscientiously chosen from the wide offer provided by the market of RF components.

The next step consists in fitting the group delays for the signal subtractions.

3.1 Delay lines

Most of the delay components are based on coaxial or microstrip transmission lines, ceramic devices or specifically designed filters. Through the years several techniques have been proposed in order to minimize the main signal path loss caused by those delay devices [70-72]. Even in some ultra-high efficiency applications, the delay elements are discarded from the second cancellation loop.

The standard components – ceramic and filter devices – will never fit with the delay requirements of the cancellation loops. The coaxial cables provide a high accuracy in the group delay adjustment with enough flexibility throughout the prototyping process. Furthermore, they can be fabricated with custom specifications in serial productions, reducing considerably the assembly costs.

A comparison has been made between a standard ceramic device (*Table 6-2*) and a custom coaxial semi-rigid delay line (*Table 6-3*). Both of them are characterized for the same UMTS downlink channel used in the simulation cases. Note the path loss difference for the same accuracy rates.

Table 6-2. LDHA25N00BAA-300 fixed delay (Murata) for UMTS downlink

Frequency (MHz)		2110	2140	2170
S_{21}	Loss (dB)	4.74	4.76	5.26
	Phase shift (º)	-185	-240	-295
	Group delay (ps)	5122	5075	4950

Table 6-3. Quickform 81 coaxial cable for UMTS downlink

Frequency (MHz)		2110	2140	2170
S_{21}	Loss (dB)	1.3	1.28	1.17
	Phase shift (º)	-174	-227	-281
	Group delay (ps)	4910	4922	5093

Some recommendations are summarized from one of the principal transmission lines provider (Micro-Coax®), which design considerations are specifically focused on Feedforward amplifiers.

3.1.1 Configuration

Circular coils are the most economical and consistent delay line packages. Other configurations such as oval or square coils are too expensive and usually require an invasive service loop to compensate the cable physical length variations. Overall delay line performance is further enhanced with circular configurations because precise connector locations can be easily achieved through the coil relaxation.

3.1.2 Length tolerances

Length should be specified electrically in lieu of physically, since the physical length for a specified electrical length may vary. Tolerances of ±15 to ±20ps are recommended at frequencies between 800 and 2000MHz. Furthermore, since mechanical forming may induce electrical length changes, it is recommended that the final electrical testing and trimming be performed on fully coiled delay lines.

3.1.3 Packaging

It is recommended solder tacking as the most effective method to ground and maintain configuration in lieu of tie wrapping and tape. Delay lines may

be configured with surface mount technology (SMT) ready to solder directly onto printed circuit boards.

3.1.4 System design

It is recommended that the delay lines should be considered through out the system design process. Delay lines with long and complicated leads or pigtails are expensive, difficult to ship and ungainly to assemble. Also, any long leads are easily damaged, potentially affecting system performance. Finally, multiple delay lines can often be nested together for maximum utilization of space.

3.2 An adjustment method for delay lines

A practical guideline is described for the adjustment of the two delay lines, consisted in equalizing the group delays of the two paths that participate in each signal subtraction.

The *Figure 6-9* can help to visualize the adjustment procedure for the first cancellation loop.

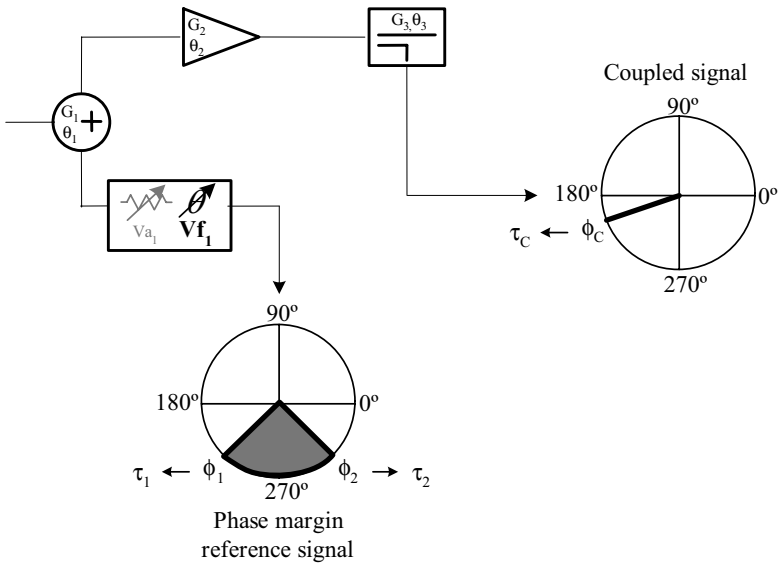


Figure 6-9. Standard adjustment method - Step 1

First of all, it is measured the group delay (τ_1-τ_2) margin for the entire phase margin (ϕ_1-ϕ_2) of the reference signal. This margin must be compared with the phase and delay values (ϕ_C-τ_C) of the coupled signal.

The group delay of the reference signal is considered higher than the coupled signal group delay, so the first delay line must be placed in the coupled path (6.6).

$$\tau_C + \tau_{dl} = \frac{\tau_1 + \tau_2}{2} \tag{6.6}$$

The new group delay of the coupled signal (τ_C) and the delay line (τ_{dl}) must be equal to the mean value of the reference signal group delay. The equation (6.6) can be also written in terms of phase shifts (6.8) with the well-known relation provided by a transmission line (6.7).

$$n \cdot \phi_{dl}[^o] = 360^o \cdot \tau_{dl}[\text{sec}] \cdot f[\text{Hz}] \quad \text{with } n = 1, 2, 3, \ldots \quad 0^o \le \phi_{dl} \le 360^o \tag{6.7}$$

The ‘n’ term is introduced in order to get phase values (ϕ) between 0º and 360º.

$$n \cdot \phi_{dl}[^o] = 360^o \cdot \left(\frac{\tau_1 + \tau_2}{2} - \tau_C \right)[\text{sec}] \cdot f[\text{Hz}] \quad \text{with } n = 1, 2, 3, \ldots \tag{6.8}$$

Secondly, it must be guaranteed the phase conditions for the signal subtraction: the reference and coupled signal must differ in 180º (6.9).

$$\phi_{dl}[^o] = \frac{\phi_1[^o] + \phi_2[^o]}{2} \pm 180^o - \phi_C[^o] \tag{6.9}$$

Usually, the (6.8) and (6.9) equations will not coincide. In *Figure 6-10* these phase margins are distinguished, one for each equation.

Ideally, both phase margins, should be superimposed, but sadly they are. The only way to match them is relaxing one of the identities, mainly the one which fixes the group delay (6.8). In other words, a slight group delay error is allowed in order to make the signal cancellation possible. In *Figure 6-11* this group delay error is added (τ_e).

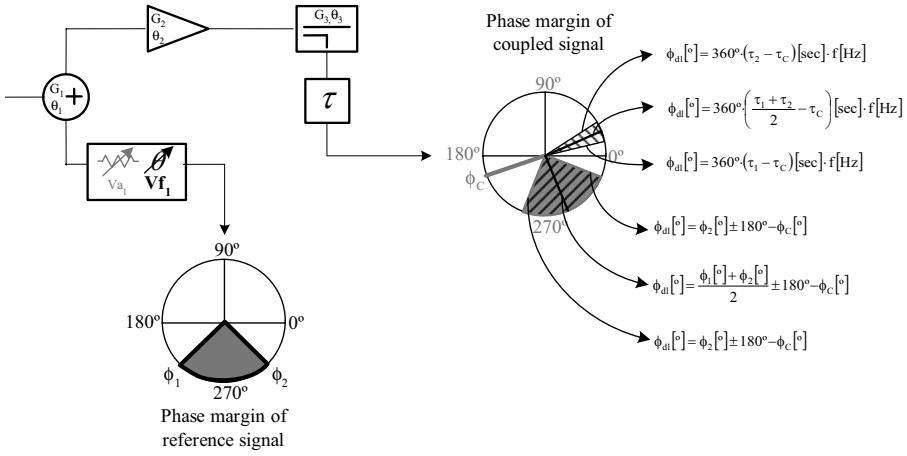


Figure 6-10. Delay element phase shifts according to delay and phase conditions

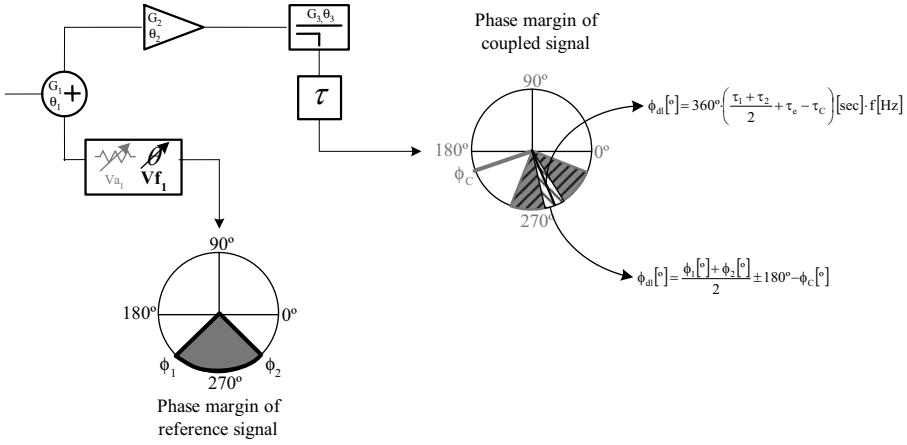


Figure 6-11. Phase margins adjusted for main signal cancellation (Loop 1)

The resultant electrical parameter (ϕ) must be transformed into a physical specification – the coaxial cable length – with (6.10).

$$L[\text{cm}] = \tau[\text{ps}] \cdot c[\text{km/sec}] \cdot \varepsilon \cdot 10^{-7} \tag{6.10}$$

ε is the specific factor of each coaxial cable that represents the relative value with respect to the speed of the light (c) (299792 km/sec).

The *Figure 6-12* shows the cancellation level of the main signal component of the error loop in a prototype implemented with this adjustment method.

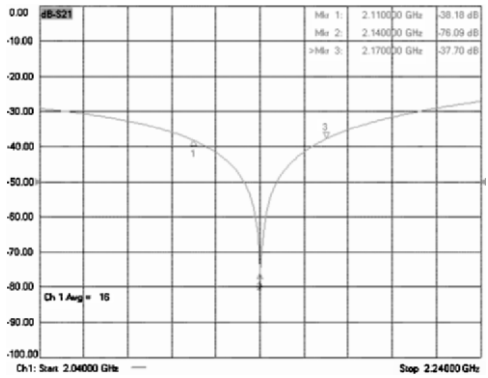


Figure 6-12. Error signal cancellation level

The delay line of the distortion cancellation loop is calculated in a similar way. The only difference lies in the test fixture, shown in *Figure 6-13*.

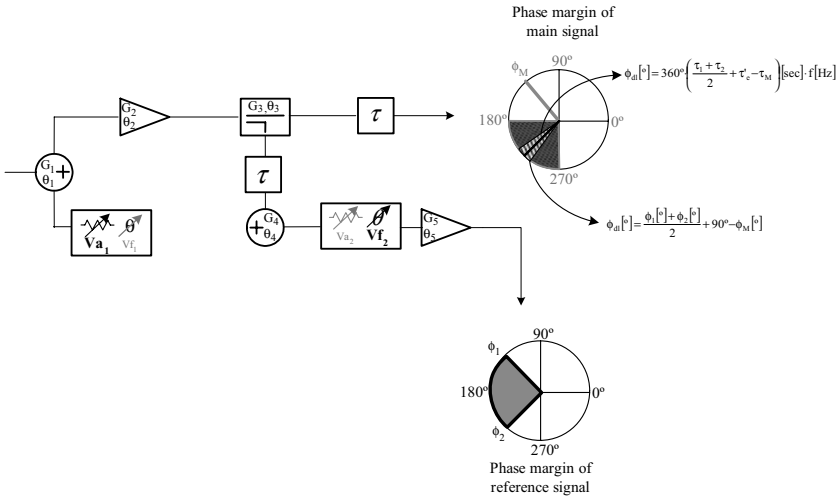


Figure 6-13. Phase margins adjusted for distortion signal cancellation (Loop 2)

This delay line must equalize only the group delays of the second cancellation loop, so the first loop attenuator has been set to its maximum level.

The signal subtraction now is done in a directional coupler instead of in a power combiner. This fact changes the phase condition of equation (6.9) into (6.11).

$$\phi_{dl}[^o] = \frac{\phi_1[^o] + \phi_2[^o]}{2} + 90^o - \phi_C[^o] \tag{6.11}$$

The group delay error introduced in order to superimpose both phase margins is represented with the τ'_{ε}.

4. DISTORTION ENHANCED MEASUREMENT TECHNIQUES

It is the moment of evaluating the improvements obtained with the Feedforward linearization technique. The measurement techniques used in obtaining the OIP_3 and ACLR parameters are as important as the implementation itself, because any mistake could mask all the improvements achieved in the implementation process.

The distortion generated by any digital modulation that employs both amplitude and phase shifts is known as spectral regrowth. As depicted in *Figure 6-14*, the spectral regrowth falls outside the main channel into the lower and upper adjacent channels [73].

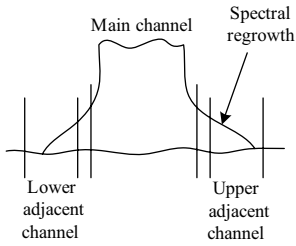


Figure 6-14. Spectral regrowth of a digitally modulated signal

Like other distortion measurements, the spectrum analyzer generates its own internally generated spectral regrowth distortion. In most cases, this spectral regrowth distortion is due to the third order intermodulation products. In addition to spectral regrowth, the phase noise and broadband

noise of the spectrum analyzer also limit the dynamic range of this type of distortion measurements.

As it has been introduced in Chapter 2, the Adjacent Channel Leakage Ratio (ACLR) is the ratio of the main channel power to the power in either of the adjacent channels. Some modulation formats require a spot measurement where power measurements are made at specific frequency offsets in the main and adjacent channels. Other formats require an integrated power measurement where the spectrum analyzer individually computes the total power across the entire main channel and each of the adjacent channels. In either case, the designer must properly set the mixer level of the spectrum analyzer in order to minimize the internally generated spectral regrowth.

However, minimizing the internally generated spectral regrowth comes at the price of increasing the broadband noise; therefore, a trade-off must be reached between both. Another complicating fact is that the mixer level must be set based on the peak to average ratio of the modulated signal.

4.1 Spectrum analyzer mixer level optimization

The distortion measurement techniques detailed in Chapter 2 considers that the spectrum analyzer distortion products fall below the distortion being measured. The techniques that guarantee accurate measurements by ensuring that the spectrum analyzer distortion does not mask the device under test (DUT) distortion, do not allow full use of the available dynamic range of the spectrum analyzer.

In order to make distortion measurements on highly linear devices which distortions are already very low, the designer must override the auto-couple features of the spectrum analyzer, allowing more flexibility in optimizing the dynamic range of the spectrum analyzer.

4.1.1 Mixer level

The first step is controlling the amount of power present at the first mixer of the spectrum analyzer. Optimizing this mixer level maximizes the dynamic range of the spectrum analyzer.

When a mixer level is set too low, the spectrum analyzer noise floor limits the distortion measurement, and when it is set too high, the distortion products of the spectrum analyzer limit the distortion measurement.

The dynamic range charts found in many spectrum analyzer data sheets show the dynamic range plotted against the mixer level. This is an extremely useful tool to understand how to set best the mixer level for the second harmonic and the third order intermodulation distortion measurements.

Learning how to construct this chart not only assists in understanding how to use them, but it also allows flexibility so that the designer can customize the chart for the actual spectrum analyzer performance. Signal to noise ratio (S/N), signal to distortion ratio and the phase noise contribute to the construction of this dynamic range chart. All of these individual terms will be discussed individually.

4.1.2 Signal to noise ratio versus mixer level

The spectrum analyzer can be characterized by an output vs. input power transfer function (*Figure 6-15*). Pin is the power at the RF input port and Pout is the signal as it appears on the spectrum analyzer display.

The noise floor of the spectrum analyzer, known as Displayed Average Noise Level (DANL), places a limitation on the smallest amplitude that can be measured.

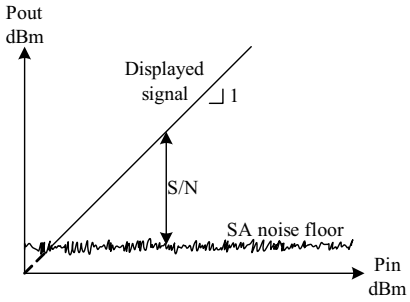


Figure 6-15. Spectrum analyzer Pout vs. Pin

The input power can be reduced in one of two ways: either the power level is decreased externally or the spectrum analyzer input attenuation is increased.

Another way of presenting the *Figure 6-15* is to plot the S/N vs. the mixer level (*Figure 6-16*).

At the 0dBc point, the mixer level is the DANL for 1Hz resolution bandwidth (RBW) and 0dB input attenuation. For example, in *Figure 6-16*, the spectrum analyzer DANL is -155dBm in a 1Hz RBW, measured with 0dB input attenuation.

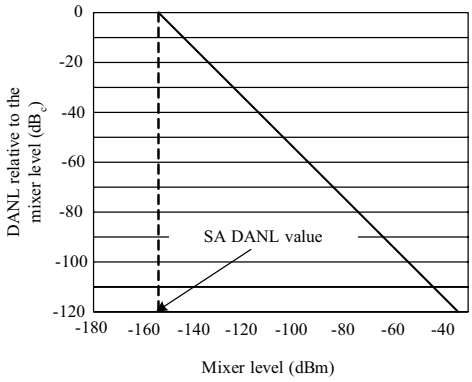


Figure 6-16. Signal to noise vs. power at the input mixer

The noise floor of the spectrum analyzer can be affected in two ways. The noise floor rises with the RBW setting according to 10log (RBW[Hz]). Increasing the RBW by a factor of 10, the noise floor increases by 10dB. *Figure 6-17* shows the noise floors for 1Hz, 10Hz and 1kHz RBW.

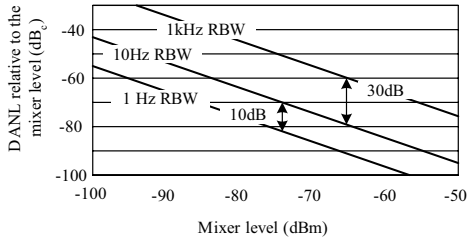


Figure 6-17. Noise floor with different RBW settings

The other mechanism that affects the displayed noise floor is the averaging scale. The spectrum analyzers have two average scales for power measurements: logarithmic power (Video) and root mean square power (RMS). *Figure 6-18* shows the relationship between both.

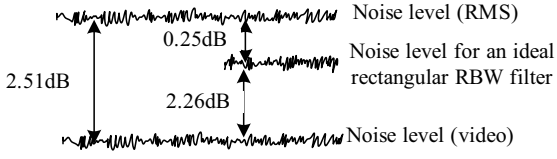


Figure 6-18. Averaging scale effect on the displayed noise

4.1.3 Signal to noise ratio with external noise

The S/N vs. mixer level graph also is useful when the noise at the input port is greater than the noise floor of the spectrum analyzer. This excess noise can stem from devices with relatively low S/N ratios comparing with the spectrum analyzer.

Figure 6-19 shows the situation where the external noise from the DUT is greater than the noise floor of the spectrum analyzer. At high signal power levels the external S/N stays constant. As the input signal power decreases, the external noise falls below the noise floor of the spectrum analyzer.

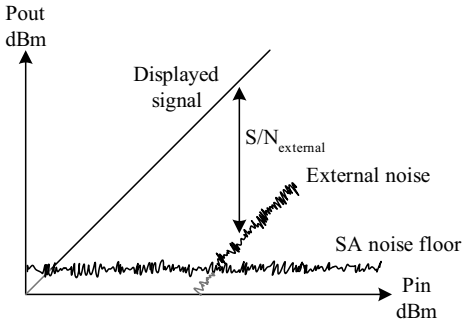


Figure 6-19. Spectrum analyzer Pout vs. Pin with external noise

Figure 6-20 adds the external noise on the S/N vs. the mixer level graph. At high powers, the DANL relative to the mixer level stays constant and at lower mixer levels, the S/N curve shows the familiar slope of -1.

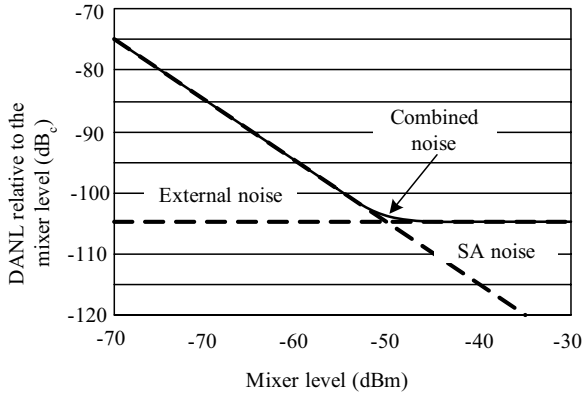


Figure 6-20. Signal to noise vs. mixer level with external noise greater than SA noise

4.1.4 Signal to distortion ratio versus mixer level

It can also be plotted the signal to distortion ratio (S/D) versus the input mixer level. *Figure 6-21* shows the distortion relative to the mixer level in dBc units vs. the input mixer level for the second and the third order distortion components.

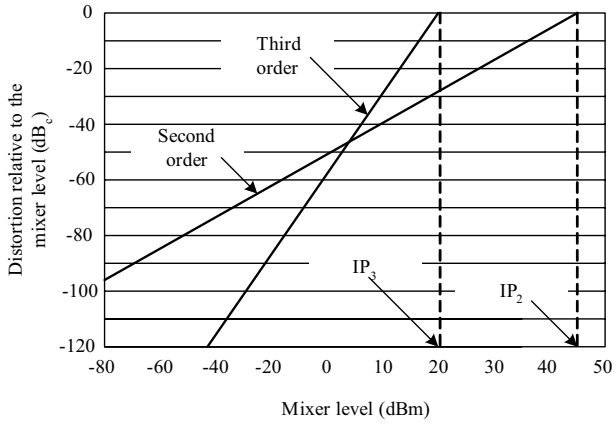


Figure 6-21. Signal to distortion versus power at the input mixer

For the second order distortion the slope of the S/D vs. the mixer level curve is +1, and for the third order +2. For the second harmonic curve the 0dBc intersection point corresponds to the second order intercept point (IP_2). In *Figure 6-21*, the IP_2 is 45dBm. For the third order curve the 0dBc intersection point corresponds to the third order intercept point (IP_3). In this example, the IP_3 of the spectrum analyzer is 20dBm.

4.1.5 The dynamic range chart

Combining the signal to noise and the signal to distortion ratios vs. the mixer level into the same graph yields the dynamic range chart (*Figure 6-22*).

The dynamic range chart allows a visual means for determining the maximum dynamic range and the optimum input mixer level for the maximum dynamic range.

Simple geometry yields practical equations for the maximum dynamic range and the optimum mixer levels for the second order harmonic (6.12) and the third order intermodulation (6.13) distortions.

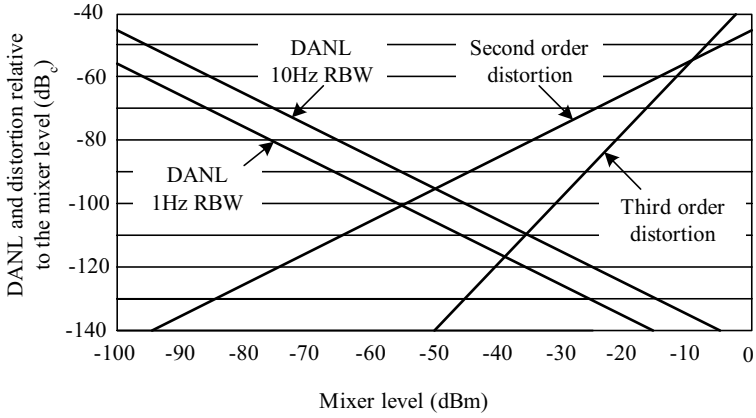


Figure 6-22. Dynamic range chart

$$\begin{aligned}&\text{Maximum Dynamic Range} = 1/2\,[IP_2 - \text{DANL}]\,\text{dB}\\&\text{Optimum Mixer Level} = 1/2\,[IP_2 + \text{DANL}]\,\text{dBm}\end{aligned} \tag{6.12}$$

$$\begin{aligned}&\text{Maximum Dynamic Range} = 2/3\,[IP_3 - \text{DANL}]\,\text{dB}\\&\text{Optimum Mixer Level} = 1/3\,[2 \times IP_3 + \text{DANL}]\,\text{dBm}\end{aligned} \tag{6.13}$$

These parameters are summarized in *Table 6-4* for the example of *Figure 6-22*. Remember that the DANL is -145dBm in 10kHz RBW, the IP_3 is 20dBm and the IP_2 is 45dBm.

Table 6-4. SA limits for second and third order distortion measurement

10kHz RBW	Maximum Dynamic Range	Optimum Mixer Level
Second order distortion	95dB	-50dBm
Third order distortion	110dB	-35dBm

4.1.6 Spectrum analyzer distortion

If distortion products of the device under test fall close to the internally generated distortion products, it will be an uncertainty in the displayed distortion. The error due to the addition of two coherent sinusoidal tones is bounded by the values shown in equation (6.14).

$$\text{Amplitude error} = 20 \cdot \log\left(1 \pm 10^{\frac{d}{20}}\right)[\text{dB}] \tag{6.14}$$

Where "d" is the relative amplitude of two tones in dB (a negative number).

+ If the DUT and SA distortion products add in phase

- If the DUT and SA distortion products add in out of phase (180º)

For a two tone procedure with ≤1MHz tone spacing, the displayed amplitude error is given by the positive sum of equation (6.14). To ensure that the measurement error due to the combination of the DUT and SA distortion products falls below a given threshold, the optimum mixer level requires a readjustment. Unfortunately, this readjustment has an adverse effect on the maximum dynamic range. The following procedure helps computing the readjusted of the dynamic range and the resulting optimum mixer level needed to ensure that the distortion measurement uncertainty falls below a desired error level.

First of all, the equation (6.14) must be solved for a specific threshold (d), which is the relative amplitude between the external and the internal distortion product amplitudes. This relative amplitude value is then used to determine how to offset the distortion curves in the dynamic range chart by -d dB.

For the second order distortion, the effective IP_2 is offset by "d" and for the third order intermodulation the effective IP_3 is offset by 0.5d. Offsetting the intercept points, instead of offsetting the curves, the equations (6.12) and (6.13) can be used to calculate the optimum mixer levels and the maximum dynamic ranges.

4.2 Enhanced ACLR measurements

Three characteristics, inherent to any spectrum analyzer, limit the dynamic range when measuring the out of channel distortion on wideband communications systems: the noise floor, the phase noise and the intermodulation distortion. The noise floor limit is the same as seen for the single wave distortion measurement in the previous subsection. Phase noise and intermodulation, however, are limits that depend on such parameters as channel separation and modulation bandwidth. In other words, phase noise or intermodulation will usually limit the dynamic range of the spectrum analyzer depending on which modulation type is being used.

4.2.1 Signal to noise ratio

The spectrum analyzer displays the main channel power spectral density (PSD) $10\log(BW_m)$ below the amplitude of a continuous wave (CW), where BW_m is the modulation bandwidth. Therefore, the displayed S/N of a digitally modulated signal is reduced by the same factor (*Figure 6-23*).

Furthermore, the displayed amplitude and the displayed broadband noise are functions of the RBW setting which, unlike the CW case, renders the S/N independent of the RBW.

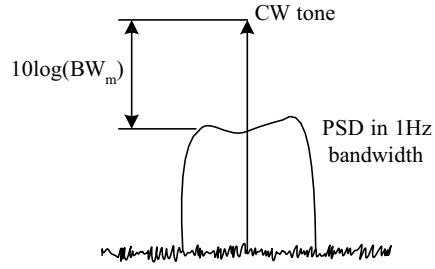


Figure 6-23. The power spectral density (PSD) of a digitally modulated signal

The displayed main channel PSD is given by (6.15).

$$P_{ch} - 10\log\left(\frac{BW_m}{RBW}\right) + 10\log\left(\frac{NBW}{RBW}\right) \tag{6.15}$$

P_{ch} is the total signal power and NBW is the noise power bandwidth of the RBW filter used for the measurement. The NBW/RBW ratio is a characteristic of each spectrum analyzer.

The equation (6.16) shows the average displayed amplitude of a -10dBm WCDMA signal measurement with a 30kHz RBW filter setting in a performance spectrum analyzer (PSA) of Agilent Technologies.

$$\text{Average displayed amplitude} = -10 - 10\log\left(\frac{3.84\,\text{MHz}}{30\,\text{kHz}}\right) + 10\log(1.06) = \text{-30.82dBm} \tag{6.16}$$

On the other hand, the broadband noise on the same RMS averaging scale is given by (6.17) with a 10dB input attenuator. The signal to noise ratio then computes to 66.88dB.

$$\text{Broadban noise} = \text{DANL} + 10\log(\text{RBW}) + 2.51\text{dB} + \text{input attenuator} = \text{-97.7dBm} \tag{6.17}$$

In order to plot the S/N on the dynamic range chart, it is best to consider the channel power (P_{ch}) in terms of the input mixer power of the spectrum analyzer (ML_{ch}).

$$ML_{ch} = P_{ch} - \text{input attenuation} \tag{6.18}$$

Thus, S/N as a function of the input mixer level is given by the equation (6.19).

$$S/N = ML_{ch} - 10 \cdot \log(BW_m) + 10 \cdot \log(NBW/RBW) - DANL - 2.51[dB] \quad (6.19)$$

Equation (6.19) makes it evident that for digitally modulated signals, the S/N is a function of the modulation bandwidth. Wider bandwidths lead to a lower S/N. *Figure 6-25* shows the dynamic range chart with the S/N plotted vs. the input mixer level for a WCDMA signal (3.84MHz). The S/N curve of the modulated signal is offset by (6.20) from the DANL curve.

$$S/N = ML_{ch} - DANL - 68.1[dB] \quad (6.20)$$

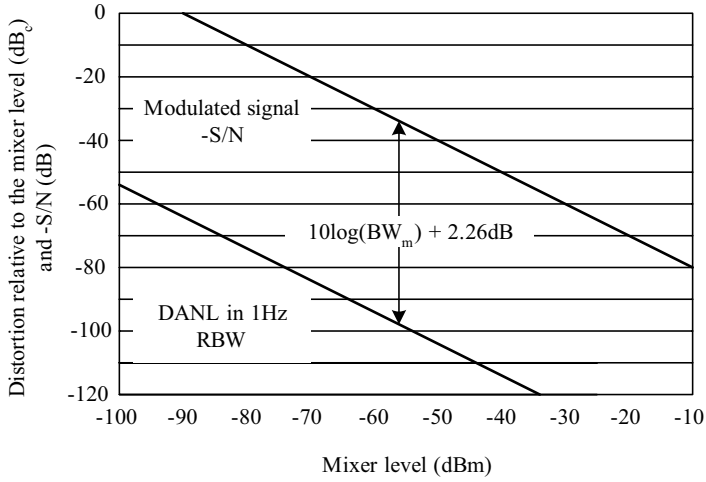


Figure 6-24. S/N vs. mixer level for a WCDMA signal in a PSA (Agilent Technologies)

4.2.2 Spectral regrowth

Third, and in some cases fifth order intermodulation distortions of the spectrum analyzer fall outside the main channel. Unlike power amplifiers, and especially Feedforward power amplifiers, the spectral regrowth of the spectrum analyzer can easily be approximated with simple algebra. The analysis of the spectral regrowth generated by the intermodulation distortions of the spectrum analyzer relies on the premise that when the power at the input mixer is far below the gain compression power level (by at least 15 dB), the spectrum analyzer behaves as a weakly-nonlinear device.

Spectral regrowth due to the intermodulation distortion is noise-like, so the displayed main channel PSD and the spectral regrowth PSD both vary by the 10log(RBW) term. Another implication is that when the distortion approaches the system noise floor. In those cases the distortion and noise add as uncorrelated powers using the (6.21) equation.

$$\text{Total power} = 10 \cdot \log\left(10^{\text{noise power}/10} + 10^{\text{distortion power}/10}\right)\text{dBm} \tag{6.21}$$

The noise power is the system noise in dBm and distortion power is the intermodulation spectral regrowth in dBm.

For measurements that rely on an integrated power measurement across the adjacent channel, predicting the level of the spectral regrowth due to third order distortion depends on the modulation bandwidth and the channel spacing. WCDMA is one important class of wideband signals that uses this kind of measurement.

For a WCDMA signal with 3.84MHz symbol rate and 5MHz channel spacing, the effective IP_3 offset is given in *Table 6-5*.

Table 6-5. Effective TOI offset for WCDMA signals

Peak t average (dB)	IP_3 offset (dB)
5.5	7.5
11	4
14.5	2

This table shows how to calculate the integrated adjacent channel power due to the third order distortion. For example, suppose 11dB of peak to average ratio and a 20dBm IP_3 of the spectrum analyzer. Then, the effective IP_3 with a WCDMA modulated signal will be 24dBm. Manipulating the math, the spectral regrowth in the adjacent channel, assuming -10dBm at the input mixer, is obtained by (6.22).

$$\Delta = 2 \cdot (\text{TOI - P}) = 2 \cdot (24 - (-10)) = 68\text{dB} \tag{6.22}$$

That is, the power integrated across the adjacent channel is 68dB below the power integrated across the main channel.

4.2.3 Phase noise influence

Phase noise also places a limitation on the dynamic range when measuring digital signals. The model used to calculate the intermodulation distortion of a digitally modulated signal also proves useful in showing how the phase noise adds to the spectral regrowth in the adjacent channel. In

Figure 6-25, the main channel is divided into segments of equal frequency widths with a single wave tone representing the power in each segment.

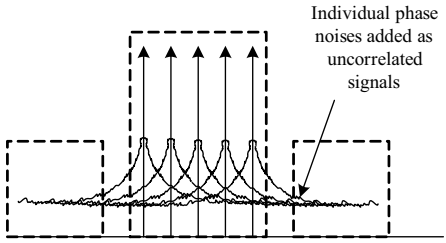


Figure 6-25. Phase noise effect in digitally modulated signal

The result is a reduction in the dynamic range of roughly $10\log(BW_m)$. For example, the phase noise specified by the spectrum analyzer is -118dBc/Hz (100kHz offset). With a modulation bandwidth of 3.84MHz, the phase noise power at 100kHz away from the edge of the main channel relative to the PSD of the main channel would be given by (6.23).

$$-118\text{dBc/Hz} + 10\log\left(3.84 \cdot 10^6\right) + \log\left(\frac{\text{NBW}}{\text{RBW}}\right) \tag{6.23}$$

4.2.4 Dynamic range chart for wideband signals

Now there are all the ingredients to create the dynamic range chart for the distortion measurements on wideband communication systems. It is important to remember that the goal is to determine the optimum input mixer level, which is a natural outcome of the dynamic range chart.

Unfortunately, the dynamic range chart for digitally modulated signals highly depend on the format parameters like modulation bandwidth, channel spacing and the peak to average ratio of the main channel signal. Though it is impossible to generalize the dynamic range chart, it is possible to get a feeling for the considerations involved in constructing a suitable dynamic range chart by showing a specific example.

Figure 6-26 shows the dynamic range chart for a WCDMA signal. The peak to average ratio is 11dB. The measurement uses an integrated power across both the main and the adjacent channels. The parameters used to construct this chart are given in *Table 6-6*.

Table 6-6. Parameters used in the dynamic chart design

Parameter	Value
Symbol rate	3.84MHz
Channel spacing	5MHz
DANL	-155dBm/Hz
TOI	21dBm+4dB offset (11dB peak to average ratio)
5^{th} order Intercept	12dBm
Phase noise	-150dBc/Hz (1.17MHz offset)

The frequency offset used to estimate the phase noise contribution results from the fact that the frequency difference between the edge of the main channel and the edge of the adjacent channel is 1.17MHz.

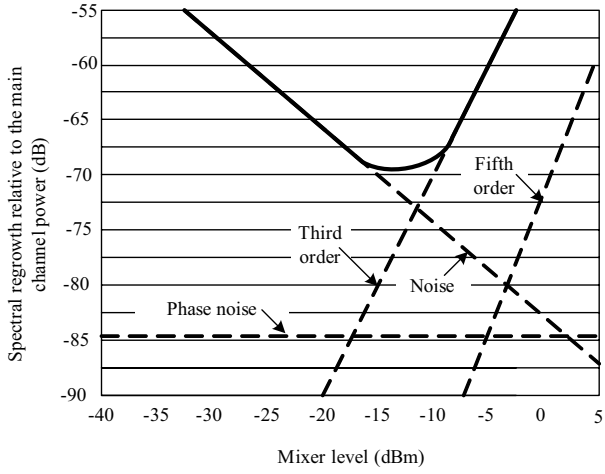


Figure 6-26. Dynamic chart for WCDMA signal

5. IMPROVEMENTS OF FEEDFORWARD AMPLIFIER

A Feedforward prototype has been implemented, following the design considerations proposed in this work for a UMTS downlink transmitters (2110-2170MHz) [32].

The enhancement achieved with this prototype is presented as OIP_3 and ACLR improvements. Obviously, the measurement guidelines of this chapter have been also conscientiously carried out.

First of all, the output power of the none-linearized amplifier is fixed. Then, the Feedforward amplifier is linearized for this specific output power with the basic adjusting method seen in Chapter 4.

All measurements are done with a performance spectrum analyzer (PSA) of Agilent Technologies (E4440), which main specifications are summarized in the *Table 6-7*.

Table 6-7. PSA (E4440) specifications

Frequency coverage	3Hz to 26.5Hz
DANL	-153dBm (10MHz to 3GHz)
Absolute accuracy	±0.27 (50MHz)
Frequency response	±0.40dB (3Hz to 3GHz)
Display scale fidelity	±0.04dB total (below -20dBm)
TOI (mixer level -30dBm)	±16dBm (400MHz to 2GHz)
	±17dBm (2GHz to 2.7GHz)
	±16dBm (2.7GHz to 3GHz)
Noise sidebands (10kHz offset)	-113dBc/Hz (CF = 1GHz)
1dB gain compression	+3dBm (200MHz to 6.6GHz)
Attenuator	0-70dB in 2dB steps

5.1.1 OIP_3 improvement

A two tone signal is used, settled in 2142.5MHz and 5MHz apart. However, in *Figure 6-27*, only the third order intermodulation product is shown, as recommended in [73].

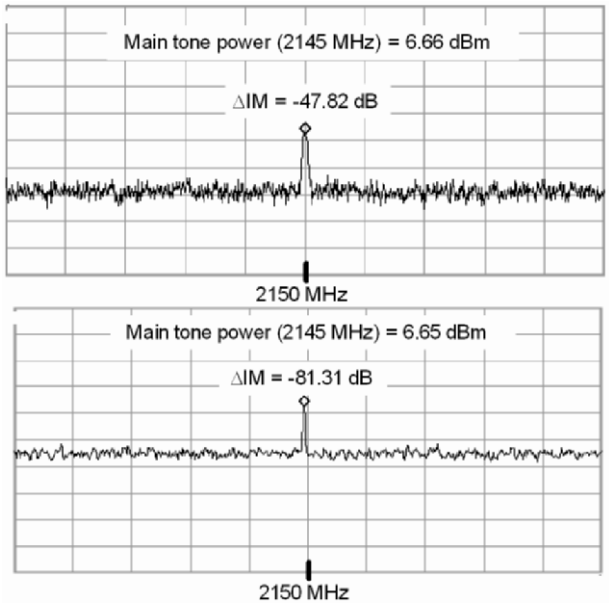


Figure 6-27. Non-linearized (above) and Feedforward (below) amplifier third order intermodulation distortion distance (I_{2T}) respectively

The Feedforward amplifier has increased the third order intermodulation distortion distance (I_{2T}) by 33.5dB for the same output power level, which implies a 16.7dB improvement on the OIP_3.

5.1.2 ACLR improvement

The high crest factors of WCDMA signals, which usually are between 8dB and 18dB, reduce the cancellation levels achieved with the two tone procedure [74]. The WCDMA signal used is the test model 1 –64DTCH– defined by the UMTS standard [61]. The centre frequency is 2142.5MHz, the channel bandwidth is 3.84MHz, the adjacent channels are at 5MHz, the ACLR is measured in 3.84MHz bandwidth and a Peak/Avg. is 10dB for 0.01% probability on the CCDF curve.

The Feedforward amplifier is again adjusted with the same basic method. The maximum ACLR improvement for the same output level is 15dB (*Figure 6-28*).

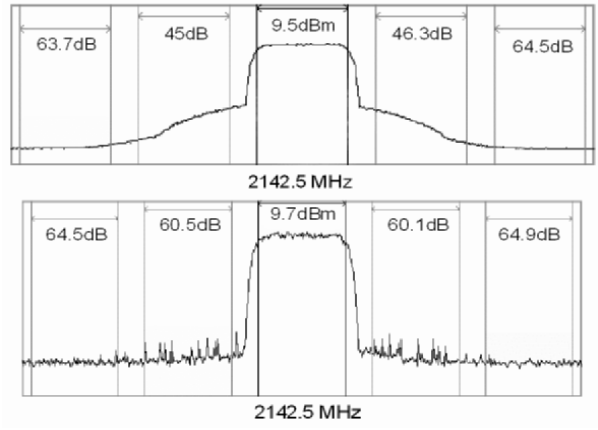


Figure 6-28. The Feedforward ACLR level in comparison with none-linearized amplifier for the same output level

Table 6-8. Enhancement of the Feedforward amplifier

		Non-linearized	Feedforward	Improvement
2 tone procedure	Pout	6,7dBm	6,65dBm	-
	I_{2T}	47,8dB	81,3dB	33,5dB
	OIP_3	30,6dBm	47,3dBm	16,7dB
Digitally modulated	Pout	9,5dBm	9,7dBm	-
	ACLR (worst)	45dB	60,5dB	15,5dB

All these measurements are summarized in *Table 6-8*, and they can be compared with the state of the art [75-82].

Chapter 7

ADAPTIVE FEEDFORWARD AMPLIFIERS

The performance of the output signal sensed in the characterization of Chapter 5 led to determine alternative adjustment methods to the basic one that try to achieve the desired trade-off between the linearity and efficiency rates of Feedforward amplifiers.

The basic Feedforward linearization technique is based on minimizing the intermodulation distortion emitted over the adjacent channels. However, some unwanted non ideal effects cause interesting variations over the main component of the output signal, making the efficiency improvement possible.

This chapter attempts to design the adjusting method that best adapts to the wideband digital communication systems with specific distortion specifications and maximum efficiency requirements.

Some traditional control methods are reviewed for the distortion detection in Feedforward amplifiers, and an alternative one is proposed adapted to the control method previously designed.

Finally, a comparison is made, between the basic and the adaptive control method proposed in this book, for the prototype implemented in Chapter 6.

1. ADAPTIVE ADJUSTING METHODS FOR FEEDFORWARD AMPLIFIERS

The basic adjusting method of Feedforward amplifiers minimizes, ideally eliminates, the distortion emissions coming from the power amplifier. As it has been confirmed in Chapter 5, all the adjusting parameters have some influence over the distortion cancellation, and all of them do it in a

convergent way. This adjusting method is the most used in the Feedforward amplifiers, and it is going to be named as the *Maximum Cancellation* (MC) method.

However, the influence of the adjustment parameters over the output power level makes possible the deliberated cancellation worsening, up to a predetermined limit, in order to maximize the output power as much as possible. This represents an alternative adjustment method that maximizes the output power level while keeps the distortion level under a specific value.

The distortion specification of each communication standard limits the unwanted emissions immediately outside the channel bandwidth. In those cases, the distortion minimization is not practical at all. From now on, this control method will be named the *Maximum Output* (MO) method [32].

1.1 Maximum cancellation method

The maximum cancellation (MC) method is based on the cancellation of the power amplifier distortion, so the Feedforward standard adjustment mechanism is used as during the circuit assembly, divided in three steps.

1.1.1 Error signal minimization

The adjustment parameters of the first cancellation loop (Va_1, Vf_1) minimize the main signal component of the error signal (*Figure 7-1*). Contrary to observed in the two tone signal analysis, the main signal component is not cancelled completely and reaches, at most, the same level as the distortion component.

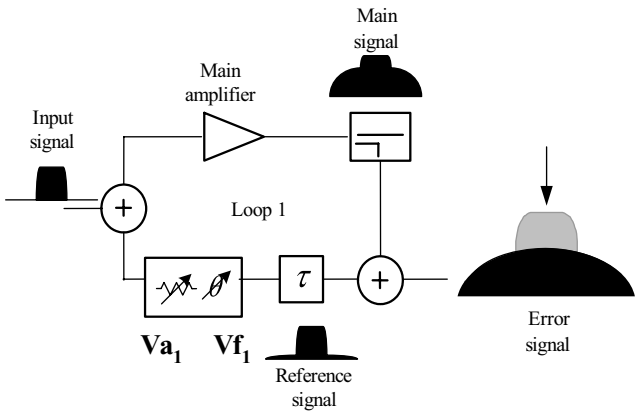


Figure 7-1. Error signal minimization

1.1.2 Distortion cancellation

The distortion component of the output signal is minimized with the second loop adjustment parameters (Va_2, Vf_2), as shown in *Figure 7-2*. As it occurs with the error signal, the distortion component converges uniformly to a unique minimum (*Figure 5-10*) so simple heuristic methods like Hill-Climbing could be employed in the control algorithm [83].

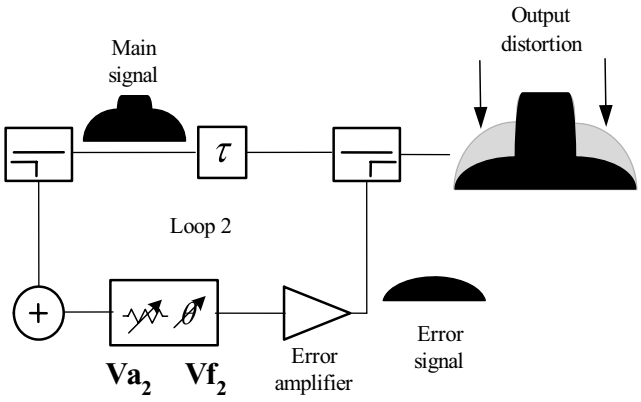


Figure 7-2. Distortion signal minimization

It is expected a total main signal cancellation in the error loop, so the main component of the output signal ought not to suffer any variation.

1.1.3 Maximum cancellation

The first loop adjustment parameters (Va_1, Vf_1) have a slight influence over the cancellation component of the output signal (*Figure 5-11*).

When the second loop adjustment parameters minimize the distortion component of the output signal, all the adjusting parameters start minimizing it simultaneously (*Figure 7-3*).

Fortunately, the distortion cancellation according to the first loop adjustment parameters performs uniformly convergent to a unique minimum, so the same control algorithm, based on any simple heuristic method, could be used for all the maximum cancellation (MC) method.

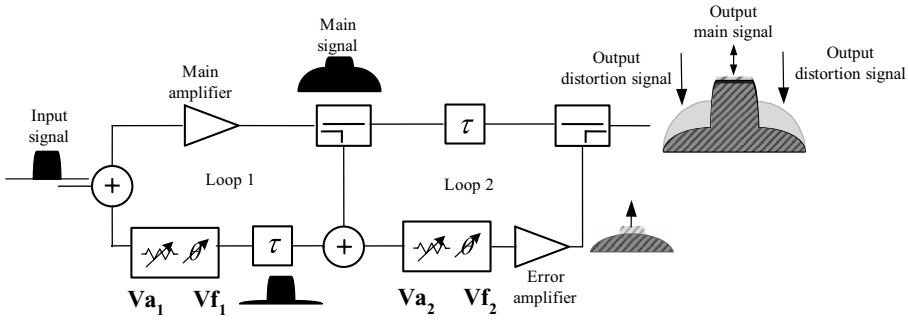


Figure 7-3. Distortion signal minimization with all the adjustment parameters

1.2 Maximum output method

The maximum output (MO) method is based on the influence of the adjustment parameters over the output main signal (*Figure 5-12* and *Figure 5-13*).

Two considerations must be done regarding the MO method. On the one hand, any change of the adjustment parameters from the MC method, inevitably, makes the distortion cancellation worse. This means that while the output main signal is deliberately modified, the distortion level is also changed, so it must be permanently checked in order to not exceed the limits fixed by each standard specification.

On the other hand, the output distortion is also limited by the error amplifier distortion. Remember that if the main signal suppression level (M_C) is decreased, the linearity requirement of the error amplifier is increased (Chapter 6 – Section 2).

Now in the MO method, in contrast to the MC method, two conditions must be fulfilled for the final success: the distortion level specification and the maximum output power.

The maximum cancellation can be represented with a single variable: the distortion component of the output signal. Furthermore, this variable is uniformly convergent to a unique minimum according to the four adjustment parameters.

The two conditions of the MO method, however, can be enclosed in an objective function (F) which maximum coincides with the success of the control criterion (7.1).

$$F = \text{Output power} - K \left| \text{Distortion Specification} - \text{Actual Distortion} \right| \quad (7.1)$$

In this function, the priority between efficiency and linearity is given by the K constant, which depends exclusively on the input power level. For a given K value the function decreases the farther the system is from the distortion specification, and increases the greater the output power is. An example of this function is given in *Figure 7-4*.

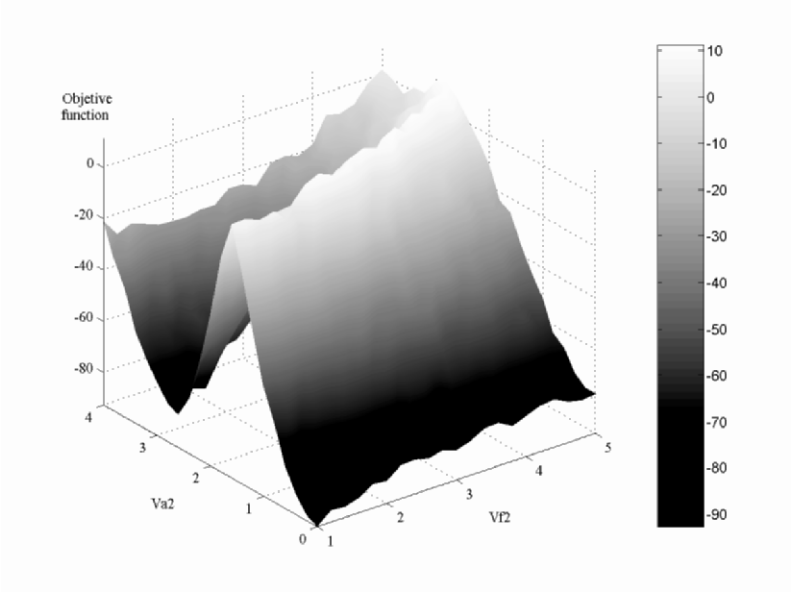


Figure 7-4. Objective function (K=10) of an Feedforward amplifier (ACLR=45dB)

The K parameter decides if the linearity restriction or the output power is the primary objective. For high input powers, and therefore high distortion values, the K value should be increased in order to give priority to the distortion specification. Nevertheless, if the distortion signal is far away from the standard specification, the K value should be decreased in order to give priority to the output power.

The MO method starts once the MC has been finished. After minimizing the error signal and the distortion signal has been minimized, the four adjustment parameters (Va_2, Vf_1, Va_2, Vf_2) try to maximize the objective function (F) with consecutive individual variations.

Unfortunately, the objective function (F) does not converge to a unique maximum value, so the heuristic methods used in the MC case should be used carefully.

This Feedforward amplifier is adapted to guarantee any prefixed distortion level specification and simultaneously to maximize the output power level, or the efficiency rate, of the transmission system.

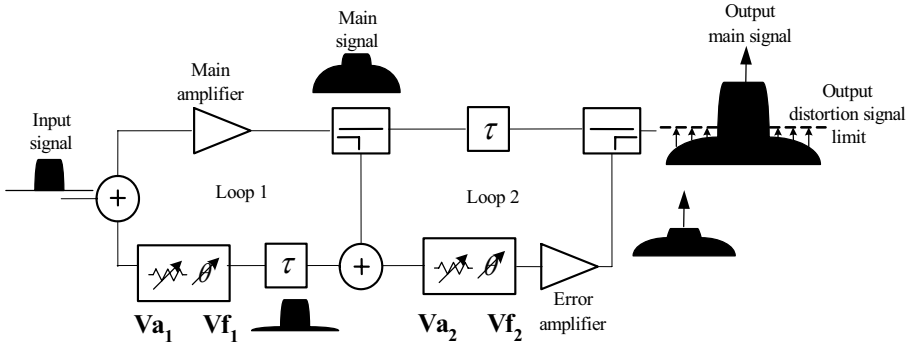


Figure 7-5. Maximum Output control method. The output power is maximized while the distortion level is limited

This control method lies in the precise measurement of the distortion signal level that emerges from the linearization circuit. The following section reviews the most frequently used distortion detection and measurement techniques, and an alternative one will be proposed adapted to the MO adjusting method.

2. DISTORTION MONITORING ARCHITECTURES

Any linearization technique must have the capacity to monitor the emerging signals from the linearization system in order to evaluate how far the system is from the required situation.

In the last 20 years, many patents have been released related to those distortion monitoring techniques, in fact representing the best information source.

In this section the operation basics of three of the most representative distortion monitoring architectures are described. Besides, some of the most recent patents are referenced with many of their significant proposals.

Those control methods, widely described in the literature [68], can be divided into three groups: signal correlation, pilot signal detection and power minimization.

2.1 Signal correlation

This method checks the system status with the analog or digital correlations between the emergent signals from the cancellation loops.

The most important difficulty lies in the masquerading effect between signals with high amplitude differences, for example, the distortion component of the output signal and the error signal.

Ideally, a null correlation would suppose a total distortion cancellation. Nevertheless, the main component of the output signal is outstandingly higher than the distortion component, disguising it completely with respect to the error signal. The *Figure 7-6* tries to visualize the masquerading effect between those signals.

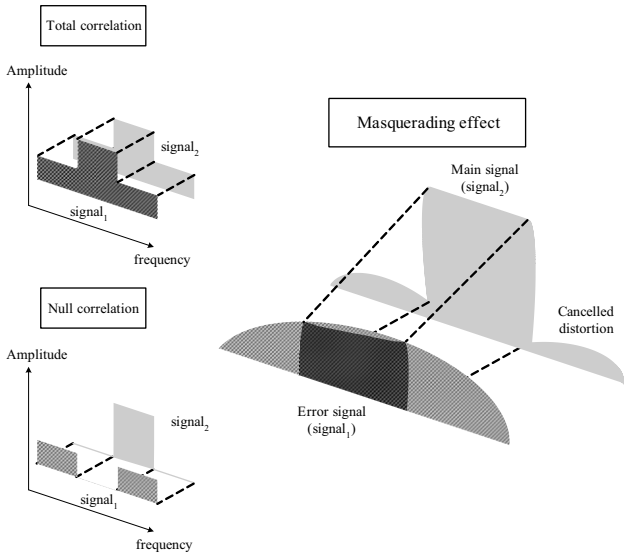


Figure 7-6. Correlation examples and the masquerading effect between the error and the output distortion signal

Any of the patents published since 1995 could be used as an excellent documentation source [84-88].

Some authors, like Bauman and Kenington, propose the use of secondary error loops in order to avoid the masquerading effect, subtracting the error signal from the output signal, forming a new error loop, and correlating the resulting signal with the error signal obtained from the first loop.

Furthermore, alternative techniques have been presented with delay adjustment components, like the delay gain and phase adjuster (DGPA), formed with two complex gain adjusters (CGA), a delay line and two power splitters. If no delay control is necessary each DGPA is reduced to a CGA.

The architecture proposed by Cavers in 1996 [84] is shown in *Figure 7-7*. For simplicity purposes, it is assumed that the delay variation introduced by the amplifiers is insignificant.

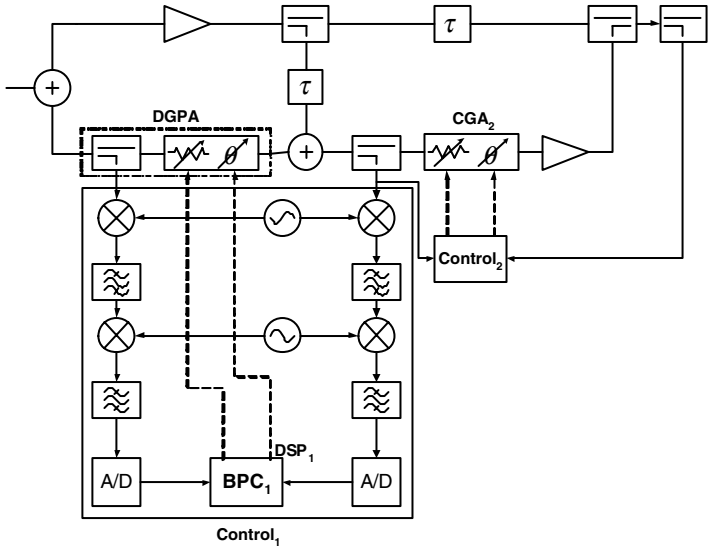


Figure 7-7. Feedforward amplifier with two signal correlators, one for each cancellation loop

Basically, the adjustment parameters are fixed by minimizing or maximizing the digital correlation between the input, error and output signals with two band-pass correlators (BPC). This operation is performed with two frequency translations, a band-pass filtering and an analog to digital conversion.

Many enhancements have been added to this scheme, such as alternative DGPA configurations with CGA and additional delay lines that increase the adaptability to the frequency nonlinearities.

2.2 Pilot signal detection

The pilot signal detection methods are based on introducing, previously to the main amplifier, a pure tone signal with the same frequency as the third order intermodulation products. The control system tries to cancel this pilot signal, so simultaneously the intermodulation products are also cancelled.

One of the architectures most frequently used is based on a narrow band receiver which is tuned to the pilot signal frequency. The main disadvantage of this technique is the impossibility to eliminate completely this pilot signal from the output signal.

Alternative solutions have been developed based on widespread modulated pilot signals, providing with more control capacity. Myer referred in 1986 [89] and 1989 [90] to this basic control method. In 1992 Shoichi [91] developed several architectures, clearly explaining the migration process of the single pilot signals to the modulated and widespread ones.

The architecture shown in *Figure 7-8* makes possible the control of both cancellation loops and the reduction of the residual pilot signal in an acceptable manner.

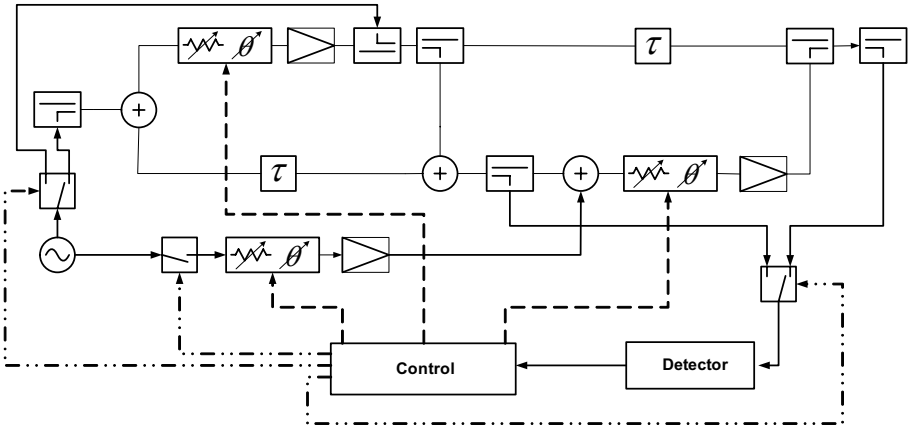


Figure 7-8. Pilot signal control architecture

The pilot signal control methods have several disadvantages in CDMA or WCDMA systems [92]. The pilot signal is masqueraded with the modulated signal unless it is widespread with the same spread code.

In the architecture proposed by Kenington [88] the pilot signal was generated from the input signal with nonlinear elements, and some complicated hardware, like image rejection mixers combined with DSP.

2.3 Power minimization

As its name suggests, this technique is based on directly minimizing the distortion components of the output signal. The basic power minimization technique uses two detection points: the main signal component of the error signal and the distortion component of the output signal. Remember that both signals must be minimized adjusting the respective adjustment parameters, as in *Figure 7-9*.

The first relevant patents based on this detection method came out at the end of the 80's and in the beginning of 90's. Depending on the application, the converging time may be too slow, the noise sensitivity in the error signal detection may be excessive and the masquerading in the output signal distortion detection may limit the profits of the Feedforward linearization technique.

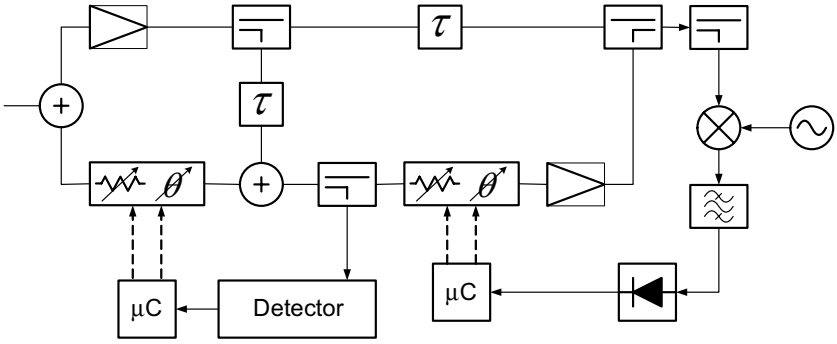


Figure 7-9. Power minimization basic architecture

The masquerading effect is faced up in [90] and [93] with a RF receiver tuned to the frequency of the distortion components. In [94], however, a second error loop is added for the input and output signal subtraction.

Some of the enhancements applied to the power minimization architectures are exhaustively studied in [95], based on the scheme of *Figure 7-10*.

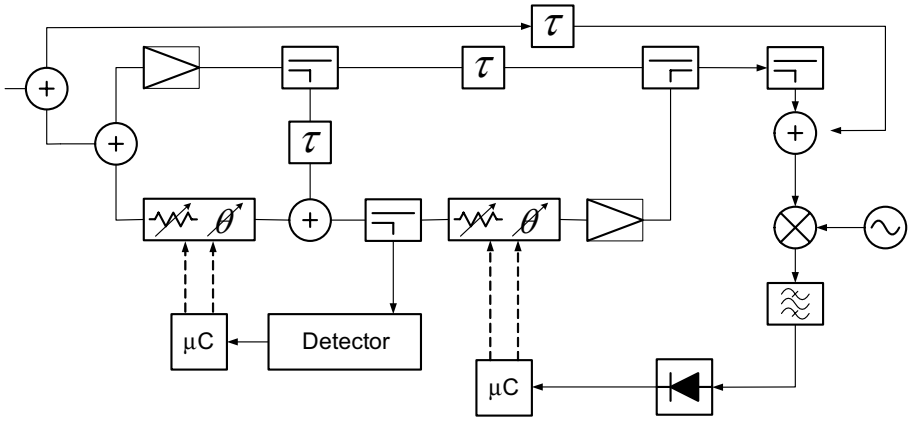


Figure 7-10. Feedforward control with output main signal cancellation

The *Figure 7-11* shows the most complete solution with a single mixer for all the output detection circuit.

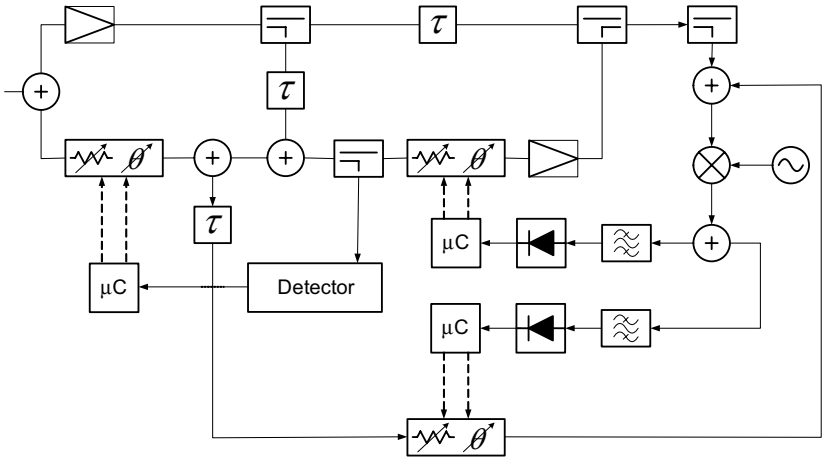


Figure 7-11. Enhanced power minimization architecture

Simultaneously, alternative designs (*Figure 7-12*) were proposed with all the monitoring done at low frequency.

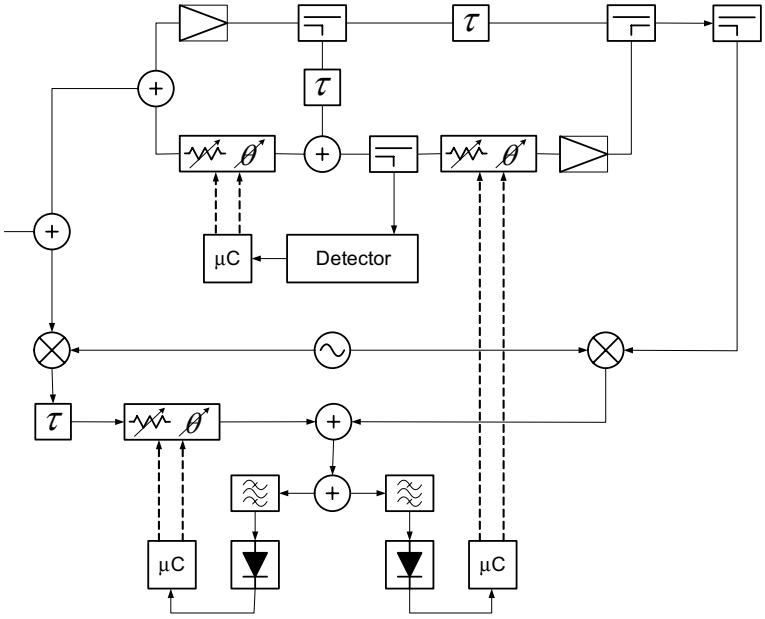


Figure 7-12. Power minimization at low frequency

Other patents [96] directly use an input detector and two tuned receivers, one for each loop. The control system executes the control algorithm from the initial adjustment parameters chosen according to the input power level.

All those detection architectures are basically designed to the reduction of the distortion components from the output signal. However, they are not focused on the measurement of its real value, so they can not be used for the deliberated limitation of the spurious emissions of a standard specification.

Among all the detection architectures, the power minimization is the only one capable of measuring signal levels as the MO method requires.

3. AN OUTPUT SIGNAL MONITORING ARCHITECTURE

The maximum output method is based on the capacity to limit and to control the spurious emissions introduced in the adjacent channels. The power minimization technique is the only one that can perform a precise enough measurement of both the distortion and the main signals.

In many wideband communication systems, the specification of the spurious emissions in the adjacent channels is given as the adjacent channel leakage ratio (ACLR). The handicap lies in the high power difference and frequency closeness of both output signal components: the main and the distortion signals. In addition, the measurement must be done with the precision and reliability demanded by the specification of each standard.

In this section some design rules are offered for the implementation of an output signal monitoring architecture for adaptive Feedforward amplifiers [97].

3.1 Switched RF receiver

The solution proposed resolves this limitation with a switched RF receiver (*Figure 7-13*).

The RF receiver only monitors the output signal as the error signal adjustment is only used as a start point.

A unique receiver is used in order to minimize the error uncertainty. The power difference of the main and distortion components can diverge between 40 to 60dB, so their amplitudes must be equalized in order to maximize the dynamic range of the power detector.

A directional coupler is used for the output signal sampling in order to minimize the main signal path loss.

The first attenuator fixes the displayed average noise level (DANL), which should correspond with the minimum distortion signal power level.

The RF signal is then down converted to an intermediate frequency (IF). The down conversion also selects which component is settled at the IF frequency, either the main or the distortion signal.

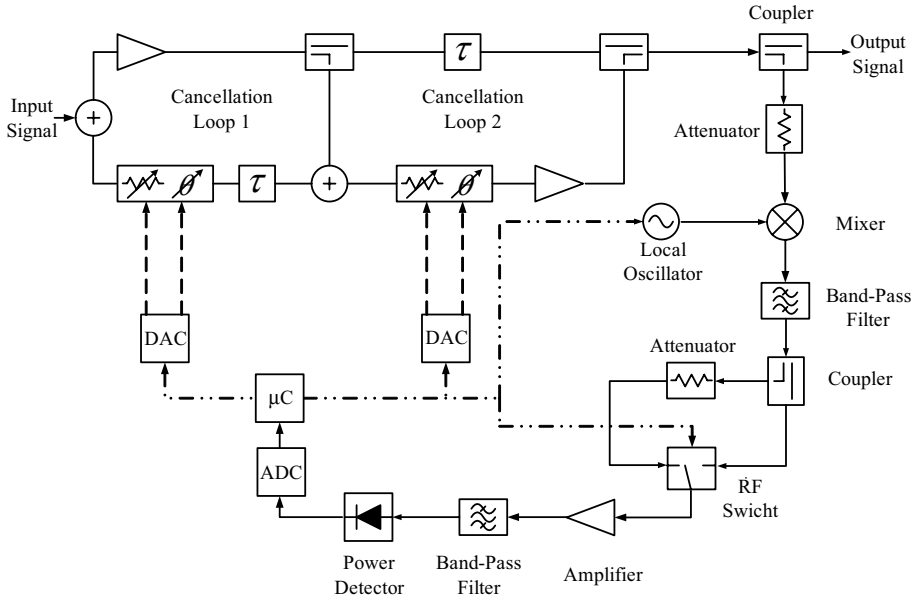


Figure 7-13. Output signal monitoring architecture with a switched RF receiver

There are two filtering steps. In first instance, a band pass filter rejects the most part of the unwanted signal. The IF signals may be either the main or the distortion component. In any case, any signal outside the IF filter will be considered as interference.

A switching circuit is in charge of interposing, or not, a second attenuator depending on which of both components is measured, the main or the distortion signal respectively, equalizing both amplitudes. The RF switch selects the direct or coupled output of a second directional coupler. The coupled output is then attenuated with the second attenuator.

The total attenuation, which results from the combination of the coupling factor and the second attenuation, must coincide with the amplitude difference between the main and the distortion signals. This point forward the IF signal will be similar for the rest of the receiver components, regardless of which signal component is been measured.

In subsequent steps, the IF signal is amplified to compensate all the transmission losses of the predecessor and subsequent components, and it is again filtered completing the rejection of all the unwanted signals.

The last stage is completed with a power detector that converts the IF signal into an analog voltage that is digitally converted (ADC). These digital signals are the inputs of the microcontroller (µC) that executes the MO algorithm.

Those steps are now discussed in more detail.

3.2 Frequency down-conversion

The frequency down conversion, together with the switching process, is one of the two most critical steps. *Figure 7-14* shows the down conversion process. The local oscillator (LO), controlled by the μC, decides whether the main or the distortion component signal is down converted to the IF frequency.

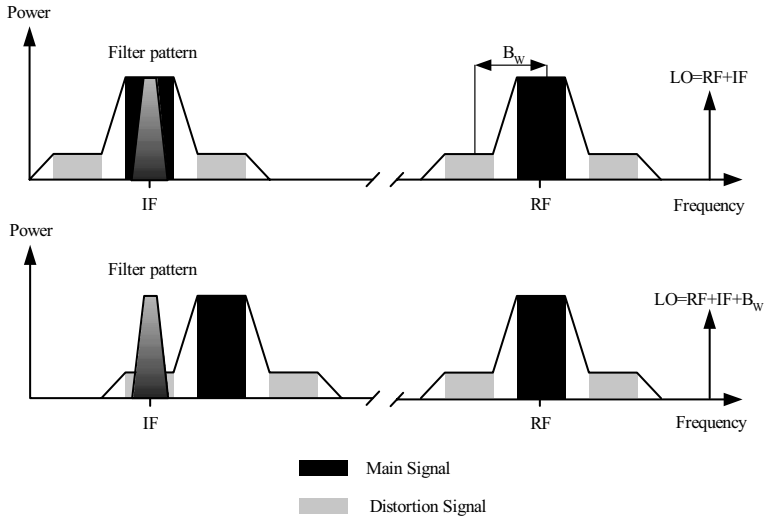


Figure 7-14. IF translation

Two aspects must be taken into account in the design of the frequency down conversion. On the one hand, the mixer distortion might be critic for the accuracy of the measurement system. It is recommended that the input third order intercept point (IIP_3) should be 20dB higher than the mixer input nominal power.

$$IIP_3[\text{dBm}] \approx P_{in}[\text{dBm}] + 20\text{dB} \tag{7.1}$$

On the other hand, the phase noise of the local oscillator (LO) can cause a reciprocal mixing. This effect is important in the measurement of the distortion component, which in the example of *Figure 7-15* corresponds with the "RF signal". The interference signal corresponds with the main component. If the LO components due to the phase noise mix with the interference signal, it will be also down converted to the IF band.

The power of the IF interference signal relative to the interfered signal is given by (7.2).

$$P_{rm} = L_{\phi}(f_m) \cdot B_w \tag{7.2}$$

Where P_{rm} is the interference signal power relative to the interfered signal in the IF (W), $L_{\Phi}(f_m)$ is the LO phase noise at f_m (W/Hz) and B_w is the IF filter bandwidth (Hz).

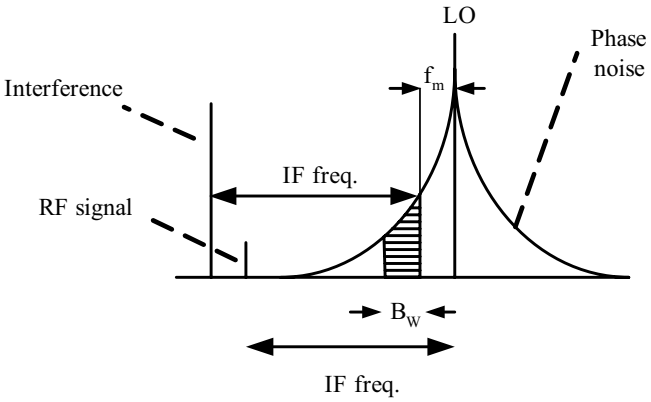


Figure 7-15. Reciprocal mixing effect

If the main signal component, as interference signal, and the distortion component, as RF signal, are equal in magnitude before the frequency down conversion, their difference (watts) in the IF will be P_{rm}.

The real difference, however, before the down conversion is precisely the ACLR level (Δ_{ACLR}) so, the difference at the IF in dB (Δ_{IF}) will be defined by (7.3).

$$\Delta_{IF}[dB] = 10\log_{10}(P_{rm})[dB_W] + \Delta_{ACLR}[dB] \tag{7.3}$$

Figure 7-16 shows a graphic example of the reciprocal mixing effect with a wideband signal.

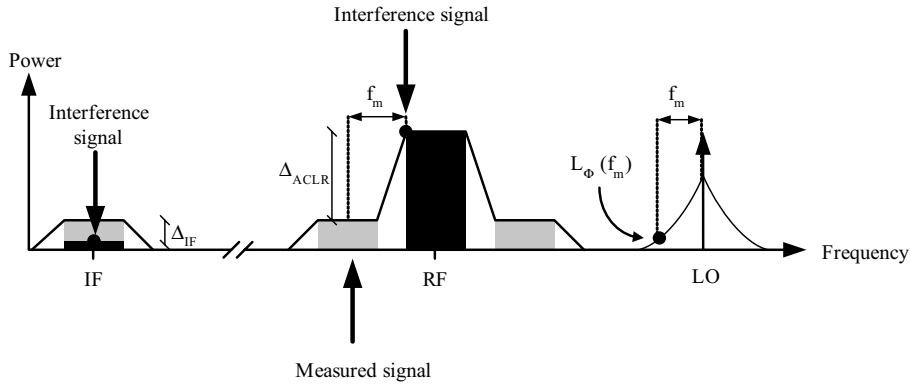


Figure 7-16. Reciprocal mixing in frequency down-conversion

The maximum Δ_{IF} value depends on the error uncertainty specification (ε). Remember that the error due to the addition of two coherent sinusoidal tones is bounded by the values shown in equation (7.4).

$$\varepsilon = 20 \cdot \log\left(1 \pm 10^{\frac{\Delta_{IF}[dB]}{20}}\right)[dB] \tag{7.4}$$

+ If both signals add in phase
- If both signals add in out of phase (180º)

Combining the equations (7.2), (7.3) and (7.4), the LO phase noise specification could be expressed in terms of the IF filter bandwidth (B_w), the actual ACLR level (Δ_{ACLR}), the frequency offset used in the ACLR measurement (f_m) and the error uncertainty (ε).

$$\begin{aligned} L_\Phi(100kHz)[dBc/Hz] &= -10\log_{10}(B_w[Hz]) + 20\log_{10}\left(\frac{100[kHz]}{f_m[kHz]}\right) \\ &- \Delta_{ACLR}[dB] - 20\log\left(10^{\frac{\varepsilon[dB]}{20}} - 1\right)[dB] \end{aligned} \tag{7.5}$$

3.3 Amplitude equalization

Despite the ACLR specification, between the main and the distortion components, varies with each communication system, it is about several orders of magnitude. The dynamic range of the power detector must be optimized in order provide the maximum resolution, so this amplitude difference must be equalized (*Figure 7-17*).

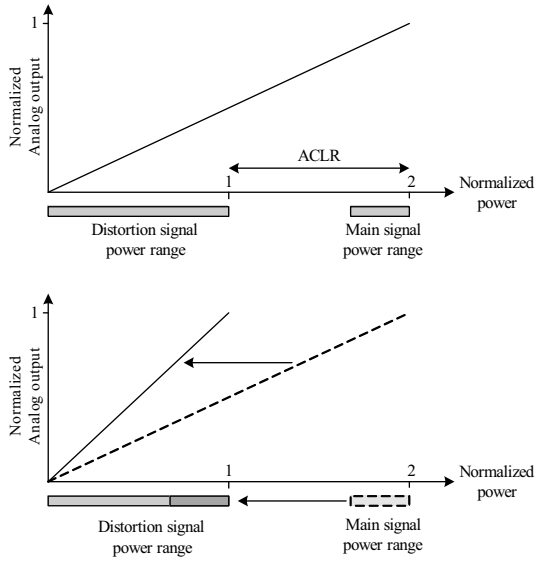


Figure 7-17. Amplitude equalization

With the amplitude equalization the dynamic range is adjusted and the resolution increased. This equalization is done with the architecture proposed in *Figure 7-18*.

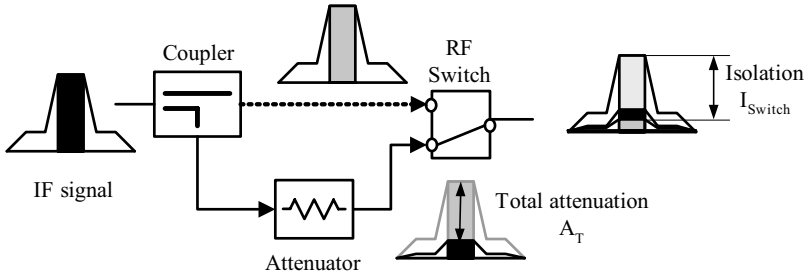


Figure 7-18. Switching mechanism in the main signal measurement

A RF switch interposes an attenuator if the main signal component is being measured, and the total attenuation (A_T) should correspond with the actual ACLR specification.

On the other hand, *Figure 7-19* shows the switching mechanism for the measurement of the distortion component.

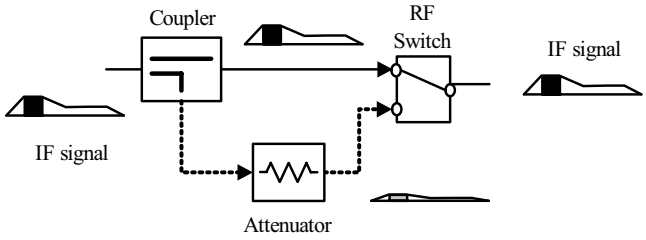


Figure 7-19. Switching mechanism in distortion signal measurement

Notice how after the equalization the IF signal keeps similar order whether the main or the distortion component is measured.

The two design parameters are the total attenuation (A_T) and the minimum switching isolation (I_{Switch}). The former depends on the actual ACLR level and the latter limits the error uncertainty (ε) due to the absence of a total isolation (*Figure 7-18*).

The difference between the interfered and interference signal can be put in terms of such error uncertainty of a coherent sum.

Thus, the minimum isolation of the RF switch can be obtained with the equation (7.6) according to this error uncertainty (ε) and the total attenuation factor (A_T).

$$I_{switch}[\mathrm{dB}] \approx A_T[\mathrm{dB}] + 20\log\left(10^{\frac{\varepsilon[\mathrm{dB}]}{20}} - 1\right)[\mathrm{dB}] \tag{7.6}$$

4. THE ADAPTIVE FEEDFORWARD AMPLIFIER

The Maximum Output (MO) control method, designed in the preceding section, guarantees the linearity specification of any communication standard and maximizes the output power level and hence the efficiency rate.

This method, however, requires a specific monitoring system, based on the power minimization technique, capable of measuring the emerging output signal from the linearization system.

In this section the MO and MC methods are applied to the Feedforward prototype implemented in Chapter 6 with the monitoring system designed in this chapter (*Figure 7-13*) for the linearity specification fixed by the UMTS standard (ACLR ≥ 45dB) [61].

Figure 7-19 shows the ACLR level of the Feedforward amplifier for both control methods (MC and MO) vs. the input signal power.

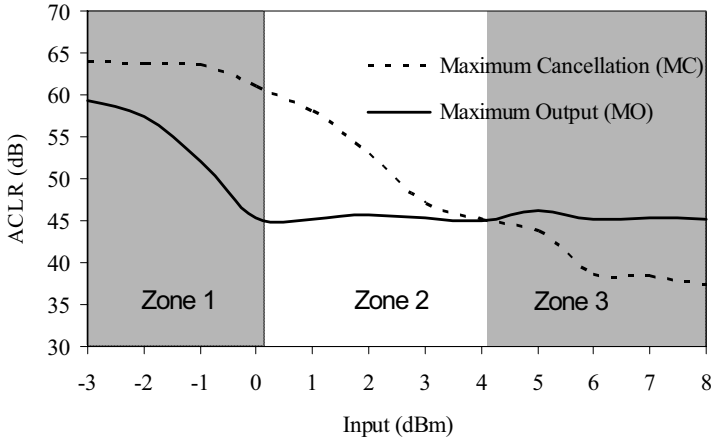


Figure 7-20. Output ACLR (dB) vs. Input signal level

Three different zones are distinguished.

The first one covers the small signals with high ACLR levels. The MO method tries to *unlinearize* the Feedforward amplifier, increasing the distortion until the ACLR level fixed by the standard. On the other hand, the MC method keeps on maximizing the ACLR level as much as possible.

The second zone works with medium power signals and moderate distortion levels. In this case the MO method keeps on making the ACLR level worse deliberately but guaranteeing the specified value (45dB). The MC method however, achieves the highest ACLR level for each input signal.

Finally, in the third zone, the power amplifier begins to operate in the saturation region and distortion grows considerably. The MO method keeps the ACLR level under the specified limit but the MC method does not.

Figure 7-20 shows the output power level for both control methods vs. the input signal.

In this case again, three different zones are distinguished coherently with the ones detected in the ACLR analysis.

With small signals the MO method takes advantages of the ACLR worsening in order to increase the output power.

As the input signal increase the MO method begins to give priority to the ACLR specification fulfillment while the output main signal power level remains practically constant.

The MO method offers two advantages with respect to the traditional MC method. The former is the improvement of the power amplifier efficiency in small signal conditions. The second one is the capacity to limit and to control dynamically the distortion specification fixed by any communication standard.

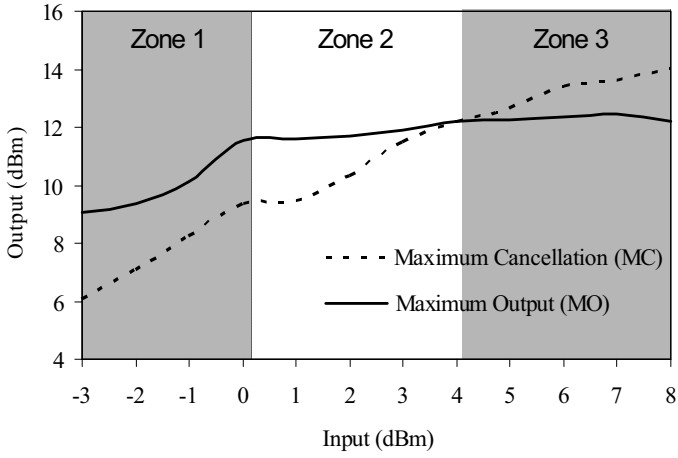


Figure 7-21. Output signal level vs. Input signal level

Now, both control methods are compared with the none-linearized amplifier. As it has been done before, the enhancement is given in terms of the ACLR and the output power level. *Figure 7-22* shows the ACLR level improvement shadowed in a different way for each method.

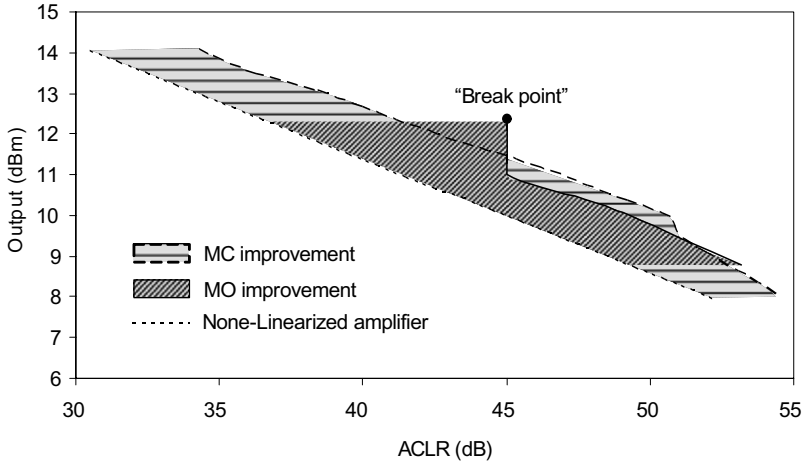


Figure 7-22. ACLR improvement according to the none-linearized amplifier

The ACLR improvement is practically similar for any output power level with the MC method.

Nevertheless, the MO method truncates this tendency once the distortion specification value is reached. From this point forward, the ACLR

improvement increases until the linearization system can not keep on linearizing any more. It has been reached the natural "break point".

Something similar occurs with the output power, as shown in *Figure 7-23*, vertically and differently colored for each method.

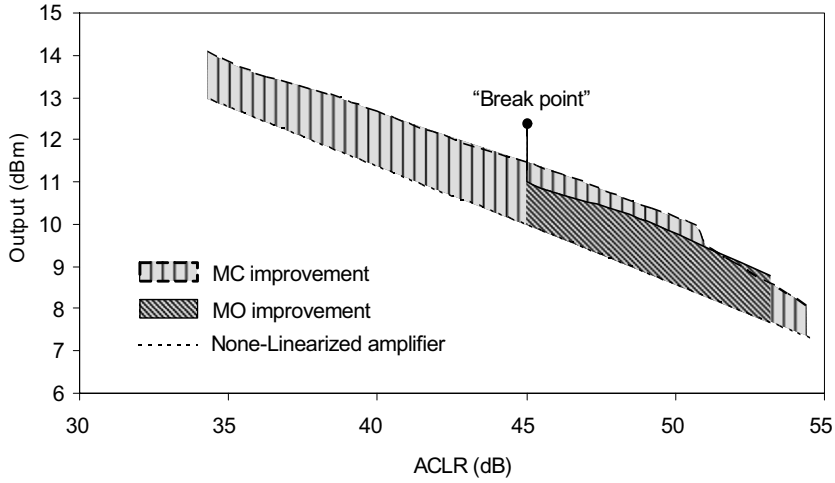


Figure 7-23. Output power improvement according to the none-linearized amplifier

The MO method improves the output power, according to the MC method, for the ACLR levels fixed by the present telecommunication standard.

5. CONCLUSIONS

The efficiency of digital communications systems is mainly affected by the power amplifier own efficiency. As it has been seen throughout this work, the digital modulations, on the one side, and the multiple access techniques on the other, force power amplifiers to adopt back-off strategies in order to fulfill the distortion specifications. Automatically, efficiency is drastically reduced, impacting on an increase in price.

The linearization techniques are used in order to minimize the effects derived from the new wideband communication systems related to the linearity or the efficiency rates.

The potential of Feedforward linearization technique is based on the distortion measurement effectiveness. The spurious emissions are usually fixed by the respective communication standards, so it is not necessary a total suppression of the distortion introduced by the power amplifier.

Throughout these lines, a Feedforward amplifier has been designed with an adaptive control system. The designed Maximum Output (MO) adjusting method has been able to achieve a trade-off between the linearity and the efficiency parameters, based on the measurement effectiveness of a RF receiver that monitors the emerging output signal from the linearization system.

In general, this adaptive Feedforward amplifier responds to the telephonic traffic and the environmental variations guaranteeing the distortion requirements fixed by each communication standard, at the same time that maximizes the transmission system efficiency.

References

[1] A. Muñoz, "Sistemas de Transmision," TECNUN (University of Navarra), 2006, pp. 330.

[2] "Directive 2004/108/EC of the European Parliament and of the Council of 15 December 2004 on the approximation of the laws of the Member States relating to electromagnetic compatibility and repealing Directive 89/336/EEC," vol. L 390: Official Journal of the European Union, 2004, pp. 24-37.

[3] "Testing and Troubleshooting Digital RF Communications Transmitter Designs," Agilent Technologies, 2002, pp. 64.

[4] "Digital Modulations in Communications Systems - An Introduction," Agilent Technologies, 2001, pp. 46.

[5] H. Pretl, L. Maurer, W. Schelmbauer, R. Weigel, B. Adler, and J. Fenk, "Linearity Considerations of W-CDMA Front-Ends for UMTS," *IEEE MTT-S Microwave Symposium*, pp. 433-436, 2000.

[6] V. Petrovic and C. N. Smith, "Reduction of intermodulation distortion by means of modulation feedback," presented at IEE conference on Radio Spectrum Conversation Techniques, 1983.

[7] M. Johansson and L. Sundstrom, "Linearization of RF multicarrier amplifiers using Cartesian feedback," *Electronics Letters*, vol. 30, pp. 1110-1112, 1994.

[8] M. Boloorian and J. P. McGeehan, "Twin-loop cartesian transmitter (TLCT)," *Electronics Letters*, vol. 32, pp. 971-972, 1996.

[9] M. A. Briffa and M. Faulkner, "Stability analysis of cartesian feedback linearization for amplifiers with weak nonlinearities," *IEE Proceedings on Communications*, vol. 4, pp. 212-218, 1996.

[10] S. I. Mann, M. A. Beach, and K. A. Morris, "Digital baseband Cartesian loop transmitter," *Electronics Letters*, vol. 37, pp. 1360-1361, 2001.

[11] J. Yi, Y. Yang, M. Park, and W. Kang, "Analog predistortion linearizer for high power RF amplifier," presented at IEEE MTT-S International Microwave Symposium Digest, 2000.

[12] C. Potter, "A 3G Base-Station PA Design Using Load Pull and Adaptive Digital Predistortion," *Wireless Design Conference*, pp. 141-144, 2002.

[13] S. C. Cripps, *RF power amplifiers for wireless communications*: Artech House Publishers, 1999.

[14] C. Haskins, T. Winslow, and S. Raman, "FET diode linearizer optimization for amplifier predistortion in digital radios," *IEEE Microwave and Guided Wave Letters*, vol. 10, pp. 21-23, 2000.

[15] K. A. Morris and J. P. McGeehan, "Gain and phase matching requirements of cubic predistortion systems," *Electronics Letters*, vol. 36, pp. 1822-1824, 2000.

[16] S. C. Cripps, "RF power amplifier linearization techniques for multichannel communications systems." Dublin: CEI-Europe, 2000.

[17] M. Ghaderi, S. Kumar, and D. E. Dodds, "Adaptive predistortion linearizer using polynomial functions," presented at IEE Proceedings on Communications, 1994.

[18] S. Kusunoki, K. Yamamoto, T. Hatsugai, H. Nagaoka, K. Tagami, N. Tominaga, K. Osawa, K. Tanabe, S. Sakurai, and T. Lida, "Power amplifier module with digital adaptive predistortion for cellular phones," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, pp. 2979-2986, 2002.

[19] C. G. Rey, "Predistorter linearized CDMA power amplifiers," *Microwaves & RF*, pp. 114-123, 1998.

[20] K. J. Muhonen, M. Kavehrad, and R. Krishnamoorthy, "Adaptive baseband predistortion techniques for amplifier linearization," presented at Thirty-Third Asilomar Conference on Signals, Systems and Computers, 1990.

[21] J. K. Cavers, "Amplifier Linearization Using A Digital Predistorter With Fast Adaptation And Low Memory Requirements," *IEEE Transactions on Vehicular Technology*, vol. 39, pp. 374-382, 1990.

[22] S. P. Stapleton, "Amplifier Linearization Using Adaptive Digital Predistortion," *Applied Microwave & Wireless*, vol. 13, pp. 72-77, 2001.

[23] J. Kim and K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory," *Electronics Letters*, vol. 37, pp. 1417-1418, 2001.

[24] J. K. Cavers, "Optimum indexing in predistorting amplifier linearizers," presented at IEEE Vehicular Technology Conference, 1997.

[25] J. J. Cavers, "Optimum table spacing in predistorting amplifier linearizers," *IEEE Transactions on Vehicular Technology*, vol. 48, pp. 1699-1705, 1999.

[26] L. Sundström, "Digital RF Power Amplifier Linearisers," in *Department of Applied Electronics*. Lund: Lund University, 1995, pp. 64.

[27] F. Zavosh, M. Thomas, C. Thron, T. Hall, D. Artusi, D. Anderson, D. Ngo, and D. Runton, "Digital predistortion techniques for RF power amplifiers with CDMA applications," *Microwave Journal*, pp. 8, 1999.

[28] H. S. Black, "Translating System," U. S. P. Office, Ed., 1928, pp. 6.

[29] H. Seidel, "A Feedforward Experiment. Applied to an L-4 Carrier System Amplifier," *IEEE Transactions on Communications Technology*, vol. 19, pp. 320-325, 1971.

[30] T. Bennett and R. F. Clements, "Feedforward - An alternative approach to amplifier linearization," *RF, Radio and Electronic Engineer*, vol. 44, pp. 257-262, 1974.

[31] N. Pothecary, *Feedforward Linear Power Amplifiers*. Norwood: Artech House Inc., 1999.

[32] J. Legarda, J. Presa, E. Hernandez, H. Solar, J. Mendizabal, and J. A. Peñaranda, "An adaptive Feedforward amplifier under "Maximum Output" control method for UMTS downlink transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, pp. 2481-2486, 2005.

[33] F. H. Raab, P. Asbeck, S. Cripps, P. B. Kenington, Z. B. Popovic, N. Pothecary, J. F. Sevic, and N. O. Sokal, "Power amplifiers and transmitters for RF and microwave," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, pp. 814-826, 2002.

[34] L. R. Kahn, "Single sideband transmission by envelope elimination and restoration," *Proceedings of the IRE*, vol. 40, pp. 803-806, 1952.

[35] F. H. Raab, B. E. Sigmond, R. G. Myers, and R. M. Jackson, "L-Band transmitter using Kahn EER technique," *IEEE Transactions on Microwave Theory and Techniques*, vol. 46, pp. 2220-2225, 1998.

[36] D. Rudolph, "Khan EER technique with single carrier digital modulations," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, pp. 548-552, 2003.

[37] M. Suzuki, T. Yamawaki, T. Tanoue, Y. Ookuma, R. Fujiwama, and S. Tanaka, "Proposal of transmitter architecture for mobile terminals employing EER power amplifier," presented at IEEE Vehicular Technology Conference.

[38] G. Baudoin, C. Borland, M. Villegas, and A. Diet, "Influence of time and processing mismatches between phase and envelope signals in linearization systems using envelope elimination and restoration," presented at IEEE MTT International Microwave Symposium Digest, 2003.

[39] M. J. Koch and R. E. Fisher, "A high efficiency linear power amplifier for digital cellular telephony," presented at IEEE Conference on Vehicular Technology, 1989.

[40] F. M. Ghannouchi, J.-S. Cardinal, and R. Hajji, "Adaptive linearizer using feedback and dynamic biasing techniques for SSPAs," presented at IEEE MTT-S International Microwave Symposium, 1995.

[41] G. Hanington, P.-F. Chen, P. M. Asbeck, and L. E. Larson, "High-efficiency power amplifier using dynamic power-supply voltage for CDMA applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, pp. 1471-1476, 1999.

[42] Y. S. Noh and C. S. Park, "Linearised InGaP/GaAs HBT MMIC power amplifier with active bias circuit for W-CDMA application," *Electronics Letters*, vol. 37, pp. 1523-1524, 2001.

[43] J. H. Kim, J. H. Kim, Y. S. Noh, and C. S. Park, "MMIC power amplifier adaptively linearized with RF coupled active bias circuit for WCDMA mobile terminals applications," presented at IEEE MTTT-S International Microwave Symposium Digest, 2003.

[44] H. Chireix, "High power outphasing modulation," *Proceedings of the IRE*, vol. 23, pp. 1370-1392, 1935.

[45] D. C. Cox, "Linear amplification with nonlinear components," *IEEE Transactions on Communications*, vol. 22, pp. 1942-1945, 1974.

[46] D. C. Cox and R. P. Leck, "Component signal separation and recombination for linear amplification with nonlinear components," *IEEE Transactions on Communications*, vol. 23, pp. 1281-1287, 1975.

[47] S. A. Hetzel, A. Bateman, and J. P. MacGeehan, "A LINC transmitter," presented at IEEE Vehicular Technology Conference, 1991.

[48] F. J. Casadevall and A. Valdovinos, "Performance analysis of QAM modulations applied to the LINC transmitter," *IEEE Transactions on Vehicular Technology*, vol. 42, pp. 399-406, 1993.

[49] X. Zhang, L. E. Larson, and P. M. Asbeck, "Calibration scheme for LINC transmitter," *Electronics Letters*, vol. 37, pp. 317-318, 2001.

[50] W. H. Doherty, "A new High-Efficiency power amplifier for modulated waves," in *Annual Convention of the Institute of Radio Engineers*. Cleveland, Ohio: Bell Telephone Systems, 1936.

[51] A. Bateman, "The combined analogue locked loop universal modulator (CALLUM)," presented at IEEE Vehicular Technology Conference, 1992.

[52] K. Y. Chan, A. Bateman, and M. Li, "Analysis and realization of the LINC transmitter using the combined analogue locked loop universal modulator (CALLUM)," presented at IEEE Vehicular Technology Conference, 1994.

[53] K. Y. Chan and A. Bateman, "Linear modulators based on RF synthesis: realization and analysis," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 42, pp. 321-333, 1995.

[54] D. J. Jennings and J. P. McGeehan, "Hardware implementation of optimal CALLUM transmitter," *Electronics Letters*, vol. 34, pp. 1816-1817, 1998.

[55] D. J. Jennings, A. Bateman, and J. P. McGeehan, "Adjacent channel power and error-vector magnitude performance of reduced complesity CALLUM systems," *IEE Proceedings on Communications*, vol. 146, pp. 297-302, 1999.

[56] K. J. Parsons, P. B. Kenington, and J. P. McGeehan, "High efficiency power amplifier design for mobile satellite earth stations," *IEE Colloquium (Digest)*, pp. 3/1-3/6, 1995.

[57] K. Muhonen and M. Kavehrad, "Amplifier linearization for the local multipoint distribution system application," presented at IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 1998.

[58] J. Dixon, "A solid-state amplifier with Feedforward correction for linear single-sideband applications," presented at IEEE International Conference on Communications, 1986.

[59] V. Steel, D. Scott, and S. Ludvik, "A 6-18GHz high dynamic range MMIC amplifier usign a feedforward technique," presented at IEEE MTT-S Digest, 1990.

[60] W. T. Thornton and L. E. Larson, "An Improved 5.7GHz ISM Band Feedforward Amplifier Utilizing Vector Modulators for Phase and Attenuation Control," *Microwave Journal*, pp. 96-106, 1999.

[61] r. G. P. Project, "Base Station (BS) conformance testing (FDD)," vol. 3GPP TS 25.141: Technical Specification Group Radio Access Network, 2006, pp. 32-40.

[62] A. A. M. Saleh, "Frequency-Independent and frequecny-dependent nonlinear models of TWT amplifiers," *IEEE Transactions on Communications*, vol. 29, pp. 1715-1720, 1981.

[63] R. Blum and M. Jeruchim, "Modelling nonlinear amplifiers for communication simulation," presented at IEEE Intrenational Conference on Communications, 1989.

[64] S. Narayanan, "Transistor distortion analysis using Volterra series representation," *Bell System Technical Journal*, vol. 46, pp. 991-1024, 1967.

[65] C. Law and C. Aitchison, "Prediction of wideband power performance of representation," *IEEE Transactions on Microwave Theory and Techniques*, vol. 34, pp. 1308-1317, 1986.

[66] M. Steer, P. Khan, and R. Tucker, "Relationship between Volterra series and generalized power series," *Proceedings of the IEEE*, vol. 71, pp. 1453-1454, 1983.

[67] G. Rhyne, M. Steer, and B. Bates, "Frequency-domain nonlinear circuit analysis using generalized power series," *IEEE Transactions on Microwave Theory and Techniques*, vol. 36, pp. 379-387, 1988.

[68] P. B. Kenington, *High-Linearity RF Amplifier Design*. Norwood: Artech House Inc., 2000.

[69] R. J. Wilkinson and P. B. Kenington, "Specifications of error amplifiers for use in feedforward transmitters," *IEEE Proceedings on Circuits, Devices and Systems*, vol. 139, pp. 477-480, 1992.

[70] P. B. Kenington, M. A. Beach, A. Bateman, and J. P. McGeehan, "Apparatus and method for reducing distortion in amplification," N. R. DEV, Ed., 1992.

[71] K. J. Parsons and P. B. Kenington, "Effect of delay mismatch on a feedforward amplifier," *IEE Proceedings: Circuits, Devices and Systems*, vol. 141, pp. 140-144, 1994.

[72] C. L. Larose and F. M. Ghannouchi, "Optimization of feedforward amplifier power efficiency on the basis of drive statistics," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, pp. 41-54, 2003.

[73] A. Technologies, "Optimizing Dynamic Range for Distortion Measurements," 2002, pp. 40.

[74] A. H. Coskun and S. Demir, "A mathematical characterization and analysis of a feedforward circuit for CDMA applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, pp. 767-777, 2003.

[75] J. Legarda, J. Presa, H. Solar, J. Melendez, A. Muñoz, and A. G. Alonso, "An adaptive feedforward power amplifier for UMTS transmitters," in *IEEE International Symposium in Personal, Indoor and Mobile Radio Communications*. Barcelona, 2004.

[76] Q. Cheng, C. Yiyuan, and Z. Xiaowei, "1.9GHz adaptive feedforward power amplifier," *Microwave Journal*, vol. 41, 1998.

[77] Y. Shcherbelis, D. S. G. Chambers, and D. C. McLernon, "Microwave feed-forward low noise amplifier design for cellular base station," *Electronics Letters*, vol. 37, pp. 359-361, 2001.

[78] J. W. Huh, I. S. Chang, and C. D. Kim, "Spectrum monitored adaptive feedforward linearization," *Microwave Journal*, vol. 44, pp. 160-166, 2001.

[79] Y.-C. Jeong, "A feedforward power amplifier with loops to reduce RF band noise and intermodulation distortion," *Microwave Journal*, vol. 45, pp. 80-91, 2002.

[80] Y.-C. Jeong, Y.-J. Song, I.-J. Oh, and C.-D. Kim, "A novel adaptive feedforward amplifier using an analog controller," *Microwave Journal*, vol. 46, pp. 76-85, 2003.

[81] S. Kang, U. Park, K. Lee, and S. Hong, "Adaptive feedforward amplifier using digital controller," *IEEE Vehicular Technology Conference*, vol. 57, pp. 2076-2079, 2003.

[82] Y. Y. Woo, Y. Yang, J. Yi, J. Nam, J. Cha, and B. Kim, "An adaptive feedforward amplifier for WCDMA base stations using imperfect signal cancellation," *Microwave Journal*, vol. 46, pp. 22-44, 2003.

[83] S. H. Jacobson and E. Yücesan, "Analyzing the performance of generalized hill climbing algorihms," in *Journal of Heuristics*: Springer Netherlands, 2004, pp. 387-405.

[84] J. K. Cavers, "Adaptive Feedforward linearizer for RF power amplifiers," U. S. Patent, Ed., 1996.

[85] R. M. Bauman, "Adaptive Feedforward sytem," U. S. Patent, Ed., 1983.

[86] T. E. Olver, "Adaptive Feedforward cancellation technique that is effective in reducing amplifier harmonic distortion products as well as intermodulation distortion products," U. S. Patent, Ed., 1985.

[87] R. H. Chapman, "Feedforward distortion cancellation circuit," U. S. Patent, Ed., 1991.

[88] P. B. Kenington, M. A. Beach, A. Bateman, and J. P. McGeehan, "Apparatus and method for reducing distortion in amplification," N. R. DEV, Ed. United States, 1992.

[89] R. E. Myer, "Automatic reduction of intermodulation products in high linear amplifiers." United States, 1986.

[90] R. E. Myer, "Feedforward linear amplifier," A. T. TELEGRAPH, Ed. United States, 1989.

[91] N. Shoichi, N. Toshio, M. Makoto, and M. Kazuaki, "Feed-forward amplifier," N. T. TELEPHONE, Ed. United States, 1992.

[92] N. Shoichi, N. Toshio, and S. Yasunori, "Feedforward amplifier," N. T. TELEPHONE, Ed. Europe, 2000.

[93] W. H. Lieu, "Linear amplifier with automatic adjustment of feedforward loop gain and phase." United States, 1991.

[94] M. G. Obermann and J. F. Long, "Feedforward Distortion Minimization Circuit," M. Inc, Ed., 1991.

[95] H. Kenichi, I. Yukio, N. Junichi, S. Yuji, S. Harayasu, and N. Masatoshi, "Feedforward Amplifier," M. E. CORP, Ed. Europe, 2001.

[96] K. S. Yoo, S. G. Kang, J. I. Choi, and J. S. Chae, "Optimal control method for adaptive feedforward linear amplifier," K. E. TELECOMM, Ed. United States, 2001.

[97] J. Legarda, J. Presa, A. Muñoz, and J. Ibarguren, "Monitoring and measurement system for the nonlinear distortion of linearization devices," S. A. Angel Iglesias, Ed. Spain, 2005 (patent pending).

Index

1dB compression point 44
Adjacent Channel Leakage Ratio
  ACLR 55
amplification modes 58
amplification topology 58
  complementary 58
  symmetric 58
  unilateral 58
amplitude modulation 10
analog signal 6
analog to digital converter 7, 9
attenuation 12
attenuator 158
automatic gain control 79

band pass filter 159
bias point 58, 60, 66
bit rate 17
bits 8
broadband noise 132, 140

cancellation loop 99
CCDF 112
channel coding 14
chip rate 112
choke 64
code division multiple access
  CDMA 19
codification 8
coil 127
collector 61, 63
  collector capacitance 71
complementary configuration 65, 67, 70
Complementary Cumulative Distribution
  Function
  CCDF 52
composite second order beat
  CSO 34
conduction angle 63, 66, 67, 95
continuous wave 60
control method 116
cordless telephone 2
  CT2 20
cross modulation 49
cross modulation coefficient 40
cross-modulation 36
crossover distortion 66
cut-off point 61

DAC 15
digital enhanced cordless telephone
  DECT 20
digital signal 7
Digital Video Broadcasting - Cable
  CVB-C 21
digital video broadcasting-satellite
  DVB-S 19
digitalization 7
directional coupler 100, 158
discrete signal 7
discrete transistors 5
Displayed Average Noise Level 134
distortion 12
distortion cancellation loop 102
dynamic range chart 133

efficiency 14, 21, 23, 59, 60, 62, 64, 65, 67, 68, 70, 71, 72, 77, 89, 92, 95, 110, 115, 116, 124, 126, 147, 151, 164, 165, 167
  average efficiency 62
  collector efficiency 59, 62
  instantaneous efficiency 60, 62, 64
  mean efficiency 60, 67
  power added efficiency 59
  total efficiency 59
electromagnetic signal 9
electromagnetic spectrum 9
equalization 104
equalize 106
error amplifier 109
error loop 101
error signal 101
Error Vector Magnitude
  EVM 53
external noise 136

finite impulse response filter 15
flatness 104
frequency dependence 85
Frequency Division Multiple Access
  FDMA 24
frequency modulation 10
frequency shift keying
  FSK 20

gain 65
global harmonic distortion 32
Global System for Mobile Communications
  GSM 20
group delay 104
group delay margin 128

harmonic component 63
harmonic distortion 28, 41
harmonic distortion coefficient 31
High Frequency (HF) 57

I/Q modulation 17
Inphase 16
interleaving 14
intermediate frequency 158
intermodulation 83
intermodulation product 34
intersymbol interference (ISI) 16
isolation 110

load adjustment 94
load impedance 64
Loop 1 101
Loop 2 102

main amplifier 101
main signal suppression level 126
masquerading effect 153
microcontroller 160
modulation 10
multi-polar alternators 3

negative dynamic resistance 2
noise 12
nonlinear distortion 27

ohmic losses 68
outphasing 65

path loss 111
peak detector 80
peak to average ratio 133
Peak-to-average power 51
phase margin 128
phase modulation 10
phase noise 132
phase shift keying
  PSK 19
phase shifters 107
polar diagram 16
power combiner 100
power detector 160
power efficiency 22
power spectral density 139
power splitter 100
propagation 12
pulse amplitude modulation 11
pulse position modulation 11
pulse width modulation 11
push-pull 65, 67

Quadrature 16
quadrature amplitude modulation
  QAM 21
quadrature phase shift keying
  QPSK 19
quiescent point 58, 61

radio-electric transmitters 1
reference signal 100
root mean square 54
rotary spark exploder 2

saturation point 61
scattering parameters 119
second order intercept point 48
semiconductor technology 57
signal to distortion ratio 134

Signal to noise ratio 134
spark transmitter 1
spectral regrowth 132
spectrum analyzer 31, 41
speech coding 14
subtraction 101
surface mount technology
SMT 127
switching circuit 159
switching rate 71
switching time 69
symbol rate 17

thermionic valve amplifiers 67
third order intercept point 43
third order intermodulation distance 42
third order intermodulation product 83, 132
Time Division Duplex
TDD 24
transfer function 84
transmission parameter 103
transmitter 14

Universal Mobile Telecommunication System
UMTS 60
utilization factor 61, 65, 71

vacuum triodes 3
Variable attenuators 107
Very High Frequency (VHF) 57
voltaic arc technology 2

WCDMA 112
wideband signal 82