Minimum power design of RF front ends

Minimum Power Design of RF Front Ends

Peter Baltus

Cover: JWL Producties Eindhoven

Minimum Power Design of RF Front Ends

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Eindhoven, op gezag van de
Rector Magnificus, prof. dr. R.A. van Santen, voor een
commissie aangewezen door het College voor
Promoties in het openbaar te verdedigen
op 16 september 2004 om 16.00 uur

door

Petrus Gerardus Maria Baltus

geboren te Sittard

Dit proefschrift is goedgekeurd door de promotoren:

prof. dr. ir. A.H.M. van Roermund
en
prof. ir. A.M.J. Koonen

Printing: Printservice TU/e

Aan Christel en mijn ouders

Contents

Section

List of symbols

Main Symbols

κ	Power linearity parameter; $\kappa = P / (G\ IP3)$
C	Capacitance
f	Frequency
DR	Dynamic range
DR_0	Dynamic range of the desired signal
DR_1	Dynamic range of the desired signal with a single interferer
DR_2	Dynamic range of the desired signal with two interferers
F	Noise factor
G	Gain
G_p	Realized power gain, or power gain: $P_{out,del}/P_{in,del}$
G_t	Transducer power gain: $P_{out,del}/P_{in,av}$
G_{av}	Available power gain: $P_{out,av}/P_{in,av}$
G_{max}	Maximum power gain: $P_{out,av}/P_{in,del}$
G_v	Realized voltage gain; $G = G_v / \zeta$
gm	Transconductance
I	Current
$HD2$	Second-order harmonic power of the output signal
$IP2$	Second-order intercept input power
$HD3$	Third-order harmonic power of the output signal
$IP3$	Third-order intercept input power
$IP3_0$	Third-order intercept input power measured for DC input signals
$IP3_1$	Third-order intercept input power measured for single-tone input signals
$IP3_2$	Third-order intercept input power measured for two-tone input signals
$IP3'$	Third-order intercept single-tone input power measured for two-tone input signals
J	Current density
k	Boltzmann constant; *k approximately equals 1.380658e-23*
L	Length of a device, or inductance value
NF	Noise figure; $NF = 10\ log(F)$
$OIP2$	Second-order intercept output power
$OIP3$	Third-order intercept output power
P	Power
PAE	Power added efficiency
R	Resistance
T	Absolute temperature

V	Voltage
W	Width of a device
$VHD2$	Second-order harmonic voltage of the output signal
$VIP2$	Second-order intercept input voltage
$VHD3$	Third-order harmonic voltage of the output signal
$VIP3$	Third-order intercept input voltage
Z	Impedance
ζ	Ratio of load and source impedance, $\zeta = Z_l / Z_{src}$

Subscripts

av	available (as in available power)
del	delivered (as in delivered power)
if	if circuit
rf	rf circuit
I	running index, e.g. indicating the i^{th} stage in a cascade
tot	total
in	input
out	output
I	running index for e.g. cascaded subcircuits
l	load
src	source
eff	effective
b	base
c	collector
e	emitter
bc	base-collector
be	base-emitter
cs	collector-substrate
dist	distortion component
fund	fundamental component (desired signal)
base	signal frequency component corresponding to the input frequency
third	signal frequency component corresponding to 3 times the input frequency

Section

Abbreviations

AM	Amplitude Modulation
AMPS	Advanced Mobile Phone System
BT	Bluetooth, a wireless personal area network standard
CDMA	Code division multiple access
CDMAOne	DSSS CDMA system, also called IS-95
CT0	Cordless Telephony System 0
CT1	Cordless Telephony System 1
CT2	Cordless Telephony System 2
DECT	Digital European Cordless Telecommunications system
DCS	Digital Communication System (GSM system in 1800MHz band)
DSSS	Direct sequence spread spectrum
EDGE	QPSK-based modulation for GSM with higher spectrum efficiency
EFOM	Equivalent Figure of Merit
EM	Electromagnetic
EMF	Electromagnetic Field
ETSI	European Telecommunications Standardization Institute
FDMA	Frequency division multiple access
FDD	Frequency division duplex (also sometimes called “full duplex”)
FHSS	Frequency hopping spread spectrum
FM	Frequency Modulation
FOM	Figure of Merit
FSK	Frequency Shift Keying
GFSK	Gaussian Filtered Frequency Shift Keying
GPRS	General Packet Radio Services, multislot packet service over GSM
GSM	Global System for Mobile Communication
HSCSD	High speed circuit switched data, multislot extension to GSM
IEEE 802.11	Family of WLAN standards
IEEE 802.11a	WLAN standard for data rates up to 54Mbps at RF frequencies between 5GHz and 6GHz
IEEE 802.11b	WLAN standard for data rates up to 11Mbps at RF frequencies between 2.4GHz and 2.5GHz
IEEE 802.11g	WLAN standard for data rates up to 54Mbps at RF frequencies between 2.4GHz and 2.5GHz
I-mode	Web-like services across GSM and PDC systems
IS-95	DSSS CDMA system, also called CDMAOne
ISDN	Integrated Services Digital Network
LNA	Low noise amplifier
MSK	Minimum Shift Keying
NMT	Nordic Mobile Telephony system

OFDM	Orthogonal Frequency Division Multiplex
PCS	Personal Communication System (used for GSM system in the 1900MHz band, as well as for IS-95 systems)
PDC	Personal digital cellular, cellular phone system in Japan
PHP	Personal Handy Phone system (Japan)
PHS	Personal Handy System (Japan)
POTS	Plain Old Telephony System (analog phone system)
PSK	Phase shift keying
QPSK	Quadrature phase shift keying
RF	Radio Frequency
SDMA	Space Division Multiple Access
SMS	Short message service (short text messages over GSM)
TDMA	Time division multiple access
TDD	Time division duplex
UMTS	Universal Mobile Telecommunications System, a standard for the third generation cellular phone & data systems.
WAP	Wireless application protocol, providing web-like services across the GSM system
WAN	Wide Area Network
WCDMA	Wideband CDMA, 3G standard originated in Japan and now unified with UMTS
WLAN	Wireless Local Area Network

Chapter

1

Introduction

Communication, in all its guises, makes the difference between a number of lonesome individuals and a community. It enables cooperation and the development of cultures. In this sense, communication has played an essential role in the development of human civilization [1].

In recent times, electrical systems have been developed to support and augment communication. Among others, these systems allow communication across much greater distances than would normally be practical or even possible. The electrical systems that enable such communication are called telecommunication systems [2]. Telecommunication systems such as telephone, television and the Internet are changing the communication between people dramatically, and have a correspondingly big impact on communities, cultures and economies. This makes telecommunication systems a very relevant research topic. Consequently, many aspects of telecommunication systems are being studied all over the world.

Since the beginning of the previous century, propagation of electromagnetic (EM) waves through space has become a very popular basis for many telecommunication systems, since it allows electrical, but wireless, communication systems. Devices that use such EM wave propagation in the frequency range of approximately 10kHz to around 1THz are commonly called radios [4]. Since radios allow communication without having to build up a wired connection between the devices, they allow the mobile use of communication devices, for example in cars (figure 1). Even though mobile use is not quite impossible using wired connections [5], it is highly impractical.

Especially for mobile use, the energy consumption of the radio is very important, since most mobile devices depend on a battery for their operation. Their energy consumption then determines the amount of time that they can be used without replacing or recharging the battery, and thus the cost and convenience of using such a device. The time that a mobile device can be used without recharging is often split in the standby time

and the talk time. The standby time is the time in which the device is not actively communicating, but only monitoring the radio channel for any incoming calls. In this mode of operation, the rate at which the battery is discharged depends mostly on the receiver. The talk time is the time in which the device is actively communicating, and both the transmitter and receiver are active. In this mode of operation, the transceiver usually dominates the battery discharge rate.

The radio frequency (RF) front end is the part of a radio that interfaces between messages in the form of electrical signals and the EM field (through the antenna), making it an essential part of any radio. A block diagram of a typical RF front end is shown in figure 3. Even though the complexity (in terms of number of components) of most RF front ends is small compared to other parts of modern radios and mobile devices, the RF front end is usually responsible for a significant part of the cost, performance and power consumption (both in terms of talk time and standby time) of the total radio.

Figure 1 *First mobile phone attempts (Copyright © 2003 Lucent Technologies [109])*

Different telecommunication systems support different types of communication. For example, broadcast systems support uni-directional, point-to-multipoint communication, whereas cellular systems support bi-directional, point-to-point communication. The various types of communication and their messages, as well as the particular properties of radio communication, are discussed in more detail in appendix A.

Telecommunication systems are evolving quickly, and this has a significant impact on the requirements for RF front ends. The history of telecommunication systems and their properties is discussed in more detail in appendix B. The main trends, especially

with respect to the radio frequencies, will be discussed in Section 1.1.

The variety of telecommunication systems results in widely varying requirements for RF front ends. This makes a completely general study of RF telecommunication front ends impractical. The context for this research is therefore limited to the current cellular, cordless, and wireless data systems. The relevance of the main topics of this thesis within this context, namely RF front ends, low power, and fundamental limits, will be discussed in the remainder of this chapter.

1.1 Trends in RF frequencies

In the past, the range of frequencies that could be used for new radio systems was limited at low frequencies by the availability of unused spectrum, and at high frequencies by the availability of relatively cheap technology in which the radio could be implemented.

With progress in technology, and increasing scarcity of free spectrum at lower frequencies, radio frequencies for wireless devices have increased. They are currently concentrated in the 1GHz to 2GHz region. Especially the explosive growth of the cellular phone, and the resulting move from bands below 1GHz to the 2GHz region, has been driving the development of cheap IC technologies with increased bandwidths.

This is currently changing: cellular and other wireless systems that need to work over longer distances will be limited by the decreasing link budget at high frequencies. In the radio transmission equation below (eq. 1), P_{RX} is the available signal power from the receive antenna, P_{TX} is the power applied to the transmit antenna, G_{RX} and G_{TX} are the gains of the receive and transmit antennas, λ is the wavelength, r is the distance between receive and transmit antenna, and α is the propagation constant. This equation holds in the far field of the antennas, assuming that the polarizations of both antennas are perfectly matched.

The propagation constant α depends on the environment of the antennas. In free space, this constant is 2. Inside buildings it varies between 1.81 and 5.22, and tends to be higher at higher frequencies [21], because the attenuation of walls increases with frequency.

$$P_{RX} = P_{TX} G_{RX} G_{TX} \left(\frac{\lambda}{4\pi r} \right)^{\alpha} \qquad \textbf{(1)}$$

For omnidirectional antennas with constant gains, the link budget scales as (eq. 2):

$$\frac{P_{RX}}{P_{TX}} \sim \frac{1}{f^{\alpha}} \qquad \textbf{(2)}$$

Increasing transmit power to make up for this reduction in the link budget is impractical in the frequency region above roughly 5GHz, especially for portable, battery-powered systems. Another solution would be the use of high-gain antennas, but since such antennas will also be highly directional, they are not practical for portable wireless devices, unless advanced beam-steering approaches are applied and a line-of-sight path is available.

Since neither of these remedies is very practical, it seems likely that there will be a split in the frequency bands for short-range, mostly line-of-sight, high-bandwidth, high-bitrate connections, and long-range (including outdoors) connections that need to propagate in all directions and through walls and other objects. The essential distinction is in the frequency f, the propagation constant α, and antenna gains G_{RX} and G_{TX}, but for practical purposes these systems will be referred to as long-range and short-range systems.

Short-range systems will continue to move to higher frequency bands to satisfy the need for increased bandwidths, whereas long-range systems will be limited to frequencies below 5GHz. These long-range systems will concentrate on efficient use of the increasingly scarce spectrum below 5GHz through efficient modulation schemes, efficient access mechanisms and adaptive use of the available resources. This split between short-distance, high-bandwidth systems and long-range systems is shown in fig. 2.

Front ends in long-range systems will require high linearity, low noise figures and high power efficiency for most efficient use and re-use of the spectrum. Since the frequencies in these systems will not increase significantly in the future, but the bandwidths of the devices in future IC technologies will continue to increase for many years to come, the power dissipation is typically limited by the dynamic range, rather than the frequency.

Front ends in high-bitrate, high-bandwidth, short-range systems will probably

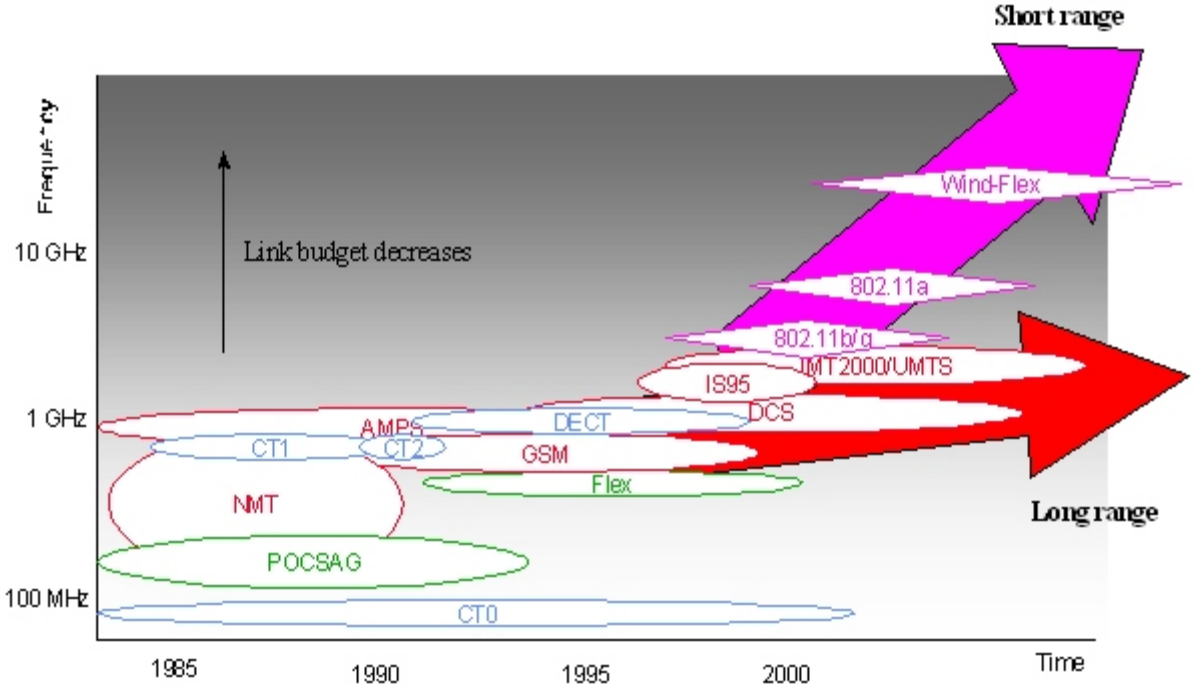


Figure 2 *Trends in wireless frequencies*

continue to move to higher frequencies. The high bandwidths for these systems will only be available at increasingly high frequencies in e.g. the 5GHz and 17GHz bands. The power dissipation in front ends for these systems will be limited by the high frequencies at which the front-end circuits will need to operate. Such systems will have relaxed requirements for sensitivity and linearity compared to long-range systems, and these requirements will typically be met at the current levels that are needed to achieve the high frequency operation of the front-end circuits. Therefore, the power dissipation is typically limited by the operating frequency rather than the dynamic range.

This divergence of specifications can already be found in current systems such as the GSM cellular system (for long range) and the Bluetooth system (for short range). Specification parameters that are especially affected are sensitivity and linearity. Sensitivity is affected, among others, by the noise figure (NF) of the receiver. The noise figure of a circuit is defined as the ratio between the input-referred noise of that circuit, and the thermal noise from the source that provides the input signal. The relation between sensitivity and noise figure is discussed in more detail in Appendix G. Linearity is affected, among others, by the third order intercept point (IP3). IP3 is defined as the input power at which the extrapolated third-order output power equals the extrapolated power of the desired signal. This is discussed in more detailed in Section H.2, and the relation between IP3 and the effect of interferers is discussed in more detail in Appendix G. Table 1 shows the NF and IP3 requirements as defined by these standards.

	GSM	*Bluetooth*
Frequency	950 MHZ	2450 MHZ
NF	9 dB	25 dB
IP3	-10 dBm	-21 dBm

Table 1 Sensitivity and linearity specs for GSM (long-range) and Bluetooth (short-range) systems

This table shows that the smaller received signals and the stronger interferers that long-range systems need to work with translate as expected in lower noise figures and higher IP3 values. Please note that competitive implementations usually need to perform better than the performance set in a standard. Still, a significant difference between both types of systems remains. A more extensive overview of system parameters can be found in appendix B.4.

Future portable wireless devices will be able to set up both short-range, high-bitrate links and long-range connections, depending on the environment and the needs of the user. This can be achieved through the use of multiple standards in one device, e.g. Bluetooth and GSM, or through a single standard that supports different trade-offs between range and data rate, such as UMTS/IMT 2000. In such multi-mode devices, the properties of the transceiver front end will depend on the particular mode that the device is being operated in. The optimizations discussed in this thesis will apply to each of these modes. Although the methods proposed in this thesis can be applied to current systems, they might need to be adapted to deal with future systems based on very different approaches, such as time-domain ultra-wideband systems.

1.2 Relevance of RF front ends

Communication is the process of exchanging information between two or more entities. In radio communication systems, this information is exchanged through electromagnetic signals. The translation of the messages into electromagnetic signals and vice versa is carried out in the transmitter and receiver parts of radio communication systems. An important part of the transmitter and receiver is the RF front end, which converts between the RF signals and the low or intermediate frequency signals. (see also appendix A)

An RF front end carries out three important signal processing steps, both in the transmitter and receiver signal path:

1. Frequency translation, in order to obtain a signal with a frequency appropriate for further transmission or processing;
2. Signal amplification, in order to obtain a signal with an amplitude appropriate for further transmission or processing;
3. Signal filtering, in order to remove unwanted signal components generated by other transceivers or by the signal-processing circuits in the transceiver itself.

A simple block diagram of a front end is shown in figure 3.

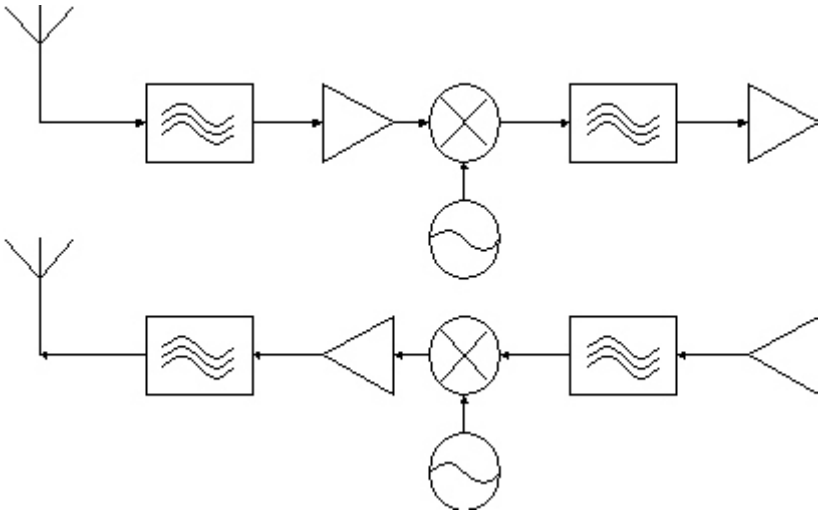


Figure 3 *Block diagram of a typical front end*

This block diagram shows that both gain and filtering are typically distributed throughout the front end. The frequency conversion, here shown in a single conversion step using mixers and local oscillator signals, can also be implemented as two or more conversion steps, since this can reduce the requirements on the performance of the individual filter, gain, and mixer blocks. A simple antenna interface is shown here, with separate, single, transmit and receive antennas. In practice, it is often desirable to share the antenna between transmitter and receiver, in order to save cost and space. This requires switches or duplex filters between the circuits and the antenna. In addition, many systems today

use multi-band radios, requiring multiple antennas or multiple antenna feeds, which in turn require extra switches or, in some cases, varicaps. Finally, antenna diversity is becoming more practical, but also more necessary, at higher frequencies, again creating the need for more antennas and related electronics.

The oscillators used for frequency conversion are often tunable to allow channel selection already in the front end of the radio, reducing the dynamic range requirements of the filters, IF amplifiers, and data converters. These tunable oscillators are typically controlled through a synthesizer loop, adding considerable complexity in terms of number of transistors to the front end.

Even so, an RF front end consists of only a very limited number of active devices, often in the order of hundreds or thousands. The silicon area of the front end decreases with improvements in IC technology that cause devices to shrink in physical dimensions. Current RF front end ICs measure just a few square millimeters. Also, the desired functionality, and therefore the number of active devices, decreases with advanced transceiver architectures such as single-conversion receivers, and further digitization of the radio. Compared to the millions of transistors in the digital parts of a handset and the millions of bytes in the software, it might be assumed that the design of an RF front end is a trivial job, with little relevance for the overall handset. From such a viewpoint, it would be hard to justify any research in this area. Fortunately for RF researchers, the RF design problem is highly relevant. Figure 4 shows the inside of a modern cellular (GSM) terminal.

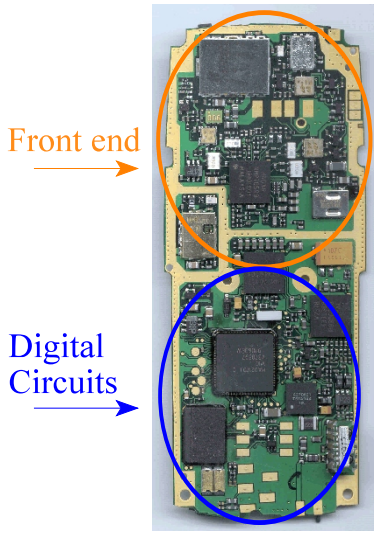


Figure 4 *Inside of a GSM handset*

As can be seen in this picture, the front-end part contributes significantly to the size of the total handset. It also has considerable influence on other properties, such as power dissipation and bill of material. Depending on the system, and trade-offs in the design, the impact of the front end on the cost, size, and power dissipation of the total system can be between 10% and 50%. Finally, the quality of the signals, and therefore the accuracy of the messages, depends very much on the properties of the RF front end. This makes

the RF front end a highly relevant part of most telecommunication systems.

1.3 Relevance of low power

RF front ends for the systems described in the previous chapters typically run from a battery that is often the largest and most expensive component in the device. The main design problem is to keep such devices small and cheap while achieving the desired functionality and performance.

The total power consumption is often dominated by the radio as much as by the digital circuits. This has an additional impact on size and cost through the battery size: in a typical cellular handset, up to 50% of the total power is consumed by the radio, both in transmit and in receive mode. Therefore, the RF front end has a major impact on achieving the design goals of a small and cheap phone. Much effort has already gone into reducing the power dissipation of RF front ends for this type of application, and significant improvements have been made over the past years [8]-[17]. Nevertheless, there seems to be no reason why, ultimately, the power dissipation should be constrained by other limits than the law of conservation of energy[1] and imperfections in implementation technology[2].

The efficiency of a circuit is defined as the ratio between the output power and the power consumed from the power supply. Transmitter efficiencies of over 40% are currently found in publications [18][19], suggesting that there is little room for improvement in this area. However, when taking into account losses in the antenna matching, filtering and switching circuits, and losses between battery and power amplifier, the overall efficiency quickly drops below 25%. Even worse, the best efficiencies are usually achieved for maximum power output levels. When power levels are decreased to more typical output levels, efficiencies often drop below 10%. In the future, we will see the emergence of more variable envelope modulation schemes, especially for the long-range systems, in order to use the available spectrum more efficiently. Also, in addition to FDMA and TDMA, CDMA will be used to allocate the available spectrum flexibly and efficiently. Variable envelope modulation schemes and CDMA access methods will increase the requirements for linearity, power control range, and accuracy of the power amplifier, resulting in overall efficiencies below 5% for output power levels 20dB or more below the maximum output power level [20]. Improved efficiency of the transmitter, at lower power levels, is therefore expected to become important. A higher efficiency of the antenna interface is likely to become even more important, since this affects performance at all power levels.

[1] The law of conservation of energy relates to electronic circuits, and thus to RF circuits, through the relation between power consumed from the power supply and signal inputs, and power delivered to signal outputs. The law of conservation of energy requires that, for any typical RF circuit that only exchanges energy with its environment through electrical signals and through heat dissipation, the sum of the average signal powers delivered at the output terminals is less than, or equal to, the sum of the average signal powers received at the inputs, and the power consumed from the power supply.

[2] Although eliminating the imperfections in the implementation technology might be physically impossible.

The power dissipation of the receiver tends to be much lower than the power dissipation of the transmitter. However, since the receiver will usually be switched on for more extended periods of time than the transmitter, e.g. to check whether new calls or messages are coming in, or whether a hand-over to a different base station is needed, the power dissipation of the receiver is nevertheless very relevant.

Receiver front ends tend to consume power far in excess of the bound set by conservation of power. The output of a typical front end might drive a sigma delta analog-to-digital converter (ADC) with a real part of the input impedance of e.g. 100kΩ, at signal levels below 1Vpp. The maximum power delivered by the front end is then $0.5V^2/100k\Omega=1.25\mu W$. A well-optimized front end consumes around 50mW, resulting in an efficiency below 0.01% in best case conditions. For lower input signals, resulting in lower output power, the efficiency further decreases by several orders of magnitude. Since the power efficiency of current receiver front ends is this low, there should be many opportunities for power reduction.

1.4 Relevance of fundamental limits

The complexity of RF circuits, when measured by the number of components, is much lower than the complexity of e.g. digital circuits. This is, of course, due to the simpler intended functionality of an RF front end circuit: it only needs to convert signals to an appropriate frequency and signal level, and filter, to some extent, some undesired signal components. Even so, the time required to develop an RF product is not proportionally shorter than that of a digital product, when taking the number of components as a reference. This has undoubtedly resulted in many debates in development groups, and in various approaches to improving the current RF IC design methods, which are sometimes perceived as inefficient and old-fashioned. Thus far, none of these approaches have resulted in RF IC development efficiencies comparable to digital IC design. Therefore, it is relevant to investigate the reasons for this lower efficiency, and to develop more efficient design methods, if possible.

Common design methods for RF front ends are based on experience, insight, and creativity of the designer. After the specifications for the product have been set, a first estimate of the distribution of gain, noise, linearity and selectivity for each of the subcircuits is made, based on experience with previous designs and/or simple calculations. The circuits are obtained from earlier designs, or from colleagues who have designed similar circuits in the past, and adapted to the new specifications. If no suitable circuits are available, they are developed from scratch.

Since the complexity of the desired functionality of an RF front end is rather low, this approach seems fairly obvious and is easy to carry out. However, mistakes made at this point in the development process can turn out to be very expensive ones indeed. They often require a redesign almost from scratch, or many iterations, to achieve acceptable results.

Subsequently, the subcircuits need to be designed and simulated. This is where an important difference between digital and RF IC design becomes relevant. RF front end circuits tend to work close to the limits of what can be achieved in a given technology, in terms of frequency, linearity and noise. In order to accurately predict the performance

of a circuit this close to the limits of a technology, one needs to take into account a large number of parasitic effects in the circuit and the package, such as parasitic capacitance, resistance, and inductance, of interconnect, as well as electrical and magnetic coupling of the interconnect and the package, crosstalk through the substrate, thermal effects, etc. Even for simple circuits, it is not practical to model and simulate all these parasitic effects, given the current state of computer hardware and software.

As a simplified example, figure 5 shows a circuit diagram of an amplifier as intended by the designer.

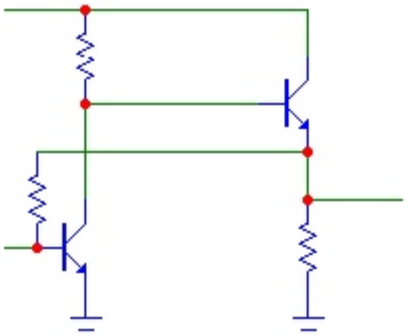


Figure 5 *Amplifier circuit as intended by the designer*

When this circuit is implemented on an IC, parasitic effects occur that are not taken into account in the schematic of figure 5 or in the models of the transistors and resistors. Figure 6 shows a simplified layout of this amplifier (in a simple, fictitious IC process): Using the layout and knowledge of the IC process, the parasitic effects in the circuit can

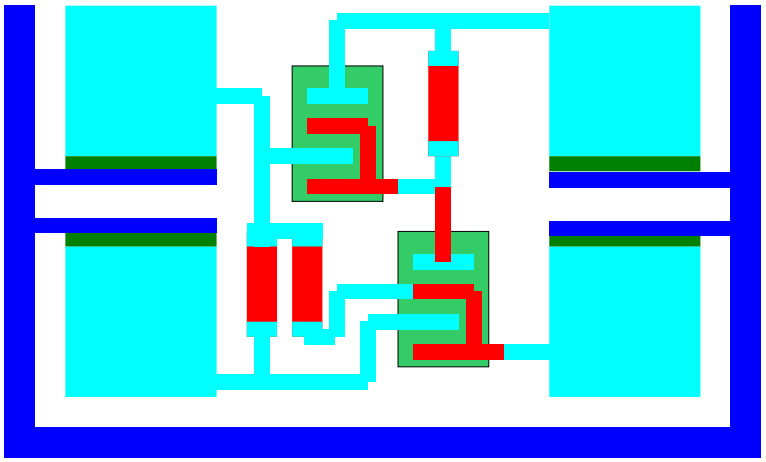


Figure 6 *Layout of amplifier circuit*

be calculated. In the schematic diagram of figure 7, a small subset of the most significant parasitic effects has been annotated.

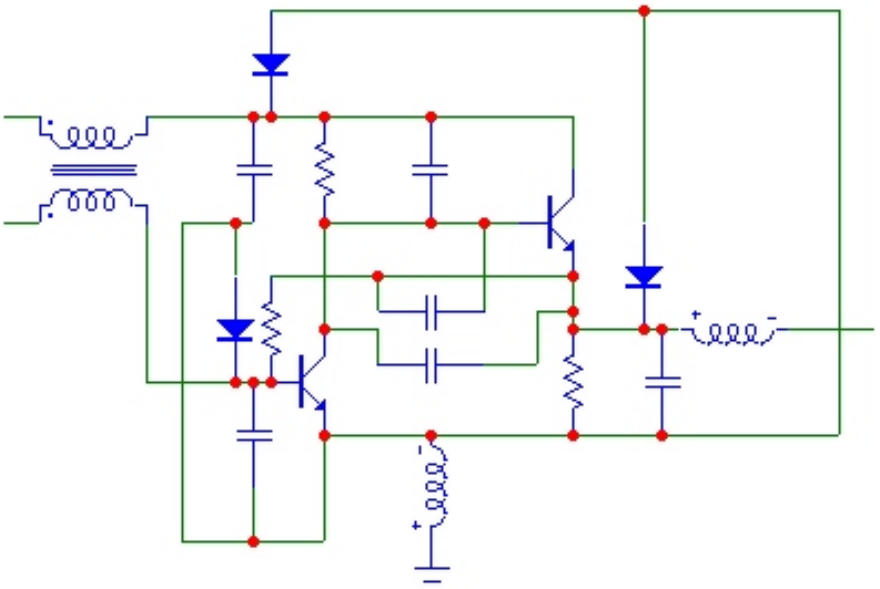


Figure 7 *Amplifier with some parasitic effects.*

As can be seen in this simple example, making a selection of parasitic elements to be modeled and simulated is essential to keep the problem manageable and the results accurate. Except in the most simple cases, even experienced RF designers cannot always make the right selection on the first try, which leads to measurement results that differ significantly from the simulation results. As a consequence, several iterations through the IC fab are needed to achieve the desired circuit performance. This product development process is shown in the flowchart of figure 8. The times in the flowchart are indications only: they depend very much on the complexity of the product, the size of the development team, their experience, changes in specifications during the development, fab arrangements and performance, stability of the IC process, accuracy of the models, stability of the IC design tools, etc. In many cases, most of the product development time is spent in the iteration loop through fabrication.

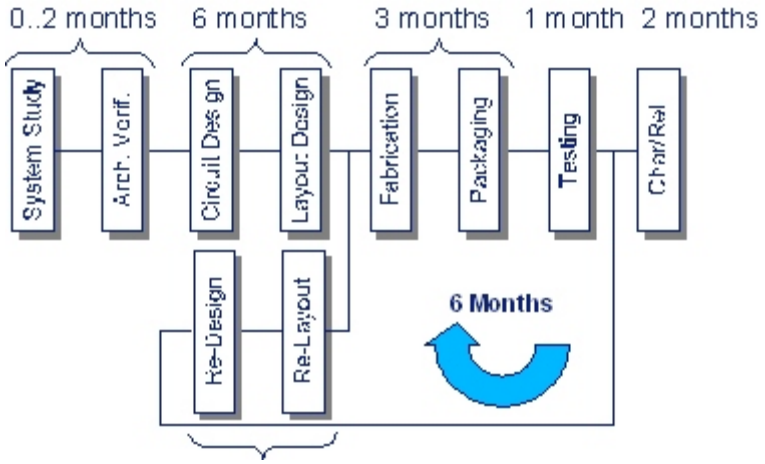


Figure 8 *IC design process*

This process is basically an implementation of an algorithm that searches for the global minimum of a function *P,* which represents the cost as a function of all the design variables v_1, v_2, v_3, ... , v_n. In general, this cost includes all aspects of a circuit that need to be optimized, such as size, technology cost, number of external components that are required, etc. In this thesis, we will focus on the power dissipation, and therefore *P* will represent the power dissipation of the circuit. The design variables include not only the parameters of a circuit (such as transistor sizes, resistor values, power supply voltages, etc.), but also architectural choices, circuit topology selection, and other degrees of freedom in a design process that are not as easy to quantify as the circuit parameters. In principle, the global optimum can be found using a suitable algorithm and sufficient iterations in the design. In practice, however, RF product development projects often need to be completed in a (very) limited amount of time, reducing the number of iterations that can be carried out. Especially changes in the early choices of a design, such as the architecture, can result in the need to start the iteration loop all over again, and, consequently, in a considerable loss of time.

Thus far, there has been no systematic way to take power dissipation into account this early in the design process. This results in higher than desirable power dissipation of the final product, and/or a large increase in development time.

Fundamental limits, e.g. for power dissipation as a function of gain, linearity and noise, can help in the development process by giving an indication for the lowest possible power dissipation that can be achieved with a given technology for a required performance level. This insight can be used to develop efficient design methods for RF front ends with a power dissipation close to these fundamental limits. Following such design methods will greatly limit the time needed for designing minimum-power RF front ends.

1.5 Thesis overview

This thesis will investigate the design of RF front ends for telecommunication systems that consume a minimum amount of power. The state of the art in low power design, especially with respect to design methods, achieved power dissipation, figures of merit, and trends, will be discussed in Chapter 2.

The problem of designing a minimum power RF front end starts with understanding the lower limits of power dissipation that can be achieved. These limits are imposed by physics and technology. In an RF front end, several different operations need to be carried out. The physics and technology imposed limits for each of these operations will be discussed in Chapter 3.

A traditional DECT RF front end is used as the basis for a discussion of low power design. Although this particular design is from 1993, the design approach is still representative for the way in which low power designs are currently carried out. For the remainder of this thesis it also provides some of the system background, and a benchmark for low power design. The DECT front end will be discussed in Chapter 4.

To design minimum power front ends, a new design method is needed. In Chapter 5, such a new design method will be introduced. It is based on a library of circuit

topologies that have been selected for the best low power performance. The selection, in turn, is based on a special figure of merit from the class of equivalent figures of merit, which will also be introduced in Chapter 5. Furthermore, a set of transforms that can be applied to any circuit will be developed. Such transforms are called "structure-independent transforms", or SITs. Using SITs, the optimum trade-off between the performance parameters of individual circuits within a front end can be determined. This optimum trade-off determines which transforms will be applied to the circuits from the library, in order to assemble a minimum power front end. A special tool has been developed to support this design method.

The design method can also be used to identify the boundary conditions at system, circuit and technology level that limit further power reduction. These boundary conditions, as well as the implications for the development of low-power systems, circuits, and devices, will be discussed in Chapter 6.

The application of several of the insights of this work to actual radio front-end circuits will be demonstrated in Chapter 7. The 3.5mW receiver includes a low-power RF front end with advanced antenna diversity, as well as a new low-power non-linear signal processing method for the IF signal. This new signal processing method is described in more detail in appendix E. It is based on the reconstruction of the phase trajectory of the received signal from the non-equidistant samples obtained from the transitions of the limiter outputs.

For many markets, the time-to-market is a very important consideration in addition to the power dissipation of the RF front end. One approach that is aimed at improving time-to-market is the use of a platform as the basis for new products. Chapter 8 discusses how a platform approach can be adapted to RF design, and how this can be combined with minimum power design.

The results of this research work is expected to help to improve the performance, reduce the power dissipation, and speed up the design of future RF front ends. Although this thesis focuses on RF front ends in wireless telecommunication systems, the results can be applied in a much wider range of systems and circuits, e.g. wired RF systems such as cable.

Chapter

2

Status and trends

Many RF design methods are currently being used, both in product development and in research. The variety is too large to allow an exhaustive treatment of all individual methods. Instead, they will be grouped in categories according to their main properties:

- Automatic design methods, as opposed to manual design methods
- Iterative design methods, as opposed to single-shot design methods
- Algorithmic design methods, as opposed to heuristic design methods
- Model-based design methods, as opposed to reality based methods
- Exact design methods, as opposed to approximate design methods
- Explicit design methods, as opposed to implicit design methods
- Bottom-up design methods, as opposed to top-down design methods
- Custom design methods, as opposed to reuse design methods
- Dedicated design methods, as opposed to general-purpose design methods

In addition, design methods might cover one, multiple, or all abstraction levels in a design. The differences between the design method categories are briefly explained in appendix D. In this section, the state of the art in low power design and design methods will be discussed in Section 2.1. Trends in low power design and RF design methods will be discussed in Sections 2.2 and 2.3.

2.1 State of the art in low power literature

Table 3 shows a small subset of the current design methods found in literature, categorized according to the classification system introduced in the previous section. It is obvious that even with this small subset of current design methods, all design method categories are already covered.

Name	Design Method properties																	Abstraction level			Literature reference
	Automatic	Manual	Iterative	Single-Shot	Algorithmic	Heuristic	Model	Reality	Exact	Approximate	Explicit	Implicit	Top Down	Meet in the middle	Bottom up	Custom	Re-cycle	Re-use	System	Circuit	Layout	
Custom design		✓	✓			✓	✓			✓	✓				✓	✓			✓	✓	✓	[67]
Adapt	✓			✓	✓		✓		✓		✓		✓			✓	✓			✓		[64] [65]
Trial and Error		✓	✓			✓		✓		✓		✓		✓		✓			✓	✓	✓	
Module generators	✓			✓	✓		✓			✓	✓		✓			✓	✓				✓	[68] [69]
Augmented mod. Gen.	✓			✓	✓		✓			✓	✓		✓			✓	✓			✓	✓	[70]
Exhaustive search	✓			✓	✓		✓		✓		✓		✓			✓				✓		[72][73][74]
Genetic algorithms	✓			✓			✓			✓	✓		✓					✓	✓			[76]
Analytical (receiver)		✓	✓		✓		✓			✓	✓		✓			✓				✓		[77]
Large signal behavioral		✓		✓	✓		✓			✓	✓		✓			✓				✓		[78]
Analytical (mixer)		✓	✓		✓		✓			✓	✓		✓			✓				✓		[75]

Table 3: RF design methods and their classification

Of all current design methods, *custom design* is still the most popular. It is the main RF design method supported by the popular CAD tools such as Cadence from Cadence Design Systems (San Jose, CA, U.S.A.), and ADS from Agilent (Palo Alto, CA, U.S.A.).

The main drawbacks of custom design are:

- *low effectiveness*. It is not clear how much margin is left between the performance that has been achieved in a design, and the optimum performance that could be achieved within the given boundary conditions (specifications, technology). The results depend very much on the experience of the designer.
- *low efficiency*. The design is assembled from scratch in a (mostly) bottom-up approach, often requiring iterations across multiple abstraction levels (system, circuit, layout, hardware) to achieve the desired performance.

These drawbacks drive the research into alternative design methods which improve upon, or even eliminate, these shortcomings. Many of these alternative methods address only one or a few abstraction levels (system, circuit, layout, hardware), or can be applied only to specific subcircuits of an RF front end.

An ideal design method should cover all applicable abstraction levels, as well as all subcircuits of an RF front end, and achieve optimum performance in an effective and efficient way. Such a design method for low power RF front-end circuits does not yet exist, but will be developed in this thesis.

Figure 10 shows the gain and power dissipation of RF low-noise amplifiers as reported in recent literature. This figure is representative of the performance and power dissipation of other RF circuits: there is a wide variety in performance parameters and power dissipation of individual circuits.

Comparing these numbers is difficult, however, since they relate to different circuit topologies, different implementation technologies, and different specifications for performance parameters such as noise, linearity, and bandwidth. A more accurate portrayal of current LNA performance would show the performance of these circuits in an n-dimensional space. The various dimensions would represent the relevant performance parameters, and points in this space would represent individual designs.

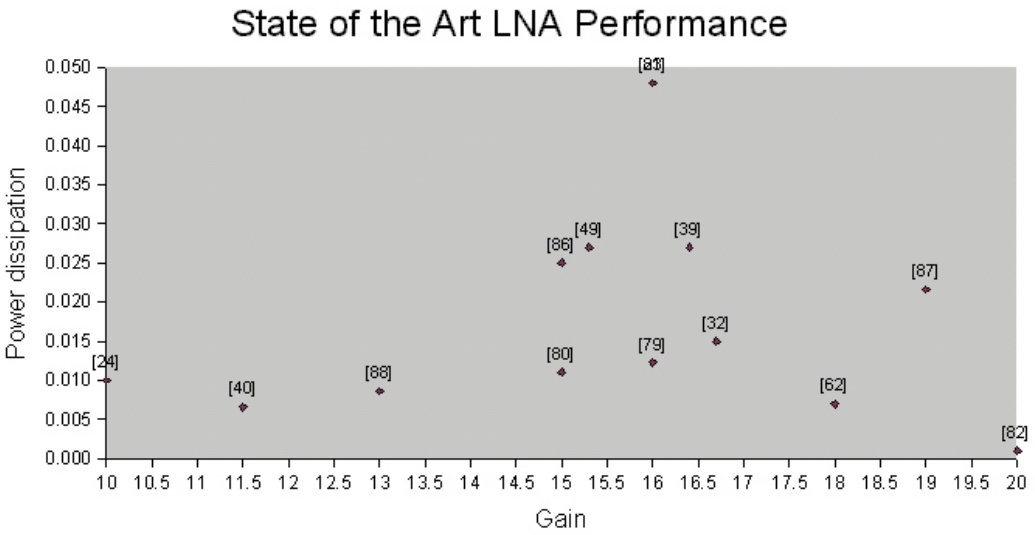


Figure 10 *Power dissipation of recently published LNA circuits versus their gain*

Since the number of relevant dimensions is significantly higher than 3, a drawing of such an n-dimensional space would not be very useful to most people. Instead, it is customary to map the points in this *n*-dimensional space onto a 1-dimensional space through a *figure of merit*.

A figure of merit, or FOM, is a function that maps any point from an *m*-dimensional space unequivocally onto a point in a 1-dimensional space. Since most FOMs do not take all performance parameters into account, *m* is usually smaller than *n.* This allows for easy identification of the relative performance of individual circuits, and therefore also of the "best" circuit. However, the relative position, and therefore the "best" circuit, depends on the FOM used for this mapping. Many different FOMs have been proposed in literature. Four examples from two publications are shown in table 4.

Name	Function	Reference
FOM1	Gain/(NF*Pdc)	[84]
FOM2	Gain/Pdc	[84]
FOM3	$(IIP3/(F\ k\ T\ B))^{2/3}$	[85]
FOM4	$\beta\ Vip3^2/\varepsilon$	[85]

Table 4 Some figures of merit for amplifiers

Considering the variety in figures of merit, it would be natural to ask for the "correct", or "best", or "most fair" FOM. This question cannot be answered, since the relevance of different performance parameters depends on the boundary conditions of the system in which a circuit will be used. In some systems, such as hearing aids or pagers, power dissipation will be a dominant parameter, whereas in other systems, such as wireless data, linearity might be more important.

2.2 Trends in low power design

The large number of performance parameters, and the large number of FOMs, make it very difficult to distinguish any trends in RF power dissipation. In figure 11, the absolute power dissipation of recent LNAs has been plotted versus their publication year. This graph does not show any obvious trends.

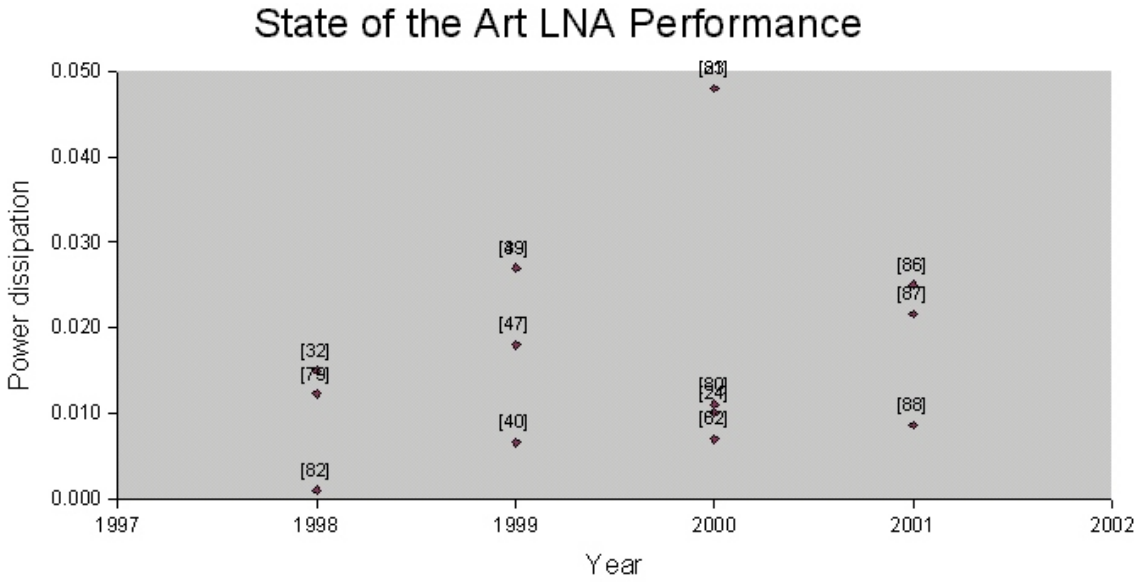


Figure 11 *Absolute power dissipation of recent LNAs over time*

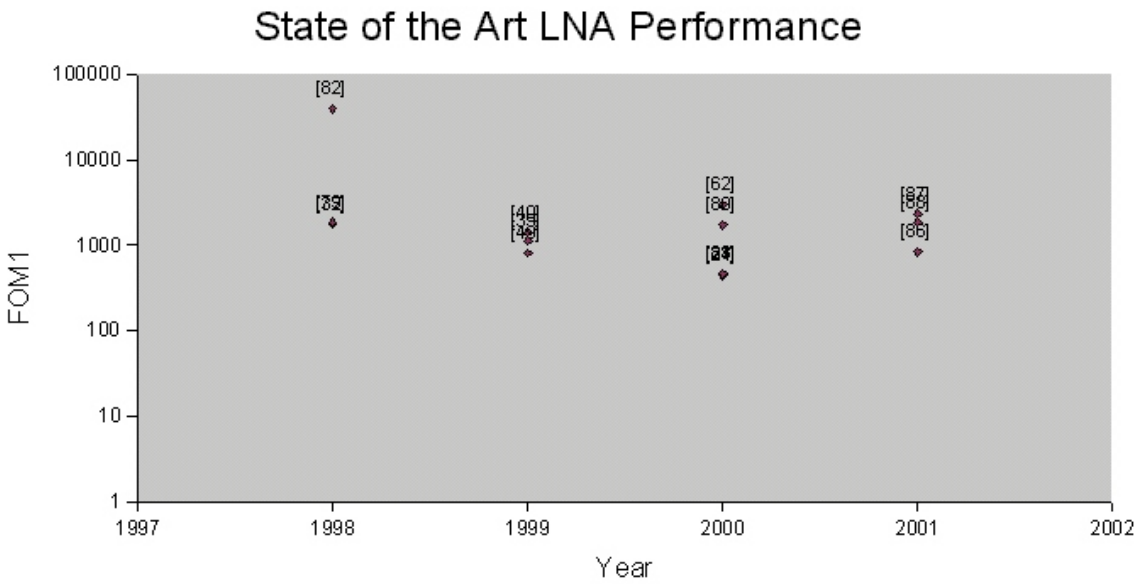


Figure 12 *FOM1 versus publication year of recent LNAs*

This makes sense, since the large variation in other parameter values of LNAs probably hides any trends that might be present in these numbers. Therefore, plotting a FOM, rather than the absolute power dissipation, might make trends easier to identify. Figure 12 shows the FOM1 of recent LNAs versus publication year.

As with the absolute power dissipation, there seems to be no obvious trend in the performance of recent LNAs. By trying many FOMs, a FOM might be found that shows a clear trend. The interpretation of such a trend would be difficult, since the selection of this FOM would be rather arbitrary. Therefore, a specific class of FOMs is required that provides a basis for comparing low power performance of RF circuits in a less arbitrary way. Such a new class of FOMs will be introduced in Chapter 5. At that point the trends in power dissipation of RF circuits will be reconsidered, and a trend will be demonstrated using this new FOM.

2.3 Trends in RF design methods

As mentioned in the previous section, the dominant design method for RF circuits is currently still custom design. Nevertheless, a push towards design methods with higher effectivity and efficiency can be detected in both literature and industry. Although the many different design methods that are being proposed initially seem very confusing, a trend can be distinguished from current methods that result in heavily cost-optimized, dedicated products, using all available degrees of freedom during the design stage, towards future methods that will result in more generic products with significant configuration flexibility. This trend can be represented in a graph depicting the trade-offs between design versus configuration flexibility, and cost versus value optimization [93] (fig. 13).

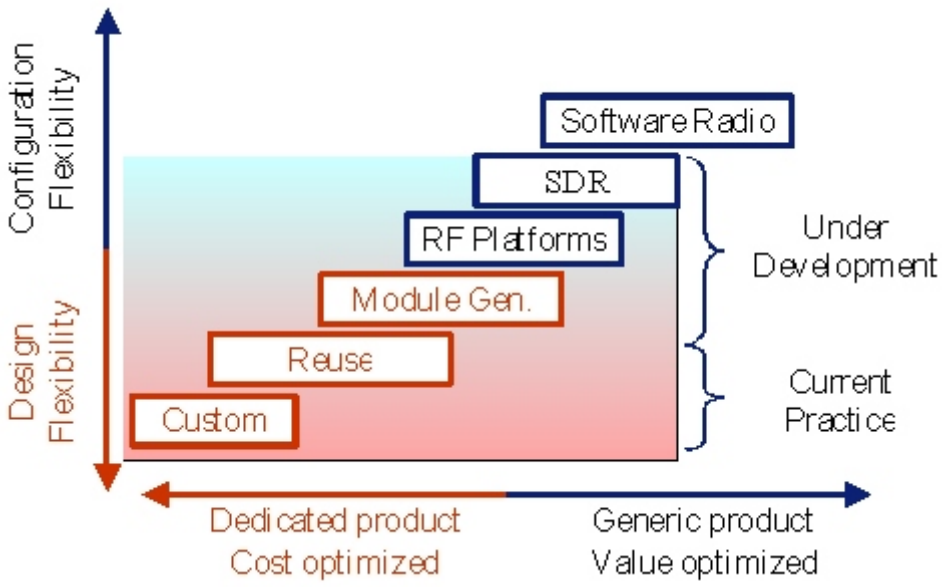


Figure 13 *Trends in design methods*

The methods that can be distinguished in this figure are:

- *full custom design*, the most popular design method at this moment, in which a product is optimized for low cost, and dedicated for a single application. Towards this end, all available degrees of freedom are used in the design of the product, and there is little, if any, flexibility in the configuration of the finished product.
- *reuse*, in which parts of a design, typically subcircuits, are not developed from scratch. Instead, they are more or less modified copies of parts of other, existing, designs. Pure reuse assumes no modification of the reused parts at all: the design is used "as is", without the need to even understand how it is constructed internally. This is still relatively rare in RF design. More often, the designs are modified, for example in order to work with other IC processes, or to achieve slightly different specifications. Such forms of reuse are often called "recycling". Recycling is closer to custom design than pure reuse, but for the purpose of this thesis, we consider them to be parts of the same category.
- *module generators*, in which (part of) a design is generated automatically, using an algorithm that takes the required specifications as an input, and that produces a representation of the design at a pre-defined abstraction level, often a schematic or layout. Module generators offer even less flexibility during the design phase, but also very little flexibility (if any) after fabrication.
- *RF platforms*, in which a product is built from pre-existing, configurable components that are assembled in a module or multi-die package. This is the first design method that offers significant configuration flexibility after fabrication; but it also offers very little flexibility during product development.
- *SDR*, or software-defined radio, in which the transceiver performance can be dynamically reconfigured during operation. This offers more configuration flexibility than RF platforms, since in RF platforms the configuration is mostly fixed during the assembly of the pre-fabricated blocks.
- *Software radio*, which offers the ultimate configuration flexibility by carrying out all signal processing in software.

These design methods will probably not be developed simultaneously. Instead, it seems likely that they will be developed over time from the bottom-left corner towards the top-right corner.

Reuse and recycling is already being used in many designs, and is currently more formally organized in many companies. A next step will be the standardization of the format in which the reusable elements are stored and transferred. If such standardization is achieved between multiple companies, it will become practical to trade RF IP blocks. This also requires a better standardization of IC processes between companies, since otherwise the transfer of IP, even if it is in a well-standardized format, would be difficult. RF CMOS processes, based on standard digital CMOS processes, seem to be the most likely candidate to be used for this purpose.

Module generators and RF platforms are still in the advanced development stage, and are not widely used yet for product development. Software-defined radio and software radio are mostly still research projects.

Newer design methods will probably not replace older ones immediately. Instead, it seems likely that they will coexist for a long time, similar to the design method developments in e.g. digital design. Which design method will be used for a specific design in this timeframe will depend on the specific requirements with respect to the

parameters on the axes of fig. 13. Therefore, it seems likely that successive generations of one design will follow a reverse timeline compared to the development of the design methods. A first product generation might be designed using a method near the top-right corner, because the high configuration flexibility will enable fast time to market, without the need for a fully cost-optimized product. Later generations of the same product might be developed using design methods towards the lower-right corner, since the markets are likely to be larger and more mature by this time, requiring dedicated, cost-optimized product designs.

Chapter

3

Low power problem

Based on the background provided in the previous chapter, it is now possible to more accurately define the central research question of this thesis:

"**What are the fundamental limits for the power dissipation of telecommunication front ends, and what design procedures can be followed that approach these limits and, at the same time, result in practical circuits?**"

under the following boundary conditions:

- the "practical circuits" need to be competitive to existing circuits with respect to cost and performance;
- the design procedures should not take significantly more time and/or effort than current design procedures;
- the application area is limited to consumer telecommunication systems, such as cellular, cordless, and wireless data systems in the range of 1GHz to 5GHz;
- it should be possible to extend and/or generalize the design procedures to other RF design areas.

This chapter will focus on fundamental limits in RF low-power design. Two types of fundamental limits can be distinguished:

1. Fundamental limits imposed by physics
2. Fundamental limits imposed by technology

This distinction is relevant, since our understanding of the laws of physics tends to remain valid for long periods of time, whereas technology limits are improved upon on a regular basis in the semiconductors industry. Therefore, limits of RF circuit and system performance can be expected to improve in parallel with technology limits, and to approach limits imposed by physics in an asymptotic manner.

3.1 Elementary operations and signal processing stages

As discussed in the introduction to this thesis, an RF front end typically carries out three elementary operations:

1. frequency conversion
2. amplification
3. filtering

Architectures and systems that don't require some or all of these elementary operations are conceivable, but they generally don't require less power dissipation. For example, architectures that use analog-to-digital converters at RF frequencies may avoid the frequency conversion operation (at least in the analog domain), but the power dissipation of the analog-to-digital converter is almost always much higher than that of a mixer. Notable exceptions are direct-detection AM broadcast receivers and near-field transponders used for e.g. theft detection and road tolling.

To carry out the elementary operations, an RF front end consists of a cascade of signal-processing stages. Each stage carries out at least one of the operations, but there is no one-to-one mapping of signal-processing stage types and elementary operations. In the receiver part of the front end, the following signal processing stage types can usually be distinguished:

- amplifiers, such as low-noise amplifiers (LNA) and voltage-controlled amplifiers (VCA), providing the gain required to increase the antenna signals to levels more convenient for further processing, especially where this further processing adds a significant amount of noise to the signal;
- mixer(s), which, in combination with oscillators and synthesizers, provide the frequency conversion (translation). Mixers are sometimes combined into pairs that mix the input signal with two independent local oscillator (LO) signals, often with the same frequency and 90° relative phase difference. Also, the frequency conversion is often carried out in several steps by cascaded mixing stages, usually with filters in between, such as in double-conversion superhet and sliding-IF architectures. In addition to frequency conversion, mixers often provide some gain as well;
- filter(s), providing frequency selectivity, and, in case of active filters often also gain.

The transmitter part of the RF front end also includes filters, mixers and amplifiers. The amplifiers in this case are typically pre-drivers and power amplifiers rather than low-noise amplifiers, but the basic principle is the same.

Please note that, while the three elementary operations are orthogonal, the signal-processing stage types (amplifiers, filters and mixers) are independent but non-orthogonal combinations of the elementary operations, as shown in table 5.

	Gain	*Frequency Conversion*	*Selectivity*
Amplifier	✓	×	×
Mixer	✓	✓	×
Filter	✓	×	✓

Table 5 Signal-processing stages and elementary operations

This table shows that all signal-processing stage types can have gain, but only the mixer can provide frequency conversion and only the filter can provide (significant) selectivity, since mixer and amplifier implementations on an RF IC are almost inevitably wideband.

In principle, it is possible to build up RF front-end circuits without amplifiers, since gain can also be obtained from mixer and filter circuits. This is not entirely unrealistic: for less demanding applications, receivers in which the mixer provides sufficient gain at sufficiently low noise levels are conceivable, although not necessarily practical (e.g. because of local oscillator leakage). Almost all current RF front-end circuits use a cascade in which all signal-processing stage types occur at least once.

To investigate the fundamental limits of power dissipation for RF front ends, the limits imposed by laws of physics and technology for each of the elementary operations will be discussed in the following sections. Since technology is constantly in development, scaling of IC technology and progress in related areas has to be taken into account (Appendix C). Subsequently, other elementary operations which are required for building a complete RF transceiver front end, but which are not usually considered part of RF IC design, will be discussed, as well as elementary operations needed for specific systems.

3.2 Gain

Gain stages have a fundamental lower limit in power dissipation that is imposed by the law of conservation of energy, which, in this case, can be formulated as:

$$\forall P_{in}: \quad P_{supply} \geq P_{out} - P_{in} \tag{3}$$

or, in words: the power provided by the supply should at least make up for the difference between input and output power. Please note that the condition "$\forall P_{in}:$" needs to be included to make this limit a property of the circuit itself, independent of the actual value of the input power.

Practical circuits often do not come close to this limit. Power amplifiers with low linearity requirements (for example, for constant-envelope signals) can have efficiencies in excess of 50%, but small signal amplifiers such as LNAs, in which the linearity is important, can have efficiencies below 0.000001%[3]. This already shows that there might be a lot of room for improvement of the power dissipation in mobile equipment.

One reason for low efficiencies is that all current active devices dissipate power, which is caused by the voltage drop across the device in combination with the output current. This is relevant for both long-range and short-range systems (Section 1.1).

[3] This is based on a DECT LNA which will be discussed in Chapter 1. This LNA consumes 3.7mA at 3V, when generating (with an antenna signal at the -90dBm sensitivity level) an output signal of -70dBm, corresponding to PAE=9.01 10^{-9}.

Another reason is that circuits often need to achieve a pre-defined RF performance with a given technology. Since these requirements are different for long-range and short-range systems, they will be investigated separately.

3.2.1 Gain in long-range systems

Long-range systems normally have strong requirements for linearity, since they are optimized for highest bandwidth efficiency. Even with ideal devices and interconnect, creating gain with low distortion will result in a reduction of the power added efficiency (PAE), for three reasons:

1. All known active devices (vacuum tubes, all kinds of bipolar and field effect transistors, MEMs) have a non-linear transfer function, resulting in distortion of the input signal.
2. Some active devices (e.g. bipolar transistors) do not have symmetric transfer functions for negative and positive output currents.
3. Parasitic non-linearities, such as the non-linear capacitors in and around active devices, cause additional distortion.

All three reasons are technology related, rather than imposed by laws of nature. However, not all three are equally likely to reduce in impact or even disappear with technological improvements. The parasitic non-linearities mentioned in number 3 usually decrease with new generations of an IC process. Number 1 and number 2 are caused by basic properties of the active device, and are therefore less likely to improve significantly in future IC process generations. All three effects can be reduced at the cost of power consumption:

1. All circuits can be transformed into circuits with higher linearity (when expressed through e.g. IP3) by connecting multiple instances of the original circuit in parallel (fig. 49). This results in a proportional decrease of resistance and inductance values, and a proportional increase in capacitance values and active device sizes. As a consequence, all voltages stay the same, all impedance levels decrease proportionally, and, therefore, all currents increase proportionally. The lower input impedance results in a higher IP3 at the same input voltage, and, therefore, in higher linearity. This is an example of a transform that will be further investigated in 5.3.1, and can be expressed as:

$$P = b_i IP_i \qquad \textbf{(5)}$$

with P the power dissipation, I the order of the non-linearity that dominates the power dissipation, b_i a circuit and technology dependent constant, and IP_i the intercept point of order i.

2. Circuits using asymmetric devices often use bias currents which exceed the maximum signal current to avoid bidirectional device currents (class A amplifiers). This DC current results in additional power dissipation, but it improves linearity. Different trade-offs between bias current and linearity result in the familiar class AB, class B, and class C amplifiers. This results in higher values for b_i in equation (5).
3. Parasitic devices that are not related to the circuit components will have a smaller impact when the circuits are scaled up in W and/or L, and therefore in current.

This results in a less than proportional increase of the power with the linearity, making (5) a worst case approximation for scaling circuits up to higher linearities.

There are other techniques that improve linearity without necessarily increasing power consumption by as much as is the case in the three approaches outlined above. These will be discussed in 6.2.

3.2.2 Gain in short-range systems

Short-range systems normally have strong requirements for gain at high frequencies, since they are optimized for the highest bandwidths, which are mostly available at high frequencies. Therefore, short-range systems require high bandwidths at low currents. This is a fundamental problem in IC design, because simple scaling does not work. As discussed in Chapter 1, scaling device lengths and widths down at high frequencies results in a relative increase of parasitic currents compared to the intended currents. Depending on the circuit topology and parameters, these parasitic currents will result in either a decrease or an increase in gain at high frequencies. A decrease in gain is most common; an increase will occur in cases where the parasitic currents effectively provide a positive feedback from output to input of one or more gain stages. Although this second case might sound relatively attractive compared to the first, in practice, it is difficult to achieve this effect without causing instability in the circuit at some frequencies.

The first case, in which the gain decreases, often creates a single dominant pole, since the ratio between the intended and parasitic currents of all lateral and vertical devices scales with the reciprocal value of the frequency. This results in the familiar constant gain bandwidth products for amplifiers. Since the ratio between intended and parasitic currents improves with larger intended currents (and larger device dimensions), the gain bandwidth product will also improve with larger intended currents. For lateral devices, the following expression can be derived, based on the expressions in appendix C:

$$G\,BW \propto \frac{\alpha_1}{2\,\alpha_5 L\left(\alpha_2\left(1+\frac{\Delta W}{W}+\frac{L}{W}+\frac{\Delta L}{W}\right)+\alpha_3\left(1+\frac{\Delta W}{W}\right)(L+\Delta L)\right)} \qquad \textbf{(6)}$$

and for vertical devices:

$$G\,BW \propto \frac{\alpha_6}{2\,\alpha_9\left(\alpha_2\left(\frac{1}{L}+\frac{\Delta W+\Delta L}{WL}+\frac{1}{W}\right)+\alpha_8\left(1+\frac{\Delta W}{W}\right)\left(1+\frac{\Delta L}{L}\right)\right)} \qquad \textbf{(7)}$$

both of which increase with the dimensions that control intended device currents (W for the lateral device, and L and W for the vertical device). These expressions are easily understood by choosing ΔW and ΔL equal to zero. The numerator and denominator in the

expression for lateral devices can then be multiplied by W , and the term W/L can be isolated as being proportional to low frequency gain. In the denominator, two sub-expressions remain. The first one, preceded by α_2, equals $L+W$, and is related to the perimeter component of the parasitic capacitance. The second one, preceded by α_3, equals $L \cdot W$, and is related to the area component of the parasitic capacitance. The numerator and denominator of the expression for vertical devices can be multiplied by $L \cdot W$, and this term can then be isolated as being proportional to the low frequency gain. Again, two sub-expressions remain in the denominator, with the first one, α_2 $(L+W)$, related to the perimeter component and the second one, α_8 $(L \cdot W)$, related to the area component of the parasitic capacitance. The terms ΔW and ΔL introduce the second order effects in the perimeter and area scaling.

The gain bandwidth product for lateral devices increases with a decrease in the length L, which again results in an increase in intended device currents. Therefore, the gain bandwidth product scales with power dissipation and vice versa, but, in general, there is no proportional relationship between them:

$$P = f_i(GBW) \tag{8}$$

It is interesting to note that the gain bandwidth products for lateral and vertical devices scale differently. For larger devices, in which ΔW and ΔL are small compared to W and L, and in which the perimeter contribution is small compared to the bottom area (especially in IC processes with trenches), the gain bandwidth product for vertical devices stays approximately constant with W and L, whereas the gain bandwidth product for lateral devices is proportional to $1/L^2$, and therefore increases quadratically with improvements in lithography. Therefore, bandwidth improvements in vertical devices need to be achieved through other means than scaling of width and length, such as scaling of vertical dimensions.

The gain bandwidth product is usually interpreted as the bandwidth of a low-pass transfer curve of an amplifier, caused by a single dominant pole, typically formed by a resistor and parasitic capacitances (fig. 14). The bandwidth of such a circuit is obviously

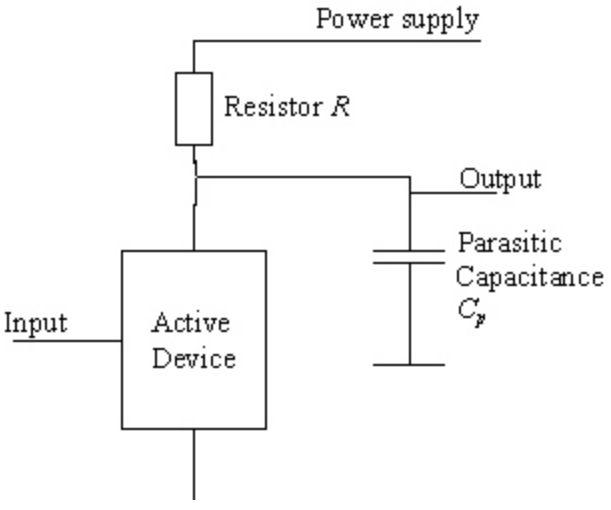


Figure 14 *Amplifier with single pole*

$1/(2\,\pi\,R\,C_p)$. However, telecommunication systems tend to use relatively narrow bands at high frequencies. The relative bandwidth varies from 1% (for example in the case of DECT) to 14% (for example in the case of IEEE 802.11a), but, inevitably, the largest part of the passband of an amplifier with a low-pass characteristic will not be used.

Therefore, it is often attractive to "tune out" the parasitic capacitance by adding an inductor L in parallel to the parasitic capacitance C_p (fig. 15). The value of the inductor L should create an admittance that has the opposite sign but the same value as the admittance of the capacitance at the center of the signal band f_c.

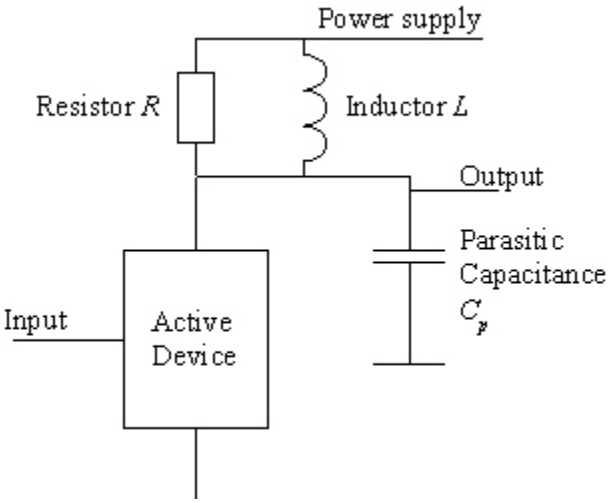


Figure 15 *Amplifier with tuned-out pole*

The bandwidth of this circuit is determined by f_c in combination with the quality factor Q of the R, L, C_p combination:

$$BW = \frac{f_c}{Q} = \frac{\dfrac{1}{2\pi\sqrt{L\,C_p}}}{\dfrac{R}{\sqrt{\dfrac{L}{C_p}}}} = \frac{1}{2\pi\,R\,C_p} \qquad \textbf{(9)}$$

This shows that tuning out parasitic capacitances does not increase the bandwidth of a circuit, but it does shift the bandwidth to another center frequency which can, in fact, be outside the pass band of the low pass configuration. For narrowband telecommunication systems, this can result in a significant reduction of the power dissipation. This model is fairly accurate for modern RF circuits, which can be demonstrated using the circuits shown in figure 16.

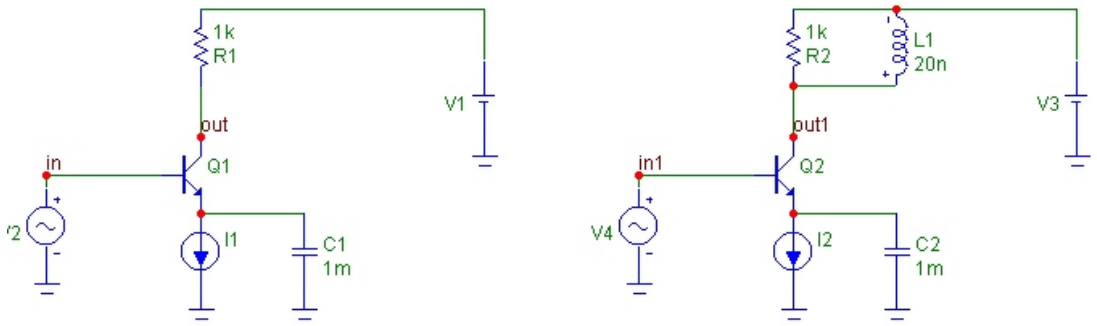


Figure 16 *Low pass and band pass class A bipolar amplifier circuits*

In this figure, both a low pass and a band pass version of the same circuit are shown, based on a bipolar class A amplifier implemented in a 0.7µm double poly BiCMOS process. The bandwidths obtained from simulations using these circuits are shown in figure 17.

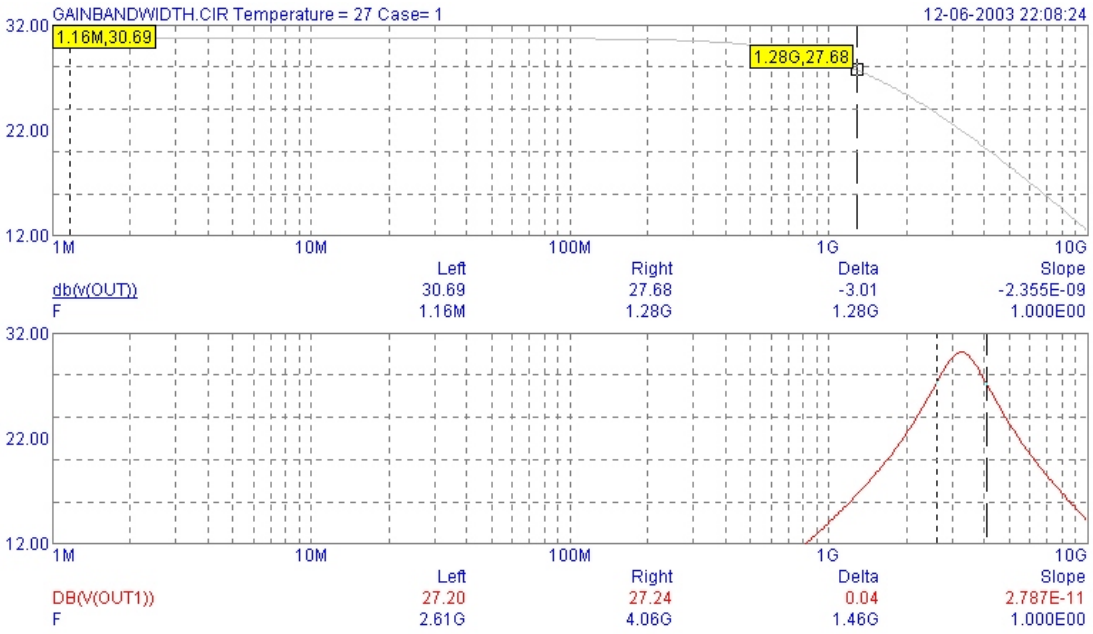


Figure 17 *Gain versus frequency of low pass (top) and band pass (bottom) amplifiers*

As shown by these simulation results, the bandwidths in both cases are comparable (1.28GHz for the low pass and 1.46GHz for the band pass circuit), even though the center frequency of the band pass circuit (at 3.33GHz) is well outside the pass band of the low pass circuit (1.28GHz). The gains in the centers of the pass bands are close as well (30.7dB for the low pass, 30.1dB for the band pass). The differences are caused by the complex higher order frequency selective parasitic elements of the bipolar transistor, which violate, to some degree, the assumption of a single dominant pole.

3.3 Frequency conversion

The fundamental limits imposed by physics on the power dissipation which is required for frequency conversion, are determined by two factors:

- the energy required to generate a signal with a sufficiently accurate frequency;
- the energy required to change the gain of an element periodically with a frequency derived from this signal.

These conditions hold for a frequency conversion by an arbitrary offset. In special cases, for example to convert a signal to a signal with an integer multiple of the original signal frequency, special solutions exist (e.g. non-linear transfer functions) that do not require a frequency reference or even an element with periodically changing gain. In telecommunication systems, there is usually a need for the more general frequency conversion by an arbitrary frequency shift, and, therefore, the limits for this type of frequency conversion will be further investigated.

The limit imposed by physics on the first factor (generation of an accurate frequency reference) is similar to the amplifier limit defined in the previous section: the output power of the signal generated by the frequency reference circuit (P_{out}) needs to be at least compensated for by the power supply (P_{supply}), that is, the conversion efficiency from power supply to reference frequency signal is 100% at most:

$$P_{out} \le P_{supply}$$

$$\frac{P_{out}}{P_{supply}} \le 100\% \qquad \textbf{(10)}$$

Practical oscillators around 1GHz usually achieve efficiencies above 1%, within two orders of magnitude from the physics imposed limits for power efficiency [24], [25], [26], [27], [28], [29].

The limit imposed by physics on the second factor (time-variant changes in gain) is set by the second law of thermodynamics, similar to the energy needed to open or close a switch. In the same way as with a switch, information about the value of the gain needs to be transported across a noisy channel. In the most simple case of a switching mixer, only two values for the gain will be used, and a single bit of information needs to be transferred to set the gain to the appropriate value. This transfer occurs through the reference frequency signal across a noisy channel. This results in a minimum energy per transition of [23]:

$$E_{switch} \ge kT\ln(2) \qquad \textbf{(11)}$$

With two switching operations per cycle of the reference frequency (f_{lo}), the power dissipation of such a switch is:

$$P_{switch} \geq 2 f_{lo} kT \ln(2) \tag{12}$$

Current circuits exceed this limit by many orders of magnitude. Mixers operating at frequencies around 1GHz usually require local oscillator (*LO*) signals with power levels between 10μW and 1mW, whereas formula (12) predicts a minimum power of 5.6pW, more than 6 orders of magnitude below this level.

In most systems, interfering signals are present at frequencies near that of the desired signal. Such interfering signals can be stronger than the desired signal. The switch signal, generated by the local oscillator, contains at least thermal noise, with a noise density of *kT*. At high frequencies, the switch signal will have finite rise and fall times, limited by the bandwidth of the technology used, and can be approximated by a sine wave. This causes mixing to occur with frequency shifts different from the frequency of the LO signal. This effect is called reciprocal mixing. As a result, the wanted output signal of the mixer decreases, and mixing products from the interfering product increase.

A very simple model for this effect assumes that the output signal of the mixer, V_{IF}, can be represented as the product of the LO signal V_{LO} and the input signal of the mixer V_{RF}:

$$V_{IF} = V_{LO}\, V_{RF} \tag{13}$$

This can be rewritten as:

$$\begin{aligned} V_{IF} &= V_{RF} \left(A_{LO} \sin(2\pi f_{LO} + \phi) + \int_{f=f_l}^{f_h} A_f \sin(2\pi f t + \phi_f)\, df \right) \\ &= V_{RF}\, A_{LO} \sin(2\pi f_{LO} + \phi) \; + \; V_{RF} \int_{f=f_l}^{f_h} A_f \sin(2\pi f t + \phi_f)\, df \end{aligned} \tag{14}$$

The noise of the local oscillator signal is represented by the integral part of this equation. A_f is the noise density, and ϕ_f is the corresponding phase of the sine wave element of the LO noise at frequency *f*. The parameters f_l and f_h define the lower and upper boundaries of the frequency band in which the noise is applied to the mixer LO input. The amplitude of the wanted output signal is not affected by the addition of the noise, and the output noise level scales proportionally with the RF signal level and LO noise level. This is not obvious: in other non-linear circuits such as e.g. oscillators, the amplitude of the wanted output signal is in fact affected by the addition of noise.

Although this model is simple and easy to use, it does not take into account that most mixer circuits are closer to switching circuits than linear multipliers. The reason for this is that common mixer circuits exhibit various undesirable properties at high

frequencies, such as strongly increased noise levels and significant distortion, when they are operated as linear multipliers.

A more realistic model would therefore introduce non-linear switching behavior. A simple non-linear switching model assumes an ideal switching stage that operates on the sign of the combined LO signal:

$$V_{IF} = \begin{cases} V_{RF} & V_{LO} > 0 \\ 0 & V_{LO} = 0 \\ -V_{RF} & V_{LO} < 0 \end{cases} \qquad \textbf{(15)}$$

In this model, the amplitude of the output signal of the mixer no longer depends on the amplitude of the LO signal. As a consequence, increasing the noise level of the LO signal not only results in an increase of the noise level of the output signal, but also in a decrease of the wanted output signal. This effect is simulated with the HP VEE tool, using a model implementation as shown in fig. 18.

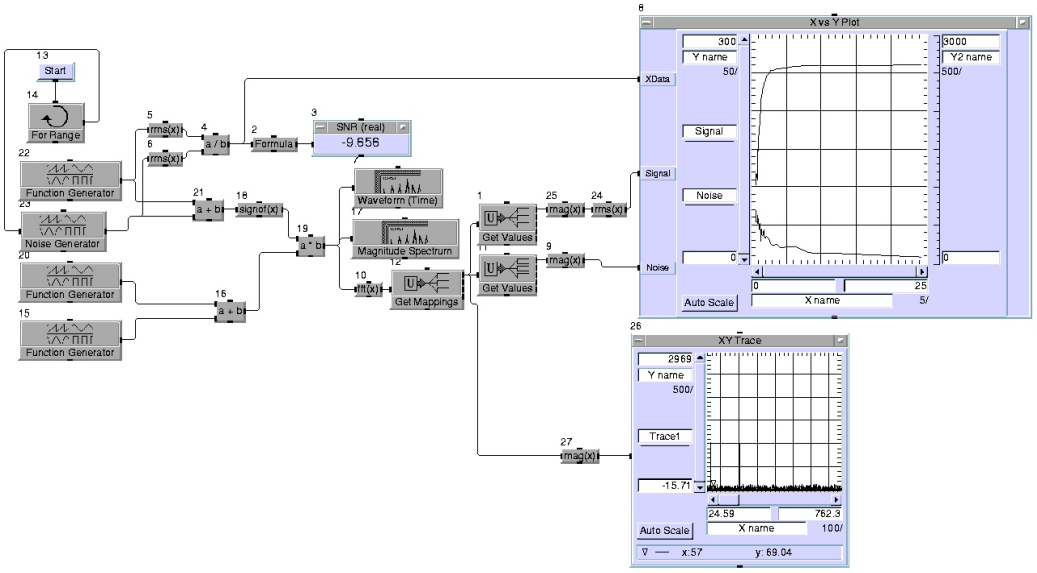


Figure 18 *Model of reciprocal mixing with non-linear switching mixer*

The simulation result of this tool is shown in the graph in the upper right corner of fig. 18. Fig. 19 is an enlarged picture of this graph. It shows the level of the wanted output signal and the output noise as a function of the noise level in the LO signal. The simulation uses a pseudo-random number generator to model the noise. In combination with the limited simulation time, this results in a random fluctuation of the signal and

noise levels in the simulation, as can be seen in fig. 19. Even so, the trend of increasing output noise levels and decreasing wanted signal levels can be clearly distinguished.

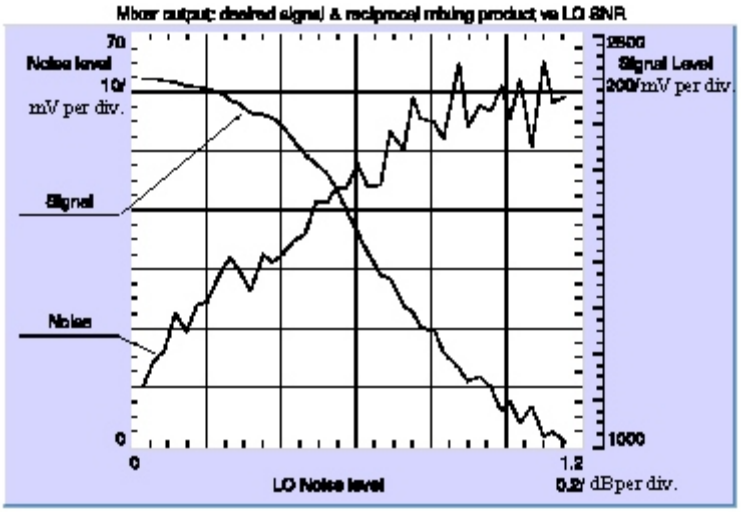


Figure 19 *Wanted output signal level and output noise level versus LO noise level*

Although fig. 19 is a convenient representation for demonstrating the close to proportional relationship between output noise level and input noise level, it is more common to show signal and noise levels as a function of the SNR of the LO signal (fig. 20). This shows more clearly that the level of the wanted output signal is only affected at very low levels of the LO signal SNR.

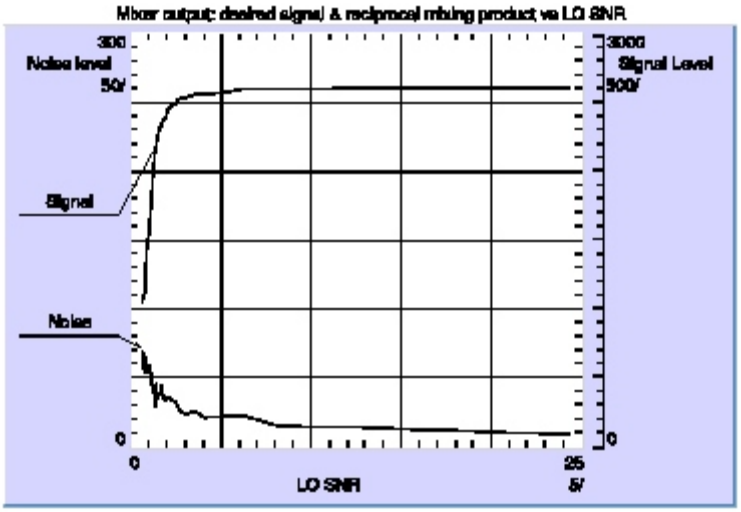


Figure 20 *Simulation results of reciprocal mixing model*

The impact of the LO signal SNR on the level of the wanted output signal and the output noise is really a technology-imposed limitation, not a physics-imposed one. To demonstrate this, the following graph (fig. 21) shows the result of the same simulation, except for the bandwidth of the LO signal, which has now been increased by two orders

of magnitudes, until the signal approximates a square wave, without increasing the amplitude of the signal.

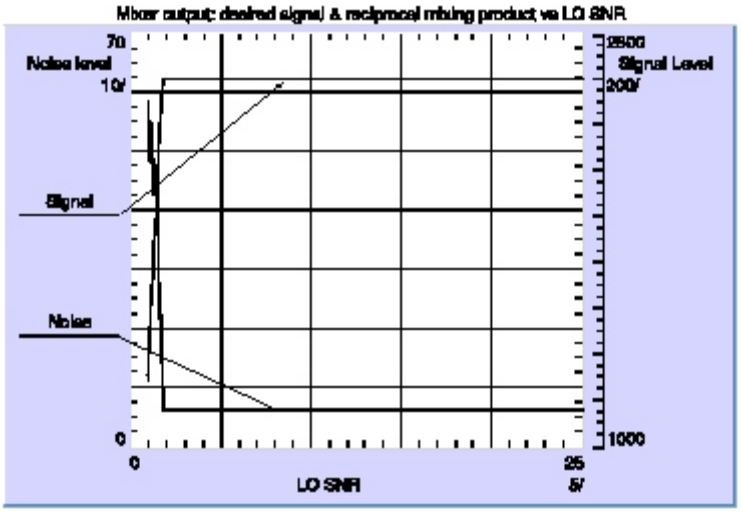


Figure 21 *Simulation result of reciprocal mixing model with increased LO bandwidth*

As is clearly shown in fig. 21, the wanted output signal level and noise level are much less affected by the SNR of the LO signal, and only for very low SNR values. To compare the technology limit and physics limit, a value of R_0 for the signal impedance level in the simulation is assumed in equation (16):

$$SNR = \frac{S}{\sqrt{4kTR_0BW}} \tag{16}$$

From equation (12), an expression for the minimum signal level S_c of the LO signal can be derived:

$$S_c = \sqrt{4R_0kTf_{LO}}\sqrt{2\ln(2)} \tag{17}$$

Substituting (17) into (16) results in an expression for the critical SNR of the LO signal, SNR_c:

$$SNR_c = \frac{S_c}{\sqrt{4kTBWR_0}} = \frac{1}{\sqrt{2\ln(2)}}\sqrt{\frac{f_{LO}}{BW}} \tag{18}$$

This SNR level is, in fact, higher than the SNR level at which the wanted output signal level starts to decrease because of reciprocal mixing, as shown in fig. 22:

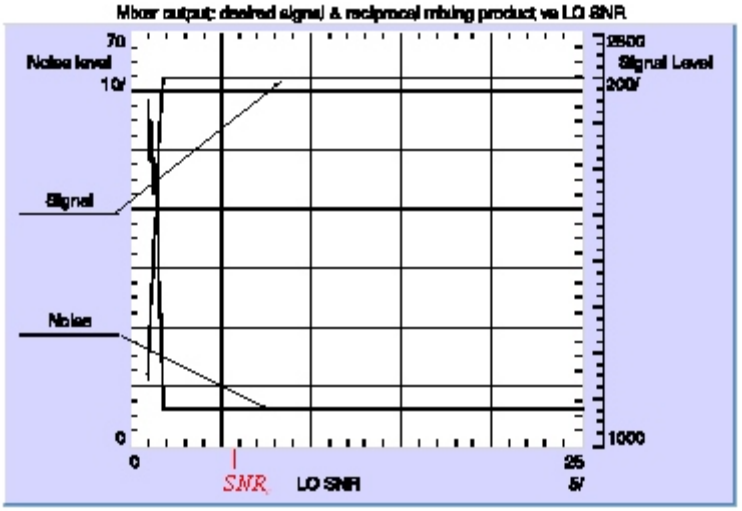


Figure 22 *Simulation results of reciprocal mixing model with increased LO bandwidth, and SNR_c limit indicated*

This demonstrates that reciprocal mixing is a technology limit rather than a physics-imposed fundamental issue.

3.4 Frequency selectivity

There is no known physics-imposed limit to frequency selectivity. As long as the technology provides ideal passive components (including zero losses, and therefore infinite quality factors), any filter function that can be built from passive elements can be realized with zero power dissipation. Implementation of on-chip frequency selectivity with current technologies has several limitations. The most significant ones are:

- Passive components (especially inductors and capacitors) are large, and therefore expensive, compared to active components. This ratio will be worse with new generations of IC processes, in which the active devices usually shrink more than the passive components. Moreover, new generations of IC processes tend to be more expensive per unit of area. This often results in a net increase of the cost of passive components. Therefore, there is a push towards active filters, especially for low frequencies, where the passive components would be even larger (and therefore more expensive) than at high frequencies.
- Even if the size and cost are not prohibitive, losses in conductors and substrate reduce the quality of passive components. This limits the achievable selectivity and introduces signal loss that needs to be compensated for by a gain stage. The extra power dissipation required for this additional gain is at least equal to the signal power lost in the passive components of the filter (see also 3.2).

- The limited accuracy of on-chip components requires special techniques for circuits with high selectivity, especially when the relative bandwidth is close to, or even smaller than, the relative frequency spread caused by the frequency-determining components. In these cases, tunable components and calibration circuits are often needed.
- To overcome the limitations of passive devices, active filters are used when significant selectivity at lower frequencies is required. These filters inherently dissipate power in the gain, transconductance, and/or gyrator elements that are used in such filters. For transconductor filters, the power dissipation P of a single transconductor-capacitor (gm-C) element depends on the required signal-to-noise ratio SNR, the required quality factor Q_0, the resonance frequency f_0, the noise factor F and the efficiency η [35]:

$$P \geq SNR\, 4\, \pi\, k\, T\, Q_0^3\, f_0\, \frac{F}{\eta} \qquad \textbf{(19)}$$

Selectivity is especially an issue for long-range systems, since, in these systems, spectrum efficiency is an important concern, and interference is a significant problem in the system design. The selectivity requirements for short-range systems are often more relaxed, with one important exception: several short-range systems, such as Bluetooth and IEEE 802.11b WLAN, operate at frequencies in Industrial, Scientific and Medical (ISM) bands. These bands are intended as "garbage areas" in which industrial, scientific and medical equipment can (within limits) radiate electromagnetic power that is generated as part of their intended operation. For example, consumer microwave ovens use RF energy in the 2.4GHz to 2.5GHz ISM band to heat food. A very small fraction of this power leaks out of the shielded cavity in which the food is heated. Since the power of such a microwave oven is very high (in the order of 700W to 1000W) to allow it to heat food very quickly, even that very small fraction is still several orders of magnitude higher than, for example, the 1mW of transmit power allowed for a class III Bluetooth device. Therefore, selectivity for short-range systems in the ISM band has a significant impact on the performance of these devices.

3.5 Other operations

In addition to the elementary operations, i.e. gain, frequency shift, and frequency selectivity, many other operations are required to complete the RF transceiver:

- electro-magnetic field to electrical signal conversion at the antenna
- analog-to-digital conversion at the ADC
- digital-to-analog convertsion at the DAC
- RF switching
- Received signal strength indication (RSSI) extraction
- channel switching
- power management

- isolation of the electronics from impedance changes of the antenna
- directional coupling

These operations have not been discussed, either because they are not part of the RF signal chain (e.g. RSSI extraction, channel switching, power management, analog-to-digital and digital-to-analog conversion), or because they are not usually implemented inside the IC (EMF to signal conversion, RF switching, isolation, directional coupling). The functions implemented by these operations have a limited interaction with the elementary operations defined before. The main interactions are shown in table 5.

	Gain	*Frequency conversion*	*Selectivity*
Antenna	✓		✓
ADC	✓		✓
DAC	✓		✓
RF switches	✓		
RSSI			
channel switching			
power management			
isolator	✓		
directional coupler	✓		

Table 5 Interaction between functions and elementary operations

As table 5 shows, there is no interaction between any of these operations and frequency conversion. The interaction with gain is, in many cases, just a compensation for losses in operations like the antenna, switches, isolator, and directional coupler. The antenna provides a significant contribution to the band selectivity in some systems. Since the functions of these elements cannot be implemented using the three elementary operations, and since the interactions between these operations and the elementary operations are rather weak, the elementary operations can, for the most part, be investigated independently of these functions.

There is, however, a strong interaction between the elementary operations gain and selectivity and the operations ADC and DAC. The performance of the ADC, as well as that of the DAC, determine the point at which the signal processing can be converted from the analog to the digital domain and vice versa. Due to improvements in data converter performance, the point at which this transition occurs is shifting more and more towards the antenna. Currently, the systems which are the focus of this research are usually implemented with digital demodulation and analog frequency conversion. The data converters are therefore located somewhere in the channel selectivity part of the transceiver. Moving them closer to the mixers, with less frequency selectivity in between, increases the performance requirements of the data convertors. There are two reasons for this:

- In the receive chain, the reduced frequency selectivity results in stronger interferers at the ADC input, requiring both higher sampling rates and a larger dynamic range of the ADC.
- In the transmit chain, the reduced selectivity results in stronger spurious signals from the quantized and time-discrete DAC output, requiring both higher sampling rates and larger dynamic range of the DAC.

With the power dissipation of a data converter at least proportional to both the sample rate and the number of quantization levels, in many systems it is not yet feasible to move the data converters all the way to the mixers without any channel selectivity in between, since this would result in prohibitive power dissipation of the data converters. On the other hand, the performance of current data converters is often sufficient to allow low power implementations without full channel selectivity between the mixers and the data converters.

This could in principle be a gradual process, in which the data converters shift towards the mixers pole by pole over time. System partitioning considerations often prevent this scenario. Many companies have developed implementations that are partitioned between a digital CMOS part and an analog CMOS, BiCMOS or even GaAs RF part. If the partitioning is such that one part of the channel selectivity is implemented in one IC, and the other part in the other IC, then the only way to build a transceiver with the correct performance is by using these two IC's in combination. At first, this might seem an advantage for the companies, since it will encourage customers to buy their complete radio as a chip set. However, it also requires both IC's to be competitive, or customers might prefer to mix and match RF and digital IC's from different companies. Such customers will require interfaces that are compatible with other vendors' IC's, thereby limiting the freedom in partitioning the transceivers between the analog and digital domain. As a consequence, some attention to the impact of the elementary operations on the data converters is justified, but since not all partitionings are practical, this discussion will be rather limited.

3.5.1 Other systems

Thus far, it was implicitly assumed that all three elemental operations are essential for all systems relevant to this research. Recently, however, there has been a lot of interest in systems that are rather different from most other systems. Especially some time-domain ultra-wideband systems (UWB), which use pulses to transmit their information, rather than a modulated carrier, could result in very different transceiver topologies and functions. For example, frequency conversion might not be required, or even feasible, for channels that span from 3GHz to 10GHz. Such systems will require different approaches at both the system level and circuit design level.

On the other hand, some of the newer proposals in the UWB standardization committees are moving in the direction of multiple sub bands and even more traditional modulated-carrier approaches, such as OFDM and wideband spread spectrum. For such systems, the approaches in this thesis do apply.

3.6 Summary and conclusions

In this chapter, three elementary operations in an RF front end have been identified: gain, frequency conversion, and selectivity. For each of these operations, the fundamental and physics imposed limits have been identified. Table 6 shows the fundamental limits, imposed by physics and technology, that have been identified for each of the three elementary operations.

	Gain	*Frequency Conversion*	*Selectivity*
Physics limits	$PAE \leq 100\%$	$P_{LO} \geq 2\,k\,T\,f_{LO}\ln(2)$	$P = 0$
Technology limits	$P = f_1(G\,BW)$ $P = b_i\,IP_i$	$P = f_2(G\,BW)$ $P = b_i\,IP_i$	$P \geq SNR\,4\,\pi\,k\,T\,Q_0^3\,f_0\,\frac{F}{\eta}$ $P = f_3(G\,BW)$ $P = b_i\,IP_i$

Table 6 Overview of fundamental and technology limits for elementary operations

The technology limit equations at the top for gain and frequency conversion are valid for short-range systems, and the equations at the bottom for long-range systems. The technology limits expression for selectivity at the top is valid for active filters, the middle one for passive filters in short-range systems, and the bottom one for passive filters in long-range systems. The linearity and gain referred to in these last two expressions refer to the gain stages that compensate for the losses in the passive filters.

Reciprocal mixing is not limited by physics, but by technology, as has been shown in 3.3. Increasing the bandwidth in the LO input stages of the mixer will reduce reciprocal mixing to arbitrarily low levels.

For practical circuits, the minimum power dissipation set by laws of physics is exceeded by quite a large margin, often multiple orders of magnitude, and determined mostly by technology limitations. This is very encouraging, since further reductions of power dissipation are likely with improvements in technology. A reduction in power dissipation might also be achieved with current technologies, by using improved system and circuit designs and design methods.

Chapter

4

A case study: the DECT front end

This chapter discusses the design of a low-power, zero-IF, DECT RF front end implemented in BiCMOS. This design was carried out around 1993, using a conventional custom design approach, which is still the most popular design approach. This makes it a very suitable benchmark for the other designs and design methods introduced later in this thesis. It also provides the systems background needed in later chapters.

4.1 Introduction

The current European cordless phone system is called the Digital European Cordless Telecommunications (DECT) system, which extends the cordless phone concept in the direction of large pico cellular systems for both speech and data communication. DECT is a short-range system, and therefore the emphasis for this design was on low power dissipation at high frequencies.

Section 4.2 introduces the DECT system and design targets, and Section 4.3 explains the principles of zero-IF receivers with their advantages and limitations. The design of the receiver is discussed in Section 4.4. Implementation and results are shown in Section 4.5.

4.2 The DECT system

In this section we will introduce some aspects of the DECT physical layer [90] that affect the design of DECT receivers. DECT implements wireless ISDN-compatible speech and data links using a network of overlapping base stations not unlike a cellular system such as GSM. However, the range of the base stations, and hence the distance between them. is much smaller than in traditional cellular systems. A typical DECT system can have cell sizes in the order of several hundred meters outdoors, and tens of meters indoors. This allows for very high traffic densities in a traditional cordless environment such as the home and office. Such small cells make DECT unsuitable for traditional cellular systems with nationwide coverage. Applications for DECT range from fully wireless PBX (private branch exchange) systems in large office complexes, spanning several sites with fully transparent operation, to the traditional use as a single cordless phone around the house. Local area networks (LAN) or connections from the public phone network to extensions in existing neighborhoods are other areas in which DECT offers an efficient and flexible alternative to other systems.

To optimize the use of the available spectrum, the DECT standard divides both the frequency band and the time into chunks that can be flexibly allocated to a number of simultaneous connections.

4.2.1 Frequency allocation

DECT systems use a frequency band in the 1.88GHz to 1.90GHz range. This range is divided into 10 channels with a channel spacing of 1.7MHz per channel. Modulation is Gaussian Minimum Shift Keying (GMSK) with a bandwidth-time product (BT) of 0.5 [91]. This provides a compromise that is less bandwidth efficient, but easier to implement than the modulation chosen for e.g. GSM. Transmission power is 0.25W, which is significantly higher than other cordless systems (typically 10mW), but much lower than traditional cellular systems (typically several Watts).

4.2.2 Time slot structure

Each channel is divided into frames. Each frame consists of 24 time slots (fig. 23). One frame takes 10ms, and a slot about 0.4ms. A standard full-duplex connection consists of two slots per frame. One slot is used for communication from handset to base station, while the other slot is used for communication in the opposite direction. Therefore, the transmitter and receiver in a handset are alternately activated with a duty cycle of $1/24^{th}$, resulting in a time-division-duplex (tdd) operation.

Each slot contains 420 bits, which are used for synchronization (32 bits), control (64 bits) and data (324 bits). The raw bit rate of a channel is 1.152Mbit/s, and the net bit rate of a standard full-duplex connection is 32kbit/s. Slots can be split into two half slots, and subsequent slots can be linked to provide flexible assignment of communication capacity.

The channel and slot which a connection uses are dynamically allocated to provide optimum reception and least interference. This also allows overlapping of base stations without complicated static channel assignments in advance. Transparent hand-over between base stations is also provided for.

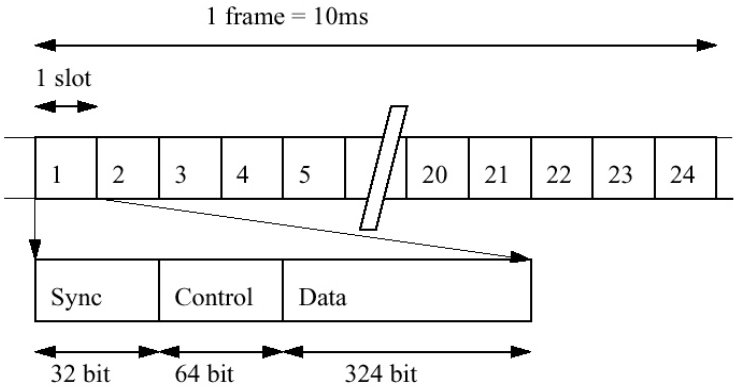


Figure 23 *Timeslot structure of a DECT signal*

4.2.3 DECT transceiver architecture

In fig. 24 a block diagram of a typical DECT transceiver is shown. It consists of a transmitter and a receiver, connected to a common antenna through a switch. The receiver and transmitter are provided with the required oscillator signals from a synthesizer block, which can be implemented with either one or two synthesizers. The output of the receiver and the input of the transmitter are connected to the burst mode controller, a digital IC responsible for the real-time parts of the protocol such as bit, slot and frame synchronization. The higher protocol levels and the user interface are implemented in software on a micro controller, which controls synthesizer, burst mode controller, antenna switch, and transmit and receive modes. The micro controller is also connected to a keyboard and display.

The codec is an ADPCM speech encoder/decoder which interfaces the compressed digital representation of speech to and from the analog signals which go to the loudspeaker and come from the microphone. In this section we will concentrate on the receiver aspects of DECT. As is obvious from fig. 24, this is only a small part of the total transceiver system.

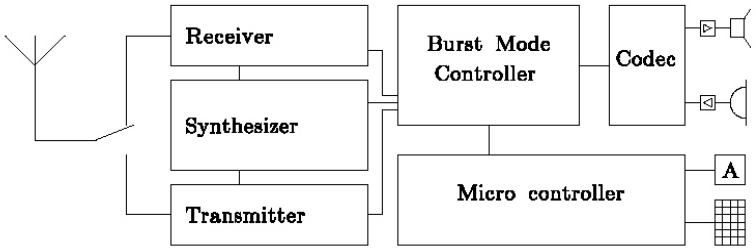


Figure 24 *System diagram of a DECT transceiver*

4.2.4 Requirements

The DECT standard defines a relatively complicated communication protocol compared to other cordless systems. This will require rather complex digital circuits (codec, burst mode controller, micro controller). Some advantages of this protocol (high traffic density,

ISDN compatibility, wide application area) have already been discussed in the previous sections. The features of the DECT standard also affect the design of the receiver front end. Some of them make the design simpler, others more complicated. These features and their impact are listed below.

Features that make the design simpler:

- The relatively large transmission power, small cell size, and low SNR required for good quality digital transmission allow for relatively relaxed receiver sensitivity.
- Distortion requirements are also moderate. Even though a DECT system in an office environment might be interference limited, the dynamic slot selection avoids many problems.
- The low duty cycle of the transmitter allows a high transmission power of 250mW with an average power dissipation of only 10mW, which is comparable to other cordless systems. The receiver also benefits from this low duty cycle, although, unlike the transmitter, it also has to be activated occasionally to check for incoming calls, even when the phone is not in a call.
- The antenna duplex filter typically found in cordless and cellular systems is not required in DECT, since the transmitter and receiver are never active simultaneously. In fact, a duplex filter is not even possible, since transmission and reception frequency are typically the same. Some kind of antenna switch might still be required.

Features that make the design more complicated:

- The RF frequency is higher than in most other cordless systems (typically 900MHz or lower). This requires more emphasis on RF design, and components and IC processes with improved RF behavior.
- The digital circuits are relatively complex compared to other cordless phones. Since the power dissipation of these circuits is expected to decrease with improvements in technology, their contribution to overall power dissipation will eventually become negligible. For now, the additional dissipation of these circuits has to be compensated for by lower power dissipation of the receiver. The relative importance of the receiver dissipation obviously depends on the ratio between active and stand-by operation of the transceiver.
- Since the digital circuits are relatively complex and add significantly to the component count, a minimum number of components and adjustments is as important in DECT as it is in other portable systems.
- To take advantage of the low duty cycle, the receiver must be capable of being powered up and switched off very quickly, within a fraction of a millisecond.

The next section will investigate the suitability of conventional and zero-IF receivers for DECT applications.

4.3 Zero-IF receivers

Conventional receivers mix the antenna signal (RF) with a local oscillator signal (LO) to obtain an intermediate frequency (IF), which can then be filtered for channel selectivity and be demodulated (fig. 25). In such a receiver, a trade-off exists for the frequency of the IF signal. High IF frequencies require difficult-to-implement channel-selectivity filters and demodulators, while low IF frequencies require difficult-to-implement image-rejection filters in front of the mixer. Often, this trade-off does not yield the required results, and a double-conversion receiver has to be used. Such receivers require more filters, mixers and oscillators, resulting in increased space, weight, power dissipation, and cost.

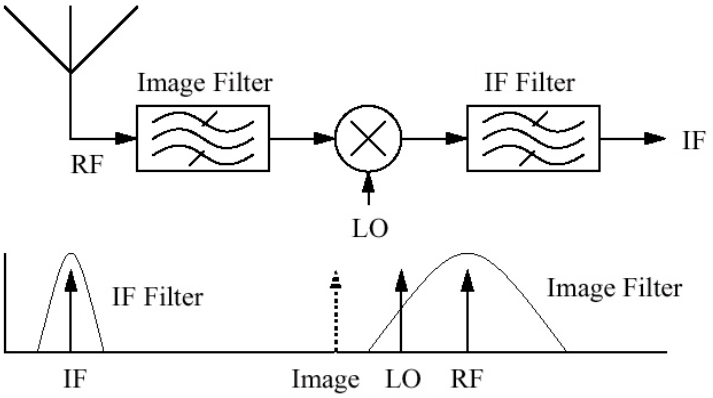


Figure 25 *Conventional receiver architecture and signals*

4.3.1 Zero-IF receiver architecture

These problems can be avoided by choosing the IF frequency at, or very close to, zero. The IF frequency is in this case much lower than the bandwidth of the signal. Receivers which apply this architecture are called zero-IF receivers.

To distinguish between positive and negative IF frequencies which correspond to the upper and lower sideband of the RF signal, a second mixer and filter branch is used in the receiver. In this second branch, an LO signal of the same frequency but different phase is used to drive the mixer. The two IF signals are commonly labeled I and Q (fig. 26).

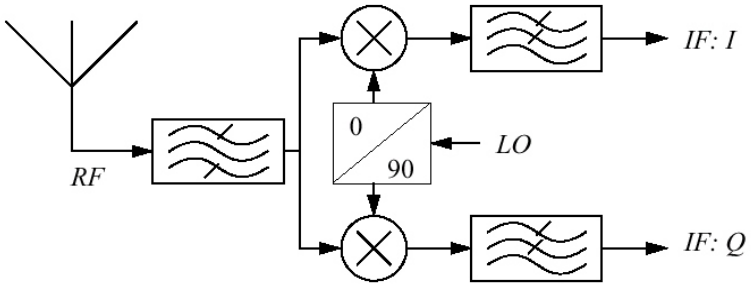


Figure 26 *Zero-IF receiver block diagram*

The phase difference between the LO signals is ideally 90 degrees. The sign of the phase difference between I and Q can be used to distinguish between positive and negative frequencies.

4.3.2 Advantages

Zero-IF receivers have several obvious advantages:

- The channel selectivity filter can be fully integrated since it can be implemented as a low-pass filter. Such a filter does not require high-quality or very accurate components, since the components only determine the bandwidth of the filter. The "center" frequency (0Hz!) is correct by construction.
- Interference problems related to IF signals do not occur, since the signal frequencies are very low.
- Only a single LO signal is required, and in the case of DECT, this signal has the same frequency as the transmitter. Usually, it can be derived from the same source as the transmitter signal. In a conventional receiver, this could also be achieved, but it is much more difficult because of the short frequency switching time required of the LO source when switching between send and receive.
- An image rejection filter is unnecessary, although some antenna selectivity might be required in a practical implementation.
- Since it is possible to integrate the receiver completely, it is in principle possible to achieve much better control over crosstalk than with a conventional receiver which requires external components. Moreover, it allows higher impedance levels to be used at high frequencies, reducing the power dissipation.
- The low frequencies in the IF part of the receiver can decrease the power dissipation as well.

4.3.3 Disadvantages

The disadvantages and limitations of a zero-IF receiver are not always apparent at first. They include:

- The dual receiver branches of a zero-IF receiver might consume more power than the single branch of a conventional receiver.
- The accurate 90 degrees phase-shift circuit and well-matched filters can only be implemented easily on an IC. For example, even a wire length difference of just 200 micrometers will introduce a phase error of about 1 degree at 2GHz. Therefore,

integration is not just possible, it is virtually mandatory.

- A fully integrated receiver implemented in a typical silicon IC process cannot apply high-quality inductors or strip lines. This will result in reduced performance and/or increased power dissipation.
- Crosstalk from the LO signal to the RF signal causes loss of sensitivity and interference. Loss of sensitivity occurs because the receiver converts the crosstalk to a DC level at IF. This DC level is in the middle of the IF band. Removing it by filtering will remove part of the received signal as well, and thus reduce sensitivity. Changes in frequency and environment will change the amount of crosstalk and the resulting DC level. Permanent adjustment is therefore not possible, and the lowest cut-off frequency of the filter is limited by the rate at which the crosstalk changes. In a DECT receiver, the quick changes in frequency and environment (body effect) will make this a difficult problem. The choice of an appropriate filter type, optimized for this application, can provide an acceptable solution to this problem. LO to RF crosstalk will cause interference, since the antenna filter and the antenna are by definition designed to effectively interface electrical signals of the LO frequency into radio waves. These radio waves can then interfere with other DECT receivers.
- A zero-IF receiver requires a special demodulator which can derive the modulation signal from two IF signals around zero. Since the baud rate of the IF signal is high compared to its center frequency, zero crossings of the IF signal do not occur frequently enough to derive the modulation signal from them. This also implies that limiting the IF signals is not possible without losing information (appendix E).

4.3.4 Comparison

A zero-IF architecture is an attractive choice for a DECT receiver because of the small size, low weight, few adjustments, simple antenna filter, single simple synthesizer, and relaxed DECT requirements for sensitivity and distortion. Power dissipation could end up comparable to a conventional receiver design because of the combined advantages and disadvantages of zero IF in this respect. This does require that the disadvantages mentioned in the previous section can be overcome by careful design, as will be discussed in the next section.

4.4 Design

The main specifications for this DECT receiver are shown in table 7. The noise figure of 9dB is in line with the low emphasis on sensitivity expected from a short-range system, as is the relaxed linearity specification of -28dBm for the third-order intercept point (IP3). The main challenges are the combination of a high RF frequency of 1.9GHz and the low power dissipation of 90mW. These specifications have been divided into specifications for the RF front end and the IF part of the receiver. This allowed for the concurrent design of these two parts as two separate ICs. Ultimately, these two designs could be combined into a combined RF/IF IC. The specifications for the RF front end are shown in table 8.

Specification parameter	*Value*
Noise Figure	9dB
IP3	-28dBm
Input Sensitivity	-90dBm for BER < 10^{-3}
Power Supply	+2.7V
Power Dissipation	90mW

Table 7 Specifications of the complete DECT Receiver

4.4.1 Front End design

The front end circuit includes a voltage-controlled low noise amplifier (VCA) with automatic gain control circuit (AGC), dual mixers, phase shift circuit, output buffers, and supporting circuits for generating reference voltages and currents, and for powering the circuit down (fig. 27).

The VCO has not been included, since the LO signal is derived from the transmitter. The input LO level has been chosen at -20dBm to achieve sufficiently low LO cross-talk. The wideband AGC circuit has been included to reduce the dynamic range requirements of the subsequent circuits, which reduces the power dissipation of the receiver.

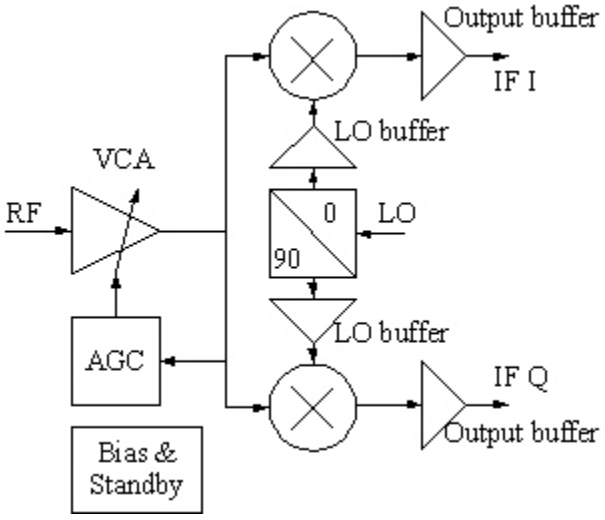


Figure 27 *Block diagram of the DECT RF front end*

A simple mathematical model was used to derive specifications of the subcircuits from the front end specs. This model linked the gain, noise figure, and distortion of cascaded stages to the overall gain, noise figure, and distortion. The model was implemented in a spreadsheet, and the subcircuit parameters were derived by manual iteration and trial and error, using specifications from existing circuits. Various distributions of gain, linearity and noise across the different circuits were tried. In the IC process used, the best choice was based on a low-noise, high-gain VCA. This reduced the noise behavior requirements of the remaining subcircuits. Distortion requirements for the subcircuits became more

severe because of the high gain, but since DECT requirements for third-order distortion are rather moderate, this was not a problem.

Second-order distortion was reduced by implementing the subcircuits with balanced signals. This also reduced noise injection and sensitivity to ground, supply and substrate noise. By means of several iterations, the specifications of the subcircuits as shown in table 9 were arrived at.

Specification parameter	*Value*
Voltage gain:	8dB to 37dB
Noise figure:	6dB
IP3:	-25dBm
IP2:	+14dBm
LO input level:	-20dBm
LO to RF leakage:	-60dBm
Input impedance:	50Ω
Phase matching:	4°
Amplitude matching:	1dB
Power Supply Voltage:	+2.7V
Power Dissipation:	45mW

Table 8 Specifications of the RF front end circuit

Power dissipation for the mixer, VCA, and phase shift circuits was initially set at 25% of the total power budget each. The remaining 25% could then be spent on output buffers and support circuits.

4.4.2 Package

At 1.9GHz, the package parasitics are so significant that they have to be taken into account as an integral part of the design. The package selected for this design is a plastic SO16, a small outline 16-pin package. A bondwire of such a package at 1.9GHz already has an impedance of about 70jΩ, which cannot easily be neglected for 50Ω inputs. Inclusion of the package in impedance matching is therefore mandatory.

The package also plays a significant role in the LO-to-RF crosstalk. Simulations have shown that even an empty package with unbalanced LO and RF signal pins achieves only 20dB of isolation, because it resonates around 2GHz. Since the RF input has to be single-ended to facilitate application of the front end, there was no choice but to make the LO input balanced. In this way, an empty package achieves about 45dB of isolation.

4.4.3 Voltage-controlled low-noise amplifier

The VCA input transistor is a common emitter, biased at 1.5mA, with an input impedance of 50-20jΩ (fig. 28). The capacitive part of this input impedance is more than compensated for by the bondwire of the package. The pin next to the RF input is used for RF ground. Internally, it is connected to the emitter of the input transistor. By connecting the pin through an AC-coupling capacitor to the ground of the antenna filter, the RF currents will run through these neighboring pins and bonding wires instead of running through the power and ground wires across the IC. Adding a decoupling capacitor of the appropriate value to the RF input will tune out the remaining inductive part of the input impedance. The resulting overall input impedance is now 50Ω.

This semi-balanced pin arrangement also provides additional isolation to LO

crosstalk. The gain of the input transistor is controlled by adjusting the collector current by means of an AC-decoupled MOS transistor. The gate of this transistor is controlled by the output of the AGC circuit. The output of this first stage has a source impedance of 500Ω and is still unbalanced. An output stage converts the signal to a balanced signal at a lower impedance. By choosing the resistor values R_c2 and R_e2 the same, and identical to R_c1, the load currents will only flow through the output transistor, and the AC collector current of the input transistor will compensate for the AC collector current of the output transistor. This provides reduced noise injection into the supply rails and the substrate, and also reduces sensitivity of the VCA to noise on the supply and in the substrate. Like the first transistor, the output transistor also runs at 1.5mA, and provides an output impedance of approximately 50Ω differential. The outputs of the VCA are connected to the RF inputs of the mixers. A more detailed diagram of the circuit is shown in fig. 29.

	LNA	*Mixer*	*Output Buffer*
Voltage gain:	20dB	13dB	4dB
Power gain:	20dB	-4dB	8dB
Noise figure:	3.5dB	10dB	18dB
IP3:	-20dBm	220mV	1V
Power dissipation: (design target)	10mW	10mW	10mW
Power dissipation: (realized)	13.5mW	10.5mW	6mW

Table 9 Specifications of the Subcircuits

Between the input transistor and the output stage, an emitter follower has been introduced to reduce the capacitive load of the output stage. The peak detector of the AGC loop is implemented with a differential pair (TNS_1,TNS_2) that compares the collector signal of the input transistor to a DC-shifted and low-pass filtered version of itself. The output of the threshold detector is fed through a series of amplifiers and a time constant (external to this circuit) to the gate of the NMOS transistor which sets the gain of the input transistor.

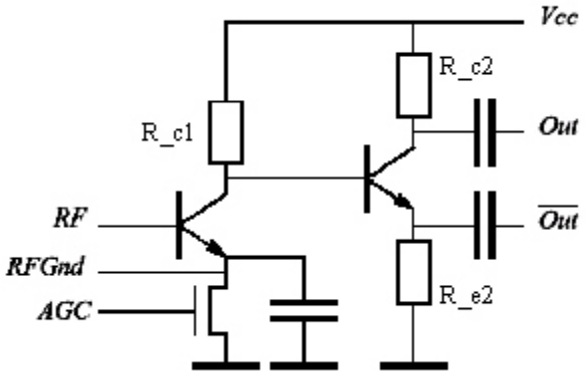


Figure 28 *Schematic diagram of a voltage-controlled low-noise amplifier*

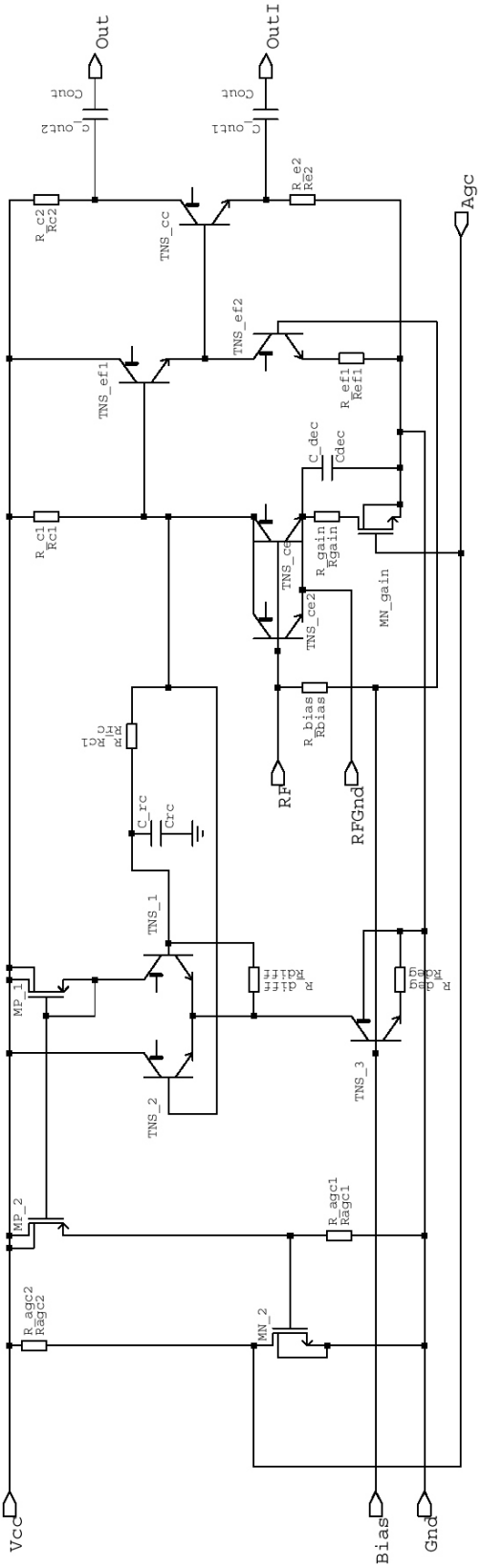


Figure 29 *Detailed diagram of VCA*

4.4.4 Mixers

The mixers are implemented as common 6-transistor double-balanced circuits (fig. 30). This provides good isolation from LO to RF inputs, and eliminates straight feedthrough of the RF signal. The noise behavior of the switching stage of such a mixer is rather complex. The noise contribution of the switching stage can be reduced by increasing switching transistor size until the noise contribution is negligible compared to the input stage. Since the output frequencies are so low, the collector-substrate capacitance of such devices is not likely to be a problem. In a design optimized for low power, however, this might not be the best solution. The input impedance of the switching transistors would decrease when the transistor sizes increase, requiring LO buffers with a lower output impedance. Since the LO buffers need a gain of about 20dB, this would result in a significant increase in power dissipation. Instead, an optimum size was found for noise behavior by fixing the power dissipation of the LO buffers and changing the size of the switching transistors.

The optimum switching transistor size was relatively small, and insufficient for the DC current of the input stage. Therefore, most of this current is routed through resistors directly to the power supply. Additional benefits from the smaller currents include smaller diffusion capacitors of the switching transistors, and higher voltage gain of the mixer because of increased output resistor values.

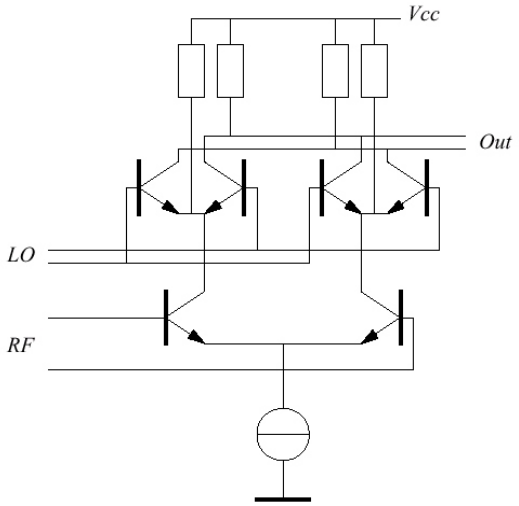


Figure 30 *Mixer schematic diagram*

4.4.5 Phase shift circuit

The phase shift circuit needs to generate two LO signals with a phase accuracy of about 2 degrees. This leaves 2 degrees for mismatches in the mixers and output buffers. Common solutions include frequency dividers and allpass filters [9]. Frequency division would require a frequency doubler on chip, since the 3.8GHz signal is not available from the synthesizer. Generating and dividing such a high frequency would have increased power dissipation significantly. Allpass filters do not need this double frequency, but the required components could not be implemented with sufficient accuracy and quality with

the available technology.

Therefore, a passive RC phase shift circuit was applied that is based on the inherent 90° phase difference between the current through, and the voltage across, a capacitor. By feeding the capacitor current through a resistor, the voltage across the resistor and capacitor will differ by 90° independent of frequency, or resistor and capacitor values. Naturally, the amplitudes of the voltages will change with frequency. The use of limiters can eliminate this problem for a very wide frequency range (fig. 31).

The main error source in this circuit is the capacitive loading of the limiter inputs on the central node of the RC network. The IC technology used for the implementation of this circuit provides poly-silicon resistors with small parasitic capacitances and metal plate capacitors with small series resistance. These parasitics of the RC network are negligible compared to the load of the limiters. The capacitor does have a parasitic capacitance to the substrate, but only from the bottom plate. By implementing the capacitor as a triple-plate structure with the middle plate connected to the central node of the RC network, the parasitic capacitance of this plate to the substrate also becomes negligible compared to the load of the limiters.

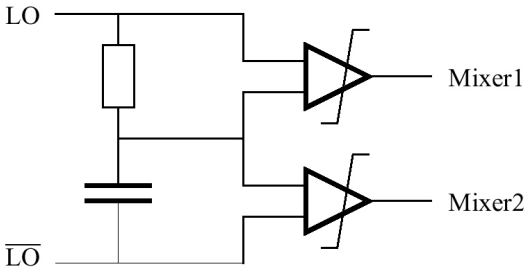


Figure 31 *Principle of phase shift circuit*

The source impedance of the LO signal does not affect the phase error directly. However, a low source impedance allows a low-impedance RC network to be used, which is less sensitive to capacitive loading. This circuit achieves power dissipation levels much lower than similar circuits that have been reported before [92].

A diagram with the implementation details of the phase shift circuit is shown in fig. 33. The RC network is followed by emitter followers to reduce the parasitic load on the central node of the RC network. The error introduced by capacitive loading of the central node is shown in fig. 32. The load of the emitter followers is mainly capacitive and amounts to approximately 20fF. In order to achieve equal delays through the emitter followers, a dummy limiter input circuit has been added across the outer two emitter followers.

Another source of phase errors is introduced by the distortion of the LO signal itself. This error can be analyzed by determining the zero-crossing positions of the output signals of the RC network when a distorted signal is applied to it. Define V_{in} as the LO source voltage with a required frequency component f_1 and a distortion component f_2 ($f_2 = n f_1$):

$$V_{in} = e^{i2\pi f_1 t} + \alpha e^{i(2\pi f_2 t + \phi)} \qquad \textbf{(27)}$$

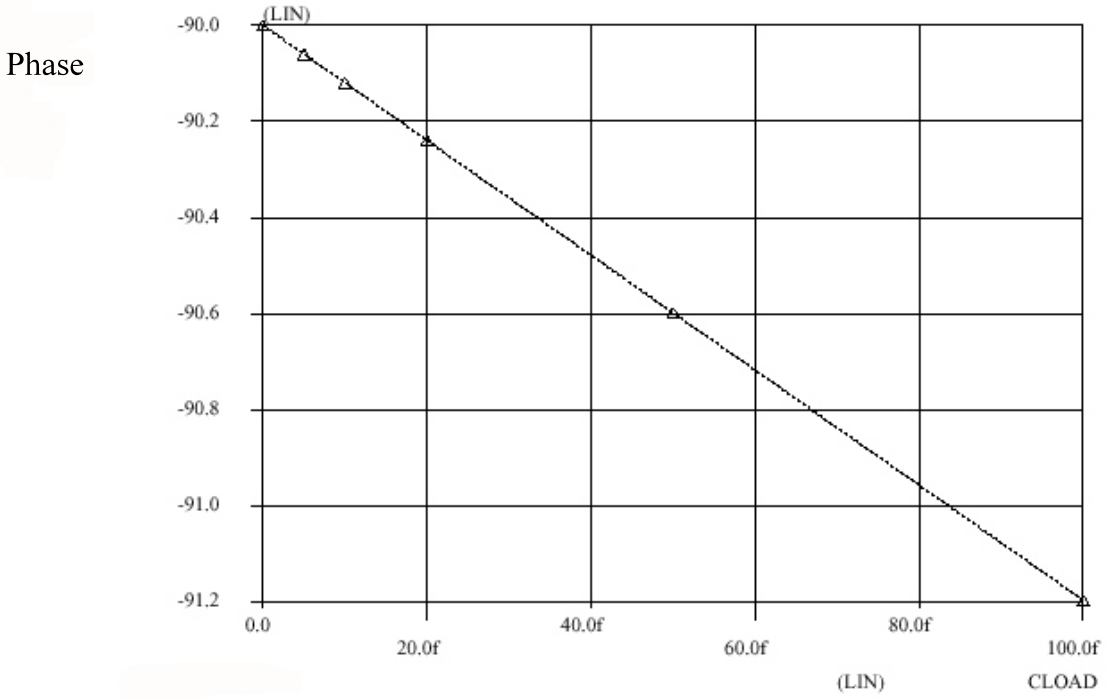


Figure 32 *Phase error introduced by loading*

Now the voltages across the resistor and capacitor can be defined as V_I and V_Q, which can be expressed in the components of the LO signal V_{in} according to the following formulas:

$$V_I = \frac{2\pi i f_1 RC}{1+2\pi i f_1 RC} e^{i2\pi f_1 t} + \frac{2\alpha\pi i f_2 RC}{1+2\pi i f_2 RC} e^{i(2\pi f_2 t+\phi)}$$

$$V_Q = \frac{1}{1+2\pi i f_1 RC} e^{i2\pi f_1 t} + \frac{\alpha}{1+2\pi i f_2 RC} e^{i(2\pi f_2 t+\phi)} \quad \textbf{(28)}$$

The relative positions of the zero crossings of V_I and V_Q define the phase difference between the signals. Solving for this phase difference yields a function $\Phi(\alpha,\varphi)$, which relates the phase difference Φ to the distortion level α and the relative phase φ of the frequency components f_1 and $f_2 = n f_1$. Fig. 34 shows a plot of this function for $f_2 = 3 f_1$.

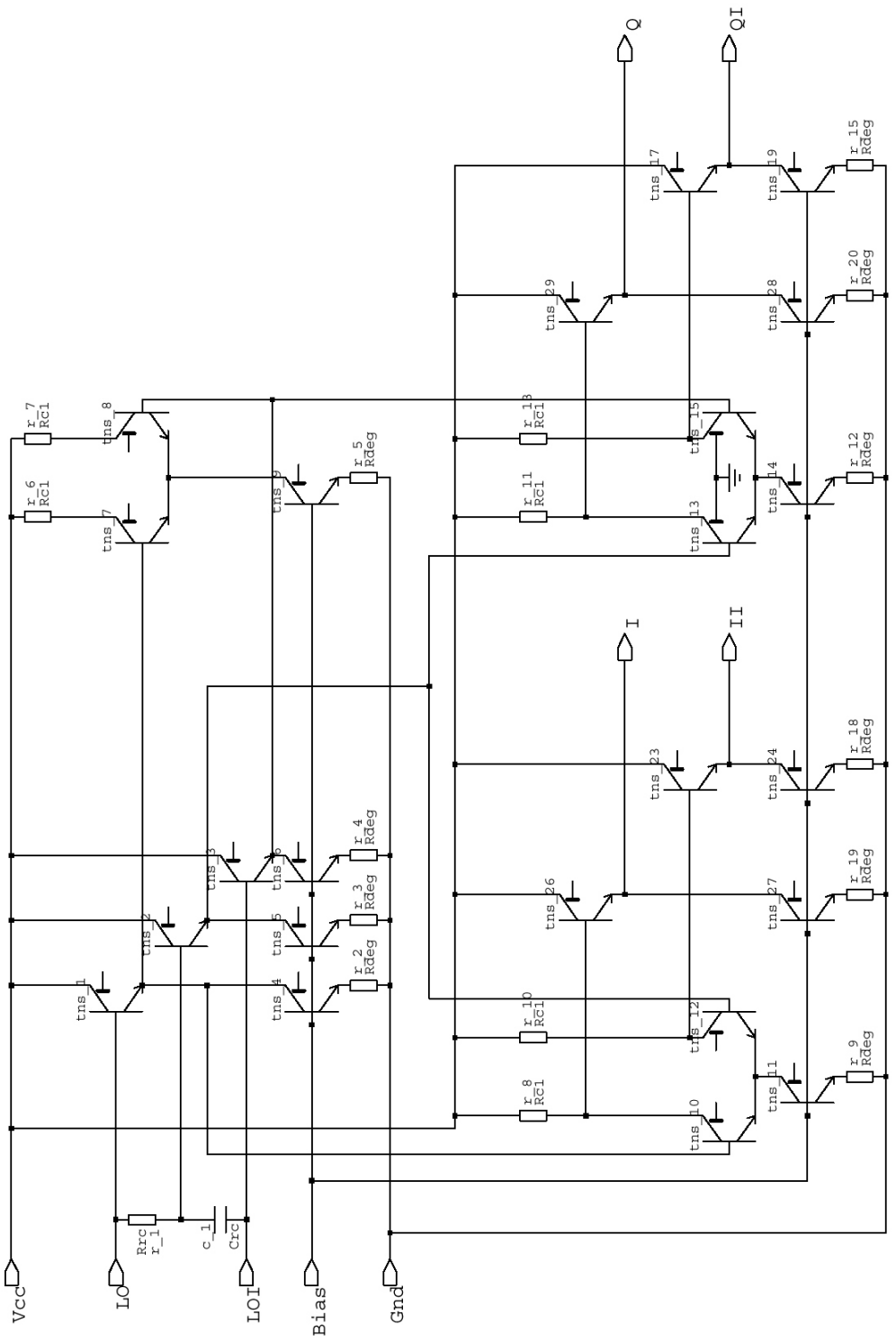


Figure 33 *Detailed diagram of the phase shift circuit*

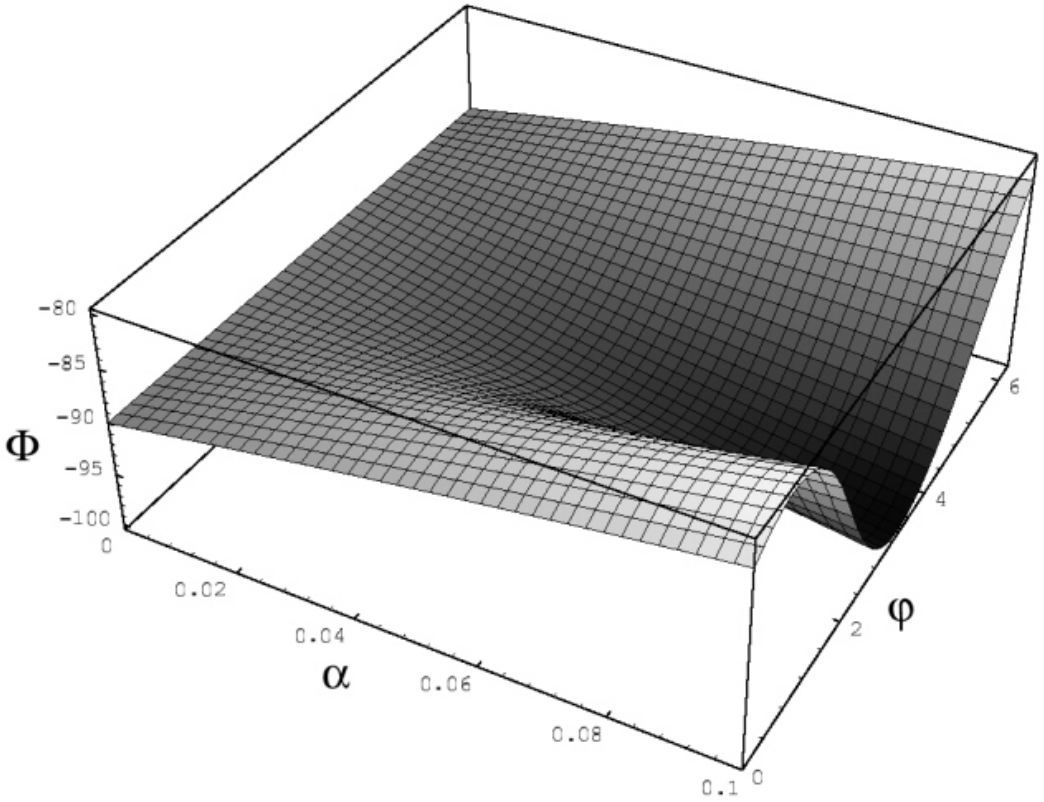


Figure 34 *Phase error introduced by third order LO distortion*

Phase accuracy has been simulated for the actual circuit at less than one degree across a frequency range from 100MHz to 3GHz (fig. 35).

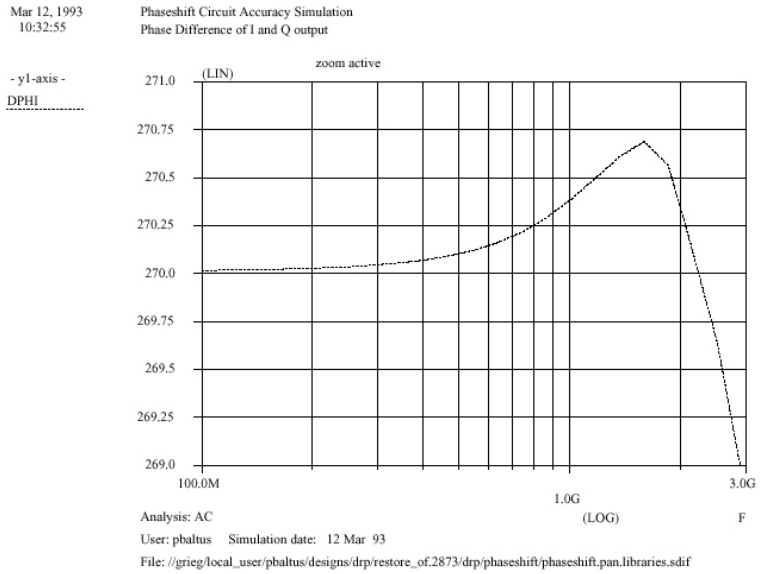


Figure 35 *Simulation results of the phase shift circuit*

4.4.6 IF circuit

A full discussion of the IF circuit (fig. 36) is beyond the scope of this thesis. The suppression of the DC component due to LO crosstalk can be optimized by an appropriate filter type at the inputs of the IF circuit. This is followed by lowpass channel selectivity filters and AGC circuits. A demodulator derives the modulation signal by combining the I and Q signals and their derivatives. The output of the demodulator is passed through a data slicer to obtain well-defined logical levels.

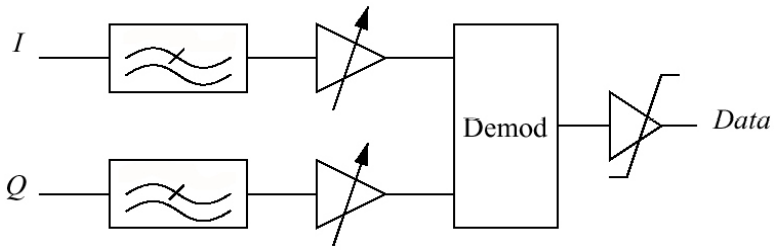


Figure 36 *IF circuit block diagram*

4.5 Implementation and results

The front-end circuit has been implemented in a 1µ BiCMOS process. The main parameters of the minimum bipolar device are shown in table 10. The layout of the circuit has been optimized for symmetry and matching, and minimization of crosstalk, as can be seen in fig. 38.

RF measurements on the wafer are not possible because of the large number of RF and DC signals that have to be applied simultaneously. Also, since the package was an integral part of the design, the performance of the circuit without package would not meet the requirements. Building a set-up which allowed quick replacement of test devices and also provided sufficiently accurate RF behavior proved to be rather involved. LO crosstalk in particular is difficult to measure, since crosstalk across a test board can easily exceed the crosstalk of the device. This was solved by building a measurement set-up as shown in fig. 37. The base of this set-up was made of metal, and had two slopes with the same angle as the pins of an SO16 package. On these slopes, two printed circuit boards (PCB) were mounted, with traces that lined up with the pins of the SO16 package at the top. These traces were designed as 50Ω micro-strip lines, and connected the signals to SMA connectors at the other end of the PCB (not shown). A mechanical clamp (not shown) was used to push the packaged front end securely down on top of the PCBs, ensuring a good contact between package pins and PCB traces. Power supply line decoupling was implemented through CMC capacitors very close to the pins of the IC. This set-up allowed measurements of RF performance at 2GHz with an isolation better than 70dB between the LO and RF traces on the PCBs.

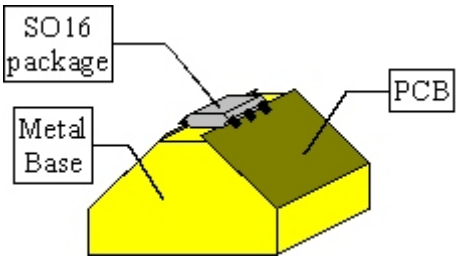


Figure 37 *Measurement set-up*

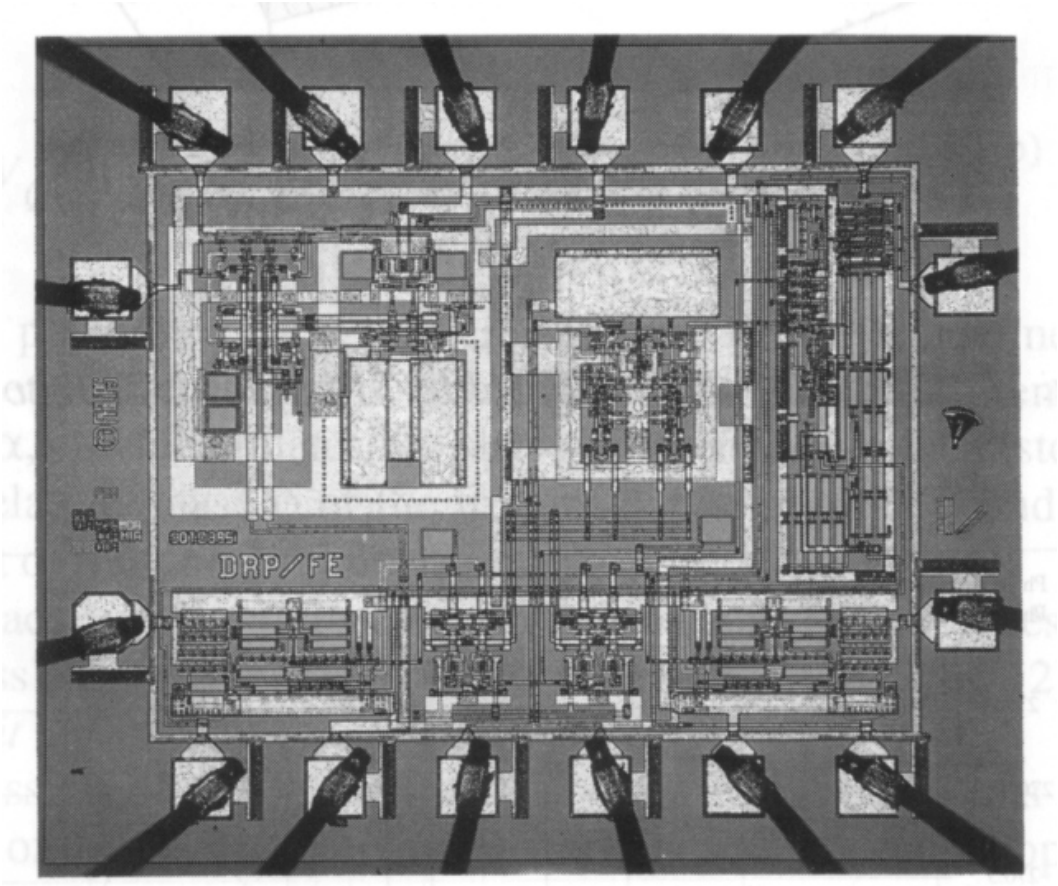


Figure 38 *Die photo of the DECT front end*

Crosstalk measurement results are shown in fig. 39. These measurement results are close to the package model simulation.

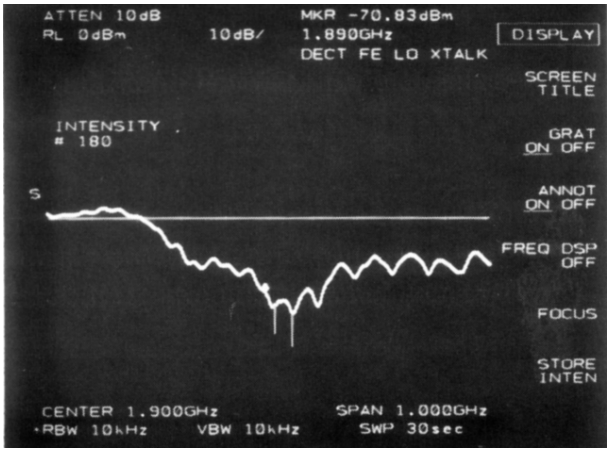


Figure 39 *LO crosstalk measured at the RF input*

This is consistent with the observation that powering down the front end actually increases the crosstalk. The phase accuracy of the phase shift circuit cannot be measured separately for the same reasons that make it impossible to implement the phase shift circuit externally. Therefore, only the phase shift accuracy of the complete front end has been measured, and it is within specifications. In fact, it is within the accuracy limits of the phase meter used for these measurements. Fig. 40 shows the I and Q signals on an oscilloscope, which gives results close to the phase meter readings.

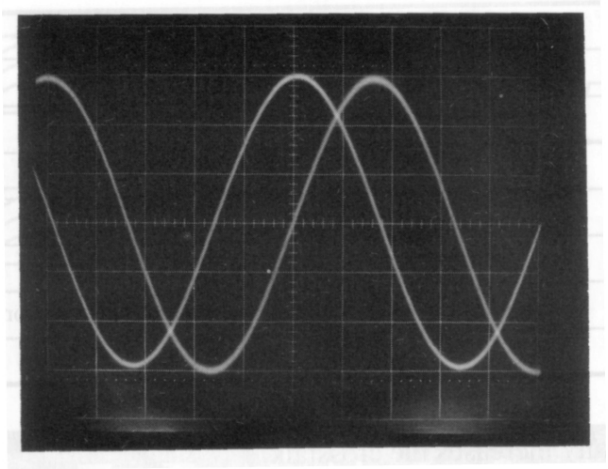


Figure 40 *I and Q output signals of the front end*

Distortion of the circuit is shown in fig. 41 below, and corresponds to an IP3 of –23dBm.

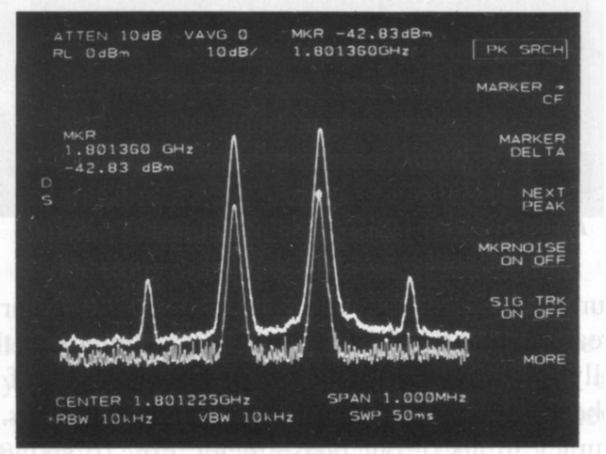


Figure 41 *IP3 measurement results*

The power dissipation for the subcircuits ended up close to the target (fig. 42). The VCA, mixers and phase shift each dissipate close to 25% of the total, and the remaining power goes to the output buffers and support circuits.

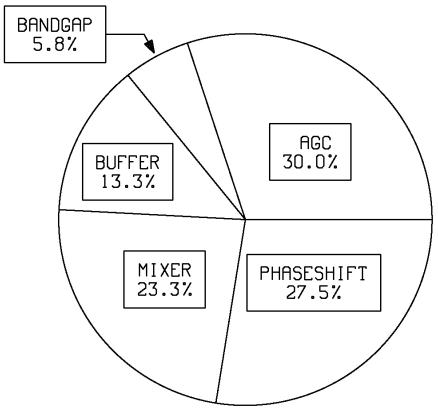


Figure 42 *Power dissipation of the subcircuits*

NPN device:	
emitter area:	$2x1\mu m^2$
Collector-substrate capacitance (Cjs)	30fF (measured at 0V DC bias)
Collector-base capacitance (Cjc)	7fF (measured at 0V DC bias)
Base-emitter capacitance (Cje)	4fF (measured at 0V DC bias)
f_T	13GHz
Resistors	low-capacitance poly resistors
Interconnect	3 metal and 1 poly layer

Table 10 BiCMOS Process Parameters

4.6 Summary and conclusions

Zero-IF receivers offer several advantages for DECT receivers, such as low weight, small size, few external components and adjustments, simple antenna filter, and single simple synthesizer. In order for a zero-IF receiver to be a real alternative to conventional receivers, several disadvantages such as phase accuracy and LO leakage need to be overcome. The receiver front end described in this section demonstrates that it is possible to solve these issues by careful design, using standard design methods. This resulted in a fully integrated receiver front end with a performance and power dissipation comparable to a conventional receiver.

In order to compare the power dissipation of this design to the fundamental limits identified in Section 3.6, we compare the LNA and output buffers to the "Gain" elementary operation, and the Mixer to a combination of the "Gain" and "Frequency Conversion" elementary operations (table 11).

	Gain	*Frequency Conversion*	*Total*	*Physics limit*
LNA at min signal	0.000001%	NA	13.5mW	100pW
LNA at max signal	1%	NA	13.5mW	0.1mW
Mixer at min signal	-0.00000061%	10µW	10.5mW	-50pW
Mixer at max signal	-0.6%	10µW	10.5mW	-60µW
Buffer at min signal	0.0000021%	NA	6mW	212pW
Buffer at max signal	2.11%	NA	6mW	251µW

Table 11 Power dissipation of LNA, mixer and output buffer compared to fundamental limits

The results in table 11 are shown for two conditions: the signal levels at minimum antenna signal, i.e. the sensitivity limit of the receiver, and the signal levels at maximum antenna signal. The reason for including both conditions is that the circuits are using class A type biasing for the active devices. This makes the efficiency very much dependent on the signal levels, as is obvious from the numbers in the table.

The minimum output level of the LNA is -70dBm, derived from the -90dBm sensitivity limit and 20dB of gain. The power dissipation is 13.5mW, resulting in a PAE of:

$$PAE = \frac{P_{out} - P_{in}}{P_{dc}} = \frac{100pW - 1pW}{13.5mW} \approx 0.0000007\,\% \qquad \textbf{(29)}$$

The maximum output level of the LNA is approximately -10dBm, the 1dB compression point. This results in a PAE of:

$$PAE = \frac{P_{out} - P_{in}}{P_{dc}} = \frac{0.1mW - 1\mu W}{13mW} \approx 0.7\,\% \qquad \textbf{(30)}$$

For the mixer, the input signal levels are 20dB higher, and the output signal level is 4dB lower, corresponding to the power gain of the mixer. This results in a PAE at minimum signal levels of:

$$PAE = \frac{P_{out} - P_{in}}{P_{dc}} = \frac{39pW - 100pW}{10.5mW} \approx -0.00000058\,\% \qquad \textbf{(31)}$$

The PAE is negative because the output power is lower than the input power, while the circuit still dissipates additional power from the power supply. The power gain of less than 1 is due to the 2.5kΩ differential output impedance of the mixers. The fundamental limit for the power level of a 1.9GHz LO signal is 10.7pW, based on formula 12 in Section 3.3. This brings the fundamental limit for the power dissipation of the mixer to 10.7pW-61pW=-50.4pW. The actual power level for the LO signal is 10μW, about six orders of magnitude higher.

The maximum input level of the mixer is about -10dBm, and therefore the maximum output level is −14dBm. This results in a PAE of:

$$PAE = \frac{P_{out} - P_{in}}{P_{dc}} = \frac{40\mu W - 100\mu W}{10.5mW} \approx -0.6\,\% \qquad \textbf{(32)}$$

The minimum input level of the output buffer is -74dBm. With 8dB of gain, the minimum output level is -66dBm or 251pW. This results in a PAE of:

$$PAE = \frac{P_{out} - P_{in}}{P_{dc}} = \frac{251pW - 39pW}{10mW} \approx 0.000002\,\% \qquad \textbf{(33)}$$

It is clear that at low signal levels, the main cause for the power dissipation is the class-A biasing of the circuits. At high signal levels, the power dissipation is still one to two orders of magnitude above the fundamental limits. This power dissipation is mainly caused by the high signal frequencies for the LNA and mixer, which require a minimum biasing of the active devices to achieve sufficient gain.

The power dissipation in the output buffer is mainly determined by the requirements for driving IF signals up to 5MHz off-chip. In the original design

specifications, an off-chip parasitic capacitance of 15pF was taken into account. The output current, and therefore the biasing, of the buffer is set by the highest slope of the output signal, in combination with this parasitic capacitance, even though no output power is delivered to this capacitor.

Chapter

5

A minimum-power design method

A new design method is required to design front ends that achieve the required performance with the minimum power dissipation possible, within a set of boundary conditions. Such a design method will be called a minimum power design method in the context of this thesis. In this chapter, such a new design method will be developed.

The method is based on circuit transforms that can be applied to any circuit. Such transforms are called “structure-independent transforms”, or SITs, and consist of a mathematical transformation of the specifications of a circuit, and a corresponding transformation of the circuit that achieves such a transformation of specifications. Based on a collection of such SITs, two orthogonal SITs (or OSITs) are derived that are used as the basis for a minimum power design method. In this design method, the signal path of a front end is considered as a cascade of signal processing circuits, such as LNAs, mixers, filters, etc. Using the OSITs, the specifications, including the power dissipation, of the individual circuit blocks can be changed, resulting in different overall specifications and overall power dissipation. Any given overall specification of the front end, if possible at all, can be achieved by an infinite number of combinations of transformed signal processing circuits. There is only one combination of transformed signal processing circuits that results in the minimum overall power dissipation, and a corresponding unique distribution of gain, linearity and noise specifications across the subcircuits. This optimum distribution of power dissipation, gain, linearity and noise has been found and can be represented as a number of analytical expressions. Based on the outcome of these expressions, the optimum transforms can be selected for the individual subcircuits.

A practical minimum design method also needs selection criteria for a library of initial signal processing circuits that will be used in an RF front-end design. A selection criterium for these circuits is a special figure of merit (or FOM), that transforms the specifications of a circuit into a one-dimensional performance criterium. The circuits with

the best figure of merit can then be used as the basis for a front end with an optimum distribution of gain, linearity, noise, and power dissipation, as described above. A special class of figures of merit, called equivalent figure of merit, or EFOM, will be introduced in Section 5.4. This EFOM is based on OSITs, which allow the specifications of any circuit with a specific EFOM value to be transformed to exactly the specifications of another circuit with the same EFOM value.

As a demonstration of the potential of this approach, the power dissipation of a DECT front end with transformed circuits is calculated, based on the specifications and circuits of the DECT front end from the previous chapter.

5.1 Design flow

The goal of RF front-end design is the development of a physical implementation that, in combination with other designated system components, can meet the specifications for a target system.

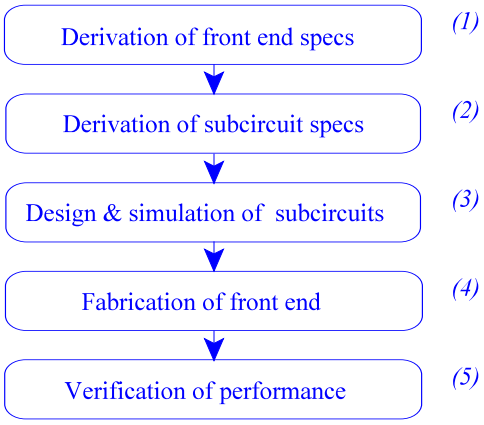


Figure 43 *Front end design process*

A typical front end design process consists of five consecutive stages, as shown in fig. 43.

Derivation of front-end specs

The transceiver specs are derived from the system specs (fig. 44). They can be found by taking into account the antenna specifications and the IF&baseband specifications. If the choice of the antenna and/or IF&baseband circuit is not yet fixed, all possible combinations can be investigated. The results of this investigation can be used to support a decision for a specific antenna and/or IF&baseband circuit. If the front end has to work with multiple antenna and/or IF&baseband circuits, the same method can be applied, but now the worst-case combination of all specification parameters can be used for the front end specification. Either way, the resulting specification is unique, and no iterations (other than for multiple antenna and IF&baseband circuits) are required within the first stage. Still, the first stage does suffer from two shortcomings that affect design productivity:

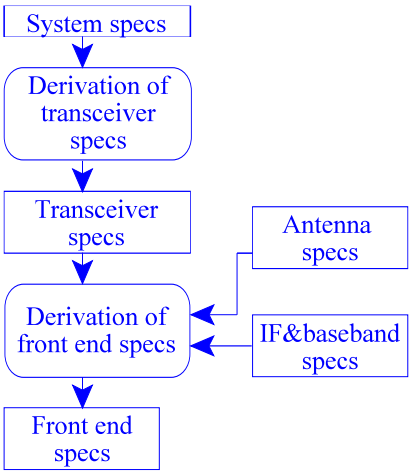


Figure 44 *Stage 1 in the front end design process*

First, if people in charge of the design process change their mind about the project targets, e.g. because of new market inputs, or because of changes in interface specifications, the design process typically needs to be restarted from stage 1. If the process has already advanced to a later stage, this results in a significant delay in the project, and reduction in productivity.

Second, the information required to derive front end specifications is not always readily available in a format that is easy to access. Also, knowledge of, and experience with, the procedure for deriving front end specs from system specifications might be insufficient.

The first shortcoming depends on many external factors. It can be reduced by formal risk assessments such as "failure mode and effect" analysis, by better project management methods, and by padding specifications to allow for some changes during the project. The second shortcoming could be minimized by storing the knowledge of relevant systems in such a way that it is easy to access for the incidental user, and by providing descriptions or even automation of the spec derivation procedure. Systems knowledge is available through e.g. ETSI CD-ROMs, books, internet sites etc. The information from these sources is seldom in a consistent format, however. In practice, it takes a lot of experience or an even larger amount of patience to understand the system requirements based on these sources. Partial derivation procedures are available through many home-brewn spreadsheets and simple software tools. These are seldom integrated into the design process, and documentation is often minimal, requiring again experience or patience from the designer who wants to use such tools.

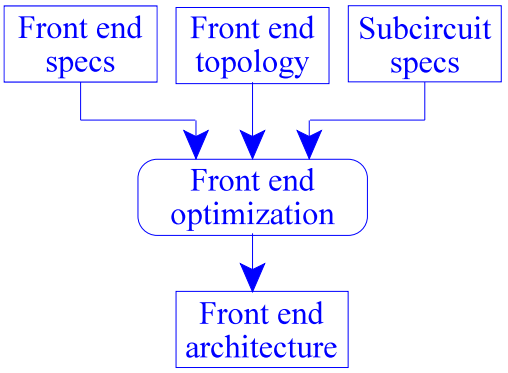


Figure 45 *Stage 2 of the front end design process*

Derivation of subcircuit specs

This is essentially an optimization problem with some fixed boundary conditions (typically specs such as sensitivity and power supply voltage) and optimization criteria (typically cost, power dissipation, etc.), as shown in fig. 45. It is therefore the same type of derivation as the first step, but at a lower abstraction level. Often, this also includes a difference in span of control. In many front end developments the standards at the system level cannot be changed. The front end product specification is often open for some debate.

If, for each of the subcircuits, the trade-offs between specs for the fixed boundary conditions and the optimization criteria are known, then the problem can be solved through mathematical analysis. However, the trade-offs for the subcircuits depend on circuit topologies and parameter values, of which an infinite number of combinations exist. Therefore, a formal optimization is impossible. Attempts have been made to generate all possible circuit topologies [73] and to automatically find optimum parameter values [65], but they cannot cope with the typical complexity of front-end subcircuits in realistic periods of time, so another approach is required.

Design, fabrication and verification

These are steps that, at an abstract level, are straightforward once stage 2 has been completed, since in this stage the desired specifications of the sub blocks are known. However, in practice, the performance of a subcircuit depends significantly on its environment because of effects like substrate noise, packaging parasitics, crosstalk, etc. This is the main reason why a full top-down approach to RF front end design is not feasible. It is not possible to cope with the level of complexity inherent in all significant effects at the higher levels in a top-down process, and therefore abstraction does not work. Until this problem is solved, there is no alternative for experience and detailed insight.For this same reason, it does not make sense to spend a lot of effort on detailed and highly accurate simulations and optimizations at a high abstraction level of an RF front end design.

An optimal design procedure is both efficient (i.e. requires the minimum amount of resources) and effective (i.e. the result is as close as possible to a predefined goal. A *minimum power* optimal design procedure should achieve both goals, and the resulting

front end should achieve the lowest possible power dissipation within a given set of boundary conditions. Typical design procedures do not meet the requirement for efficiency, because of the many iterations that are part of such a design method. Figure 46 shows the design procedure of fig. 43 with the most common iteration loops. The circuit design block has been split into the sub blocks “circuit design” and “simulation”, in the shaded area in the center of the figure, to emphasize the inner loop of the design process. For an optimum design procedure, it is necessary to eliminate these iterations as much as possible. The iterations consume time, and therefore reduce the efficiency of the design procedure. Also, it is not possible to find the best front-end and circuit topology from an infinite number of topologies with a finite number of iterations. Therefore, in practice, the effectiveness of this procedure is also less than optimum.

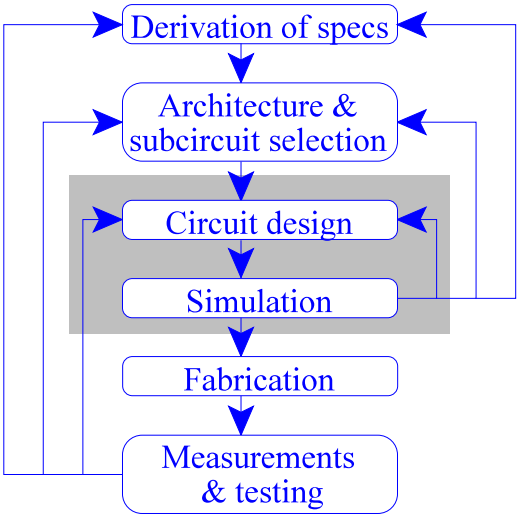


Figure 46 *Typical design procedure with iteration loops*

5.2 A SIT and EFOM based design procedure

It would be useful if the first steps in a top-down design method could be carried out more efficiently. This would result in a selection of front end topologies, subcircuit specifications and subcircuit topologies that, within the limits of the abstraction, define the global optimum solution for power dissipation. Because of the limits of the abstractions used, this global optimum will not be highly accurate, but it will be the best starting point for a front end implementation. This implies that some iterations in the implementation stages might still be required, but with the fundamental limits provided in this thesis, the margins are now well understood.

This can be achieved by defining a library with a finite number of circuit and front-

end topologies, together with a procedure that can transform these circuits into circuits with the desired specifications. By adding an extra boundary condition to the problem, requiring the use of front end and circuit topologies from this library, the problem becomes finite and therefore solvable. This only works if the "best" topologies are included in the library. Therefore, a criterium is needed to decide which topologies to include. Such a criterium can be based on an EFOM (Section 5.4). Once a library with a finite number of the "best" topologies exists, the trade-offs between the specifications (including the power dissipation) of each subcircuit are known. The performance of the total front end can now be formulated mathematically as a function of the specifications of each of the subcircuits, in combination with the front end topology. By applying all predefined boundary conditions on this function, a limited set of subcircuit specifications can be derived that meet these boundary conditions. To find the unique, "optimum", set of subcircuit specifications, an additional cost function is defined. For minimum-power front ends, this cost function is power dissipation. If there is a way to solve the subcircuit specifications within the limits of the boundary conditions, and for the minimum of the cost function, it is possible to find, for each of the subcircuits, the specifications that will result in the lowest possible total power dissipation. This will result in the optimum distribution of gain, noise figure, and linearity, for each of the subcircuits.

Once the optimum distribution is known, the subcircuits needed to achieve this distribution can be derived from the subcircuits in the library by applying the OSITs on which the EFOM is based to each of the selected subcircuits. The front end can then be assembled from these transformed subcircuits. This eliminates the need for any iterations, since the required performance will be achieved "by construction". Therefore, this procedure will eliminate all iterations from the design procedure in figure 46, and will result in the required front end performance for the lowest possible power dissipation: maximum effectiveness in combination with maximum efficiency (fig. 47).

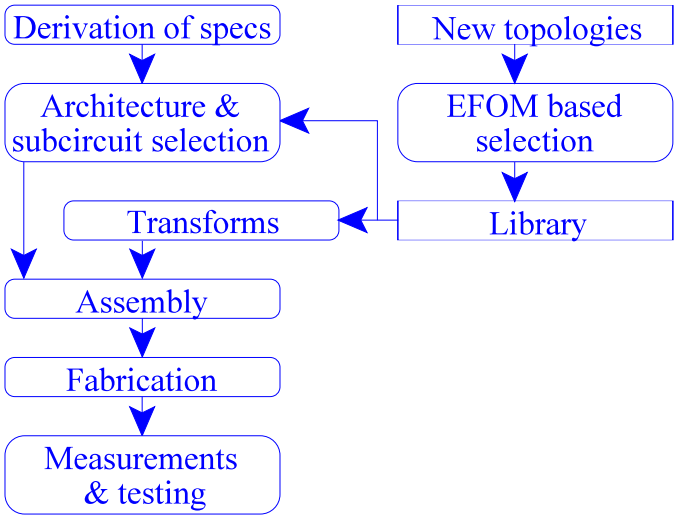


Figure 47 *Optimal design procedure*

To implement this procedure, it is necessary to define an EFOM, and the proper transforms to modify any circuit into another circuit with different specifications but with the same EFOM value. It is also necessary to find a solution for the values of the

subcircuit specifications that achieve the minimum of the cost function while meeting all boundary conditions. Finally, it is necessary to formulate the steps of the design procedure in such a way that they can be carried out within a very short period of time, e.g. by automating the individual steps.

Naturally, such a design procedure is limited by the accuracy of the models on which the EFOM is based, and by the parameters taken into account in the performance definition and transforms. All parasitic effects that are not included in these models, and all parameters lacking from the performance definition and transforms, will affect the result of the design procedure. As a result, additional work is required to arrive at an optimum front end. Nevertheless, by including the most relevant parameters in the performance definition, and the most critical effects in the models, a good first approximation of optimum subcircuit specifications will be obtained, as well as a clear method with which these specifications can be achieved through transforms of circuit topologies from the library. Therefore, this design procedure will result in a first design that is close to the global optimum. Such a design can be turned into a fully optimized circuit with relatively little effort. The design procedure will also provide a quick indication of the power dissipation that is ultimately achievable for a given set of performance requirements and boundary conditions.

5.3 Structure independent transforms (SITs)

The number of circuit topologies is so large, that it is impractical to take all possible topologies into account. For the same reason, it is not practical to consider circuit transformations that are specific for a circuit topology. Practical transforms should therefore be independent of the circuit topology or structure, in other words, they should be *Structure Independent Transforms*, or SITs. In order to find such transforms, the model from Section 3.1 will be used. In this model, the signal processing in the RF front end is implemented using a cascade of two-port signal-processing stages, that carry out one or more of the elementary operations of the RF front end. Some blocks, such as mixers, have in reality more than two ports. They can nevertheless be modeled as a two port, as long as the signals on the other ports are well-defined (e.g. the local oscillator signal in a mixer). This approach will be used throughout this section.

To assess the impact of transforms on the performance of a signal-processing stage, some assumptions about the main performance parameters are required. The main parameters in RF front ends are gain, noise and distortion, as discussed in appendix A. Since the focus of this thesis is on minimum power design, the power dissipation is, of course, an additional relevant parameter. To assess the impact of cascading the circuits, the input and output impedances are relevant as well. Therefore, the design space that will be considered for minimum power design methods has the following dimensions: gain (G) noise factor (F), linearity ($IP3$), power dissipation (P_{dc}), input impedance (Z_{in}) and output impedance (Z_{out}). Since most telecommunication systems are relatively narrow-band, all these parameters are defined at a single signal frequency.

5.3.1 A set of structure independent transforms

In this section, signal-processing stages are modeled as imperfect building blocks with a power gain G_p, noise factor F, third-order intercept point $IP3$, input impedance Z_{in}, output impedance Z_{out} and power dissipation P_{dc}. Using this model, a collection of SITs will be discussed. All of them are known transforms. What is new is the formal description as a transform, and the matching mathematical description of the transformation of the specifications. This allows the transforms to be used in a generic way, and therefore become the basis of a general design method. This collection of SITs is by no means exhaustive. Fortunately, completeness is not required for a minimum power design method. The SITs will be used to derive the orthogonal transforms needed to adjust each of the relevant performance parameters. As will be shown in the next section, this collection provides a more than sufficient basis for the required orthogonal transforms.

To simplify the expressions, it is assumed that all source and output impedances are, initially, equal. This is not essential for the transforms. If required, different impedance levels can be modeled by putting an additional impedance transformer at the input or the output of a building block.

Impedance transform

In narrow-band systems, a low loss transformer with an impedance transformation ratio α can often be implemented cheaply at the PCB level, and sometimes (depending on the IC technology) also on chip. When put in front of our building block, an impedance transform with a factor α changes the input impedance, but leaves all other performance

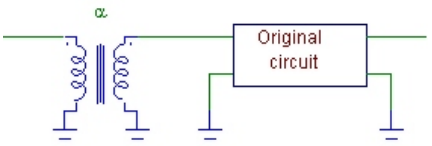


Figure 48 *Impedance transform*

parameters unchanged. However, this transform is a very useful part of other SITs. The corresponding transform of the specifications is shown in (34).

$$\begin{aligned} Z_{in} &\rightarrow \alpha\ Z_{in} \\ Z_{out} &\rightarrow Z_{out} \\ G_p &\rightarrow G_p \\ F &\rightarrow F \\ IP3 &\rightarrow IP3 \\ IP2 &\rightarrow IP2 \\ P_{dc} &\rightarrow P_{dc} \end{aligned} \qquad \textbf{(34)}$$

Parallel transform

The parallel transform puts two identical circuit blocks in parallel. These blocks are shown as amplifier symbols in fig. 49. As a result of this transform, the input and output impedances are reduced by a factor of 2, the gain and noise figure remain unchanged, and the IP3 and power dissipation are increased by a factor of 2. The input and output impedances can be restored by applying impedance transforms with α=2 at the input and α=0.5 at the output. At the circuit level, this transform is equivalent to doubling the W/L ratio of lateral devices and the active area of vertical devices, not taking into account imperfect scaling. As a result, the node voltages in the circuit will remain the same, whereas the branch currents will double, and the impedance levels will go down by a factor of 2. This transform can in principle be carried out by any factor, not just 2. Please note that the transform with a factor smaller than 1 is not always possible for technologies with vertical devices, since halving the W/L ratio of lateral devices is in principle possible, but halving the area of vertical devices is limited by lithography and possibly other boundary conditions in IC fabrication. Obviously, a process that allows accurate

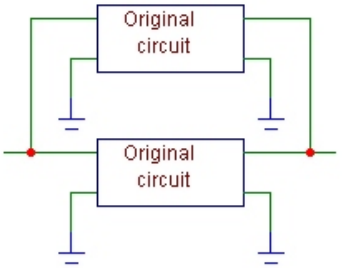


Figure 49 *Parallel transform*

scaling down to very small currents offers more flexibility with respect to this transform. The corresponding transform of the specifications is shown in (35).

$$\begin{aligned} Z_{in} &\rightarrow \frac{Z_{in}}{2} \\ Z_{out} &\rightarrow \frac{Z_{out}}{2} \\ G_p &\rightarrow G_p \\ F &\rightarrow F \\ IP3 &\rightarrow IP3 \\ IP2 &\rightarrow IP2 \\ P_{dc} &\rightarrow 2\ P_{dc} \end{aligned} \qquad \textbf{(35)}$$

Cascade transform

When cascading blocks, the gain increases exponentially as a function of power dissipation, which seems very attractive. However, both the noise figure and the IP3 deteriorate compared to a single block. If the gain is high compared to the noise factor, then the total noise figure will be only slightly higher than the noise figure of a building block. However, in this case the intercept point will decrease approximately proportionally to the increase of the gain. The corresponding transform of the specifications is shown in (36).

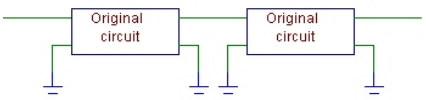


Figure 50 *Cascade transform*

$$
\begin{aligned}
Z_{in} &\rightarrow Z_{in} \\
Z_{out} &\rightarrow Z_{out} \\
G_p &\rightarrow G_p^2 \\
F &\rightarrow F + \frac{F-1}{G_p} \\
IP3 &\rightarrow \frac{IP3}{G_p+1} \\
IP2 &\rightarrow \frac{IP2}{G_p+1} \\
P_{dc} &\rightarrow 2\ P_{dc}
\end{aligned}
\qquad \textbf{(36)}
$$

Parallel-cascade hybrid transform

By combining the cascade and parallel transforms, it is possible to approximate a transform that increases the gain while keeping noise figure and IP3 constant. This is achieved by putting G_p+1 groups of two cascaded building blocks in parallel. For gains much larger than 1, the power dissipation increases proportionally to 2 times the gain

increase. For building blocks with lower gain, this effect is even worse. Obviously, gain is expensive. If the gain is much larger than the noise factor, the noise factor does not change significantly. It would perhaps be more obvious to build this hybrid transform by cascading a first stage, consisting of two parallel building blocks, and a second stage,

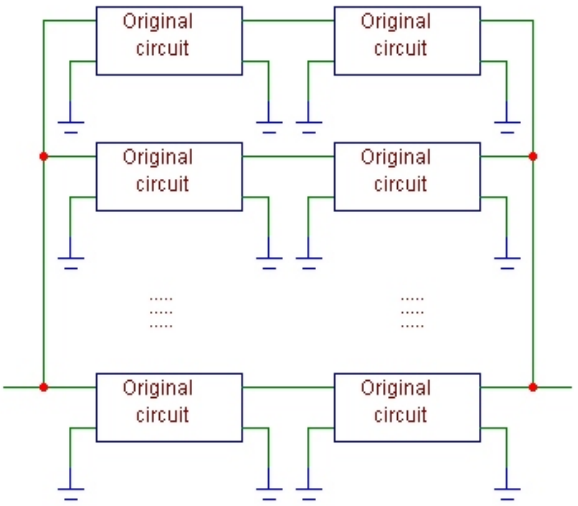


Figure 51 *Parallel/cascade hybrid transform*

consisting of *2* G_p building blocks in parallel. The stages would then be coupled through an impedance transform with $\alpha = Gp$. This is more consistent with the common practice in receivers, in which the IP3 of subsequent stages is increased about proportionally to the gain accumulated up to that point. This results in the same gain vs. power dissipation trade-off, and therefore the same SIT. The corresponding transform of the specifications is shown in (37).

$$\begin{aligned} Z_{in} &\rightarrow \frac{Z_{in}}{G_p+1} \\ Z_{out} &\rightarrow \frac{Z_{out}}{G_p+1} \\ G_p &\rightarrow G_p^2 \\ F &\rightarrow F + \frac{F-1}{G_p} \\ IP3 &\rightarrow IP3 \\ IP2 &\rightarrow IP2 \\ P_{dc} &\rightarrow 2\,(G_p+1)\,P_{dc} \end{aligned} \qquad \textbf{(37)}$$

Attenuation transform

Attenuation is typically easy to implement both on and off chip. It is assumed that the attenuator itself does not introduce additional noise. The dynamic range of the total circuit is not significantly affected, but the gain is reduced, and the noise figure is increased by the same factor as the attenuation. Note that the inverse transform, adding gain in front of the original circuit, is in principle also a valid SIT. However, it is more difficult to implement, and the performance is much more technology dependent. Therefore, it is not as convenient a SIT as the attenuation transform. The corresponding transform of the specification parameters is shown in (38).

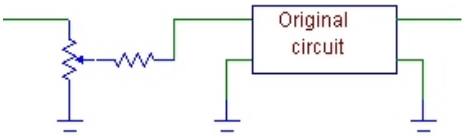


Figure 52 *Attenuation transform*

$$
\begin{aligned}
Z_{in} &\rightarrow Z_{in} \\
Z_{out} &\rightarrow Z_{out} \\
G_p &\rightarrow \alpha\, G_p \\
F &\rightarrow 1 + \frac{F-1}{\alpha} \\
IP3 &\rightarrow \frac{IP3}{\alpha} \\
IP2 &\rightarrow \frac{IP2}{\alpha} \\
P_{dc} &\rightarrow P_{dc}
\end{aligned}
\qquad \textbf{(38)}
$$

Mismatch transform

Mismatch is different from attenuation, in the sense that the excess input power is not dissipated but reflected. This is true both for signal and noise power. Since the excess input power is not dissipated, no excess noise is generated and the noise figure is not affected. The corresponding transform of the specification parameters is shown in (39). Mismatch is also different from the other transforms in this section in the sense that it changes the matching between circuits. All other transforms are based on the assumption that the match between source, load, and circuit block will not be changed. In most on-chip RF circuits, there will be significant mismatches between circuit blocks, both because matching circuits are expensive to build on-chip, and because this increases the robustness of the IC performance with respect to process and environmental changes. The mismatch transform is the only transform that changes this (mis)match. It is attractive

because it offers a trade-off between gain and linearity without affecting the noise behavior of a circuit.

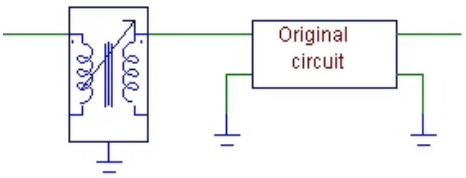


Figure 53 *Mismatch*

$$
\begin{aligned}
Z_{in} &\rightarrow \frac{1+s_{11}}{1-s_{11}} Z_{in} \\
Z_{out} &\rightarrow Z_{out} \\
G_p &\rightarrow s_{11}\, G_p \\
F &\rightarrow F \\
IP3 &\rightarrow \frac{IP3}{s_{11}} \\
IP2 &\rightarrow \frac{IP2}{s_{11}} \\
P_{dc} &\rightarrow P_{dc}
\end{aligned}
\qquad \textbf{(39)}
$$

Distortion cancellation transform

A very interesting class of transforms is the class of distortion cancellation transforms. These transforms are based on the principle that the ratio between distortion component and signal component is generally dependent on the signal level. By processing (at least two) different levels of the same signal through identical circuits, it is possible to obtain two distorted signals with different signal-to-distortion ratios. By adding these two distorted signals with appropriate gain factors, it is possible to cancel a distortion component while preserving (part of) the signal.

In principle, a perfect cancellation can be obtained for a single distortion component. In practice, the reduction is limited by the matching that can be achieved between the branches. By recursively applying this procedure, all distortion components could be eliminated. Obviously, this is impractical for more than two or three components because of the cost and power dissipation of the circuits involved. Fig. 54 shows a distortion cancellation diagram with two branches.

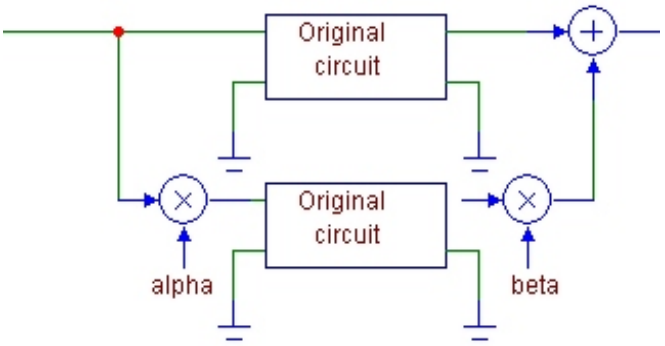


Figure 54 *Distortion cancellation*

A widely used distortion cancellation transform is balancing, which can be used to cancel second order components. In this case, alpha=1 and beta=-1. Also, third order cancellation is sometimes used, e.g. in "multi-tanh" circuits. In the case of two branches, the coefficients should be chosen such that beta=-1/alpha^3 (in order to cancel the third order component) and beta≠-1/alpha (otherwise the wanted signal would be suppressed).

Distortion cancellation transforms are special in that they trade accuracy for linearity. Accuracy is required both in terms of matching between the elements in the two branches and absolute accuracy of the coefficients alpha and beta. There is no real trade-off between distortion cancellation and power dissipation. In general, there can be an impact on the gain and noise figure. For the second-order distortion cancellation there is no such penalty, since there is no reduction of the signal level or signal-to-noise ratio. For the third-order distortion cancellation, there is a finite penalty in terms of the gain and noise figure, but this can be made arbitrarily (and infinitesimally) small by choosing large values of alpha. A large value of alpha will cause the third order distortion component to become relatively large compared to the wanted signal as well as to the noise. Only a small fraction of the output, corresponding to a small beta, is required to cancel the distortion component, and this small version of beta will result in small effects on the noise and wanted signal levels.

Therefore, there is no significant trade-off between distortion cancellation and gain, noise figure or power dissipation for second- and third-order distortion cancellation. This implies that distortion cancellation always yields improvement independent of other transforms applied to the circuit, and is only limited by practical considerations. The transform of the specification parameters for the second order distortion cancellation transform is shown in (40), and for the "infinitesimal penalty" case of the third order distortion cancellation transform in (41).

$$\begin{aligned} Z_{in} &\rightarrow 2\,Z_{in} \\ Z_{out} &\rightarrow 2\,Z_{out} \\ G_p &\rightarrow G_p \\ F &\rightarrow F \\ IP3 &\rightarrow 2\,IP3 \\ IP2 &\rightarrow \infty \\ P_{dc} &\rightarrow 2\;P_{dc} \end{aligned} \qquad \textbf{(40)}$$

$$\begin{aligned} Z_{in} &\rightarrow Z_{in} \\ Z_{out} &\rightarrow Z_{out} \\ G_p &\rightarrow G_p \\ F &\rightarrow F \\ IP3 &\rightarrow \infty \\ IP2 &\rightarrow IP2 \\ P_{dc} &\rightarrow P_{dc} \end{aligned} \qquad \textbf{(41)}$$

Feedback transform

Feedback is again a special transform, in the sense that it is not obvious to implement for processing stages that include the frequency conversion elementary operation. This would require an additional frequency conversion elementary operation in the feedback path, and this additional frequency conversion will itself have the same impairments, such as distortion and noise generation, that feedback is supposed to reduce. Therefore, feedback cannot be considered a structure independent transform in the strict sense.

Even for signal-processing stages that do not include the frequency conversion operation, implementing feedback while maintaining the power match is not trivial, and merits a separate investigation.

5.3.2 Orthogonal SITs (OSITs)

The transforms described in the previous section are somewhat arbitrarily selected from the infinite number of possible transforms. Most of the transforms also affect multiple parameters simultaneously, but not independently. For an efficient minimum power design method, a minimum set of *optimum*, *orthogonal* transforms is required.

- *Optimum* transforms are defined here as transforms that achieve any desired performance with the lowest power dissipation for any performance of the original circuit.
- *Orthogonal* transforms are defined here as transforms that have each at least one specification parameter that is invariant through the transform, and this parameter is different for all orthogonal transforms. The transforms are called orthogonal because the vectors in design space corresponding to the invariant parameter of any

orthogonal transform is orthogonal to all other invariant vectors of other orthogonal transforms.

Orthogonal structure-independent transforms (OSITs) for power, gain and linearity can be found by combining the parallel transform and the mismatch transform. This results in two orthogonal transforms:

1. The OSIT_1 transform, consisting of a parallel transform with a parameter "mult" indicating the (fractional) number of stages that should be connected in parallel, preceded and followed by impedance transforms with parameters "1/mult" and "mult", respectively. This transform only affects power dissipation and IP3, without affecting gain and noise figure. It is shown in figure 55.

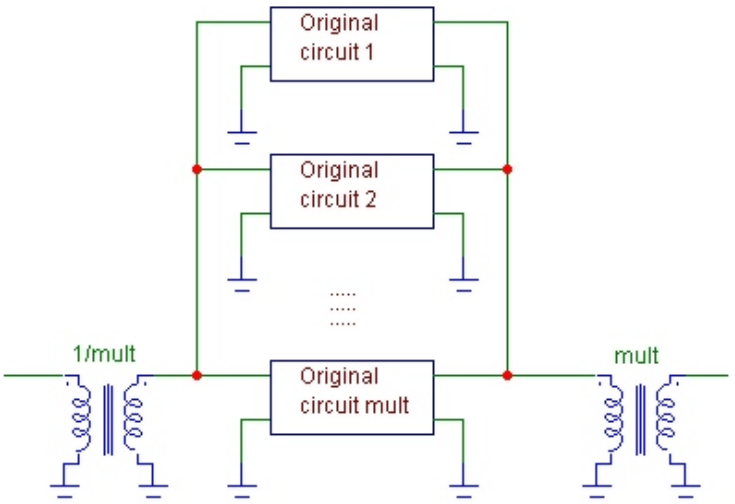


***Figure 55** OSIT_1 transform*

2. The OSIT_2 transform, a combination of the OSIT_1 transform and mismatch with parameter "s_{11}", as shown in fig. 56, under the condition s_{11}=mult. This affects gain and power dissipation without affecting noise figure and IP3. The transformation of transformer K3 corresponds to $\frac{1-s_{11}}{1+s_{11}}$, and is therefore responsible for the mismatch with factor s_{11}.

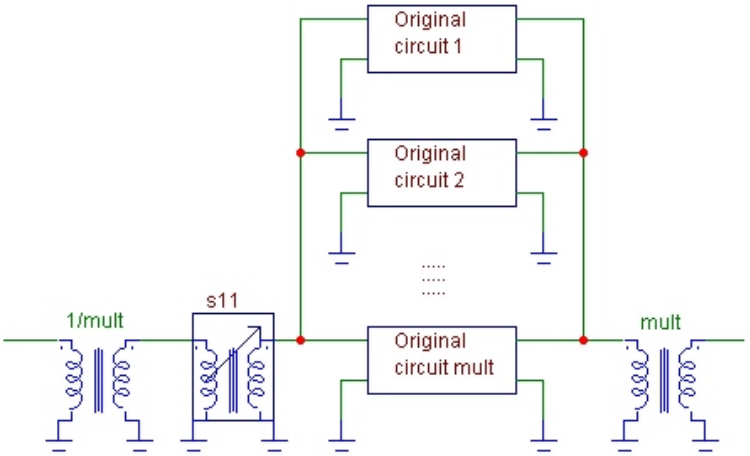


Figure 56 *OSIT_2 transform*

These transforms show that for every circuit, the gain or the IP3 can be increased independently, at the cost of a proportional increase in power dissipation. The transforms are defined in such way, that they do not affect the source and output impedance of the blocks. This allows the OSITs to he applied to subcircuits in a cascade without affecting the other subcircuits.

The orthogonality of the OSIT_1 and OSIT_2 transforms does not by itself prove that these transforms are optimum. However, none of the other transformations offer a better trade-off between power, gain and linearity. Therefore, the optimality of OSIT_1 and OSIT_2 is assumed for the remainder of this thesis. If, at some point, a better orthogonal SIT is found, it can replace OSIT_1 or OSIT_2 without affecting the proposed method for minimum power design.

Table 81 shows the effects on gain, noise, linearity and power dissipation of both OSIT transforms.

	Gain	**NF**	**IP3**	**P_{dc}**
Original circuit	G	NF	IP3	P
OSIT_1	G	NF	mult IP3	mult P
OSIT_2	s_{11} G	NF	IP3	s_{11} P

Table 11: The effects of OSIT_1 and OSIT_2 transforms

Since none of the SITs showed a trade-off between low power dissipation and noise performance, the noise figure is invariant for both OSITs. For most designers, this might seem non-intuitive, since they often have the experience that the noise figure can be improved at the cost of an increased power dissipation. In Section 5.5.4, this apparent contradiction between experience and theory will be investigated further.

5.4 Equivalent figures of merit (EFOMs)

OSITs transform any circuit, with a performance represented by a point P_1 in the m-dimensional space V, into another circuit with a different performance, represented by a point P_2. The set of points P_i that can be covered by applying all possible OSITs on a circuit, when taken together, form a hyperplane A in V. Any circuit with a performance represented by a point on A can therefore be considered to have a performance equivalent to any other circuit with a performance represented by any other point on A.

This equivalence is the foundation for a new class of figures of merit (FOMs), called "equivalent figures of merit", or EFOMs. EFOMs are a subset of FOMs, since they require the existence of transforms that can be applied to any circuit. These transforms should generate all new circuits with combinations of performance parameters that have the same EFOM value. This requirement allows EFOMs to be used directly in a design process: the circuit with the best EFOM value is always the best basis for any other set of specifications, as long as the performance parameters in the EFOM and in the specification are identical. The base circuit can then be transformed into the required circuit using OSITs. This sets EFOMs apart from other FOMs: they are much more than "just another FOM".

The OSIT_1 and OSIT_2 transforms from the previous section are an ideal basis for developing an EFOM. OSIT_1 allows for a proportional trade-off between power and linearity. OSIT_2 allows for a proportional trade-off between gain and linearity. Since they are orthogonal, they can be combined easily. The combination allows a proportional trade-off between power and the product of gain and linearity. This combined trade-off can be represented by an EFOM, which in this thesis will be called the *power linearity factor* κ:

$$\kappa = \frac{P_{dc}}{G \cdot IP3} \qquad \textbf{(43)}$$

It relates the power dissipation P_{dc} of a circuit to its gain G and linearity $IP3$. Since the *output IP3* (*OIP3*) equals the product of gain and input IP3, this is equivalent to:

$$\kappa = \frac{P_{dc}}{OIP3} \qquad \textbf{(44)}$$

This is the reason for naming κ the power linearity factor. κ is invariant under the OSITs defined in previous section. Note that κ increases with increasing power dissipation, and decreases with increasing gain and linearity. In other words, circuits with a low κ are more desirable for low power design.

Since circuits with a dominant pole also allow a trade-off between gain and bandwidth (Section 3.2), and since there is often no significant dependence of IP3 and power dissipation on the signal frequency, the power linearity factor for this class of

circuits can be extended to the following EFOM:

$$EFOM\_1 = \frac{BW}{\kappa} = \frac{G\ BW\ IP3}{P_{dc}} \qquad (45)$$

Because of the inversion of κ in EFOM_1, a higher EFOM_1 value corresponds to a better performance, which is more natural for a figure of merit. Both κ and EFOM_1 are EFOMs. The difference between EFOM_1 and κ is that EFOM_1 also takes the bandwidth of a circuit into account, and this is a relevant parameter for RF front ends, especially in the context of short-range systems.

EFOMs are a more solid base for determining trends in the power dissipation of RF circuits. Figure 57 shows the same LNAs as discussed in Chapter 2, figure 10, but this time the EFOM_1 is plotted versus the year of publication.

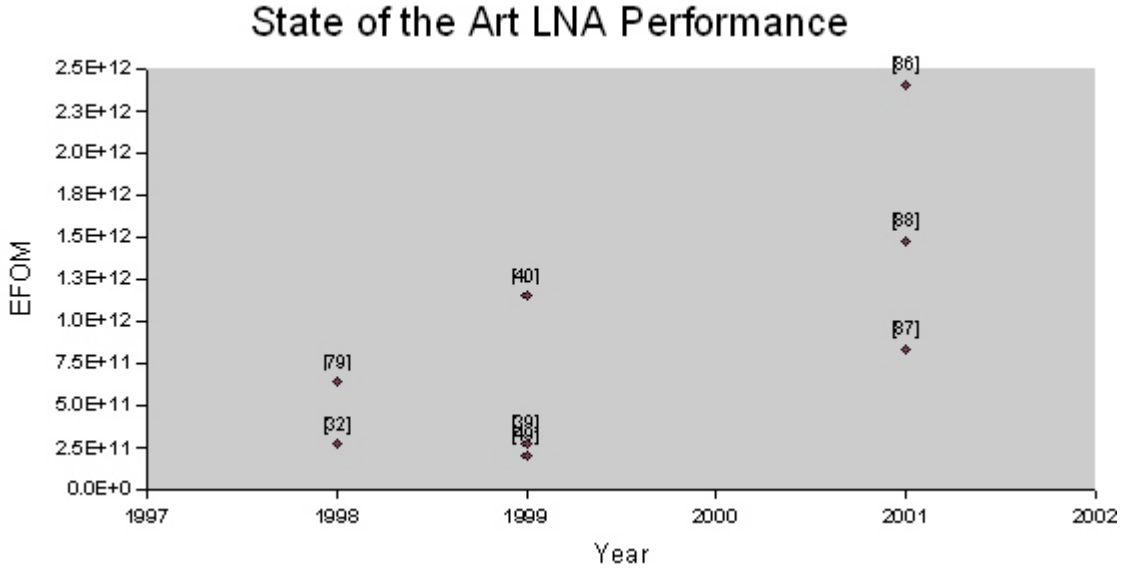


Figure 57 *EFOM_1 of recently published LNA circuits versus year of publication*

By presenting the LNA performance through EFOM_1, a trend can be easily recognized: in recent years, the performance/power dissipation ratio of LNAs, as expressed by EFOM, has significantly improved. This is based on a limited set of LNAs from literature. A more thorough verification of this trend is beyond the scope of this thesis, but could be an interesting area for further research.

It is likely that this trend is caused in part by the increasing number of applications for which low power dissipation is an essential property. There might be at least one other cause for the reduction in power dissipation. Figure 58 shows the same graph, but this time the labels for the individual points indicate the technology used, rather than the literature reference.

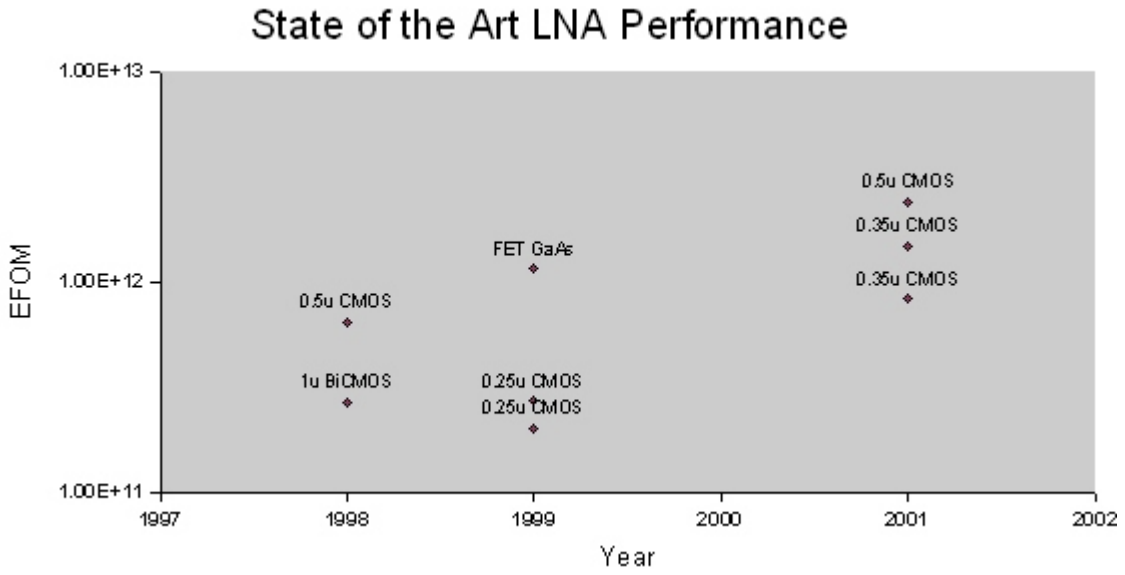


Figure 58 *FOM1 versus publication year of recent LNAs, with IC technology used*

This graph suggests a relation between EFOM and technology used. CMOS LNAs seem to outperform bipolar/bicmos and even GaAs LNAs. Therefore, the increasing use of CMOS technology for RF circuits might have influenced the trend in power dissipation of RF circuits as well. This is also consistent with the observation that silicon CMOS technologies have overtaken silicon BiCMOS technologies in terms of bandwidth in recent years (fig. 59).

For CMOS technologies, there is a high level of consensus and standardization in the industry, e.g. through the Semiconductor Industry Association SIA. This results in clear roadmaps. For BiCMOS, the situation is less clear, and there is a larger variety in processes and roadmaps in various companies. Therefore, the CMOS trend can be indicated by a line, whereas the BiCMOS trend can only be approximated by a box.

In addition to these developments at the circuit and technology level, there are also developments at the system level to enable lower power dissipation. They can be divided into the following categories:

- *Relaxed specifications for the RF part:* this approach has been followed in e.g. DECT

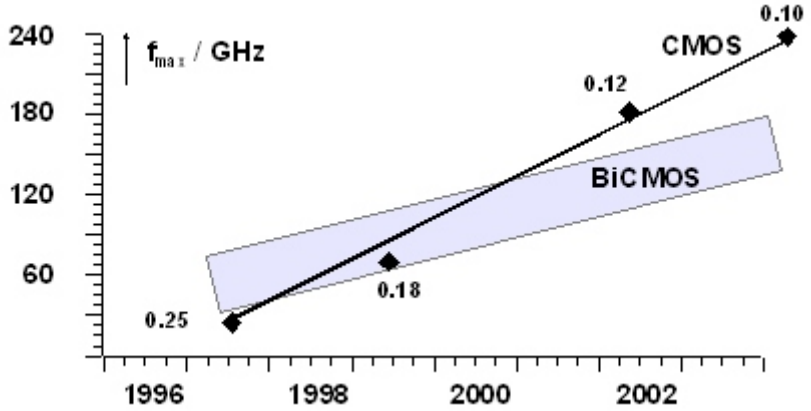


Figure 59 *Trends in bandwidth of CMOS and BiCMOS technologies*

and BlueTooth standardization. The expectation was that the more relaxed specifications of the RF circuits would result in lower power dissipation and lower cost. These results have been achieved only partially, at best, since IC vendors used "better than standard performance" as a differentiator in their competition, resulting in much higher than required RF performance and limited reduction in power dissipation and cost.

- *Low-power modes at system level:* this approach takes power dissipation into account already in the early definition phases of a system. A typical example is the Japanese digital video standard ISDB-T, in which a special mobile/low-power mode allows for the reduction of the signal bandwidth (and therefore dynamic range, bitrates etc.) already at the antenna. Another example is the low-power standby mode in BlueTooth, which is achieved through low duty-cycle sampling of the radio signal to check for incoming data traffic.
- *System-level low-power trade-offs:* in this approach, the system trade-offs are based at least in part on low-power considerations. For example, the total power dissipation of a receiver might be reduced by advanced antenna diversity systems, because such systems can significantly reduce the requirements for e.g. sensitivity, and thus more than compensate for the power dissipation of the additional diversity circuits.

Obviously, these approaches can be combined as well, and such combinations are increasingly used for systems targeted for low-power applications.

It is important to realize that equal performance, as indicated by the EFOM, does not imply equal desirability in every situation. Some combination of performance parameters, as represented by a point P_a, might be more desirable in a given application than another set of performance parameters, as represented by a point P_b. For example, an amplifier with high gain and low linearity might be more desirable for some application than an amplifier that is derived from the original one through SITs, which offers a lower gain but higher linearity. In general, the "desirability" of parameter combinations can be identified through a cost function. Performance parameter combinations with equal desirability might be represented by another hyperplane B in V. The intersection of A and B represent the circuits that are equally desirable, and can be derived from a single circuit through SITs.

The desired performance of a circuit in a front end depends on the application of the front end, but also on the performance of other circuits in the signal chain. To find the desired performance parameters of a circuit, it is necessary to find the optimum distribution of performance parameters across a signal chain.

5.5 Optimum distribution of gain, linearity and noise

An optimum distribution of gain, noise and linearity for any front-end configuration and specification can be derived, based on OSITs. In addition to the OSITs, expressions are needed to relate the front end specifications to the circuit specifications of the signal-processing stages in the front end signal chain. These expressions will be discussed first.

5.5.1 General Problem

The general problem is shown in figure 60. Assume a cascade of signal-processing stages numbered *1* through *n*. Equations (46) show the relations between the overall specifications and the performance of the individual blocks. The expression for the third-order distortion of cascaded stages is discussed in more detail in 8.7.1. The expression for total gain shows that the total gain is the product of the gain of the individual stages, and the expression for the total power represents the sum of the power dissipation of the individual stages. The power of an individual stage is related to the gain and linearity of that stage through the power linearity parameter κ. The total noise factor, finally, is represented by Friis' formula [4].

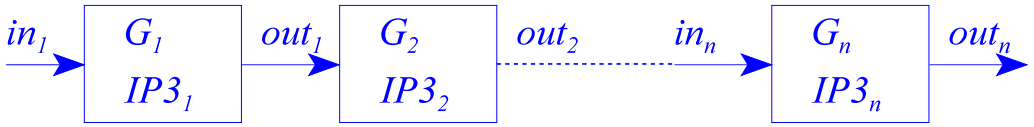


Figure 60 *Cascaded stages*

$$
\begin{aligned}
IP3_{tot} &= \frac{1}{\sum_{i=1}^{n} \frac{\prod_{j=1}^{i-1} G_j}{IP3_i}} \\
G_{tot} &= \prod_{i=1}^{n} G_i \\
P_{tot} &= \sum_{i=1}^{n} P_i \\
P_i &= \kappa_i \, G_i \, IP3_i \\
F_{tot} &= 1 + \sum_{i=1}^{n} \left(\frac{F_i - 1}{\prod_{j=1}^{i-1} G_j} \right)
\end{aligned}
\qquad \textbf{(46)}
$$

For the construction of an optimum distribution of gain, linearity, and noise in a front end, it is assumed that a library of the "best" (as selected by EFOM_1) signal-processing subcircuits is available from which such a front end will be constructed. For each of the

front-end subcircuits in the library, the parameters F_i and k_i need to be known[4]. If these parameters are not known, they can be easily derived by circuit simulation. The required performance of the total front end is defined by F_{tot}, G_{tot}, and $IP3_{tot}$. Using OSITs, the expressions for cascaded noise factor F_{tot}, linearity $IP3_{tot,}$ and gain G_{tot} can be solved for G_i and $IP3_i$. If these parameters have been solved, P_i and P_{tot} can be calculated as well.

An infinite number of solutions exist. To find a unique solution, one more constraint is needed: the solution has to achieve the global minimum value of P_{tot}. Please note that different constraints are possible, and sometimes even more relevant. For example, total chip area will in practice be a very relevant parameter for industrial designs. In order to determine the minimum power for a front end, however, this is the most relevant constraint in this case. Once the unique solution has been found, the minimum achievable power dissipation is known. In the next step, such a circuit can be assembled by applying the OSITs to the original subcircuits with the parameters determined by the solution. The resulting, new, subcircuits are then assembled into a front end circuit by cascading them according to the selected front end architecture.

There might be cases in which selecting the "best" subcircuits for the library is not obvious. For example, since the noise factor is a very relevant parameter for a front end, and since this parameter is not taken into account in the EFOM_1, the subcircuit with the best EFOM_1 value might not have the best noise factor. In such a case, multiple subcircuits of the same type can be added to the library, with different trade-offs between noise factor and EFOM_1. EFOM_1 can still be used to exclude circuits that don't offer either a better EFOM_1 value or better noise factor, thereby significantly reducing the size of the library. With multiple instances of the same type of subcircuit in the library, the minimum power dissipation needs to be calculated for every combination of subcircuits assembled according to the front end architecture. Once the minimum power dissipation for every combination is known, selecting the best combination is trivial.

5.5.2 The solution for n=2

For a cascade of two stages, the solution of these equations is straightforward. The equations (46) for this case are:

[4]Every sub block can be characterized by its (transform invariant) nosie factor F_i and its power linearity factor κ_i. Each sub block then represents a class of potential sub blocks with different G_i, $IP3_i$, and P_i, but with the same κ_i.

$$
\begin{aligned}
P_{tot} &= P_1 + P_2 \\
P_1 &= \kappa_1 G_1 IP3_1 \\
P_2 &= \kappa_2 G_2 IP3_2 \\
\\
G_{tot} &= G_1 G_2 \\
\\
F_{tot} &= F_1 + \frac{F_2 - 1}{G_1} \\
\\
IP3_{tot} &= \frac{1}{\frac{1}{IP3_1} + \frac{G_1}{IP3_2}}
\end{aligned}
\tag{47}
$$

where κ_1, κ_2, F_1, F_2, F_{tot}, G_{tot} and $IP3_{tot}$ are known. These equations can be solved for G_1, G_2, $IP3_1$ and $IP3_2$ by first solving for the gains:

$$
\begin{aligned}
G_1 &= \frac{F_2 - 1}{F_{tot} - F_1} \\
\\
G_2 &= \frac{G_{tot}}{G_1}
\end{aligned}
\tag{48}
$$

Now $IP3_1$ can be calculated:

$$
IP3_1 = \frac{1}{\frac{1}{IP3_{tot}} - \frac{G_1}{IP3_2}} \tag{49}
$$

Substituting G_1 in (49) by the value from (48) results in explicit solutions for G_1, G_2 and *e.g.* $IP3_1$. What remains is the solution for $IP3_2$. It would also have been possible to solve for $IP3_2$ first, which would of course result in the same solution. There are an infinite number of solutions for $IP3_2$ that satisfy the relations and known parameters. To find a unique solution for minimum power, P is expressed as a function of $IP3_2$ only, using the solutions for G_1, G_2 and $IP3_1$ from the previous equations. A typical graph of $IP3_1$ as a function of $IP3_2$ is shown in fig. 61.

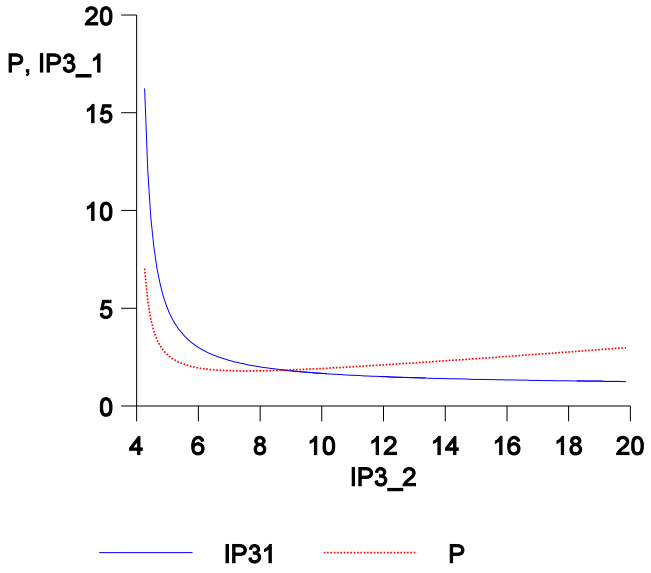


Figure 61 *$IP3_1$ as a function of $IP3_2$*

For $IP3_2$ values below, in this example, 4, no solution is found for P and $IP3_1$. Whereas $IP3_1$ is a monotonic function of $IP3_2$, P has a minimum for a specific value of $IP3_2$. This minimum can be found by determining the extrema for P:

$$\frac{\partial P_{tot}}{\partial IP3_2} = 0 \tag{50}$$

This expands to:

$$\frac{\kappa_2 (F_{tot} - F_1)\, G_{tot}}{F_2 - 1} - \frac{\kappa_1 (F_2 - 1)^2}{(F_{tot} - F_1)^2 \left(\frac{1}{IP3_{tot}} - \frac{F_2 - 1}{(F_{tot} - F_1)\, IP3_2} \right)^2 IP3_2^2} = 0 \tag{51}$$

This equation has two roots for $IP3_2$: r_1 and r_2 (52). The root r_2 has a positive second derivative and r_1 has a negative second derivative.

$$r_1 = IP3_{tot}\,\frac{(F_2-1)\left(\sqrt{\kappa_2 G_{tot}}(F_{tot}-F_1)-\sqrt{\kappa_1(F_{tot}-F_1)(F_2-1)}\right)}{\sqrt{\kappa_2 G_{tot}}\,(F_{tot}-F_1)^2}$$

$$r_2 = IP3_{tot}\,\frac{(F_2-1)\left(\sqrt{\kappa_2 G_{tot}}(F_{tot}-F_1)+\sqrt{\kappa_1(F_{tot}-F_1)(F_2-1)}\right)}{\sqrt{\kappa_2 G_{tot}}\,(F_{tot}-F_1)^2} \qquad \textbf{(52)}$$

Since the positive second derivative in combination with a zero of the first derivative corresponds to a minimum, r_2 represents the value of $IP3_2$ that corresponds to the minimum power dissipation. Please note that the power which corresponds to r_1 actually represents a total power dissipation that can be lower than that represented by r_2. However, the corresponding value of $IP3_1$ is negative, and therefore represents no valid solution. Substituting r_2 into the initial equations results in the following parameter values:

$$G_1 = \frac{F_2-1}{F_{tot}-F_1}$$

$$G_2 = \frac{F_{tot}-F_1}{F_2-1}\,G_{tot}$$

$$IP3_1 = IP3_{tot}\left(1+\sqrt{G_{tot}\,\frac{\kappa_2(F_{tot}-F_1)}{\kappa_1(F_2-1)}}\right) \qquad \textbf{(53)}$$

$$IP3_2 = IP3_{tot}\,\frac{F_2-1}{F_{tot}-F_1}\left(1+\sqrt{\frac{\kappa_1(F_2-1)}{G_{tot}\,\kappa_2(F_{tot}-F_1)}}\right)$$

The power level that corresponds to this solution is:

$$P_{tot} = IP3\,\frac{\kappa_1(F_2-1)+2\sqrt{G\kappa_1\kappa_2(F-F_1)(F_2-1)}+\kappa_2 G\sqrt{F-F_1}}{F-F_1} \qquad \textbf{(54)}$$

This is the minimum power dissipation that can be realized by any cascade of two stages. The corresponding solution for the gain and linearity (53) represents the minimum power distribution of gain, linearity and noise across two cascaded stages. It is of course desirable to extend this solution to more than two stages.

5.5.3 The General Solution

The general problem for arbitrary values of n can be formulated as follows:

Given numbers:

$$\begin{aligned} \kappa_i &> 0 \\ F_i &> 1 \\ F_{tot} &> 1 \\ IP3_{tot} &> 0 \\ G_{tot} &> 0 \end{aligned} \tag{55}$$

find the minimum of:

$$\begin{aligned} P_{tot} &= \sum_{i=1}^{n} P_i \\ P_i &= \kappa_i G_i IP3_i \end{aligned} \tag{56}$$

and the corresponding values of G_i and $IP3_i$, under the conditions:

$$\begin{aligned} \prod_{i=1}^{n} G_i &= G_{tot} \\ \sum_{i=1}^{n} \frac{F_i - 1}{\prod_{j=1}^{i-1} G_j} &= F_{tot} - 1 \\ \sum_{i=1}^{n} \frac{\prod_{j=1}^{i-1} G_j}{IP3_i} &= \frac{1}{IP3_{tot}} \end{aligned} \tag{57}$$

This general optimization problem was solved mathematically by dr. ir. A.J.E.M. (Guido) Janssen, at the Philips Research Laboratories in Eindhoven, using Lagrange multipliers. The derivation of this solution can be found in appendix F. The solution for the gain, rewritten using the same parameters as in the problem definition above, is:

$$G_1 = \sqrt[3]{\frac{(F_2-1)^2}{\kappa_1}}\ \frac{1}{F-F_1}\sum_{j=1}^{n-1}\sqrt[3]{\kappa_j(F_{j+1}-1)} \tag{58}$$

$$G_i = \sqrt[3]{\left(\frac{F_{i+1}-1}{F_i-1}\right)^2\frac{\kappa_{i-1}}{\kappa_i}} \qquad i=2...n-1 \tag{59}$$

$$G_n = \frac{G_{tot}\,(F_{tot}-F_1)}{\sqrt[3]{\frac{(F_2-1)^2}{\kappa_1}}\left(\sum_{j=1}^{n-1}\sqrt[3]{\kappa_j(F_{j+1}-1)}\right)\left(\prod_{j=2}^{n-1}\sqrt[3]{\frac{(F_{j+1}-1)^2\kappa_{j-1}}{(F_j-1)^2\kappa_j}}\right)} \tag{60}$$

The solution for IP3 can be rewritten as:

$$IP3_i = IP3_{tot}\sqrt{\frac{\prod_{j=1}^{i}G_j}{\kappa_i}}\ \frac{\sum_{j=1}^{n}\sqrt{\kappa_j\prod_{k=1}^{j}G_k}}{G_i} \tag{61}$$

The minimum power dissipation can now be expressed as:

$$P_{\min} = IP3_{tot}\left(\sqrt{\kappa_n G_{tot}} + \frac{\left(\sum_{i=1}^{n-1}\sqrt[3]{\kappa_i(F_{i+1}-1)}\right)^{\frac{3}{2}}}{\sqrt{F_{tot}-F_1}}\right)^2 \tag{62}$$

For n=2, this solution is of course identical to the solution found in the previous section.

In the next step, a minimum power front end can be assembled by applying the OSITs to the original subcircuits with the parameters determined by the solution (58)-(62).This will result in new subcircuits with the optimum gain and linearity for the front end. These new subcircuits can then be assembled into a front end circuit by cascading them according to the selected front end architecture.

5.5.4 Discussion

The general solution (62) has some interesting interpretations:

- The total power dissipation scales with the total IP3 specification. This is to be expected, since the OSIT_1 transform can also be applied at the front end level.
- Even though there is no OSIT for noise factors at the subcircuit level, both the special and the general solution show a trade-off between noise factor and power dissipation $P=f(\kappa_i, G_i, IP3_i)$ of the total front end. This is consistent with the experience of RF designers that low noise factors require high power dissipation at the front end level, and can be explained through Friis' formula [4]. The total noise figure of multiple, cascaded, stages can be reduced by increasing the gain in the first stage. Increasing this gain results in higher power dissipation. In addition, the linearity requirements for subsequent stages increases because of the larger signal levels, resulting in more power dissipation of these stages as well. Therefore, this design method does provide the link between noise factor and power dissipation for the total front end, even though this is not obvious from the descriptions and transforms for the individual subcircuits.
- The denominator of the second term in the main brackets equals the square root of the difference between the target noise factor and the noise factor of the first stage. When the noise figure of the first stage is higher than the total noise figure, no solution is found, of course. This is consistent with the situation in which a too high noise figure in the first stage cannot be compensated for by any means in subsequent stages.
- For first-stage noise factors lower than the total noise factor, the total power dissipation increases rapidly when the noise factor of the first stage approaches the total noise factor. This makes sense, since a higher noise factor in the first stage would require more gain in the first stage to reduce the noise contributions of subsequent stages. This will also result in higher signal levels in subsequent stages, and therefore higher linearity requirements, which in turn result in more power dissipation.
- The numerator of the second term between the main brackets in (62) shows that the additional power dissipation, caused by a higher noise factor in the first stage, is dependent on both the power linearity factors κ of all stages but the last one, and their noise factors. The influence of the last stage is taken into account by the first term between the brackets.

5.6 A minimum-power front-end design procedure

A minimum-power design procedure, based on OSITs and EFOMs, will follow the stages as have been shown in Section 5.2, fig. 47. The stages that are new and different compared to typical current design procedures have been shaded in figure 62. The design method uses standard stages for assembly, fabrication and measurement & testing. Also, new topologies are created using traditional methods. The hierarchy in abstraction levels that is introduced in this design method is required to build a model of reality in a reduced form, which in turn is required both for developing an insight in the problem, and for a synthesis of the solution. The most important new stages will be discussed in

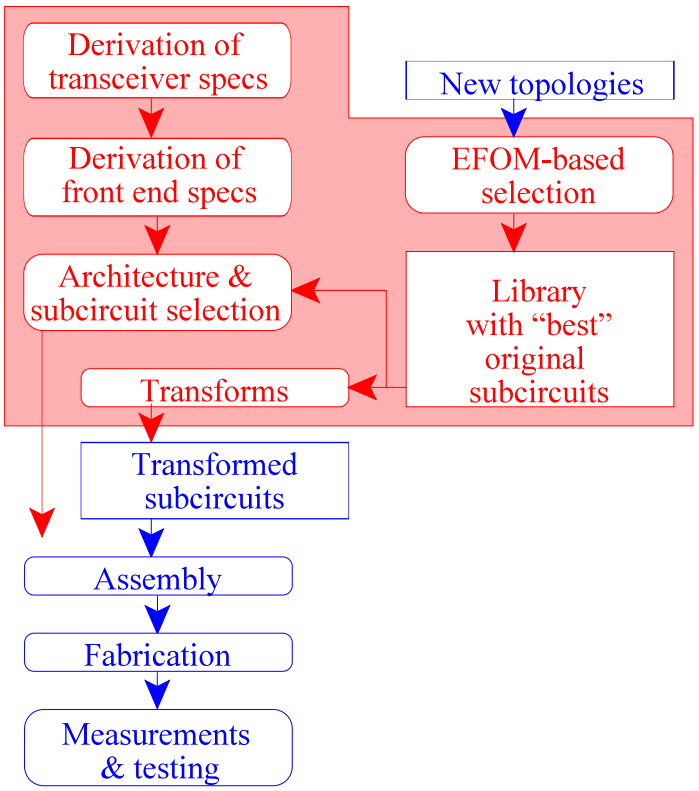


Figure 62 *New elements in the optimum design method*

the following sections.

5.6.1 Derivation of transceiver specifications

The starting point in this stage are the system specifications. To make the design method as efficient as possible, these specifications should be stored in a format that is consistent between systems, and very easy to access. From these system specifications, the specifications for a specific transceiver need to be derived. Many parameters of the system automatically apply to the transceiver, such as the RF signal frequency range, transmit power, etc. Other parameters are usually described in a different way in a transceiver specification and in a system specification. For example, in a system specification, usually the maximum interferer levels are specified, whereas in a transceiver specification, *IP3* and 1dB compression levels are specified. This makes sense, since the system specifies the conditions in which a system is expected to operate, whereas a transceiver specification specifies the properties of the transceiver which should operate as part of this system. The parameters that typically need to be converted in the receive chain include noise figure (*NF*), co-channel suppression (SNR_{min}), linearity (*IP3*), AGC range (ΔG_{AGC}), 1dB compression point (P_{1dB}), selectivity ($S(\Delta f)$), and VCO carrier-to-noise ratio (CNR_{VCO}). The noise figure is determined from the sensitivity level (P_{sens}), the thermal noise power in the channel bandwidth (*BW*), and the minimum SNR for demodulation (SNR_{min}), which is assumed to be identical to the co-channel suppression. Usually, it is assumed that the noise figure of the receiver makes up the difference between the theoretical sensitivity in case of thermal noise only, and the actual sensitivity (excluding implementation losses in the digital circuits, where appropriate).

This results in the following expression for *NF*.

$$NF = P_{sens} - SNR_{\min} - 10\log(kT\,BW) \tag{63}$$

Please note that in this formula, *NF*, SNR_{min} and P_{sens} are in dB. The SNR_{min} can be derived from the co-channel interferer level specified at an offset of 0, i.e. in the middle of the receive channel.

$$SNR_{\min} = P_{signal} - P_{interferer@\Delta f=0} \tag{64}$$

Now, define the maximum level of a distortion component to be equal to a co-channel interferer:

$$P_{dist} = P_{signal} - SNR_{\min} \tag{65}$$

The minimum IP3 value is determined from this distortion level, and the level of the interferers on adjacent channels. Adjacent channels are channels that are one channel spacing above or below the wanted channel. Alternate channels are channels that are two channel spacings above or below the wanted channel. A detailed discussion of *IP3* and its relation to interferers can be found in appendix H.2.

$$IP3 = P_{interferer@\Delta f=adjacent} + \frac{P_{interferer@\Delta f=adjacent} - P_{dist}}{2} \tag{66}$$

An AGC function in the RF front end helps reduce the power dissipation by reducing the signal levels in subsequent stages when high input signal levels are present. The reduced signal levels translate in lower distortion requirements, which in turn result in lower power dissipation as predicted by EFOM_1. The optimum AGC range of a receiver is determined from the dynamic range of the desired signal, which could be dealt with fully by AGC operation, and the dynamic range of the desired signal with one or two interferers. The dynamic range required to process the interferers cannot be reduced by AGC operation, since the front end should process both the wanted signal and the interferers without distorting the interferers, and without decreasing the signal-to-noise ratio for the wanted signal. Therefore, this settles the minimum dynamic range. The AGC range should make up the difference between the dynamic range of the signal by itself, and the dynamic range required for the correct processing of the interferers. The optimum AGC range can be defined as:

$$\Delta G_{AGC} = DR_0 - \mathrm{Max}(DR_1, DR_2) \tag{67}$$

with DR_0 being the dynamic range of the wanted signal, DR_1 the dynamic range of the wanted signal and one interferer, and DR_2 the dynamic range of the wanted signal and two interferers. To maximize the benefit of the AGC operation on the power dissipation, it is assumed that the AGC function is implemented in the first block of the signal-processing chain, typically the LNA. Therefore, the minimum 1dB compression point is calculated from the sensitivity level in combination with the dynamic range required by the interferer levels:

$$P_{1dB} = P_{\min} + Max(DR_1, DR_2) \tag{68}$$

Usually, some margin is added here to allow for imperfect AGC operation and similar imperfections. The selectivity S(Δf) is determined by the relative strengths of adjacent and alternate channel interferers compared to a co-channel interferer. Assuming that the demodulator is not frequency selective within a 5 channel bandwidth, the single tone interferers at adjacent and alternate channels need to be suppressed to a level comparable to a co-channel interferer. This results in a selectivity of:

$$S(\Delta f) = P_{interferer@0} - P_{interferer@\Delta f} \tag{69}$$

The maximum phase noise of the VCO is typically determined by the reciprocal mixing of this phase noise with single-tone interferers. The reciprocal mixing products at Δf (for interferers at adjacent channels) and 2 Δf (for interferers at alternate channels) will coincide with the wanted signal, and need to be lower than a co-channel interferer. Therefore, the minimum CNR is determined by:

$$CNR_{\min}(\Delta f) = P_{signal} - SNR_{\min} - P_{interferer@\Delta f} - 10\log(BW) \tag{70}$$

5.6.2 Derivation of front end specifications

Typically, the IF&baseband circuit has specifications for *IP3* and *NF* that, in combination with an ideal front end, achieves system specifications for a range of front end gain values. Therefore, there is some freedom in choosing the optimum gain. The noise figure and linearity of the front end depend on the selected gain. More specifically, the maximum noise factor of the front end, F_{RF}, can be expressed as a function of the front-end gain G_{RF}, the total noise factor of the transceiver F_{tot}, and the noise factor of the IF&baseband circuit, F_{if}:

$$F_{rf} = F_{tot} - \frac{F_{if} - 1}{G_{rf}} \tag{71}$$

This function results in negative, and therefore impossible, values of F_{rf} for low values of G_{rf}. In other words:

$$G_{rf} > \frac{F_{if} - 1}{F_{tot}} \tag{72}$$

This effect is shown[5] in fig. 63.

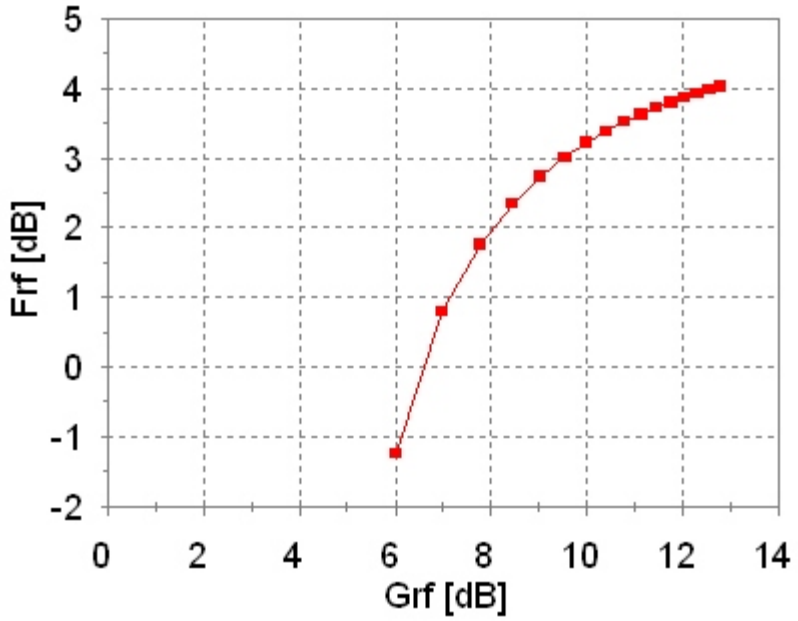


Figure 63 *Maximum noise factor of front end as a function of front-end gain*

For high values of the front end gain, the maximum noise factor asymptotically increases towards F_{tot}. This makes sense, since for high front end gain, the noise contribution of the IF&baseband circuits is negligible. In the same way, the minimum IP3 of the front end depends on the front end gain. The minimum value of the IP3 can be expressed as a voltage $VIP3_{rf}$ using the expression for the total 3rd order distortion, Section H.2.2:

$$IP3_{tot} = \frac{1}{\frac{1}{IP3_{rf}} + \frac{G_{rf}}{IP3_{if}}} \tag{73}$$

[5] This graph is based on the following example parameters: $F_{tot} = 10$, $F_{if} = 3$

Solving this equation for the minimum $IP3_{rf}$ value results in:

$$IP3_{rf} = \frac{IP3_{if}\ IP3_{tot}}{IP3_{if} - G_{rf}\ IP3_{tot}} \tag{74}$$

When $G_{rf}\ IP3_{tot}$ exceeds $IP3_{if}$, the minimum value for $IP3_{rf}$ becomes negative and no solution exists. In other words:

$$G_{rf} < \frac{IP3_{if}}{IP3_{tot}} \tag{75}$$

This effect is shown[6] in fig. 64.

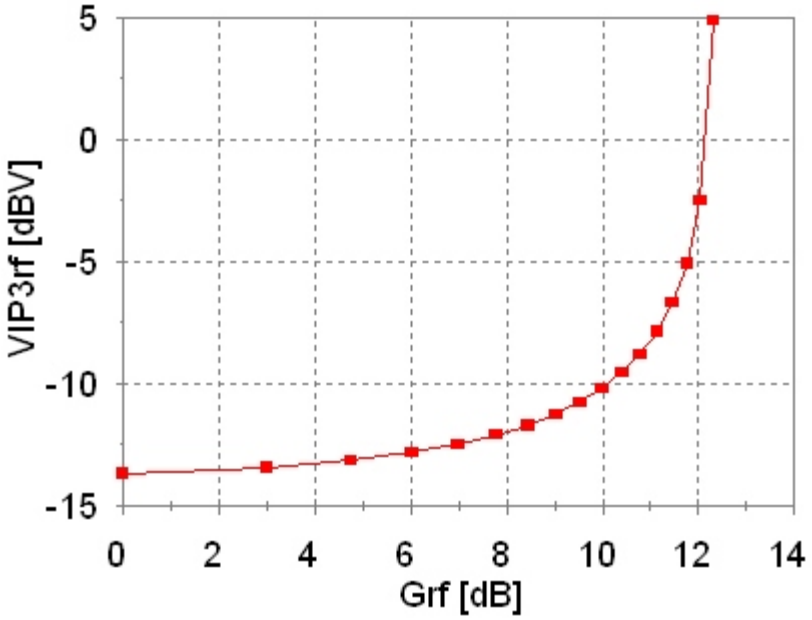


Figure 64 *Front end IP3 voltage versus front end gain*

For very low gain values, the $IP3_{rf}$ value decreases asymptotically to the $IP3_{tot}$ value. By combining (72) and (75), we find the boundary conditions for the front-end gain:

$$\frac{F_{if} - 1}{F_{tot}} < G_{rf} < \frac{IP3_{if}}{IP3_{tot}} \tag{76}$$

[6] In this graph, the following example parameter values have been used: $VIP3_{if} = 0.83$, $VIP3_{tot} = 0.2$, and therefore $VIP3_{if}/VIP3_{tot} = 12.36$dB.

From the outcome of the optimum distribution of gain, noise, and linearity, it is clear that very low noise figures or very high linearity both result in very high power dissipation. Moreover, $IP3_{rf}$ decreases asymptotically for low gain values, and F_{rf} increases asymptotically for high gain values. In the region "in between", i.e. for average G_{rf} values, neither $IP3_{rf}$ nor F_{rf} have a strong dependence on G_{rf}, so the optimum will be very shallow. Therefore, a common sense value for the front end gain is "in the middle" between these two extremes:

$$G_{rf} = \sqrt{\frac{(F_{if} - 1)\, IP3_{if}}{F_{tot}\, IP3_{tot}}} \tag{77}$$

Based on this choice for the gain, the noise figure and IP3 can be determined from equations (71) and (74).

5.6.3 Architecture and subcircuit selection

Now that the front end specs are known, the architecture and subcircuit selection can be carried out. This can be achieved through a straightforward algorithm, as shown in fig. 65. Since the number of circuits is limited by of the EFOM_1 based selection, and since the number of viable front end architecture topologies is also rather small, an exhaustive search is feasible. Because analytical expressions have been derived for the gain and linearity distributions, and for the minimum power dissipation, the compute time for this algorithm is typically in the order of a few seconds, even on a small computer.

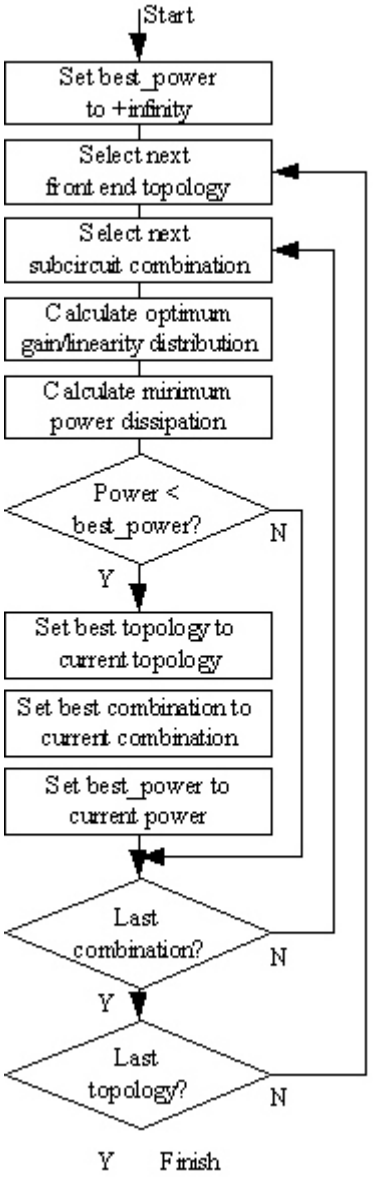


Figure 65 *Flow chart of architecture & subcircuit selection algorithm*

This algorithm will provide the best topology for the front-end architecture, as well as the best combination of subcircuits. The corresponding power dissipation is also calculated. The formulas for the optimum noise, gain and linearity distribution provide the specifications for each of the subcircuits. Since the OSITs are orthogonal, the corresponding transforms can be easily derived from these specifications. The subcircuits to be used can now be constructed from the selected subcircuit collection and the corresponding transforms, and can then be assembled according to the selected topology. This results in a minimum-power front end.

Filling of the library is a very simple process: start by putting in one circuit of each signal-processing type, and calculate their EFOM_1 value. It is now simple to decide whether a new circuit should be added to the library: if the EFOM_1 value of the new circuit is better than the one in the library, it should be added, and the existing circuit should be removed, unless the noise figure of the existing circuit is better than the new circuit. In that case, both stay in the library. In the same way, of the EFOM_1 value of the new circuit is worse than the one in the library, it should be discarded, unless the noise figure of the new circuit is better than that of the existing one. In that case, they also stay both in the library. In practice, there are additional boundary conditions that might require multiple circuits of each type in the library. For example, to facilitate single-chip integration, it is useful to have circuit implementations for different IC processes in the library. Also, the transformation range of circuits is limited in practice. Circuits with a larger transformation range, especially towards lower power dissipation, but with a lower EFOM_1 value, can still be useful parts of the library.

This procedure focuses on the effective and efficient design of an RF front end with minimum power dissipation, while taking into account only the primary specifications of an RF front end: linearity, gain, and noise. In practice, there are almost always secondary performance parameters in addition to the low power requirements, such as EMC behavior, cost, robustness against process variations, radiation hardness, temperature stability, power supply rejection, etc. In cases where such secondary considerations are important as well, or where power dissipation is not even the primary concern, the outcome of the design procedure is probably not the best design for a front end meeting all these considerations. Nevertheless, it will still provide a good starting point from which to build a front end that takes into account these other considerations, and it serves as a reference point that shows how much additional power is needed to achieve the secondary performance specifications. In many cases, the additional requirements can be taken into account by the selection of subcircuits for the library, which guarantees that the secondary requirements will be met. In that case, this design procedure can still be used to find the lowest possible power dissipation for a front end within these secondary boundary conditions.

5.7 A minimum-power front-end design tool

To achieve maximum efficiency when using the design procedure described in the previous section, a software tool was developed that implements and integrates the essential elements of this method. It consists of databases to hold the information in a consistent and easy-to-access format, and programs that implement the transformations on specifications and the selection of circuits and topologies. An overview of this tool, including the databases and their connections through the programs, is shown in figure 66.

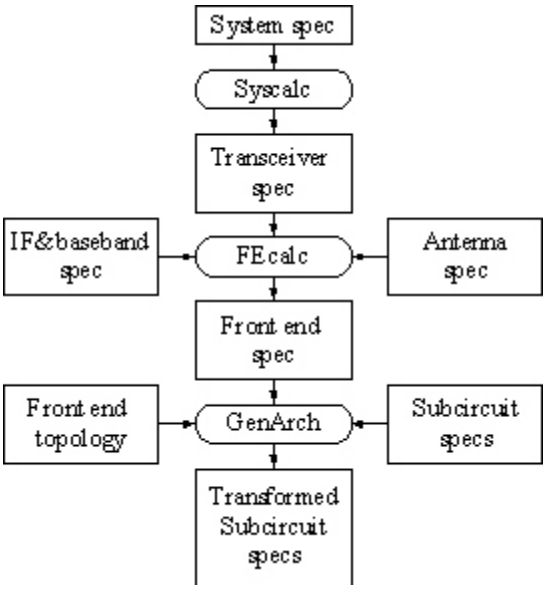


Figure 66 *Tool for the minimum power front end design method*

The rectangular boxes indicate data stored in databases. Each database has a corresponding editor program associated with it, that allows the designer direct read & write access to the database. The rounded boxes indicate programs that carry out transformations on data in one or more databases, and that store the results in another database. This tool is called FAT (Front- end Architecture Tool). The top-level interface to this tool is shown in figure 67.

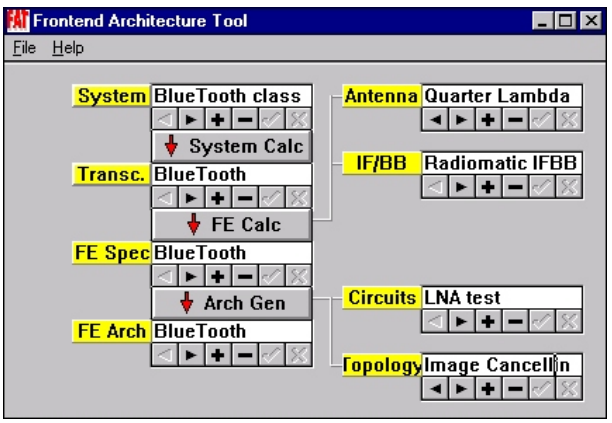


Figure 67 *Top-level interface to the FAT tool*

The interface is organized along the top-down design flow described in the design procedure in the previous section. In the left column is the straight top-down path, starting from system specifications, through transceiver and front-end specifications, to front-end architecture and subcircuit selection. In the right column, the supporting

databases that store information about antennas, IF&baseband circuits, front-end subcircuits, and front end topologies are shown. The connecting lines show where these supporting databases are used. The window shows database records in the white text boxes. The name of the database is in the label to the left of each text box. The buttons below the text boxes are for selection, addition and deletion of specific records in the respective databases. Clicking on a database label opens an editor program for the corresponding database. An example of such an editor (in this case for the system data) is shown below (fig. 68):

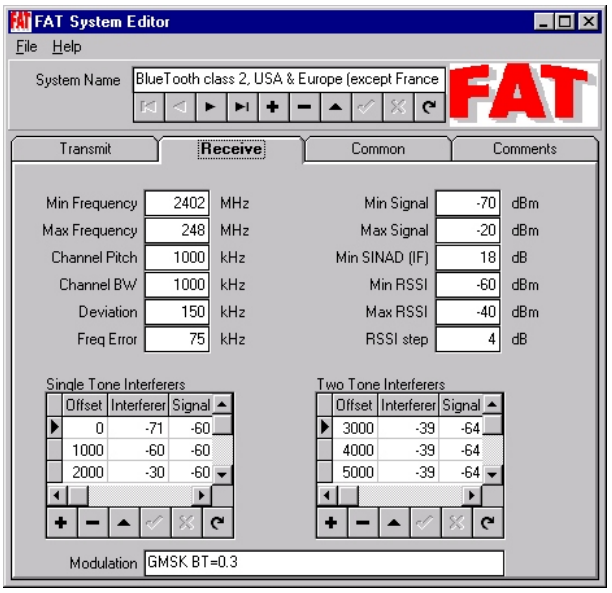


Figure 68 *System editor interface to the FAT tool*

The buttons with the arrows start programs that carry out the transformation of data from the database above to data in the database below the arrow. An example of such a transformation program (in this case the “FE Calc” module) is shown below (fig. 69):

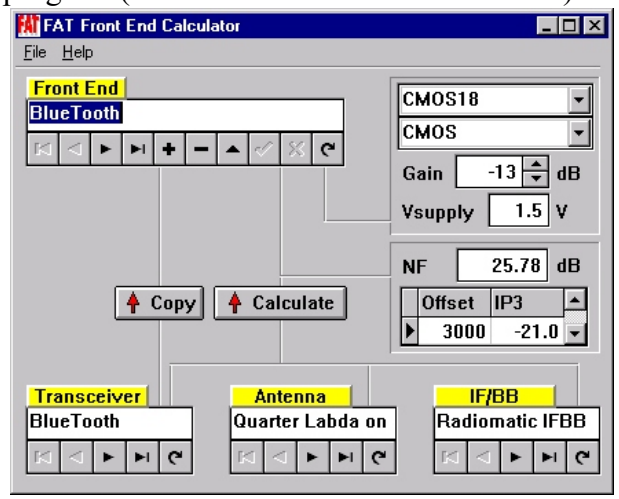


Figure 69 *Transformation module of the FAT tool*

Again, the text boxes with attached labels and buttons below allow access to related databases, and clicking the labels will start relevant database editors. The “copy” button with the arrow copies the corresponding data from one database to the other without any operations. The “calculate” button with the arrow calculates new parameters for the output database based on the data from the input databases and data provided by the user, such as (in this case) the process, technology, gain and supply voltage. The interface of the program that implements the topology and subcircuit selection, using the optimum distribution of gain, noise and linearity, is shown in figure 70.

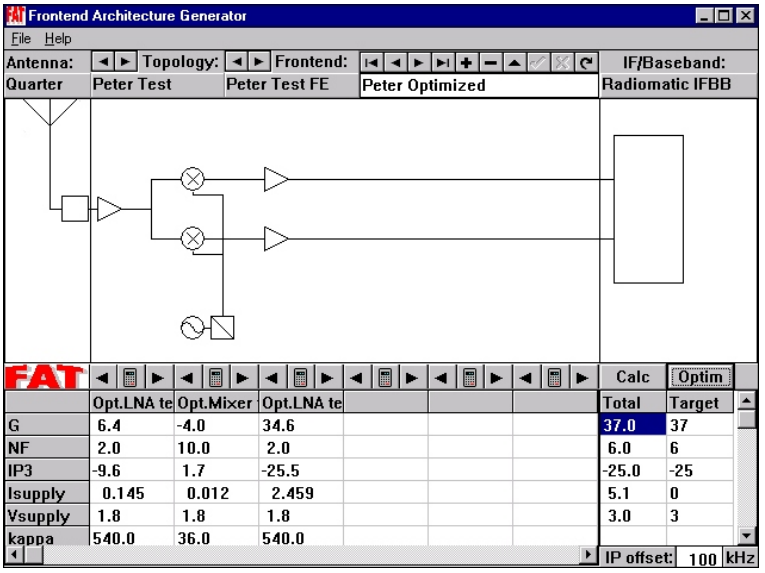


Figure 70 *Optimization module of the FAT tool*

Using this tool, the calculations for front end specifications and a minimum power implementation of the front end can be determined in a few minutes, if the information has already been entered in the databases. Often, designers might want to spend some time exploring this level of the design by entering hypothetical, potentially feasible circuits, or investigating trade-offs between the front end and other circuits (IF/baseband, antenna interface). This tool allows such explorations to be carried out efficiently. Finally, the databases are set up in such a way that they can be shared through a network between many users, allowing easy cooperation and dissemination of factual knowledge. A detailed description of the databases and the transforms used in the FAT tool is included in appendix G.

5.8 The DECT front end revisited

To get a first indication of the potential power savings that could be achieved using the design method and tools developed in the previous sections, the FAT tool was used to make an estimate of the minimum power dissipation that could have been achieved for the DECT front end if this design method would have been available at the time. Starting point is a library of the same circuits that were used in the original design:

	LNA	*Mixer*	*Output Buffer*
Voltage gain:	20dB	13dB	4dB
Power gain:	20dB	-4dB	8dB
Noise figure:	3.5dB	10dB	18dB
IP3:	-20dBm	0dBm	-4dBm
Power dissipation: (design target)	10mW	10mW	10mW
Power dissipation: (realized)	13.5mW	10.5mW	6mW

When keeping the specifications for the front end identical to the original target, as described in Chapter 4, the starting point for the optimization is represented in fig. 71.

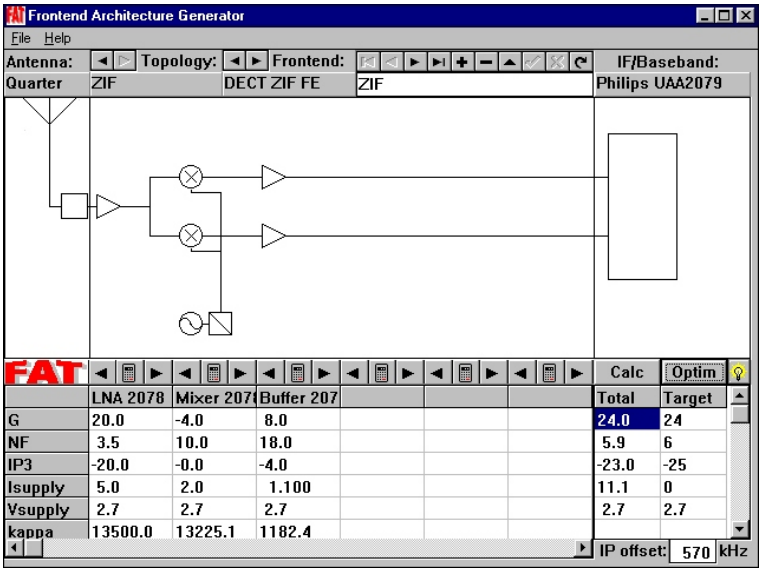


Figure 71 *DECT front end before optimization*

The optimum distribution of gain and noise figure can be calculated, as well as the predicted current consumption, by pressing the "optim" button. The result is shown in fig. 72.

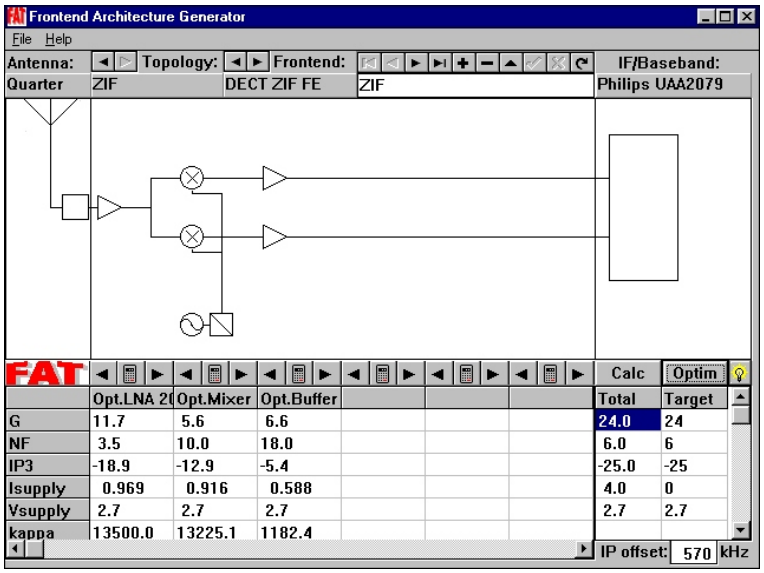


Figure 72 *Distribution of gain, noise and linearity after optimization*

The gain is distributed more towards the output of the front end. This results in lower gain and linearity requirements for the blocks near the input, and therefore a significant current reduction: from 11.1mA to 4.0mA, a factor of 2.7!

The power dissipation of the DECT front end of Chapter 4 was orders of magnitude higher than the minimum imposed by physics. This might raise the suspicion that it was a particularly bad design, not typical for the state of the art at that time, and that it therefore provides a rather optimistic view of what can be achieved through this design method. With the EFOM_1 introduced in this chapter, it is possible to put this design in the context of the designs discussed in Section 5.4. Figure 73 shows the power dissipation and performance of the DECT LNA, as expressed by EFOM_1, in this context. The EFOM_1 of the DECT LNA, although much earlier in time than the other LNAs, fits well in the trend shown in this figure, and is therefore typical for such designs. This is consistent with the observation that this design was turned into a competitive product.

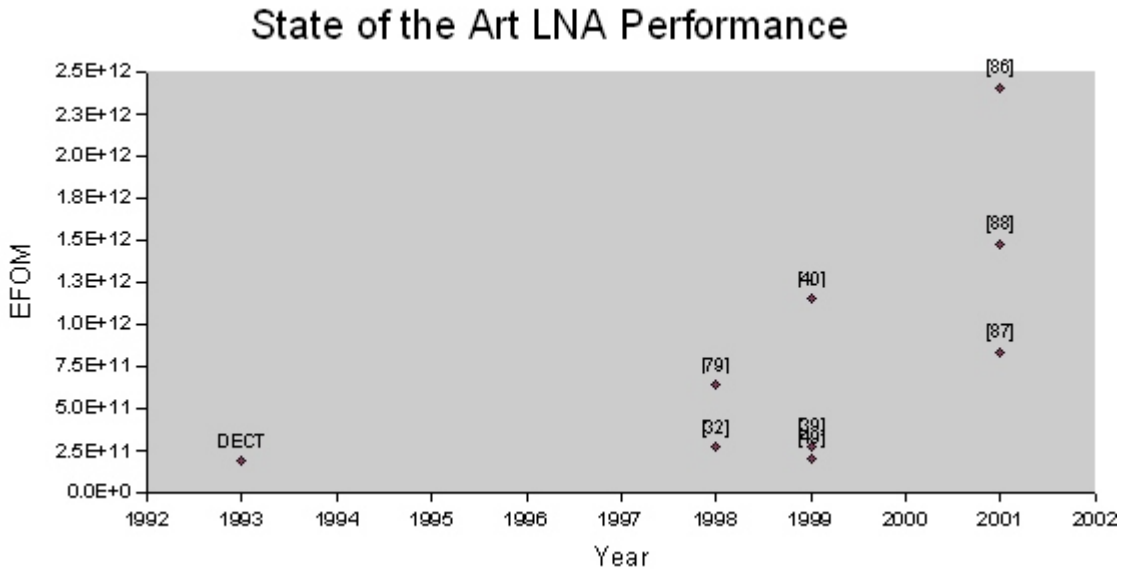


Figure 73 *EFOM of DECT LNA in the context of recent LNAs*

Optimization of the gain, linearity and noise distribution across signal-processing stages is just one aspect of this design method. The specifications for the DECT front end are significantly better than required by the DECT system standard. This results in a front end that is more robust to process variations and environmental changes, and also offers a competitive advantage because of the better performance than products from some competitors. These advantages are real and relevant, and it is acceptable that they come at the cost of some increase in power dissipation. However, until now it was not possible to provide a realistic estimate of this higher power dissipation. With the FAT tool, this question can be answered in minutes by carrying out the complete top-down design starting from the system specifications, and targeting a product that exactly meets the DECT specifications without any overkill. This results in relaxed requirements for gain (16.7dB instead of 24dB), noise figure (21.7dB instead of 6dB), and linearity (-28.7dBm IP3 instead of -25dBm). This results in a further reduction of the power dissipation, as is shown in fig. 74. The current consumption is now reduced to 0.1mA. This is a reduction by a factor of more than 100!

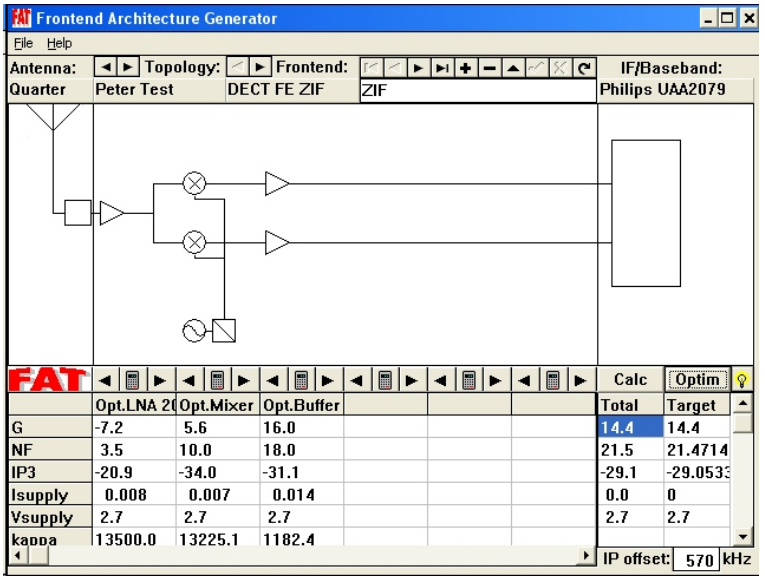


Figure 74 *Optimum distribution of gain, linearity, noise and power dissipation for DECT system specifications without design and production margins*

The accuracy of such predictions is limited, since scaling of circuits and devices over such a large range will probably introduce significant parasitic effects not included in the simple models on which this design method is based. This does not alter the relevance of this approach, though, since a power saving of even a single order of magnitude is already sensational in handheld, battery-powered devices!

5.9 Summary and conclusions

An effective and efficient minimum-power design method has been developed. The method is based on structure-independent transforms of subcircuits, which allow the transformation of the set of performance parameters of any circuit to another set of performance parameters. These OSITs have been used for two purposes: to develop a selection algorithm for building a library of subcircuits, and to calculate the optimum distribution of gain, linearity, and noise in a front end, including the corresponding minimum power dissipation.

As part of this work, a new class of figures of merit (FOMs) has been identified. Equivalent figures of merit, or EFOMs, are a subclass of FOMs, which identify the OSITs that can transform a circuit with a specific EFOM value into another circuit with any combination of specifications that result in the same EFOM value. Much more than just “another FOM”, EFOMs make the figure of merit much more meaningful: instead of only comparing performance, as ordinary FOMs do, EFOMs allow the transformation of performance to suit the needs for a specific application.

A design procedure that is based on this design method has been introduced, and covers the complete top-down trajectory from system specifications to testing. This

design procedure is supported by a special tool (FAT) that supports the storing and sharing of specifications and subcircuit information, as well as the transforms and calculations in the top-down trajectory. By using this tool, a power dissipation reduction by a factor of 2.7 has been predicted for the DECT front end from Chapter 4, using the same subcircuits as a starting point, and achieving the same specifications as the original front end. A spectacular power dissipation reduction by over two orders of magnitude has been predicted by relaxing the specification until it exactly meets the DECT system requirements.

Chapter

6

Low power boundary conditions

Thusfar, the investigation of minimum power front-end design has focused on the optimum implementation of front-ends for systems within pre-defined boundary conditions. These boundary conditions include system, circuit, and technology aspects. In this chapter, the impact of these boundary conditions on low power front-end design will be discussed. More specifically, changes in boundary conditions that result in reduced power dissipation will be investigated.

6.1 System

The design method discussed in the previous chapter considered system specifications to be a boundary condition, and focused on achieving the lowest power implementation of a system specification. However, decisions during the specification of a system can have a large impact on the power dissipation of RF transceiver components of this system. In this section, the impact of system aspects related to radio transmission itself will be discussed, and system choices that reduce the total power dissipation of the transceiver will be identified.

In any radio transceiver system, there is a relation between transmit power, receiver sensitivity and the maximum range at which reliable reception of the transmitted signal can be achieved. This relation is expressed through the radio transmission equation (cf Section 1.1). When defining a new system, the radio transmission equation points to several parameters that can be adjusted. These parameters affect system performance and transceiver power dissipation:

- *Transmit power*. The power dissipation of the transmitter depends on the transmit power, through the efficiency of the RF power amplifier.

- *Receiver sensitivity.* The minimum required power at the receive antenna depends on the receiver sensitivity. The receiver sensitivity, in turn, affects the receiver power dissipation.
- *Range.* The received power decreases with the square of the distance between the transmit and receive antennas. This decrease in received power needs to be compensated for by higher transmit power and/or improved receiver sensitivity. Both of these translate into higher power dissipation of the RF transceiver.
- *Frequency.* For constant antenna gain, e.g. when always using half wavelength dipoles, the path loss increases with f^{α}, where α also increases with frequency in most environments. Therefore, a higher frequency will have to be compensated for by higher transmit power and/or improved receiver sensitivity, both of which translate into higher transceiver power dissipation. The consequences of this effect, including the split into short-range and long-range systems, have already been discussed in Section 1.1.
- *Antenna gain.* The gain of the transmit and receive antennas can be traded for lower transmit power and/or reduced receiver sensitivity, which in turn reduce the power dissipation of the transmitter and/or receiver.

6.1.1 Transmit power and antenna interface losses

The transmit power has to be generated by the RF power amplifier. In addition, the RF power amplifier also needs to generate the power that will be lost in the antenna interface circuits. Since the efficiency of a typical RF power amplifier is already close to the physics-imposed limits (cf Section 1.3), a higher transmit power will have to result in higher power dissipation. The scaling is not fully proportional, since there is some power dissipation that is not dependent on the transmission power, e.g. for the LO generation, modulator, and mixers. This constant contribution to the power dissipation causes the total power dissipation of the transceiver to increase less than proportionally to the transmit power. In fact, at low power levels, it will asymptotically approach the constant power dissipation of these additional circuits. Also, devices do not scale accurately down to low current levels, due to parasitic effects. Qualitatively, the power dissipation of a transmitter will scale as shown in fig. 75. Therefore, decreasing transmit power to reduce transceiver power dissipation is very relevant, especially since the transmitter often consumes a significant part of the total power dissipation, but only down to the point where the contributions from other circuits become dominant.

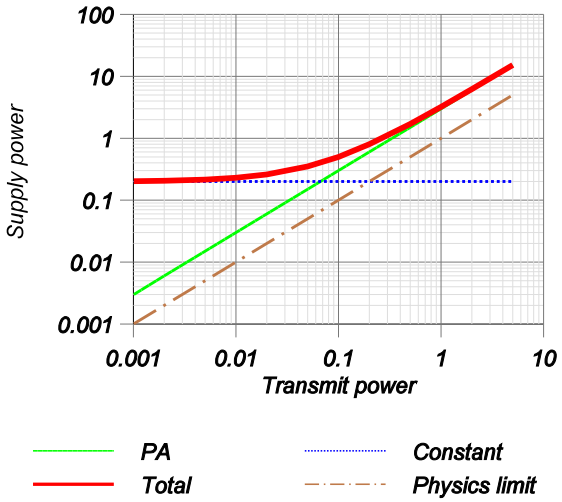


Figure 75 *Power dissipation of a typical transmitter as a function of transmit power to the antenna*

This effect is generally well-known, and usually taken into account in the definition of new systems.

6.1.2 Receiver sensitivity

There is also a strong impact of high receiver sensitivities on the receiver power dissipation. This effect can now be quantified using the theory of Section 5.5. As an example, the power dissipation for the DECT front end has been calculated using the FAT tool for various target noise figures. The results are shown in fig. 76.

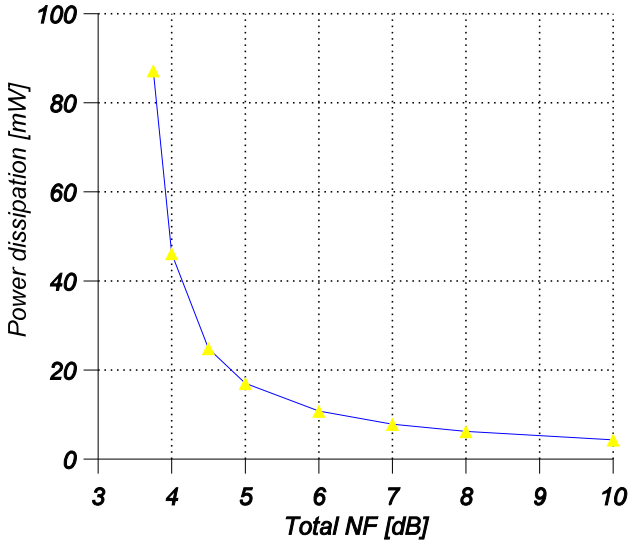


Figure 76 *Power dissipation as a function of noise figure for the DECT front end*

The impact of the total noise figure on the total power dissipation is especially strong for target noise figures within two dB from the noise figure of the first stage. The radio transmission equation shows that in a radio system, the same reliability and range can be achieved by a system in which both the transmit power and receiver sensitivity limit are increased, or decreased, by the same amount. The strong non-linear dependency of both the transmit power and the receiver sensitivity on their respective power dissipations, suggests that there is an optimum transmitter power/receiver sensitivity combination that achieves minim total transceiver power dissipation. This optimum depends on the properties of the transmitter and receiver, on the relative power dissipation of the transmitter and receiver, on the relative fractions of the time in which both are active, and many more factors. This is an area that requires further study.

6.1.3 Range

Since an increase in range results in higher power dissipation of the transmitter and/or receiver (see previous sections), keeping the range of a system limited helps to reduce the power dissipation. At the same time, a larger range is a benefit for the users in many systems. Often, it is even the main feature. Reducing range is therefore not an obvious or attractive proposition. The contradictory requirements for a long-range low-power system can be reconciled by using a cellular system in combination with roaming. In this way, the area in which a user can seamlessly use the radio link is much larger than the range of the radio link itself. This way, a system that behaves like a long-range system can be built up with short-range radio links.

6.1.4 Frequency

As discussed in Section 1.1, the need for higher data rates pushes short-range systems towards higher frequencies, despite the higher path loss at such frequencies. In fact, in combination with a cellular and roaming infrastructure, the higher path loss and resulting shorter range can be regarded as an advantage, since it allows smaller cell sizes and therefore a higher traffic density per area.

6.1.5 Antenna Gain

The effects of path loss on the power dissipation of the transmitter and receiver suggest that high antenna gains are very desirable from a low power perspective. However, the gain of an antenna is directly coupled to its directivity. For portable equipment, this seems to be a drawback, since there is no fixed orientation of the transmit and receive antenna with respect to each other. Antenna gain for portable equipment only makes sense if it is adaptive, i.e. if the antenna automatically and continuously updates its antenna pattern to have the maximum antenna gain in the direction where the signal is coming from (for the receive antenna), or the direction in which the signal should be going (for the transmit antenna).

In addition to increasing the average received signal power, adaptive antenna gain also decreases the variation of the received signal power. The received antenna power predicted by the radio propagation equation is an estimate of the average power that will be received at some distance between transmit and receive antenna. The received antenna

power will vary significantly around this average over small distances, due to multi-path interference. Indoor values in the 2.5GHz range show variations in the order of 10dB over distances of a few centimeters. A graph of a typical indoor signal strength distribution is shown in fig. 77 [97].

It is possible to significantly reduce the effect of multi path by antenna diversity schemes, in which the signals of multiple antennas are combined. These antennas can e.g. be at different locations, or they can have different polarization or different antenna patterns, etc. The signal from one of these antennas can be selected as input for the receiver, depending on the signal quality on each of the antennas, or the signals from several antennas can be combined in ways to further improve the total signal quality.

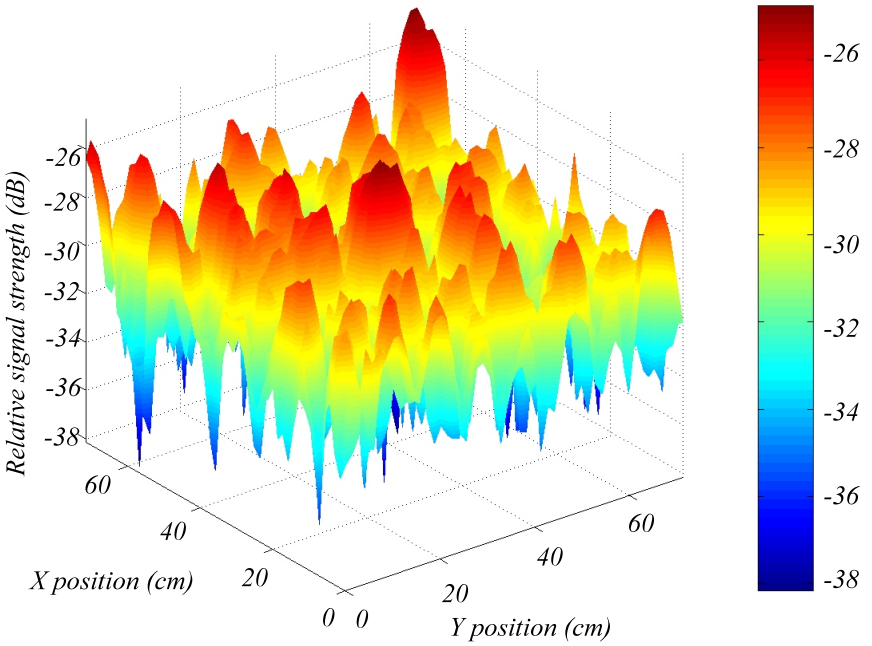


***Figure** 77 Indoor relative signal strength in a two-dimensional plane at 2.5GHz*

Because of the digital coding of the information in modern portable wireless devices, it is not the average received power that is relevant, but whether the power level is high enough to allow perfect reconstruction of the information after error correction. This is possible if the signal level exceeds a threshold. The fraction of the area inside (or outside) a building in which the received antenna power exceeds this threshold is called coverage [97]. For example, a coverage of 99% indicates that in 1% of the area the signal is too weak for reliable information transfer, but in the remaining 99% of the area, perfect information transfer can be achieved. For portable wireless devices that transmit and receive digital information, coverage is a more relevant performance parameter than average received antenna power.

Even relatively simple schemes, such as equal-gain combining of two antennas, can result in an improvement of about 10dB when targeting equal coverages in the 98% range indoor [97] (fig. 78). In equal-gain combining antenna diversity, the signals from multiple antennas are added with the same gain, but with an optimized phase difference. This

concept is shown in fig. 79. In angle-scanning diversity, the same approach is taken, but the optimized phase difference is determined periodically by scanning all possible settings of the phase difference. For speeds up to 5km/h and update frequencies of 1kHz, angle scanning performs as well as ideal equal-gain combining. Angle-scanning diversity has been patented [112].

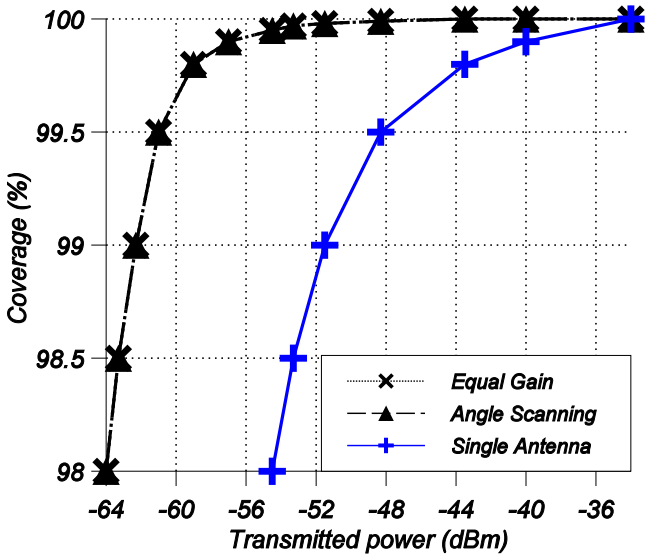


Figure 78 *Coverage of single-antenna and dual-antenna equal-gain and angle-scanning diversity*

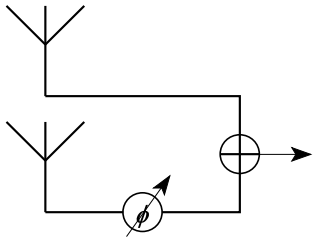


Figure 79 *Equal-gain combining concept*

One way to interpret the function of an equal-gain combiner is to model the output signal from the adder as if it was generated by a single antenna with a complex antenna pattern. In that case, the effective antenna pattern of this virtual antenna depends on the value of the variable phase shifter. Fig. 80 demonstrates some of the antenna patterns that can be generated in this way.

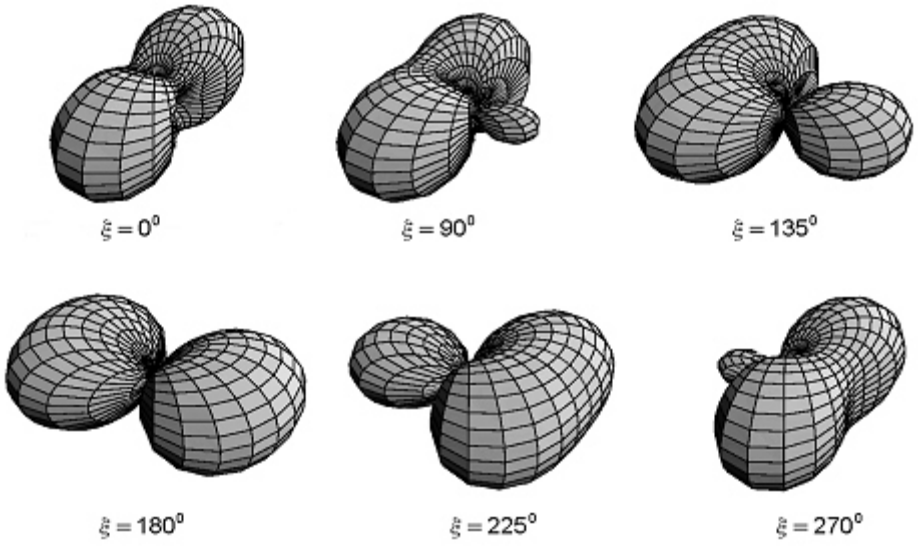


Figure 80 *Effective antenna patterns for different settings of the phase shifter [97]*

The approximately 10dB of link budget improvement that can be achieved with equal gain combining on one side of the radio link[7] can be used to increase coverage and range. However, it can also be used to reduce the power dissipation of a system, while maintaining the original coverage and range. For example, the transmit power can be reduced by 5dB and the receiver sensitivity by the "other" 5dB. In this way, approximately 3x power dissipation reduction can be achieved both in transmit and receive mode if the power savings are implemented on both transceivers in a link. An even larger power dissipation reduction can be achieved in asymmetrical systems, by reducing both the transmit power and receiver sensitivity by 10dB.

Figure 78 shows that the improvement depends on the target coverage. For higher coverage targets, the improvement is substantially more than 10dB. In figure 81, the relative power dissipation of the receiver is shown for both long-range and short-range applications, assuming that half of the improvement in link budget is allocated to reducing transmit power, and the other half to reducing receiver sensitivity. Complementary to the situation with antenna losses, the increased signal levels provided by antenna diversity can be used to optimize the front end for a proportionally increased noise figure and proportionally decreased gain, to maintain overall sensitivity and signal levels. In addition, the input IP3 of long-range systems has to increase proportionally to the signal levels to maintain overall linearity. The result is shown in fig. 81. In this figure, the relative power dissipation is the power dissipation of a receiver with equal-gain

[7] It is possible to implement advanced antenna diversity schemes on both sides of the radio link. Implementing diversity on one side of the link results in link budget improvements because of the antenna gain created by multiple antennas, and because of the suppression of multipath fading. Adding antenna diversity on the other side of a radio link as well will give additional antenna gain. The improvement because of the further suppression of multipath is, however, typically negligible. Therefore, adding diversity on the other side of a radio link will probably further improve the link budget by only about 3dB rather than another 10dB.

antenna diversity relative to a receiver without antenna diversity. A short-range system benefits more from antenna diversity, since there is no penalty (in terms of linearity) in dealing with the larger combined signals, whereas in a long-range system extra linearity is required for dealing with larger signal levels. The situation for long-range systems can be improved by using a diversity control algorithm that reduces interfering signals. This can be achieved by optimizing for a trade-off between received antenna power of the wanted signal and suppression of unwanted signals. The impact of such a scheme depends on the amount of interferers that can be expected, and is an area that needs further research.

In this simple model, the power dissipation of the overhead circuitry needed to implement antenna diversity has not been taken into account. In principle, the diversity combination can be carried out by passive components between the antennas and the receiver, without any additional signal-processing circuits. Also, with the predicted power dissipation savings of 2.5 to 30 times for this example, there would still be appreciable power savings in the brute force approach of doubling the complete receiver circuit.

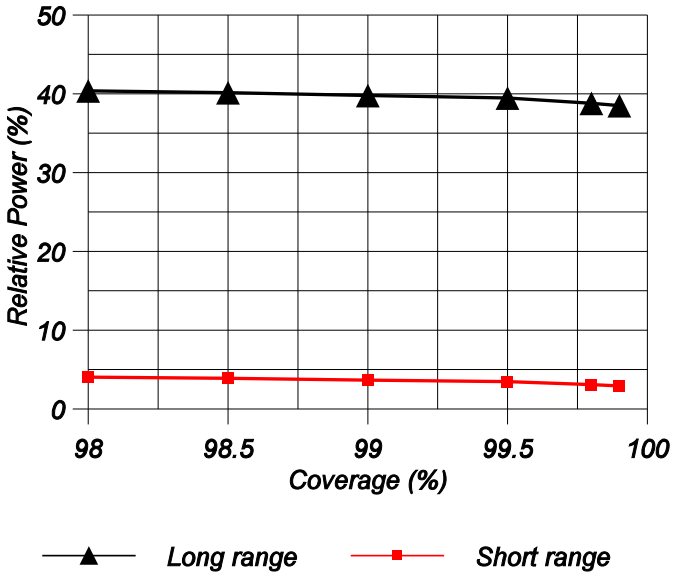


Figure 81 *Relative power dissipation of long-range and short-range systems with equal-gain antenna diversity versus coverage. A power dissipation of 100% corresponds to no antenna diversity.*

6.1.6 Antenna interface losses

Whereas antenna gain helps to reduce power dissipation, any loss in the interface to the antenna increases power dissipation. Such a loss has to be compensated for by an equal decrease in noise figure of the front end, in order to maintain the same overall noise figure (and therefore sensitivity). The decrease in input IP3 requirements of the front end is offset by an identical increase of the gain requirements. Typical losses in the antenna interface include losses in the antenna filter, duplexer and/or switches, and can add up to about 2dB or more.

Using again the DECT front end as an example, the effect of antenna interface losses on power dissipation can be estimated using the FAT tool. The result is shown in

fig. 82.

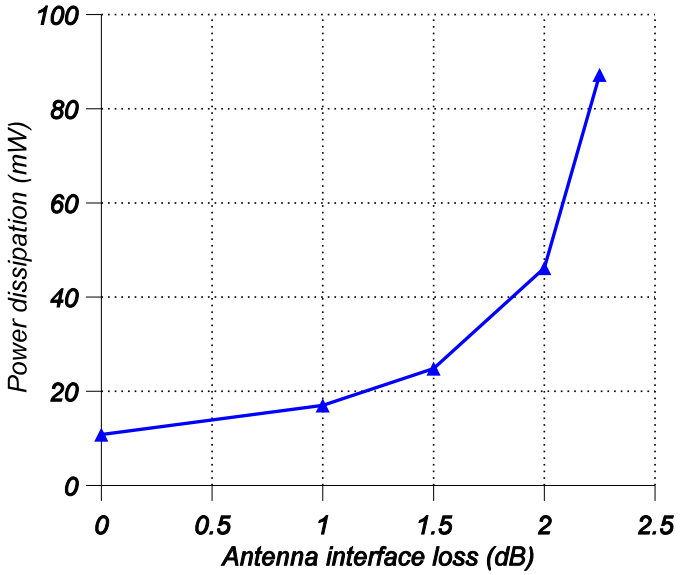


Figure 82 *The effect of antenna interface losses on the power dissipation of the DECT front end*

6.2 Circuit

In the design method described in the previous chapter, the performance of individual circuits was considered a boundary condition. The design method focused only on selecting the "best" circuit. In this section, the design of circuits that are optimized for low-power operation will be investigated.

Circuits for short-range systems operate at very high frequencies, close to the limits of the IC process. Therefore, the most effective circuit topologies are typically very simple. New low power circuit topologies, other than the scaling mentioned in this section, are not likely to be used for short-range systems. Circuits for long-range systems do benefit from improvements in technology, because the RF frequencies for these systems will remain limited to the region below approximately 5GHz. Therefore, the margin between required circuit performance and available technology performance will increase for these type of systems, and creative circuit topologies are more likely.

6.2.1 Circuits for short-range systems

In short-range, high-bandwidth systems, the current required to achieve sufficient gain at high frequencies usually results in higher linearities than needed. Therefore, the main problem will be in designing circuits with sufficient gain at high frequencies, and with low power dissipation. This can be achieved by reducing currents in the circuits (to reduce the power) while decreasing transistor sizes (to keep the available bandwidth constant), and increasing impedance levels proportionally to the current (to keep the gain constant), using OSIT_1 as the appropriate transform.

This approach is often limited by technology that does not allow further scaling of the transistors and impedances with proportionally scaled parasitics. In that case, parasitics of transistors, passive devices and/or interconnect start to dominate the achievable gain at the desired signal frequency. A technology that can scale much further towards small devices and low currents, called silicon-on-anything, will be discussed in the next section. Using such a technology, currents can be scaled down to the point where the linearity is just good enough, and further power savings will then require approaches similar to those for long-range systems as discussed in the second part of this section. The result of this approach is a circuit with high internal impedance levels.

Circuits with high internal impedance levels are often difficult to interface to external signals and components, since the characteristic impedance levels on a PCB tend to be limited to a few hundred Ohms for practical reasons[8]. This makes interfacing with the outside world awkward, and is an additional impetus for highly integrated transceivers with a minimum number of interfaces to the outside world. At these interfaces, the impedance of the signals has to be adjusted to external impedance levels either electronically or by passive components. Electronic matching will result in increased power dissipation as well as additional noise and distortion. Matching through passive components that are on-chip is costly because of the required chip area, and if the on-chip passive components are of limited quality (as e.g. on-chip inductors often are) signal losses are incurred. Matching through passive off-chip components is most common, but results in a higher overall component count. A better result could be achieved with a low cost IC technology that allows integration of high quality passive components. Fortunately, silicon-on-anything also offers high quality passive devices at low cost. An alternative solution would be the use of an optimized passive integration technology specifically for the input and output matching. Within Philips Semiconductors, such technologies are available, e.g. Passi, PICS, COSIP etc.

6.2.2 Circuits for long-range systems

Low-power circuits for long-range systems need to meet three main design goals:

- low noise figures at high frequencies
- good power/linearity (κ) values at high frequencies
- efficient output power generation

The noise figure is mostly limited by the properties of the active device and the availability of high-quality passive components for the matching networks. Good power/linearity (κ) values require high gain and/or good linearity at high frequencies and low currents. One way of achieving this is through distortion cancellation (cf. SITs, Section 5.3.1). In principle, a perfect cancellation can be obtained for a single distortion component. In practice, the reduction is limited by the matching that can be achieved between the branches. By recursively applying this procedure, in principle all distortion components can be eliminated. Obviously, this is impractical for more than two or three components because of cost, power dissipation, and accuracy of the circuits involved.

[8] The characteristic impedance Z_0 of a microstrip line on a PCB can be approximated by the impedance of a cylinder with diameter d above an infinite ground plane at distance D, in an environment with dielectric constant ε_r: $Z_0 = 138/\sqrt{\varepsilon_r} \log(D/d)$. Because of technological constraints, such as minimum track widths and maximum via heights, D/d cannot be made very large. This limits the characteristic impedance to a few hundred Ohms in practice.

In order to achieve good linearity, many RF front end circuits operate in class A. The efficiency of the generated output power for class-A circuits is rather poor. Even at maximum output levels it does not exceed 25% for a resistive load. The lowest signal levels in a front end are often five to eight orders of magnitude lower than the maximum signals, resulting in efficiencies as low as 25 10^{-10} for such signals. Achieving better power efficiency for low signal levels requires circuit topologies with power dissipation that adapts to the signal levels, for example class-A/B outputs or class-A circuits with adaptive biasing, dependent on the signal levels. B. Gilbert has introduced a mixer structure, called the "micromixer", that depends on a class-A/B balun in the bottom stage [98]. Also, emitter followers with class-A/B behavior are well-known, e.g. the topology in figure 83. These are examples of a trend towards more complex RF circuits for long-range systems, that take advantage of the increasing margin between required circuit performance and available technology performance.

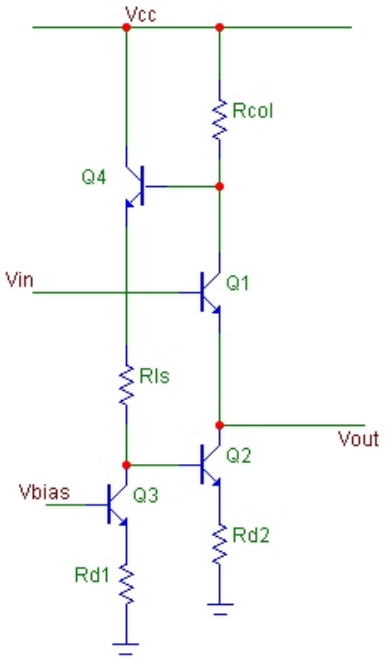


Figure 83 *Class-A/B emitter follower*

In the circuit of fig. 83, Q1 is the actual emitter follower. The biasing for Q1 is provided by Q2. The current through Q1 is sensed by Rcol. A feedback loop, consisting of Q4, Rls, and Q3, counteracts fluctuations of the current through Q1 by modulating the current source Q2. If the current through Q1 becomes smaller, the collector voltage rises. This rise in voltage is applied to the base of Q2 through the level shifter consisting of Q4 and Rls. This increase in base voltage is translated into a higher collector current of Q2, and therefore an increase in bias current for Q1. The voltage drop in the level shifter is determined by the bias current through Q3, and therefore by Vbias. Vbias also sets the quiescent current through the emitter follower. The modulation of the tail current provided by Q2 causes the class-A/B operation. It reduces current fluctuations in Q2, resulting in higher AC current gain of the emitter follower, and therefore higher input

impedance, lower output impedance, and less distortion. Unfortunately, the circuit is asymmetric, in that the current of the current source Q2 can never be negative. Therefore, if more output current is required than can be provided by the quiescent current of the device, the tail current in Q2 switches off, and the current through Q1 starts to change with the output current. The second-order distortion caused by this effect can be suppressed by using two of these emitter followers in balanced circuits.

Although the micromixer and the class-A/B differential pair provide useful building blocks for RF class-AB front ends, it is not yet practical to build a complete class-AB front end. Most RF class-A circuits are balanced circuits, because they offer better second order distortion, better power supply rejection, and because they generate less interference for other circuits. Class-AB circuits should preferably be balanced as well, for the same reasons. In addition, class-AB circuits will generate variable supply currents, which can generate significantly more interference in other circuits than the constant supply currents of a class-A circuit. This effect can be reduced by using balanced class-AB circuits, that reduce especially the supply current variations at the signal frequencies. What is left is supply current variations at double the signal frequencies, but these are often easier to deal with. The main missing circuit type is therefore a class-AB balanced amplifier, equivalent to a differential pair. Such a balanced gain stage could be constructed from two input stages of the micromixer, since one input stage converts a single-ended signal into a balanced signal. Using two such circuits with the outputs cross-coupled allows differential-input, differential-output operation. However, this topology has no common mode rejection at all. As part of this thesis, the following differential class-A/B amplifier has been investigated (fig. 84).

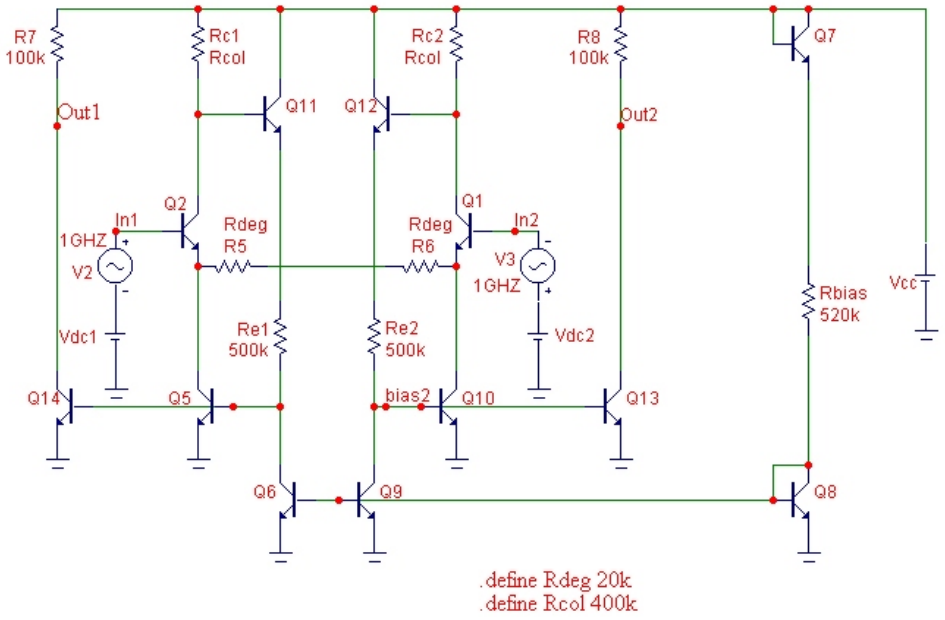


Figure 84 *Class-A/B differential pair*

In this amplifier, the transistors Q1 and Q2 form the actual differential pair. R5 and R6 are degeneration resistors to improve linearity. The usual tail current source has been replaced by two independent tail current sources Q5 and Q10. They are driven by collector-current sensing resistors Rc1 and Rc2, through level shifters created by Q11 and

Q12 in combination with Re1 and Re2. These feedback loops again counteract any changes in collector current of the input devices Q1 and Q2. In the class-A region, the current that is needed to compensate these differences in collector current is exactly identical to the current through R5 and R6. This difference current is mirrored through Q13 and Q14 to the outputs. By increasing the size of Q13 and Q14 compared to the current sources Q5 and Q10, additional power gain can be implemented.

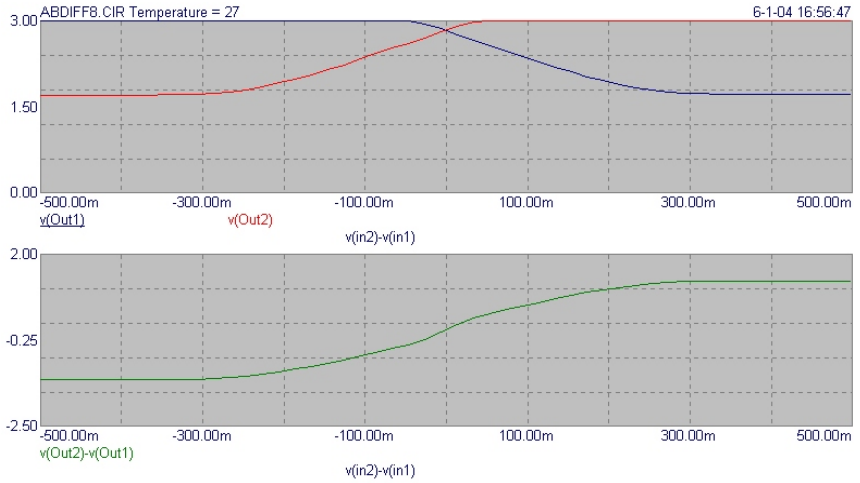


Figure 85 *DC transfer curves of the class-A/B differential pair: the top graph shows the single ended output voltages versus differential input voltage. The bottom graph shows the differential output voltage versus differential input voltage.*

In the upper graph of fig. 85, the individual output signals are shown as a function of the differential voltage across the input. Only in a small region around zero, both outputs are active and the circuit operates in class A. For input signal levels above 30mV, one output saturates and the other output continues. This provides a much higher compression point than would be possible with a class-A circuit and the same quiescent current. The lower graph of fig. 85 shows the differential output signals as a function of the differential voltage across the input. The higher gain in the class-A region is clearly visible in the center of this graph.

The modulation of the power supply current is shown in fig. 86, showing a reduction by more than a factor of 2. The small signal linearity is higher than the linearity of a class-A differential pair with the same quiescent current, since the feedback loops and output stages act as a translinear circuit that compensates for the exponential behavior of the transistors. However, there is still a significant distortion in the transition from class-A to class-B operation, since the saturation of one side of the circuit reduces the gain by approximately a factor of two. This differential pair is very useful in situations where it is necessary to occasionally deal with strong single-tone interferers. In the absence of such interferers, the circuit operates at low currents in class A, and is highly linear. When such interferers are present, the circuit switches instantaneously into class-A/B operation, and avoids the signal deterioration caused by the limiting that occurs in class-A circuits with low biasing and therefore low compression points.

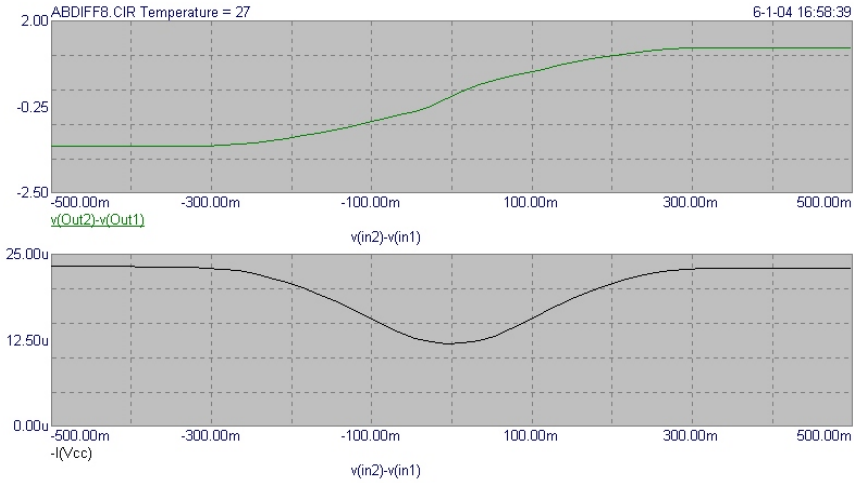


Figure 86 *The top graph shows the differential output voltage versus the differential input voltage of a class-AB differential pair. The bottom graph shows the supply current versus the differential input voltage of the same circuit.*

This circuit can be the basis for a full class-A/B front end. It does require sufficient gain at high frequencies and low quiescent currents for the control loops, and is therefore suited for the same type of IC processes required for low-power, short-range circuits. This class-AB differential pair has been patented [111].

6.3 Technology

The power consumption of an RF circuit depends, among others, on the properties of the devices in the circuit, and therefore on the technology used to implement them. The insights derived from the minimum power design method of Chapter 5 can be used to identify the properties that limit a further reduction of the power dissipation of RF circuits. This can be used as a basis for developing new technologies that are optimized for low power RF circuits.

To quantify the properties of devices and IC processes that affect the power dissipation of RF front-end circuits, various methods can be used, including:

- design, fabrication and measurement of benchmark circuits
- coupled device and circuit simulation [99]
- circuit simulation
- figures of merit (FOMs)

FOMs are an attractive and popular method, because they give a quick indication of the properties of a device and/or process, and can be calculated with a minimum amount of effort. Moreover, they allow for a very straightforward analysis of the factors in a device and/or process that can be changed to improve the results.

6.3.1 A new FOM for the gain of active devices

For short-range systems, the main performance criterium for the active device is gain at high frequencies and for low currents. For long-range systems, the performance criterium should also include linearity. In this subsection, a new FOM for short-range systems will be introduced, and in the next subsection this will be extended to an EFOM that is relevant for long-range systems.

The figure of merit for the bandwidth of active devices is currently still f_T, the (extrapolated) frequency at which the current gain of a transistor is reduced to unity. Unfortunately, f_T is a very poor indicator for the performance of low power RF circuits [110]. The main shortcoming is that f_T neglects several significant parasitic effects in a device because of the transistor configuration for which f_T is relevant is not representative for a typical RF circuit. The only configuration for which f_T accurately predicts the unity gain bandwidth would be a common emitter stage in which the output of the device is shorted, and the input is driven by an ideal current source. In low power RF circuits, such configurations are not used. Another currently popular figure of merit for active device bandwidths is f_{max}. However, f_{max} is not a very relevant FOM for most RF circuits either, since it requires perfect power matching at the input and output of the device. Since the input and output impedances of active devices at these frequencies are partially capacitive, a matching circuit requires inductors. In most IC technologies, inductors have a limited quality factor, and their large physical size makes them expensive. Therefore, most on-chip RF circuits don't use power matching, and require other figures of merit.

A new figure of merit should represent the bandwidth of a common emitter configuration with resistive load and significant gain at low currents. Therefore, the gain of such a configuration versus current and frequency will be discussed first. At low frequencies, the gain of a bipolar transistor is typically limited by the voltage drop across the parasitic emitter resistance R_e, that reduces the effective transconductance:

$$g_{m,eff} = \frac{g_m}{1 + g_m R_e} \tag{78}$$

in which g_m is the transconductance of an ideal bipolar transistor:

$$g_m = \frac{q\, I_c}{k\, T} \tag{79}$$

with q the electron charge, k the Boltzman constant, T the absolute temperature, and I_c the collector current of the active device. Another limitation for the voltage gain is the maximum value of the collector load resistor R_l:

$$R_{l,\max} = \frac{V_{cc} - V_{ce,\min}}{I_c} \tag{80}$$

in which V_{cc} is the power supply voltage, $V_{ce,min}$ the minimum collector-emitter voltage of the device, and I_c the collector current. This value of R_l is often not practical, since it only applies to very small signal levels. A more typical value of R_l would be ½ $R_{l,max}$, since this allows for maximum output signal levels. The maximum voltage gain that can be achieved by this active device can be found by combining (78) with (80):

$$G_{v,\max}(0) = g_{m,eff}\, R_{l,\max} = \frac{q}{kT}\,\frac{V_{cc} - V_{ce,\min}}{1 + q\, I_c R_e \,/\, kT} \qquad \textbf{(81)}$$

The only parameter in (81) which is significantly influenced by the device and process properties is R_e. The variation of V_{ce} is often negligible compared to V_{cc}. It might seem that R_e is therefore a good indication of the low frequency gain performance of a transistor. However, only the product $I_c\, R_e$ affects the gain - one is meaningless without the other. The effect of parasitic emitter resistance can be reduced by running the transistor at very low currents. This is already included in $G_{v,max}(0)$.

In general, the bandwidth of a common emitter stage is limited by a combination of at least 7 and possibly as many as 24 time constants [104]. In low power RF applications, two of them are dominant in most devices and processes. One time constant limits the frequencies that can get into the intrinsic transistor from the outside (input time constant), and another time constant limits the frequencies which can get out of the transistor (output time constant):

- The output time constant consists of the external load resistor R_l and the collector-base (C_{jc}) and collector-substrate (C_{js}) capacitor of the active device. This results in a definition (82) for the output bandwidth f_{out}: the frequency at which the voltage gain G_v has decreased by 3dB from an initial value $G_v(0)$.

$$f_{out} = \frac{g_m}{2\pi G_v(0)\left(1 + g_m R_e\right)\left(C_{js} + C_{jc}\right)} \qquad \textbf{(82)}$$

  It is conventional to define f_{out} for a low voltage gain of 10, since this is a typical value for the input stage of a low noise amplifier (83). Note that the output bandwidth increases with increasing collector current.

$$f_{out} = \frac{g_m}{20\pi\left(1 + g_m R_e\right)\left(C_{js} + C_{jc}\right)} \qquad \textbf{(83)}$$

- The input time constant consists of the base resistance R_b and a combination of base-emitter junction capacitance C_{je}, diffusion capacitance C_D, and the Miller-enlarged collector base junction capacitance C_{jc}. As a first approximation, the effective base resistance is assumed to be lumped in series, and all capacitors are assumed to be

connected in parallel. Using this simple model, the input bandwidth f_v is defined as the frequency at which the gain of a common emitter stage with has dropped by 3dB (84):

$$f_v = \frac{1}{2\pi R_b \left(C_{je} + \left(G_v(0) + 1\right) C_{jc} + C_D\right)} \tag{84}$$

With the usual assumption of Gv(0)=10, this becomes:

$$f_v = \frac{1}{2\pi R_b \left(C_{je} + 11\, C_{jc} + C_D\right)} \tag{85}$$

Note that the input bandwidth decreases when the collector current of the device increases, because C_D increases with the collector current.

The expression for f_v does not take into account the effects of the parasitic emitter resistor. This resistor serves as a feedback across the first gain stage, reducing the low-frequency gain, and increasing the input bandwidth by the same amount. This increased bandwidth is indicated by the symbol f_v'. The combined effect of $G_{v,max}(0)$, f_v, f_v', and f_{out} on the gain and bandwidth behavior of a bipolar CE stage is shown in fig. 87 for the case where f_{out} exceeds f_v':

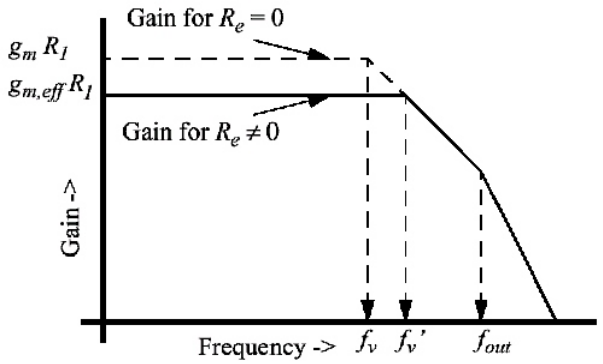


Figure 87 *Voltage gain as a function of frequency*

Whether f_v exceeds f_{out} or the other way around depends on the biasing of the transistor. In fact, a biasing condition exists at which the total bandwidth of the circuit is optimized: higher currents would reduce the input bandwidth, and smaller currents would reduce the output bandwidth. By extending fig. 87 to include the collector current as an independent variable, this optimum biasing point can be made visible. In fig. 88, such a graph is shown for a common emitter stage with a minimum size NPN device in the QUBiC process [102]. In this case, it is obvious that for biasing currents lower than the collector current for peak f_T, no optimum bandwidth: is found: increasing the collector current results in a continuously increasing bandwidth. This is also consistent with the experience

of many RF designers: “if you want higher frequencies, you need to spend more current”.

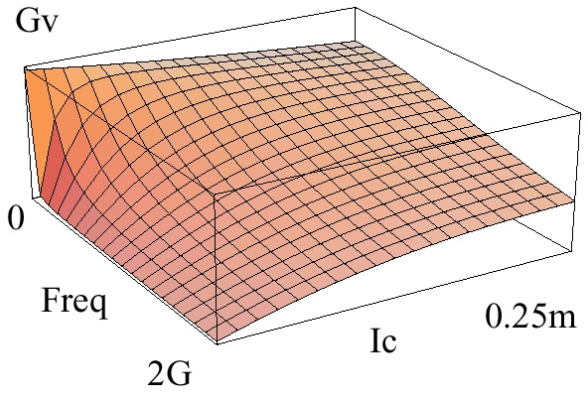


Figure 88 *Gain of a QUBiC1 CE stage versus frequency and current*

The difference between fig. 88 and fig. 87 is caused by the relatively low f_{out} of this device: f_v (1.6GHz) still exceeds f_{out} (1.2GHz) for the highest biasing condition. Therefore, performance is limited mainly by f_{out}. To obtain better performance from this circuit, the device should be optimized to improve the parameters that make up f_{out}, i.e. C_{js} and C_{jc}. In this case, the value of C_{js} is much larger than the value of C_{jc}. This is rather typical for many small bipolar RF devices, and can be solved by reducing C_{js} significantly, e.g. through trenches, such as in QUBiC4. The gain for such a device is shown in fig. 89. In this graph, an optimum biasing condition at high frequencies can easily be identified.

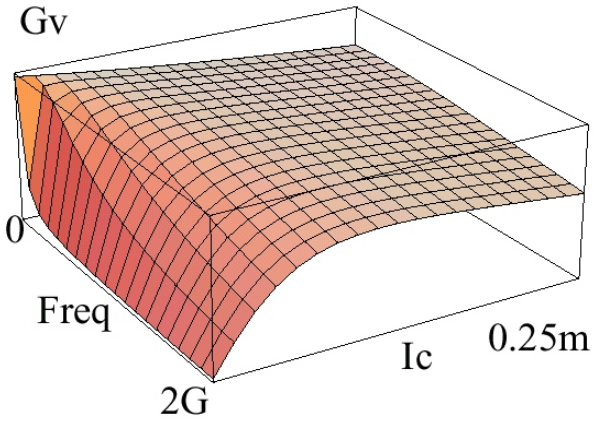


Figure 89 *Gain versus frequency and biasing for a minimum size QUBiC1 NPN device, with C_{js} reduced by 90%*

RF IC processes are often characterized by their f_T. However, f_T as a FOM for RF low-power performance is not very relevant, since f_{out} and f_v do not depend significantly on

f_T. This is further illustrated by comparing fig. 88 to fig. 90, in which the f_T of the transistor has been doubled compared to fig. 88 without any obvious effect on the high-frequency gain of the device. The negligible impact of f_T on the gain can be explained by the dominant time constants formed by relevant parasitic elements of the transistor, such as the base resistance and the collector-substrate capacitance, that are neglected by f_T. Therefore, optimizing the next QUBiC generations for f_{out} rather than for f_T is much more relevant for low power RF circuits. With the trenches in QUBiC4, this is indeed the direction that the technology is now developing.

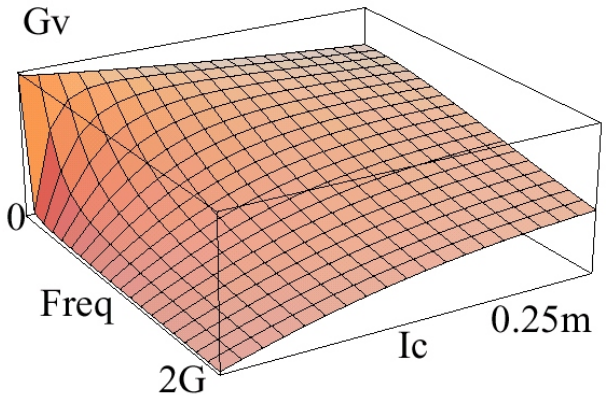


Figure 90 *Voltage gain as a function of frequency and current, of a minimum size NPN in QUBiC1 with f_T twice the original value*

The total bandwidth of the active device is sometimes defined by the figure of merit f_a. The bandwidth of a common emitter stage with resistive collector load such that $G_v(0) = 10$ is defined as the available bandwidth f_a. In this way, f_a combines the effects of f_v and f_{out} in a single FOM.

While the output bandwidth of most devices was dominated by the collector-substrate capacitance, f_{out} and f_a provided adequate indications of the RF performance limits of active devices. With new technologies, such as the trenches in QUBiC4, that significantly reduce the collector-substrate capacitance, this is no longer true, and a new FOM is needed. The bandwidth of a common emitter stage in an RF circuit is now significantly affected by the load of the next stage. This load depends of course on the specific design in which the common emitter stage is used. In some cases, it will be an emitter follower which provides a relatively low load to the CE stage, in other cases it will be another CE stage with a larger device and larger currents. As with the choice of $G_v(0)$ for f_{out}, a typical situation can be defined. In this case, such a typical situation is the loading of the CE stage by an identical CE stage, which provides a load somewhere between the two cases mentioned before. This results in a new FOM which will be named the collector bandwidth f_{coll} in this thesis. The corresponding expression for f_{coll} is given by (86).

$$f_{coll} = \frac{g_m}{2\pi G_v(0)\left(1 + g_m R_e\right)\left(C_{js} + C_{je} + \left(G_v(0) + 2\right) C_{jc} + C_D\right)} \tag{86}$$

The input bandwidth is not significantly affected by the loading of the next stage. Similar to the definition of the available bandwidth f_a of an unloaded common emitter stage, another new FOM is defined here which represents the bandwidth of a loaded common emitter stage as the combined input bandwidth and the collector bandwidth. It is named the loaded bandwidth f_l. This FOM is relevant when comparing modern technologies that use trenches or other techniques to reduce the parasitic capacitance at the output of the active device, especially the implementation of circuits for short-range systems.

For low currents, f_l is dominated by f_{coll}, and C_D will become negligible compared to the other capacitances in the denominator (C_{js}, C_{jc}, and C_{je}). Also, the term $g_m R_e$ will become negligible. Therefore, at low currents, f_l can be approximated by (87).

$$f_l = \frac{g_m}{2\pi G_v(0)\left(C_{js} + C_{je} + \left(G_v(0) + 2\right) C_{jc}\right)} \tag{87}$$

This bandwidth is proportional to the current. Therefore, the ratio between f_l and the collector current is a good measure for the bandwidths that an active device can achieve at low currents, and will be called the bandwidth-current ratio or *BCR* (88, 89, 89).

$$BCR = \frac{q}{2kT\pi G_v(0)\left(C_{js} + C_{je} + \left(G_v(0) + 2\right) C_{jc}\right)} \tag{88}$$

With the usual assumption $G_v(0) = 10$, this becomes:

$$BCR = \frac{q}{20kT\pi\left(C_{js} + C_{je} + 12\, C_{jc}\right)} \tag{89}$$

6.3.2 An EFOM for active devices

For long-range systems, *BCR* is not a sufficient indication for the low power limits of a device, since it does not take into account the linearity requirements of these systems. Another FOM is needed for this purpose, and can be based on the EFOM_1 that was introduced in Section 5.4. Since the EFOM_1 is valid for any two-port circuit, it can be applied to active devices in a two-port configuration:

$$EFOM\_1 = \frac{BW}{\kappa} = \frac{G\,BW\,IP3}{P} = \frac{G_v(0) f_l\, IP3}{I_c V_{cc}} \qquad (90)$$

At low currents, the term f_l / I_c can be replaced by *BCR*. Together with the usual assumption of $G_v(0)$ = 10, EFOM_1 for active devices can be rewritten as:

$$EFOM\_1 = \frac{10\ BCR\ IP3}{V_{cc}} \qquad (91)$$

The EFOM_1 value can be improved by the *BCR*. This has the additional advantage that the device becomes more attractive for both short-range and long-range systems. In terms of the FOMs discussed in the previous sections, this can be achieved by improving the f_{coll}, which, in turn, can be achieved by reducing the C_{je}, C_{jc} and C_{js} of the active device. Note that the diffusion capacitance C_D, and therefore the f_T, does not appear in the expression for *BCR*, and is therefore not relevant for the low-power, high-frequency performance of active devices in either short-range or long-range systems.

6.3.3 Influence of passive devices

With so much emphasis on active devices, both in the previous sections and in literature in general, it is easy to forget that other parts of an IC also affect the RF performance. In circuit simulators, bipolar transistors are often described in excruciating detail, using e.g. more than 50 parameters to describe a single transistor in the Mextram model. Parasitic effects of passive components, IC substrate, interconnect and package parasitics get much less attention, and are sometimes completely ignored, even though they can have a significant impact on the performance of high-frequency, low-power circuits [103]:

- In the DECT LNA, simulations indicate that the 15µm-long track of minimum-width bottom metal at the collector of the input transistor, reduces the bandwidth of this circuit by 10%.
- Metal plate capacitors are often required to achieve sufficiently high quality factors, but these capacitors can have up to 50% of their nominal value as parasitic capacitance to the substrate.
- Inductors in typical BiCMOS processes achieve quality factors of 10 or less, making them rather useless for low-noise oscillators or narrow-band filters. For a very low voltage LNA circuit (e.g. V_{cc}=1V), higher quality inductors would be very useful to compensate collector-substrate capacitances and achieve power matching.
- Crosstalk between different parts of the circuit through interconnect and substrate coupling are becoming more and more a cause of performance problems in high-frequency circuits.
- A combination of bondwire inductance (several nH), bondpad capacitance (about 1pF) and ESD protection diode capacitance (around 1pF) can have resonance frequencies around 2GHz, upsetting input matching and gain characteristics of high-frequency ICs.

- In some circuits, such as the DECT front end, crosstalk is already limited by the inductive coupling of the bondwires of the IC package.

The importance of these effects typically increases with frequency and impedance level. Therefore, these effects will be particularly relevant for future high-frequency low-power circuits. Also, many of these effects don't scale with device improvements, and special approaches might be required to deal with them. The process discussed in Section 6.3.4 also addresses high-quality passive components.

6.3.4 An optimized low-power IC process: Silicon on Anything

To achieve very high output bandwidths at low currents, the capacitance from collector to substrate should be reduced drastically. At the same time, the capacitance of the interconnect needs to be reduced as well, in order to allow for the higher impedance levels in the circuits. There are two ways in which the capacitance of interconnect can be reduced:

1. By reducing the dielectric constant of the material between the interconnect and the ground;
2. By increasing the distance of the interconnect to the ground relative to the line width of the interconnect.

The first approach is limited, since the relative dielectric constant of SiO_2, the most common inter metal dielectric, is around 4. A decrease to values around 2 is possible using low-k dielectric materials, but significant further improvements are unlikely. The second approach is limited, because the capacitance of interconnect will no longer scale proportionally once the width of the interconnect becomes much smaller than its height over the substrate. This effect is shown in fig. 91.

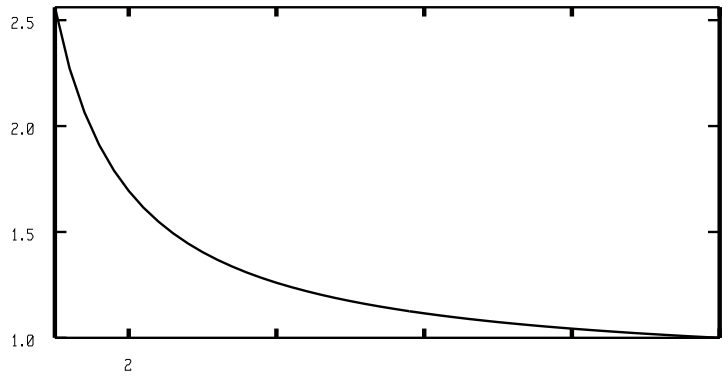


Figure 91 *Relative capacitance versus width/height ratio of interconnect lines*

The increase in oxide thickness that would be required to achieve an interconnect capacitance reduction by an order of magnitude is very impractical using standard IC-processing techniques. Therefore, a different approach is needed.

Both the scaled transistor and scaled interconnect parasitics have been addressed simultaneously in the Silicon-on-Anything (SOA) process [106]. This is a bipolar IC process in which the active layer of an (upside-down) SOI wafer is glued onto a substrate of choice (the *"anything"*) *after* processing, as shown in fig. 92.

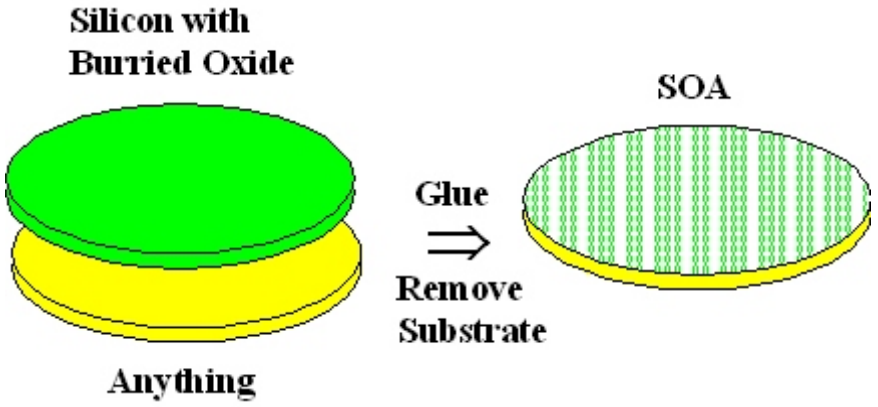


Figure 92 *The post-processing steps in SOA processing*

The substrate of the SOI wafer is now facing upwards. This silicon substrate is removed completely, using the buried oxide as an etch stop. Usually, glass is selected for the substrate because it is cheap and has low losses over a wide frequency range. From a design point of view, SOA offers five important advantages compared to a conventional bipolar silicon process:

- a lateral NPN transistor with 0.1µm^2 emitter area using 0.5µm lithography. A common emitter stage provides 2.4GHz of bandwidth at 20dB of gain with only 10µA current;
- interconnect (including bondpads) with 5 to 20 times lower parasitic capacitance to ground (depending on the line width);
- almost perfect isolation between circuit blocks;
- integrated inductors with Q values of up to 40;
- an *individual* trade-off between f_T, base resistance, Early voltage and breakdown voltages for *every individual* transistor in the design, by changing the collector drift region (parameter L_{cdr} in table 12).

L_{cdr}	*1.5*µ	*1.0*µ	*unit*
f_T	6.6	9.2	GHz
I_{fT}	14	63	µA
f_{max}	10.5	11.2	GHz
f_V	5.1	3.4	GHz
R_b	7.6	12.4	kΩ
C_{jc}	354	563	aF
C_{js}	550	550	aF
R_e	295	360	Ω

Table 12 SOA Device parameters for two sizes of the collector drift region

The process also provides 15kΩ/□ poly resistors, 1.5nF/mm^2 capacitors, PIN diodes, varicaps, PNPs and JFETS, and 2.5 metal layers, all in 14 mask steps, which together

with the 0.5µ lithography results in low fabrication costs. This allows cost-efficient integration of high performance input and output matching circuits. Fig. 93 shows the layout and cross section of an SOA transistor connected to a polysilicon resistor.

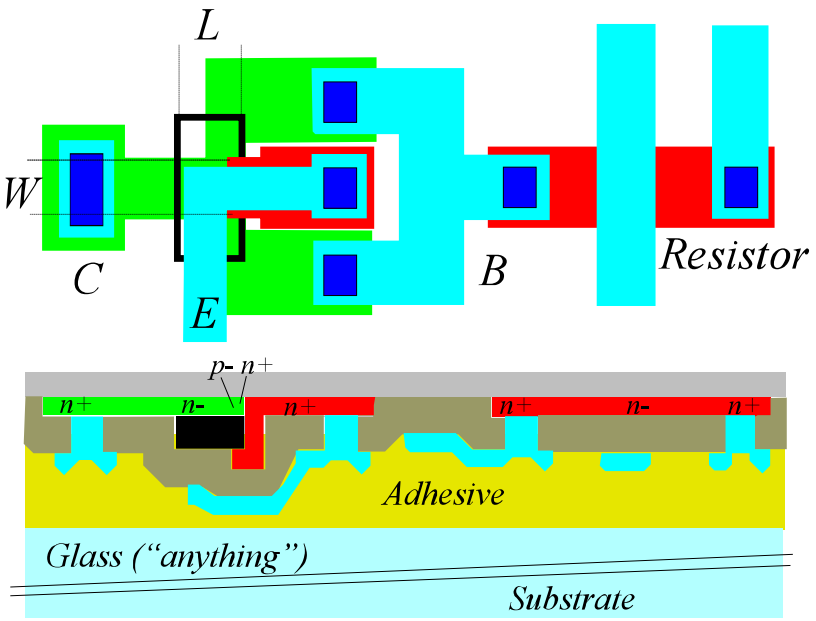


Figure 93 *Cross section of an SOA transistor*

To demonstrate that the goal of increasing gain at low power levels has been achieved in the SOA process, figure 94 shows the f_a of both a minimum-size SOA transistor and a standard 1µm QUBiC1 transistor.

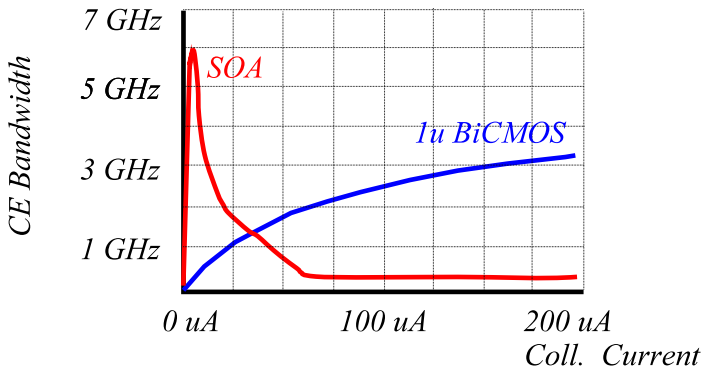


Figure 94 *Available common emitter bandwidth (fa) versus collector current for a minimum transistor in SOA and in QUBiC1*

The high bandwidth at low currents is a direct result of the scaling of the emitter area and proportional scaling of the transistor parasitics, enabled by the fully isolating substrate. The difference between the processes is also clearly captured by the newly introduced

FOM and EFOM_1 for active devices in this chapter, as shown in table 13. The *BCR* indicates about 20 times better low power performance of the SOA process compared to QUBiC1, and the EFOM_1 for active devices indicates an almost 2x better performance of the SOA process compared to QUBiC1. In other words: the SOA process is more than one order of magnitude better for short-range systems, and almost a factor of 2 better for long-range systems. This is consistent with the experience of designers who have designed circuits in this process: low power high frequency circuits work extremely well, but the advantage for circuits that need higher currents to achieve the required linearity is not nearly as big.

	SOA	**QUBiC1**
BCR	126MHz/μA	5.32MHz/μA
EFOM_1	293MHz	177MHz

Table 95 BCR and EFOM_1 figures of merit for SOA and QUBiC1 processes

These active devices, together with the enhanced passives and low-parasitic interconnect make SOA an ideal technology for low power RF design, especially for short-range systems.

6.4 Summary and conclusions

The minimum power design method that was introduced in Chapter 5 is, in addition to being useful as the basis for a design procedure, also a good basis for identifying the limits for further power dissipation reduction at the system, circuit and technology levels. At the system level, significant improvements are possible through advanced antenna diversity schemes. By using the extra margin in the link budget that is created by advanced antenna diversity for relaxing the requirements of the transmitter and receiver, the power dissipation of a system can be reduced by factors of 2 to 20, depending on the system.

At the circuit level, scaling to high impedance levels using OSIT_1 provides the most promising approach for short-range systems. Long-range systems, on the other hand, will benefit from an increasing margin between circuit requirements and technology performance. This increased margin can be used to reduce the power dissipation by more elaborate circuits than before. In time, RF circuit design for this type of system is likely to develop in the direction of analog/mixed-signal type of circuits. The first steps in this direction are already visible in class-AB circuits. In this section, a new class-AB differential amplifier has been introduced.

A convenient way of analyzing the requirements for new technologies is through FOMs, especially for active devices. A new FOM for active devices, named bandwidth-current ratio (*BCR*), has been introduced, which is relevant for low power RF circuits, especially in short-range systems. This FOM has then been extended to an EFOM_1 for active devices, which is relevant especially for long-range systems. The *BCR* has been used to identify the device parameters that are limiting further power reduction of RF

circuits in both system types. The junction capacitances C_{js}, C_{jc} and C_{je} are the dominant parameters for *BCR*. The diffusion capacitance does not have a significant influence, which confirms the lack of relevance of f_T as a FOM for low power RF circuits.

A new technology, called Silicon-on-Anything (SOA), offers exactly the improvements in *BCR* identified in this chapter. In addition to these optimized active devices, SOA offers the high quality passive devices and low-parasitic interconnect needed to build complete low power RF circuits. A comparison of BCR and EFOM_1 numbers between SOA and QUBiC1 shows an advantage of more than one order of magnitude for short-range systems, and a factor of about 2 for long-range systems. In the meantime, next generations of the QUBiC process family have been developed, but QUBiC1 is a relevant benchmark because of the comparable time frame, process complexity and cost of both processes.

Chapter

7

Low-power front end circuits

Many of the concepts and ideas described in this thesis have been used in the implementation of a new transceiver with very low power dissipation. Specifically, the OSIT_1 transform from Chapter 5 has been used to scale circuits to high impedance levels and low currents. Also, power dissipation reduction based on angle-scanning diversity, as described in Section 6.1, has been applied, as well as the short-range circuit optimizations in Section 6.2. This circuit has been implemented in the silicon-on-anything (SOA) IC process (Section 6.3). In fact, it was the first circuit developed in SOA. It was instrumental in the definition and optimization of the process. Finally, the IF IC uses the resistive interpolation, multi-phase differentiate-and-cross-multiply demodulation, and data reconstruction approach described in appendix E. The design method from chapter 5 was not completely developed yet at the time of this design, therefore only an early version could be has been used in this design. If the complete design method had been available, the total power dissipation might have been even lower than the results presented in this chapter!

The transceiver implements a proprietary[9] WLAN radio at 2.5GHz, and can be classified as a short-range system (cf Section 1.1).

[9]This transceiver is not compatible with any of the IEEE 802.11 standards, even though it operates in the same 2.5GHz ISM frequency band as the IEEE 802.11b and IEEE 802.11g systems. The data rate of this system is 100kbps, the transmit power 10dBm, and the modulation GFSK with BT=0.5.

7.1 Transceiver architecture

In a short-range system, the emphasis will be on achieving high bandwidths at low currents, rather than on linearity. Therefore, an IC process is needed that allows scaling to high impedance levels for both the devices and the interconnect, as modeled by the OSIT_1 transform (Section 5.3). Obviously, SOA (Section 6.3.1) is the most appropriate technology choice. In fact, this circuit was the first complete transceiver designed in the SOA technology. Since, at that time, the SOA process was new and not yet fully stabilized, the design was partitioned into 4 separate dies to reduce the risk, and to allow for easier testing. The partitioning chosen was:

- receiver front end
- VCO/synthesizer
- IF circuits & demodulator
- power amplifier

This closely resembles the partitioning selected for the RF platforms approach, as discussed in the next section. One advantage of this partitioning is that there are relatively few RF interfaces between the different dies, reducing the need for input and output buffers at RF to drive off-chip interconnect. Only the connections from the VCO/synthesizer die to the power amplifier and receiver front end are actually running at RF frequencies. A further partitioning into smaller circuits would have further reduced the risk, and allowed for easier measurements, but the many RF interfaces would have caused a significant increase in the power dissipation.

SOA transistors have an f_T of less than 10GHz. The success of this 2.5GHz transceiver, at more than 25% of the f_T, demonstrates once more the lack of relevance of f_T, at least for low-power RF front ends (Section 6.3).

To further reduce the power dissipation, angle-scanning antenna diversity was implemented, as described in Section 6.1. This improved the link budget by almost 10dB. These 10dB were traded for reductions in transmit power, sensitivity requirements, and interfering signals, and to improve coverage. A GFSK modulation was chosen for the system. This allowed for non-linear processing of the signals at IF and in the demodulator of the receiver, using interpolation and reconstruction as described in appendix E. GFSK also enabled direct modulation of the VCO during transmit, eliminating the need for up conversion mixers.

For the design of a low-power receiver in SOA, two areas need special attention:

- processing high dynamic range signals (weak signals and strong interferers)
- interfacing to external circuits

Processing high dynamic range RF signals is usually carried out in class A stages to minimize distortion, especially since the current for class-A operation is often needed anyway to achieve sufficient gain-bandwidth products. In SOA, gain-bandwidth products of 25GHz are achieved at collector currents as low as 10μA. Angle-scanning diversity avoids having the power dissipation of most of the circuits dictated by the dynamic range of the antenna signals. In angle-scanning antenna diversity, the signals of a number of antennas are first shifted in phase, and then added to form an effective antenna pattern with one or more beams and nulls. The beams increase the level of desired signals, and

the nulls reduce the level of interferers. This is usually implemented by programmable phase shifters at RF, but they are difficult to implement at 2.5GHz.

In this design, a zero-IF architecture has been chosen. A block diagram of a dual-beam, equal-gain antenna diversity receiver is shown in fig. 95. The phase shifting is carried out after the quadrature down conversion by adding weighted I and Q signals. Shifting the original signals I and Q over an angle ϕ to obtain the phase-shifted signals I_{ps} and Q_{ps} is implemented as:

$$\begin{cases} I_{ps} = \cos(\phi)\, I - \sin(\phi) Q \\ Q_{ps} = \cos(\phi) Q + \sin(\phi) I \end{cases} \quad \textbf{(92)}$$

The weighting is accomplished through the 4 variable-gain amplifiers in fig. 95. These variable-gain amplifiers are implemented as Gilbert cells, and are controlled by dual 8-bit DACs that generate the voltages corresponding to the cos(ϕ) and sin(ϕ) terms in eq. 92.

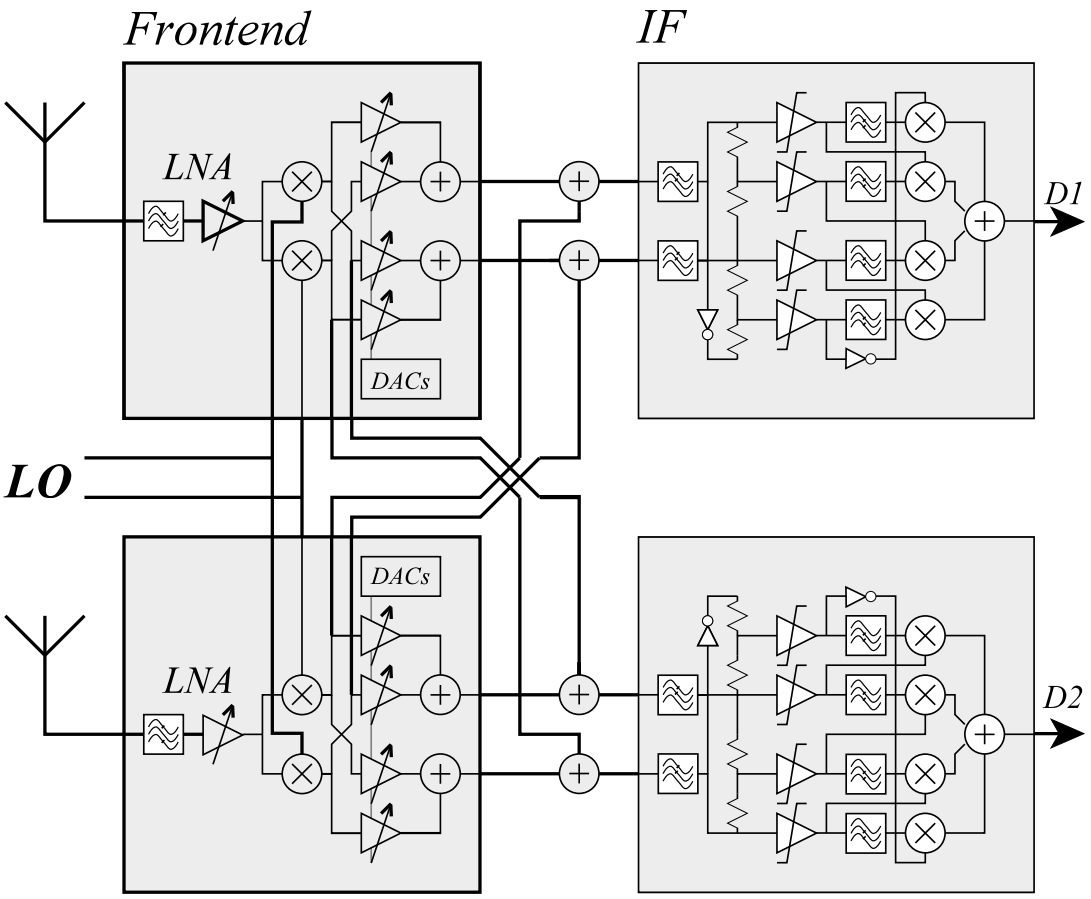


Figure 95 *Block diagram of the complete receiver*

This method of phase shifting provides the highest flexibility, since the weighting factors can be adjusted in very small steps, which is especially important for combinations of a larger number of antennas.

For two antennas, a simpler circuit is possible as well, using resistive interpolation (fig. 96). In this circuit, the I+ and I- signals are the balanced outputs of the "I" mixer in

the zero-IF receiver, and the Q+ and Q- signals are the balanced outputs of the "Q" mixer. The resistive interpolation provides phase shifting in steps of $\pi/4$. Typically, this is sufficient for a two-antenna, angle-scanning diversity system, but smaller steps can be implemented easily by additional taps in the resistive divider. This divider is linked in a diamond across the four phases that come out of the balanced outputs of the two mixers. In the schematic, this is accomplished through the wire connecting the bottom of the lowest resistor with the top of the highest resistor. In a layout, a more symmetrical approach can be achieved by folding the resistors. The series resistors of R, together with the load resistors of $(1+\sqrt{2})$ R, provide an attenuation such that the output signal has the same amplitude at every phase setting. These resistor values assume a voltage source and an infinite impedance load at the output, as well as zero "on" impedance of the NMOS switch transistors. In practice, the resistor values need to be adjusted to take imperfect source and load impedances, as well as realistic "on" impedance of the NMOS devices, into account.

The gates of the NMOS devices are controlled by digital signals in such a way, that only one device out of every group of eight is switched "on", and the other seven are switched "off". Therefore, the output signal corresponds to the signal at one of the nodes of the resistive divider. This circuit provides one balanced pair of outputs. The second pair of balanced outputs needed in a zero-IF receiver can be created in a similar way, by adding two more groups of eight NMOS devices to the taps on the resistive divider.

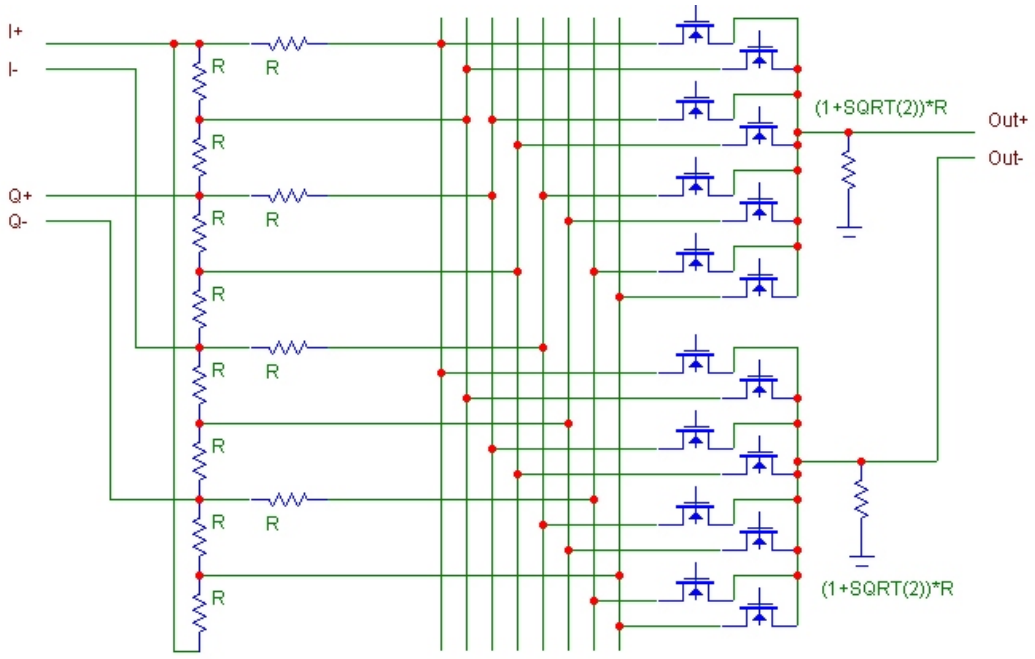


Figure 96 *Diagram of the alternative, interpolation type, phase shifter*

Resistive dividers can also be used for diversity schemes with three or more antennas. However, such schemes typically use maximum ratio combining rather than equal gain combining. For maximum ratio combining, both the amplitude and the phase of the individual antennas needs to be adapted. This requires a two-dimensional resistive network, which is more complicated to design and implement, especially if crosstalk and

parasitic phase errors need to be minimized. This implementation for angle-scanning antenna diversity, in which the phase-shifting is achieved through weighted addition of quadrature IF signals has been patented [113].

Since MOS devices where not stable yet in SOA at the time of this design, a phase shifter circuit with DACs and variable gain amplifiers was selected instead of this more simple circuit.

7.2 Receiver front end

Interfacing to external circuits is complicated by the high impedance levels on chip, in this design between 5kΩ and 50kΩ. The off-chip levels are, for practical reasons, around 50Ω.

This can be solved through passive LC-type transformers and/or electronic buffer stages. Since the first solution consumes a lot of die area, and the second a lot of current, reducing the number of external connections is very important for low-power RF ICs. This was one of the reasons to select a zero-IF architecture. Ultimately, the total receiver would be implemented as a single IC. With the current partitioning, a simple receiver can be built from 3 ICs: front end, IF, and VCO/synthesizer. A full dual-beam diversity receiver consists of 2 front-end ICs, 2 IF ICs and 1 VCO/synthesizer IC. The VCO/synthesizer IC is discussed in [83]. It consists of an LC oscillator at twice the RF frequency, followed by a traveling wave divider that generates the quadrature signals, and a programmable divider chain.

The receiver front end has an RF input that uses a combination of LC circuits and electronic buffers to achieve the required impedance transform. The schematic diagram of the LNA is shown in fig. 97. It is implemented with an integrated LC-type transformer that feeds into a common-base stage with 300Ω input impedance. The common-base transistors Q1 and Q2 each consist of 7 minimum transistors in parallel. Each minimum transistor is biased at 24μA. The measured s_{11} at the RF input is -12dB. The output of the common base stage is connected to two differential pairs which provide variable-gain operation. These feed into the RF inputs of the mixers. The LNA provides 20dB of gain at 336μA with a 4dB noise figure and -20dBm IP3. The higher noise figure compared to many other front ends is caused by the common-base input stage, but is acceptable because of the path loss improvement from the angle-scanning diversity. The common-base input stage helps to achieve good linearity performance at low currents.

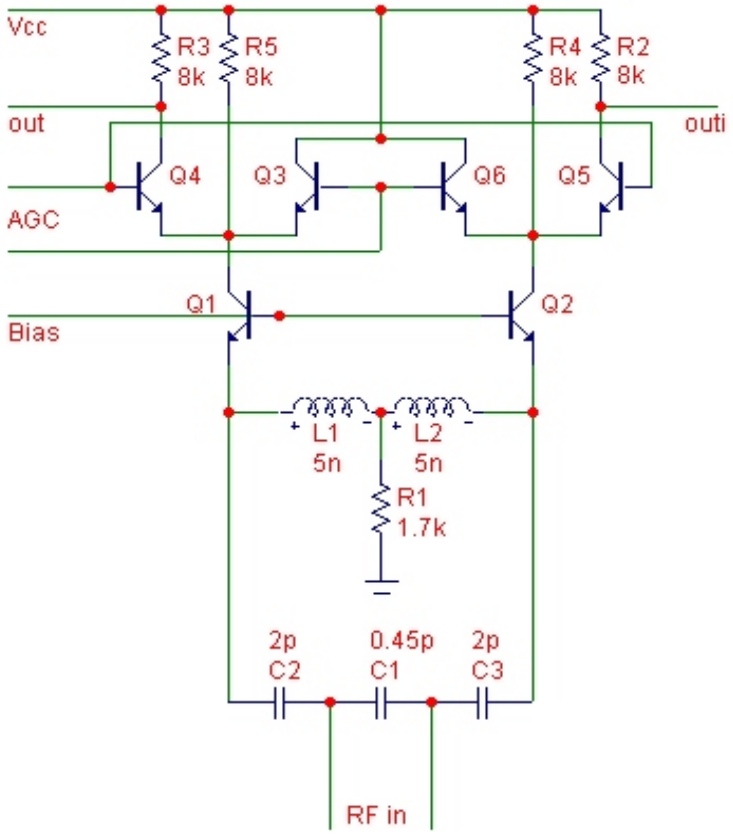


Figure 97 *LNA Schematic*

When comparing this LNA with the LNA from the DECT front end from Chapter 4 and the LNAs from literature as discussed in Section 2.1, the following EFOM values are found (fig. 98).

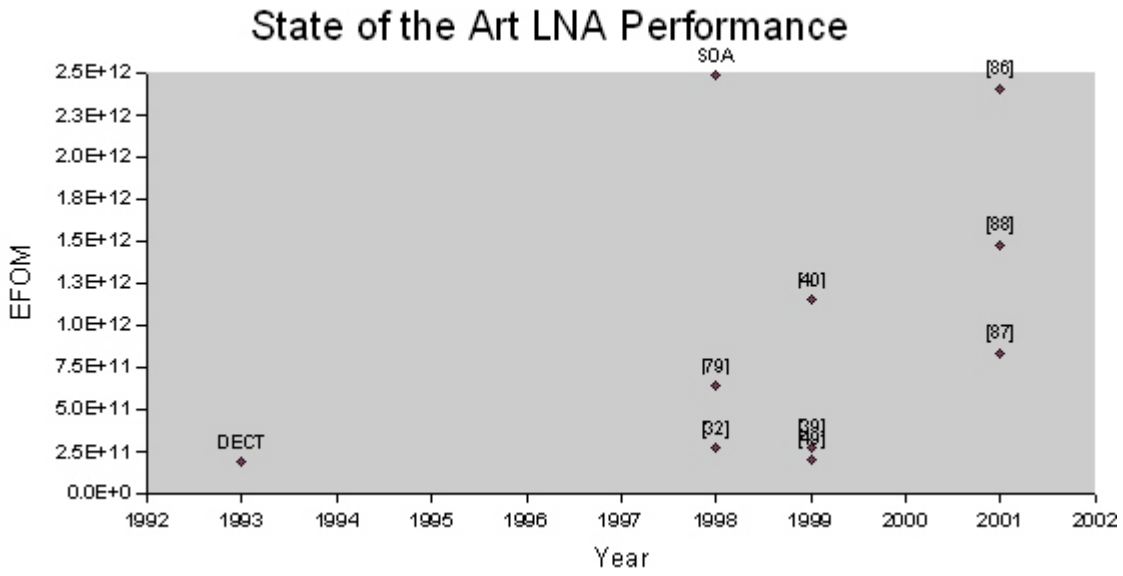


Figure 98 *EFOM of SOA LNA in context of DECT and other LNAs from literature*

As this graph shows, the low-power performance of this LNA is better than any other LNA in this graph. The fact that such a good EFOM value is achieved at such a low power consumption level, namely 1mW, demonstrates how well SOA technology allows the scaling of circuits to low currents without affecting the EFOM value.

The OSIT_1 describes the scaling of linearity and current without affecting the EFOM value. Scaling upwards, i.e. increasing the current and linearity, is very straightforward, and is accurately predicted by OSIT_1. Scaling downwards, to lower currents and linearities, is limited by the scaling properties of the technology used, as discussed in C. The high EFOM values at low currents that are achieved with SOA prove that this technology is capable of accurate scaling down to low currents, thanks to the small, lateral active devices, the high-quality passive components, and of course the low-parasitic interconnect.

The LO inputs are implemented with electronic buffers. In a single-chip version of this receiver, the LO will obviously be an internal connection which doesn't need these buffers. The buffers are implemented as differential pairs and provide 8.5dB of gain. The mixers themselves are Gilbert cell topologies and provide 17dB of gain, whereas the phase shifters have a maximum gain of 0dB. The total die area of the front end IC is 9mm^2, but a significant part is used for the digital control and DACs of the diversity circuits. The measured performance shows an overall gain of 35dB, with 6.2dB noise figure and -22dBm IP3. The IP3 measurement result is shown in fig. 99.

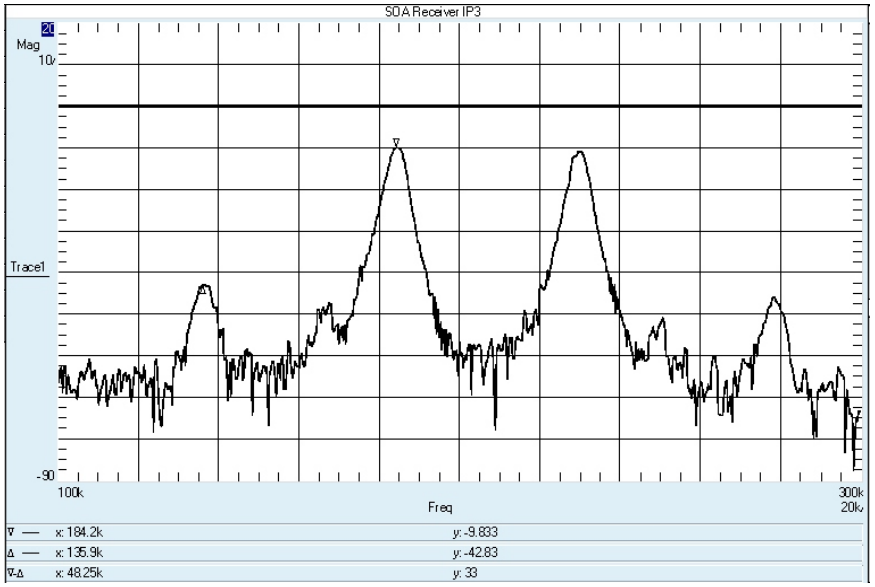


Figure 99 *IP3 Measurement result*

The supply current of the total receiver front end is 1.0mA at 3.0V, with about half that current going into the LNA and mixers, and the other half into the DACs and phase shifters, which can be powered down independently. Biasing is provided by on-chip bandgap reference sources. The complete die is shown in fig. 100. Although this die is "upside-down" after the substrate transfer, the devices are still visible because the original substrate has been removed. The replacement glass substrate below the devices

is transparent. This explains the shadows from the bondwires and large metal structures such as bondpads and inductors.

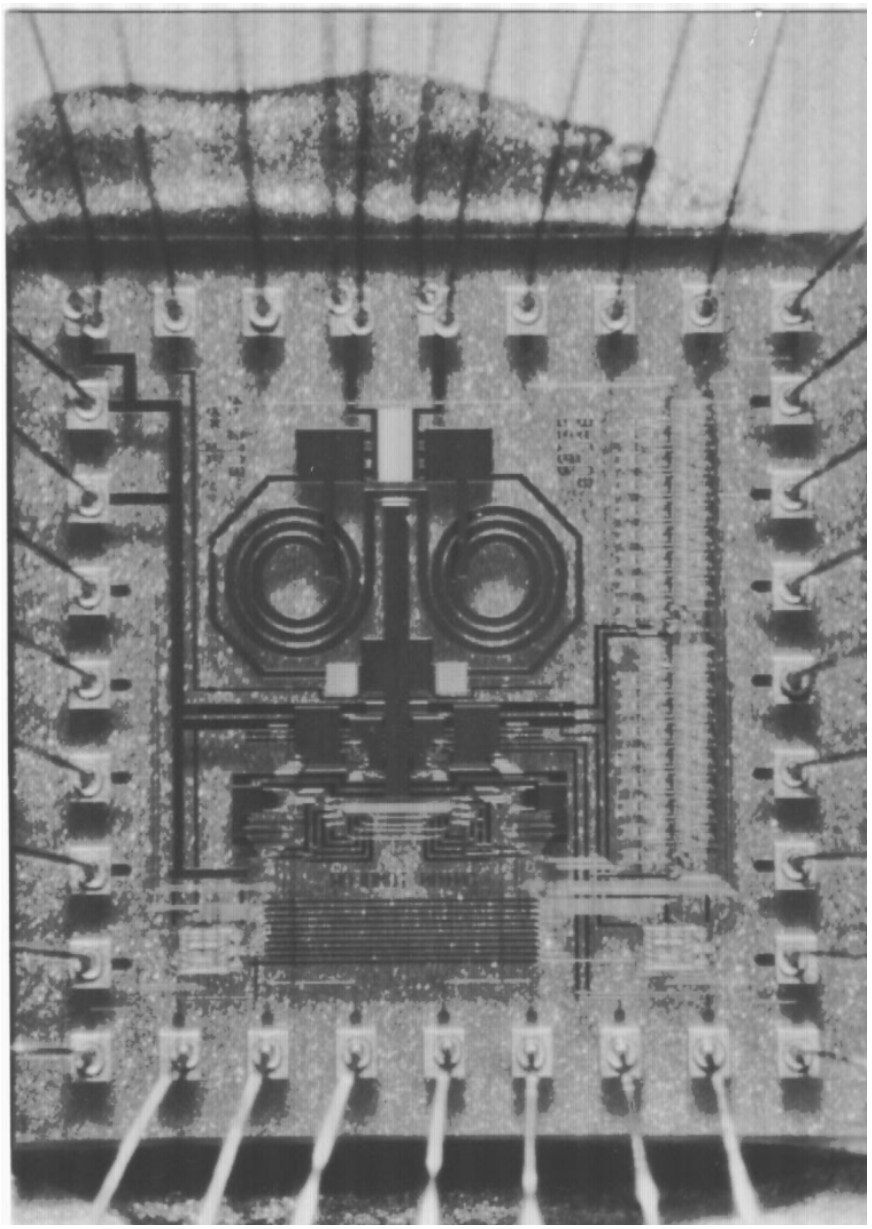


Figure 100 *Die photo of the SOA WLAN front end*

7.3 IF and demodulator

The IF IC (see fig. 95) consists of dual 4th order active lowpass filters, with a bandwidth of 100kHz, that provide channel selectivity. To fit a high-bit-rate signal into a frequency-modulated channel with a limited bandwidth, the frequency deviation is kept small, in this case at approximately half the bit rate. Since this results in few zero-crossings of the IF signal for each bit, special attention is needed for non-linear processing in limiters and demodulator.

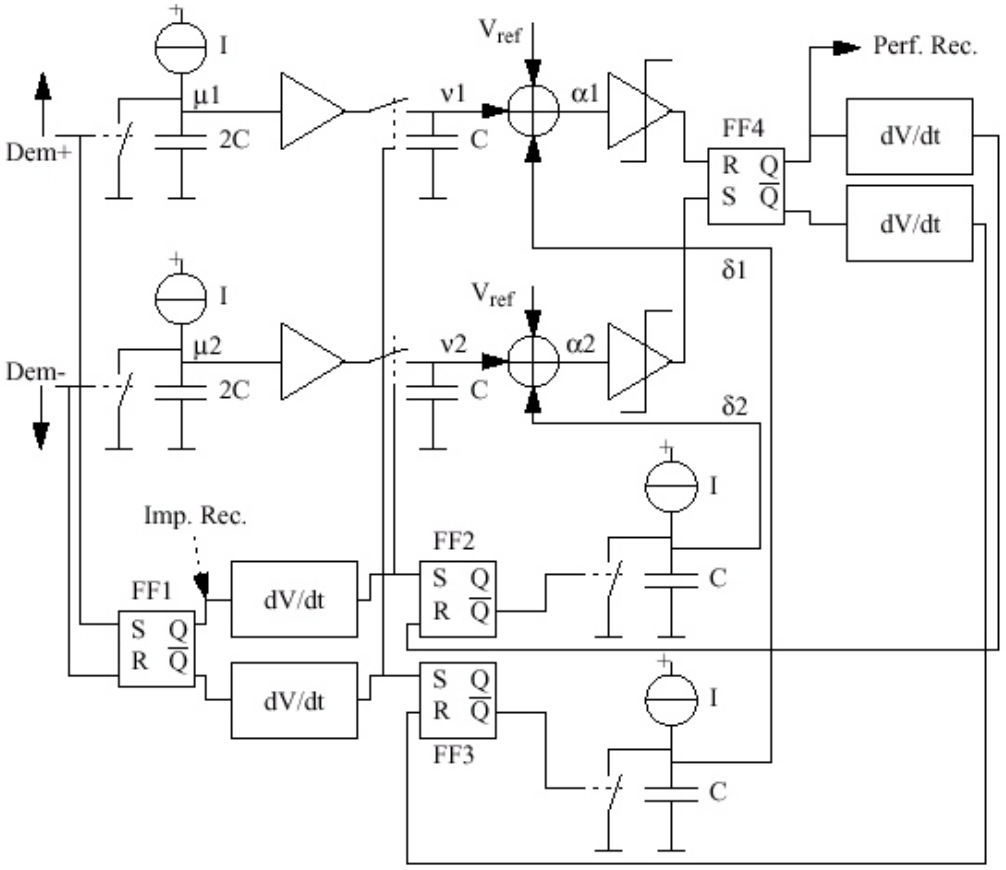


Figure 101 *Reconstruction circuit*

As described in appendix E, this can be solved by creating multiple IF signals at different phase offsets. Such additional IF signals can be created by resistive interpolation, in the same way as the phase generation for the antenna diversity (fig. 96). In fact, the same circuit could be used by adding extra switches. However, this would increase the number of switches in the overall system. Also, additional channel filters would be required. Since this would consume additional power, and since the filtering and interpolation are both linear operations, they can be done in a different order without affecting the result: first filtering, then interpolation.

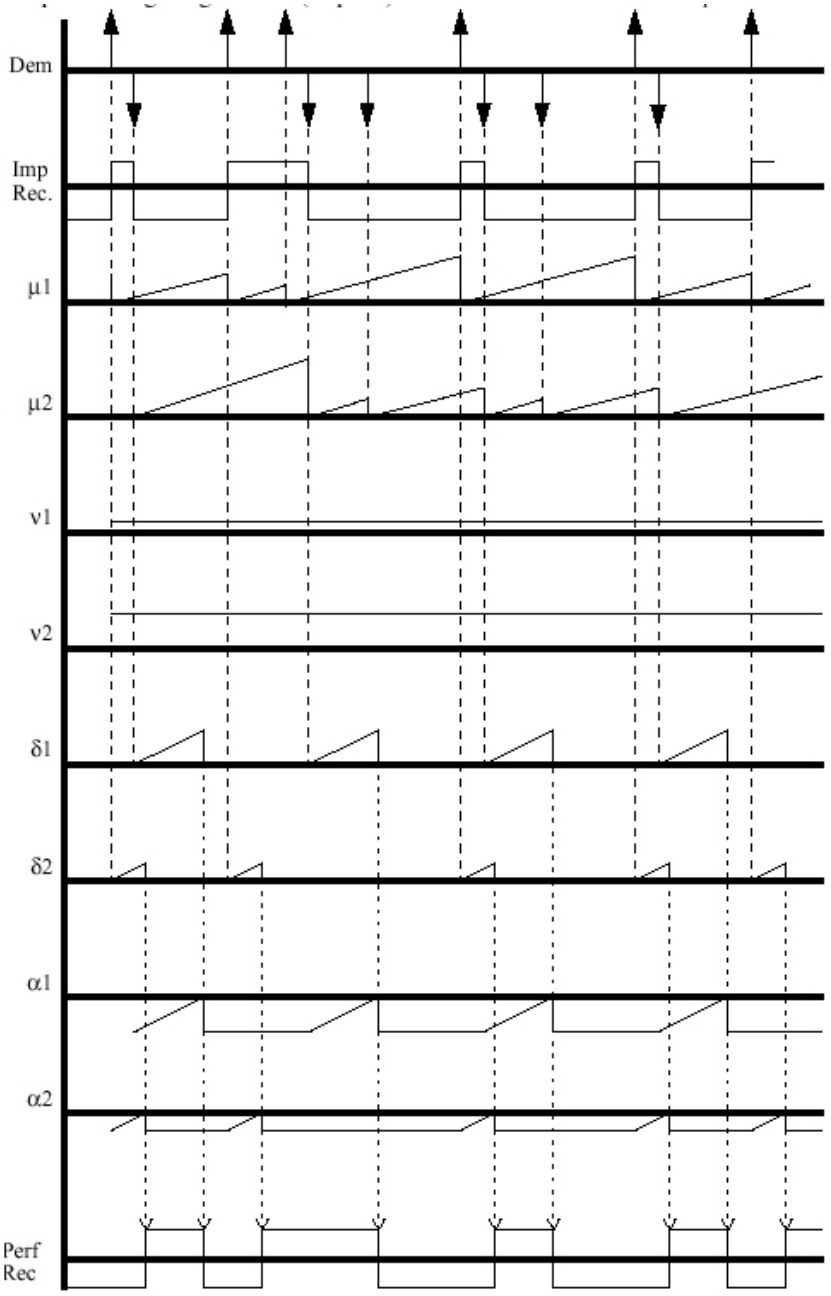


Figure 102 *Signals in the reconstruction circuit*

After the interpolation, limiters provide the high gain required to drive the demodulator with sufficiently large signals. They also eliminate any amplitude variation. The IF signals are then converted to baseband information by a 4 branch differentiate-and-cross-multiply demodulator, as shown in fig. 162 in appendix E. The demodulator provides pulses at 8 times the IF frequency, with positive pulses for positive frequencies and negative pulses for negative frequencies. This is sufficient for a reliable detection of the data. However, there will still be some jitter on the data transitions, so bit clock recovery might be affected. This is corrected by a reconstruction circuit based on the

reconstruction algorithm of appendix E.

This circuit is connected to the outputs of the differentiate-and-cross-multiply demodulator. A schematic diagram of the reconstruction circuit is shown in fig. 101. Some of the signals in this circuit are shown in fig. 102. The purpose of this circuit is to shift the transitions of the demodulated data signal to the correct position, which is a fixed offset in time after the time halfway in between the two adjacent pulses with opposite sign that caused the data transition. The principle of this circuit depends on measuring the time through the voltage of an integrator with a fixed input. By resetting these integrators when a positive or negative pulse occurs, the time since the last positive or negative pulse can be determined from the output voltages of the respective integrators. Because the time intervals are represented by voltages, a voltage representing a moment in time halfway in between two pulses can be determined by taking the average of the voltages representing the time intervals. Adding an offset in time can now be represented by adding an offset voltage to the output of an integrator. The operation of the reconstruction circuit is described in more detail in the next paragraph.

The positive and negative pulses of the four-branch derivative and cross-multiply demodulator are fed into the top part of the circuit. The demodulator output is shown as "Dem" in the signal diagram. In fact, the positive and negative pulses are provided separately to individual inputs of the reconstruction circuit, labeled "Dem+" for the positive pulses, and "Dem-" for the negative. In the top part of the circuit, two voltages $\nu 1$ and $\nu 2$ are created that are proportional to the desired variable time delay, which is half the time between successive opposite pulses. The signals "Dem+" and "Dem-" operate switches in parallel to capacitors with a value 2C. When a switch is closed by a pulse, the voltage across the capacitor will be reset to 0V. After the switch opens again, the voltage across this capacitor ($\mu 1$ resp. $\mu 2$) will start increasing because of the current I from a current source that flows into each capacitor. The voltages across these capacitors are copied as a voltage on the output capacitors ($\nu 1$ and $\nu 2$), when the "Imp. Rec." signal changes. By using the "Imp. Rec." signal for this rather than the "Dem+" and "Dem-" signals directly, the reconstruction circuit deals correctly with multiple demodulator output pulses with the same polarity.

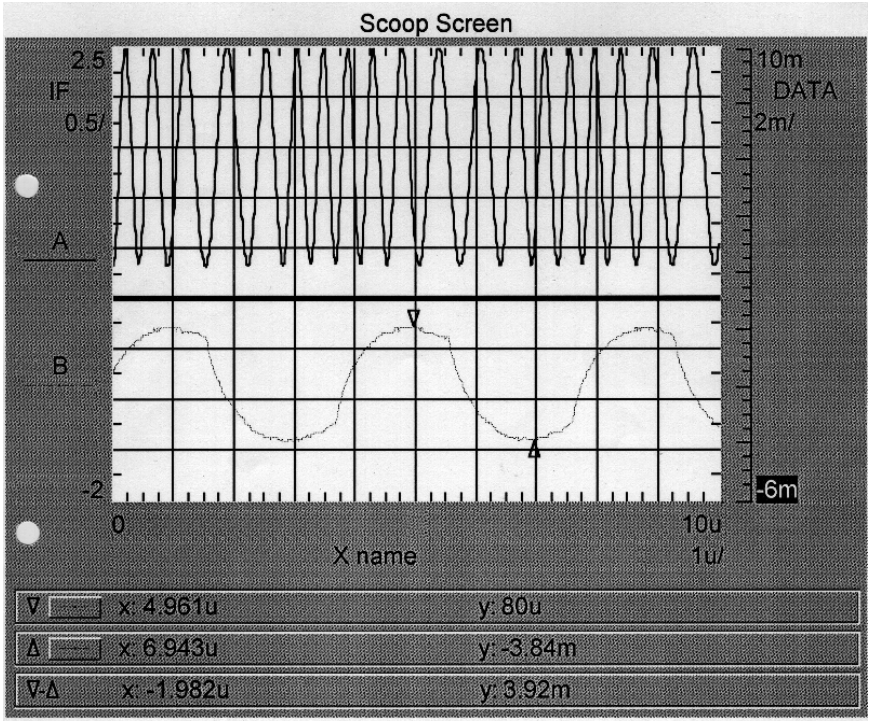


Figure 103 *IF signal & data from IF IC*

In the bottom part, the same "Dem+" and "Dem-" signals are used to create voltages δ1 and δ2, which start to rise linearly with time after positive respectively negative data transitions. The "Dem+" signal is connected to the "Set" input of a flip-flop FF1, and the "Dem-" signal to the "Reset" input of this flip-flop. Therefore, the output of this flip-flop will indicate the sign of the last known instantaneous frequency of the IF signal, "Imp. Rec.". This is a rough approximation of the received data, in which the value of the data is correct, but the moments of data transition are not yet correct. Transitions in the "Imp. Rec." signal are also used to set either FF2 (for a 0→1 transition) or FF3 (for a 1→0 transition). These flip-flops indicate that a transition of the reconstructed signal is pending, and should be carried out after a delay that corresponds to the ν1 and ν2 signals. When FF2 is set, δ1 will start increasing, and δ2 will start increasing when FF3 is set.

The signals δ1 and ν1 are added, as well as δ2 and ν2. To both of these sums, a negative threshold voltage "Vref" is added, resulting in two signals α1 and α2. The zero crossings of these signals define the exact moments on which the reconstructed data, "Perf. Rec.", should make a positive or negative transition, which is achieved through limiters (for detecting the zero crossings), and a flip-flop FF4 that represents the reconstructed output signal. The "Vref" signal adds a constant delay to the output signal. The minimum delay is set by the causality requirements discussed in appendix E. This new method of reconstructing the data from a limited low-IF FSK signal has been patented [114].

A die photo of the complete IF IC is shown in fig. 104, and the output signal of the demodulator is shown in fig. 103. The total supply current of the IF IC is 0.15mA at 3V, bringing the total power dissipation of a simple receiver to 3.5mW, or 7mW for the dual-beam angle-scanning receiver. Typical 2.5GHz WLAN receivers in literature around the

same time frame report power dissipations in the range from 40mW to 116mW [115] [116] [16] [117] [118], or about one order of magnitude higher[10]!

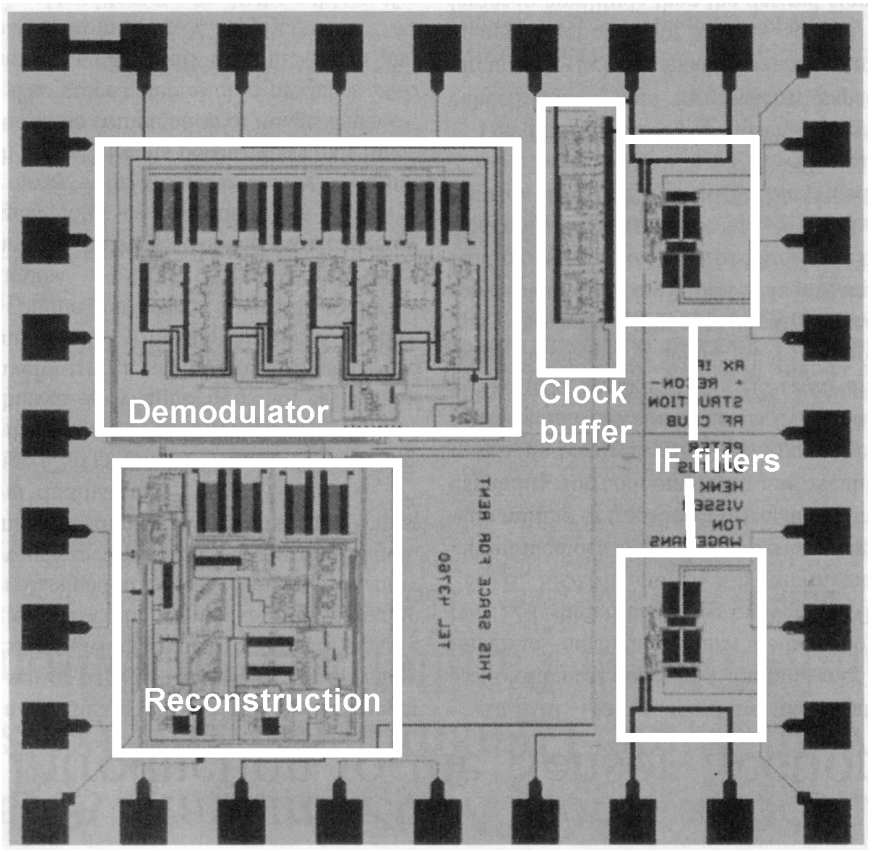


Figure 104 *Photo of the IF IC*

7.4 Summary and conclusions

In this chapter, the design of a 7mW angle-scanning diversity receiver for a proprietary WLAN system has been discussed. The angle-scanning antenna diversity method, as well as the implementation using weighted addition at IF, have been patented. The receiver has been implemented in SOA technology. The very low power dissipation, about one order of magnitude lower than most other WLAN receivers reported in literature, demonstrates the advantages of the trade-offs at system, circuit and technology level that have been discussed in the previous chapter. The receiver also employs a unique IF signal processing chain, with resistive interpolation to generate the input signals for a multi-

[10] Although some part of this factor is due to the higher data rates of most of these WLAN receivers compared to the proprietary system described in this section, the largest part of this factor is still due to system and circuit design, and technology. This is supported by the relatively small difference in power dissipation between Bluetooth and WLAN receivers, even though the Bluetooth receivers also run at a much lower bitrate. The largest part of the power dissipation in these receivers is in the RF part, which is not affected by bandwidths and/or datarates of the system.

phase differentiate-and-cross-multiply demodulator, followed by a data reconstruction circuit that has been patented as well.

Chapter

8

RF Platforms

Thus far, this thesis has concentrated mostly on the design of minimum-power RF front ends. The minimum-power design method offers an efficient way to design RF front ends with minimum power dissipation, within the boundary conditions of this method, and within the accuracy of the models used. In many product development projects, there is a strong time pressure. Some of the causes of delays are outside the boundary conditions and models used in the design method proposed in this thesis. Therefore, an additional method to speed up the RF front-end design is required, and will be described in this section.

8.1 Introduction

As discussed in Section 1.4, a significant delay in the RF design process is caused by modeling inaccuracies. This causes the performance measurement results to be significantly different from the simulation results. As a consequence, multiple iterations over the design-layout-fabrication-packaging-testing loop are required, as shown in figure 105.

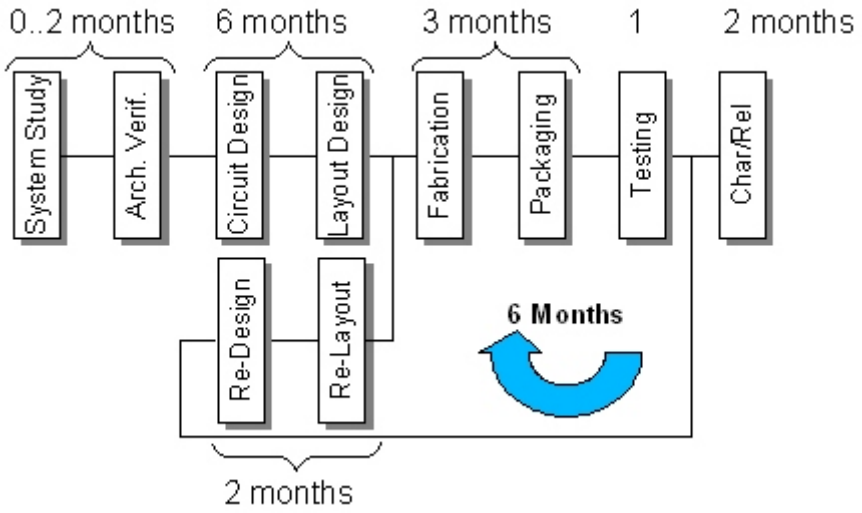


Figure 105 *A typical RF design process flow*

The minimum-power design method provides the optimum implementation of a system specification using transformed instances from a circuit library. The modeling inaccuracy is reduced if the library circuit performance is based on measurements of these circuits. In that case, only the inaccuracies of the SITs, and interactions that are not modeled in the design method, will cause differences between measurements and simulation results. Due to the limits of scaling, and the many significant parasitics in RF circuits that cause interactions between subcircuits, these inaccuracies and interactions can still cause significant problems that result in the need for additional iterations in the design flow.

The need for faster product development times is caused by the decrease of the profit margins for new products over time, as shown in fig. 106.

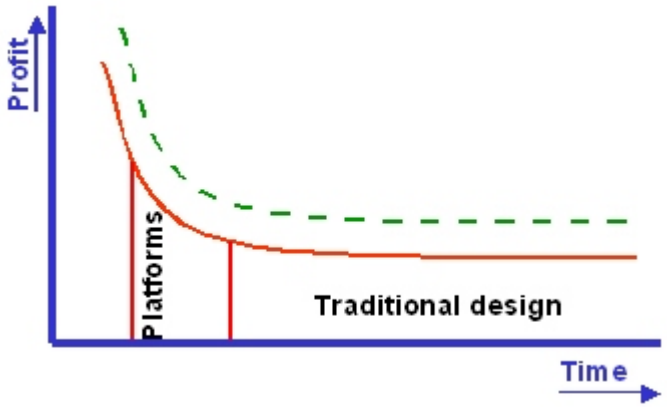


Figure 106 *Decreasing profit margins for new products*

The solid bottom curve shows the profit margin over time after product introduction. The area below this curve labeled “traditional design” represents the integral of this profit

over time that can be earned using traditional design methods. A faster design method could provide additional earnings, represented by the boxed-in area labeled “platforms”.

One might even argue that there are additional benefits to entering a market for a new product early, since this gives access to more design-in windows, and therefore a larger market share. This larger market share allows the NRE costs to be distributed across a larger number of sold products. This will reduce the cost per product, and therefore increase the profit. This is represented by the dashed top curve. In the area of digital circuit design, such a decrease in development time is achieved by switching from custom design to platform-based design. In platform-based design, a product is designed from a standard template and a library of building blocks that can be added to, or removed from, this template in a very flexible way.

8.2 RF Platform elements

The design method developed in this thesis can be the basis for a similar approach, using the library of subcircuits as building blocks, and the front-end topologies as templates. However, the development time reduction achieved in digital design will be difficult to duplicate for RF design, since adding and removing building blocks can cause parasitic interactions between such blocks that are difficult to model and predict. These uncertainties can be avoided by using models based on measurements of the same building blocks that will actually be used in the product. To make this possible, an RF product can be based on an assembly of pre-fabricated integrated circuits in a module. Each of the integrated circuits contains a building block, or subcircuit, from the library. By selecting the appropriate building blocks, and by assembling them in a module, a product can be built from very well-characterized subcircuits, and the modeling inaccuracies are eliminated. Also, interactions through substrate, power supply, package etc. can be strongly reduced or eliminated since the individual building blocks do not share a substrate or package. In addition, the power supply can be very well decoupled between the individual building blocks of the module.

This type of substrate and assembly technology is available as a mainstream, low-cost process through several companies. At Philips Semiconductors this technology is available both for prototyping and for mass production. The fabrication time for a module, based on existing ICs and substrates, is usually much lower than for monolithic integration. Typical turn-around times for prototyping are in the order of a few days.
This approach helps to bring down the development time in two ways:

1. The uncertainty in modeling accuracy is eliminated by basing the design on measurement results of the actual integrated circuits;
2. Even if the performance of the prototype is different from the predictions, an iteration loop will be much shorter, typically a few weeks rather than half a year.

Therefore, a design method that uses a library of existing building blocks in the form of physical ICs, and integrates them through assembly in a module rather than through monolithic integration, can be expected to significantly speed up product development. Several other elements are required to make this approach feasible and practical:

- A limited set of front-end architectures that cover a large application area using building blocks.

- Building blocks that can be configured to different trade-offs in performance, depending on the settings of some adjustment pins. Such configurable building blocks reduce the number of blocks required to cover a large application area.
- A set of specifications for these configurable building blocks, such that this large application area can be covered with a small number of blocks.
- A design method that supports the design of products using such building blocks in a module.
- A characterization method for extracting performance parameters from building blocks that can be used in the design method.
- A low investment platform implementation strategy.

This approach is called RF Platforms. A study project to develop this method was carried out by a team at Philips Semiconductors Advanced Systems Labs. The first application area, targeted by the RF Platforms study project, included cellular and wireless connectivity systems of generation 2.5, such as GSM/DCS/PCS, GPRS, EDGE, IS95, IS98, AMPS, DECT, Bluetooth, IEEE 802.11a, IEEE 802.11b, IEEE 802.11g, GPS, and their multi-mode, multi-band combinations.

8.3 Architecture and partitioning

The selection of a limited set of architectures for RF platforms should take into account the impact of such a choice on the time to market (TTM), as well as performance, low power consumption, and multi-band multi-mode capabilities. Based on these selection criteria, a number of transmitter and receiver architectures were investigated and rated.

For the receiver, zero-IF, low-IF and superhet were considered. Both technical and implementation issues were taken into account when rating the relative advantages and disadvantages of each architecture. The technical parameters considered include IP2, IP3, sensitivity, achievable signal-to-noise ratio, image rejection, channel selectivity, 1/f noise in signal path, wide-band suppression, single- tone interferer, self reception, RF selectivity, group delay/ripple, DC cancellation, and spurious transmissions in receive mode. Special attention was paid to multi-band/multi-mode possibilities and power consumption. The implementation issues considered include synthesizer & VCO design difficulties, flexibility in design, availability of IP and building blocks, time to market, cost, and integration level.

For the transmitter, two-step conversion, direct conversion, and closed loop/VCO modulation (TX-offset) architectures were investigated in the same way as the architectures for the receiver. Performance parameters taken into account include VCO pulling, integration capability, IF / image rejection, phase noise, spurious, power control range, and linearity. The same implementation issues as for the receiver were considered for the transmitter as well.

The architectures were rated by a team of experts on each of these parameters and issues, taking into account the boundary conditions with respect to market, technologies, and existing IP as they existed at that time in Philips Semiconductors. The outcome might be different in other environments. In this case, zero-IF and low-IF architectures for the receiver, and direct conversion for the transmitter were selected as the most appropriate architectures for this RF platform instance.

The partitioning granularity is a trade-off between flexibility and cost efficiency. A high granularity, i.e. small building blocks, allows for very generic building blocks with a high re-use count. An extreme example of this would be a granularity at the level of discrete components such as transistors and resistors. A less extreme example would be a granularity at the level of signal-processing stages as defined in this thesis, e.g. an LNA, a mixer, etc. A low granularity, with large building blocks, allows for a very efficient implementation, since the overhead of assembly costs, interface circuits, duplication of input and output buffers, bondpads, etc. will be lower. Other considerations with respect to partitioning included the need for full-duplex operation, advantages of optimum IC process selection for individual building blocks, parasitic effects such as crosstalk and pulling, and the number of interconnections between blocks.

Various partitionings have been investigated with respect to these issues. Initial estimates of assembly cost, yield, and flexibility settled the total number of building blocks for a transceiver to about 5. This number was based on the specific substrate and assembly technologies available at the time within Philips Semiconductors. A number of alternative partitionings are shown in figure 107.

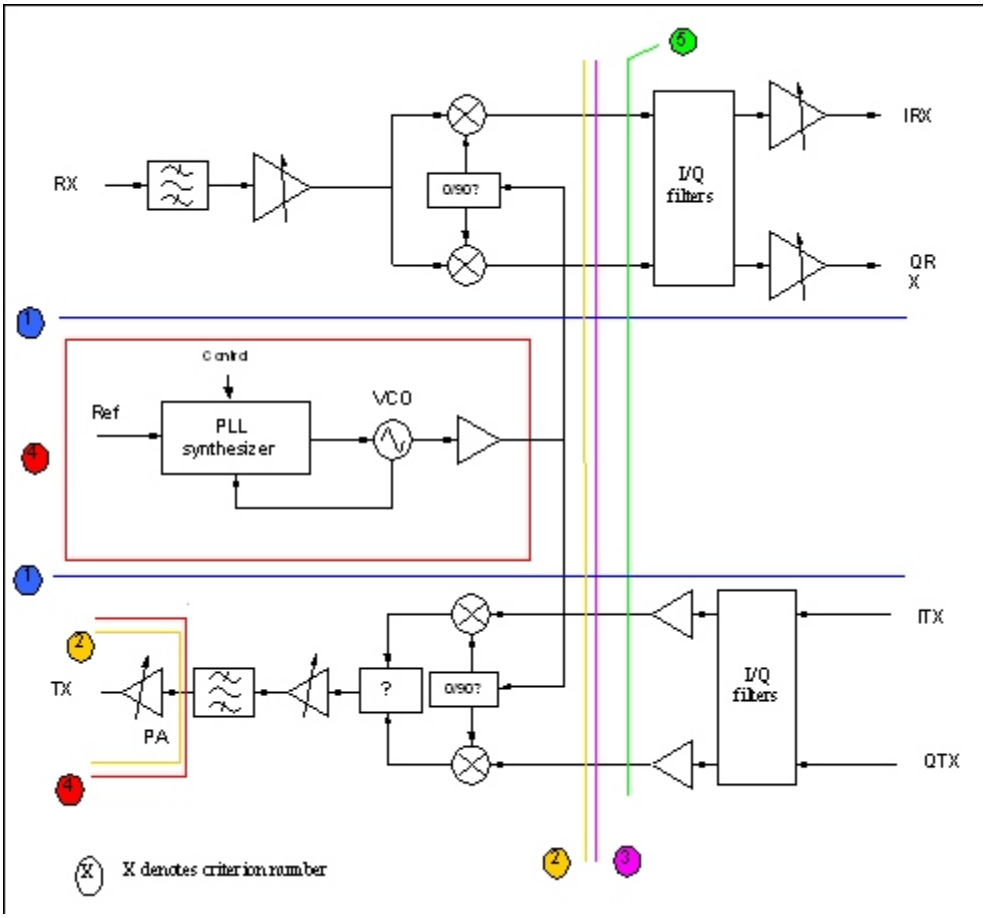


Figure 107 *Potential partitioning choices for RF platforms [119]*

The choices for each of the partitioning alternatives in figure 107 are listed below:

1. best isolation between transmit and receive parts in full-duplex operation;
2. optimum use of individual IC processes for each building block;
3. maximum flexibility and re-use possibilities;
4. reduced parasitic interactions such as crosstalk, pulling, thermal effects, etc.
5. allows for the use of superhet architectures in addition to zero-IF and low-IF.
6. minimizes the number of interconnections between the building blocks.

When combined, this results in the following optimum partitioning (figure 108).

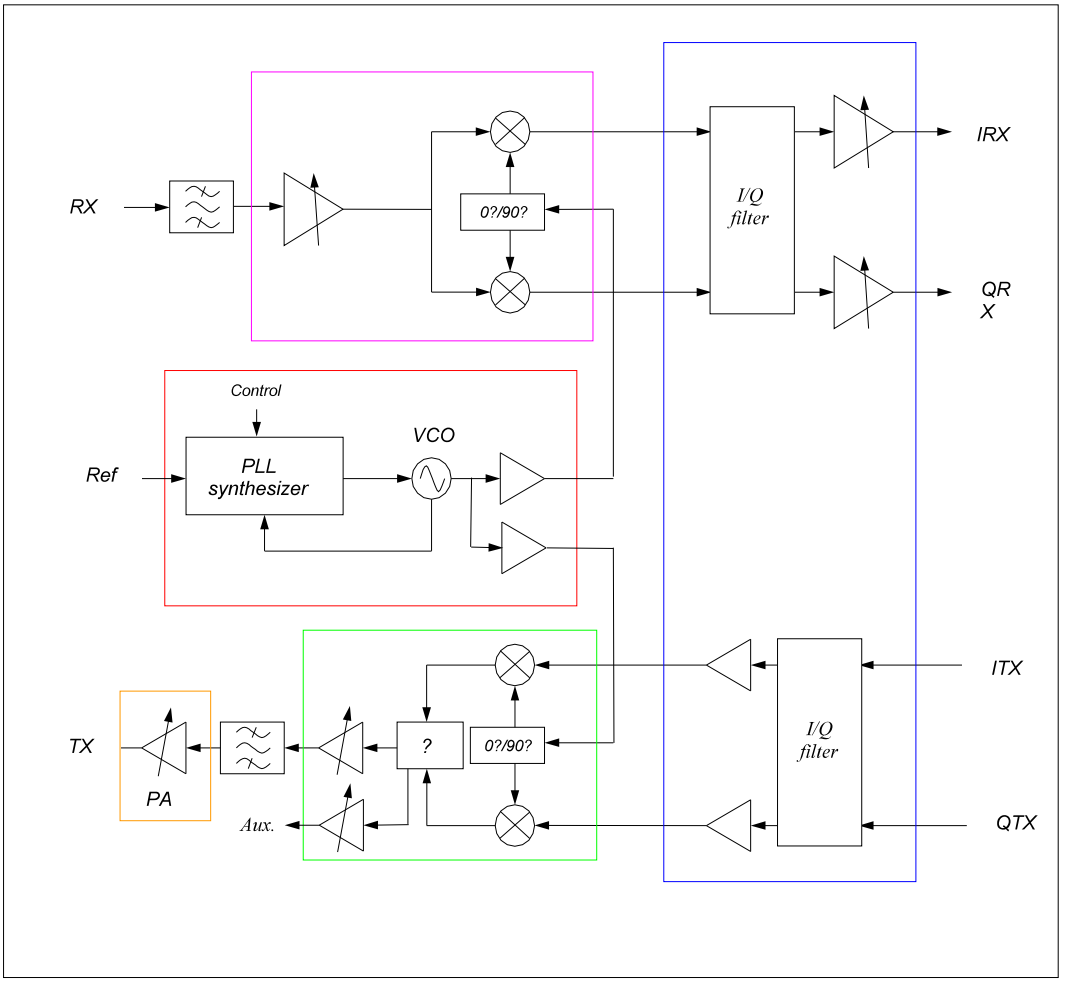


Figure 108 *Optimum partitioning of building blocks for RF platforms [119]*

A product based on this partitioning would look like fig. 109.

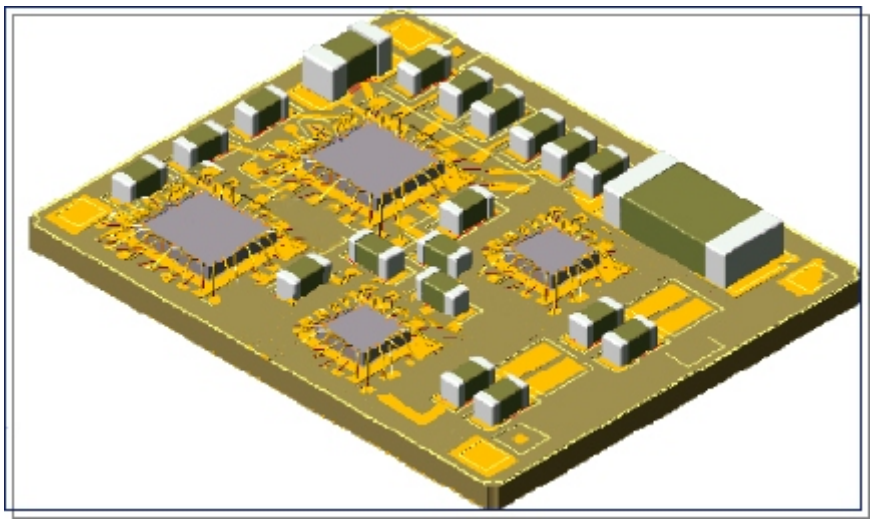


Figure 109 *Drawing of an RF platform implementation of a simple transceiver (Drawing by Martin Barnasconi, Philips Semiconductors)*

A more complex multi-band, multi-mode module, integrating a multi-band WLAN transceiver according to IEEE 802.11a, IEEE 802.11b, and 802.11g, in combination with Bluetooth, would consist of eight building blocks and measure about 11x15mm, according to first design studies at Philips Semiconductors Advanced Systems Labs.

8.4 Building block specification

Based on the architecture and partitioning choices shown in the previous section, the specifications for a limited set of building blocks need to be determined. This problem is much more complicated than the derivation of specifications for subcircuits in a single transceiver, as described in Chapter 5.

The specifications for a single transceiver can be represented by a single point in design space, where each of the dimensions corresponds to a parameter from the specifications. The area to be covered by the building blocks for RF platforms is represented by a cloud of points in the design space. A simple, three-dimensional representation of such a cloud and the design space is shown in figure 110.

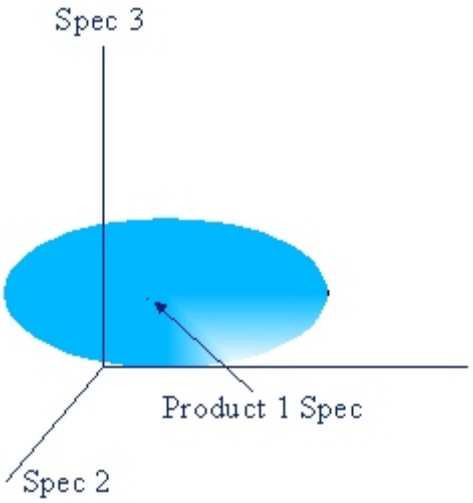


Figure 110 *Design space and specification cloud for application area*

When deriving specifications for subcircuits from a system specification, there are degrees of freedom. In this thesis, these degrees of freedom have been used to minimize the power dissipation. Similar degrees of freedom exist in translating a cloud of specifications at the product level to clouds of specifications at the building block level (fig. 111).

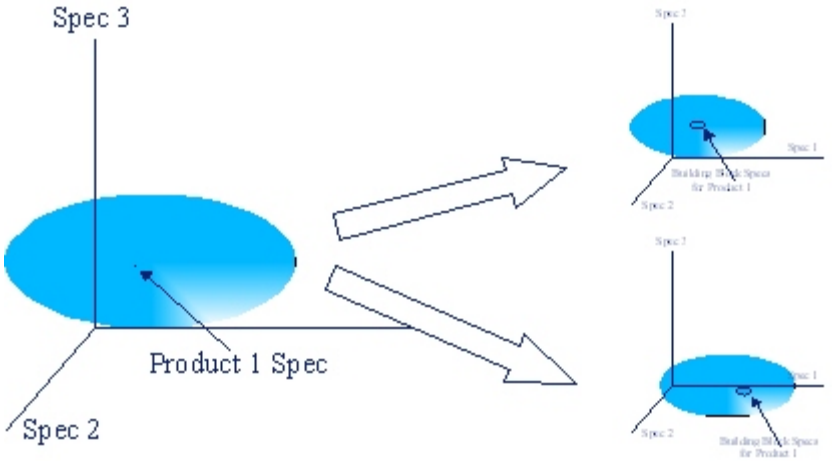


Figure 111 *Design spaces for an RF transceiver, and for individual building blocks of such a transceiver*

These extra degrees of freedom can be used to reduce the number of building blocks required to cover the whole application area. Finding the minimum number of such building blocks is a complicated problem that needs further study. It depends, among others, on the range over which the performance trade-offs of single building blocks can

be configured. The OSITs, developed in this thesis, are an interesting basis for such configurability. Investigating the limits of RF configurability is another area that needs further study, and is currently the topic of the Ph.D. study of Maja Vidojkovic at the University of Technology, Eindhoven, The Netherlands. In the RF platform project, the problem was solved manually by a small team of experts, which resulted in a set of specifications for 16 building blocks (excluding the power amplifiers) that can be combined to build any single- or multi-mode transceiver within the defined application area. For single-mode transceivers, the preferred combinations of the small-signal blocks are shown in table 112.

	Down converter block nr	***Transmit & Receive IF block nr***	***Upconverter block nr***	***VCO/ Synthesizer block nr***
GSM 900	1	1	1	1
DCS 1800	2	1	1	1
PCS 1900	2	1	1	1
DECT	2	2	1	1
Bluetooth	3	2	2	2
802.11b	3	3	2	2
802.11g	3	3	2	2
802.11a	4	3	3	3
GPS	5	4	N/A	4

Table 14 Preferred combinations of building blocks for single-mode transceivers

The GSM/DCS/PCS standards each also cover the EDGE and GPRS extensions. The numbers in each column below a building block type indicate which building block of this specific type is preferred for this application. Detailed specifications for each of these building blocks have been created, and they are currently being developed.

8.5 Design method

The design method for RF-platform based products will be very different from traditional RF design methods. Instead of designing a monolithic custom circuit starting from individual devices such as transistors and passives, an RF-platform based product is designed as a module containing a number of existing integrated circuit building blocks and some discrete components. This change in design method has both technical and human consequences.

On the human side, designers will have to change their way of working. Part of their experience and expertise will no longer be needed, especially that part relating to design and layout of circuits at device level. RF-platform based designs also offer less freedom and room for creativity at the device and circuit level. Although this is offset by a faster development time, this change is not likely to be perceived as purely positive from a designer point of view. In addition, an RF platform approach requires

development and sharing of building blocks across organizational boundaries, which is not always obvious and painless in a large organization.

On the technical side, in addition to the development of the building blocks, there is a need to adapt design flows, tools and libraries to this new approach. In fact, there will be two design trajectories:

1. The development of building blocks according to RF platform specifications in a monolithic custom integrated circuit, starting from the device level up. This design trajectory closely resembles the design of current RF IC products. The main difference is that the building blocks should be optimized for use in a module rather than in an IC, so that the environment is well-known; in fact, it will be often part of the same design trajectory as the IC itself. Also, the specifications will call for configurability across a wide range, which is unusual for typical RF products.
2. The development of products as modules containing several existing building blocks from a library. This design trajectory is rather new, and requires a different design flow, tool set, and library.

A design flow for the RF platform product and building block development trajectories is shown in fig. 112 below. The RF platform product development starts with the usual system study and verification of the architecture. The following step is the selection of the appropriate building blocks from the library, based on the measured performance parameters of these blocks. Also, the configuration parameter values need to be determined. With a sizable library, this quickly becomes a significant problem. For example, with the 16 blocks in the current library, and with 2 parameters per block, and 10 characterized values per parameter, the number of combinations in a typical single-mode, single-band transceiver is already $2.4\ 10^{10}$. For a multi-band, multi-mode product, this number increases quickly. Obviously, an exhaustive search through all combinations is not practical. Therefore, a block selection method has been developed in a masters study by Haris Duric, at the University of Technology, Eindhoven, the Netherlands [77]. This method uses the parameters from the characterization of the building blocks, in combination with a genetic programming algorithm, to quickly find the optimum combination. In initial tests, the algorithm found the optimum solution in less than one minute.

After the selection of blocks and configuration parameters, the total product is simulated. This simulation is based on behavioral models that take their parameters from the characterization of the building blocks. Since these parameters are based on measurements of the actual circuits that will be used in the product, the simulations can be more accurate than circuit level simulations. Since the block and parameter selection algorithm already provides the optimum solution, the simulation is an extra verification step that should normally not require iterations.

Once the simulation results confirm that the desired performance is achieved, the product is assembled from the building blocks, and measured. There should be few surprises at this point, because of the accurate simulations. However, given the general sensitivity of RF design to many parasitic effects, a small number of iterations might be required. Because of the fast prototyping turn-around time, the time penalty for such iterations, if any, is low. If tests show that the product does indeed meet the performance requirements, it can go through the normal characterization, qualification and reliability testing steps before being released as a product.

Since there are two development trajectories in this approach, part of the drawbacks for designers can be eliminated by assigning them to the trajectory that appeals to them most. In the building block development trajectory, a lot of the traditional freedom and creativity is needed to implement the best possible configurable building blocks. In the product development trajectory, designers get the immediate satisfaction of seeing products in the market that they helped to develop. Also, alternating between the two development trajectories might be an option for designers who don't want to overspecialize.

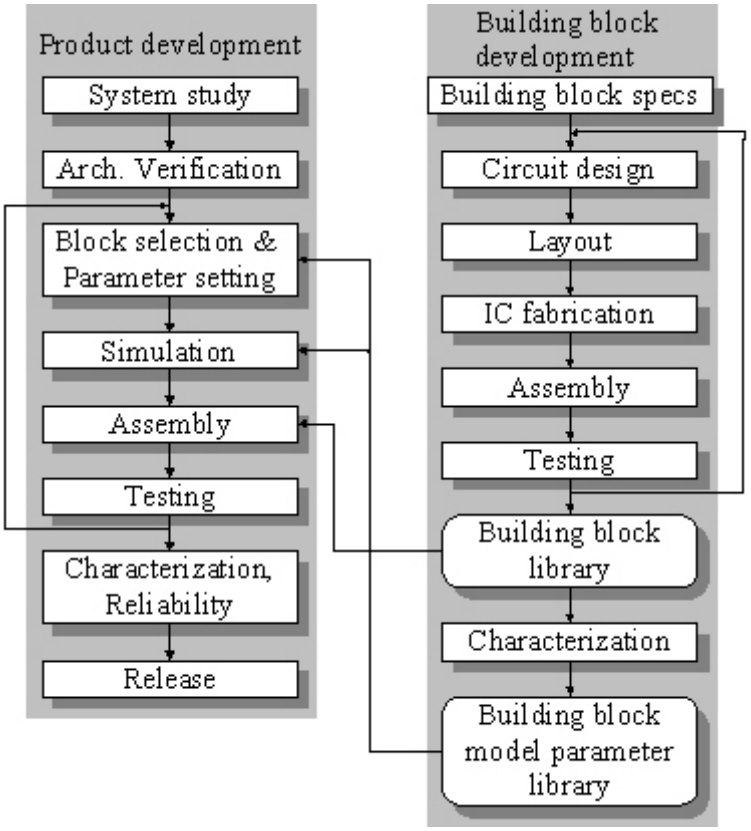


Figure 112 *Design trajectories for RF platform product and building block development*

8.6 RF Platform Implementation

Implementing a new product development method such as RF platforms in a large organization is far from trivial. The normal resistance to any change, in combination with the (perceived) drawbacks related to this approach, will make the introduction difficult. Moreover, as long as it has not yet been proven, there will be a reluctance to rely on this approach for developing new products, because of the risk involved. Finally, the required openness to sharing and supporting building blocks that are developed in one part of the

organization, for the benefit of another part of the organization, might not always be present. Without the support of a central organization, in this case the CTO, the development of this method would not have been possible. Now that the method has been developed to a point where it can be applied to the development of a first product, support from a central organization by itself is not sufficient for the next steps, since:

- the cost of developing the building blocks for the library is relatively high, comparable to the cost of developing 4 custom transceivers. Allocating sufficient resources to develop these building blocks, and later to maintain, support, and extend the library, is difficult on the basis of central funding alone. This is especially the case in a typical business environment, where the relative priorities for short-term product development projects are much higher than for investments in future design methods;
- the resistance to using a new approach for a product needs to be overcome, and this resistance is bigger of the new approach originates in a different part from the organization;
- designers in product development organizations need to get a good understanding of this approach and its consequences for their work and for them.

A new approach has been worked for this second phase in RF platform development, which resolves these issues. In this approach, projects are selected in which products are developed that contain subcircuits which are close to the RF platform building block specifications. By allocating a small number of additional designers from a central organization to such a development project, these subcircuits can be developed jointly. After the subcircuits have been developed, they are modified by the central organization to fit the needs for the RF platform building blocks. At the same time, the circuits are merged by the business developers to end up as a monolithic product.

Because of this synergy, the building blocks can be developed at a fraction of the investment that would otherwise be needed. Moreover, the normal (monolithic) product development is sped up as well. Once the building blocks exist, they can be assembled into a product in parallel with the traditional product development, with minimal additional effort. This eliminates the risk of depending only on an RF platform approach for the first products that are developed this way. It also allows a clear comparison of the cost and speed advantages of both product development methods.

8.7 Summary and conclusions

There are product-market combinations that require faster time to market than can be achieved with traditional design methods. The minimum power design method that has been introduced in Chapter 5 already speeds up the design process by reducing the number of iterations. In this chapter, RF platforms have been introduced as a complement and extension to the minimum power design method. RF platforms consist of a library of configurable building blocks. These blocks are pre-fabricated as individual dies, and are combined on an RF module substrate to form the complete product. The pre-fabricated dies can be characterized based on actual measurements. This allows the design of the product to be carried out at the block level, with behavioral models that use parameters based on the characterization of the actual dies. This is expected to increase the accuracy of the simulations because it eliminates the uncertainty inherent in transistor

level circuit simulations. Moreover, the combination of independent dies on a module substrate eliminates most of the parasitic interactions between the circuit blocks, which should reduce the number of iterations required to arrive at a working product. Finally, the assembly of building blocks on a module substrate can be carried out much faster (typically in the order of days) than the fabrication of ICs in an IC fab (typically in the order of months). These combined effects should significantly speed up the time to market of RF front ends.

The configurability of the building blocks is required to reduce the number of blocks needed to cover a significant application area. It has been shown that 16 blocks are sufficient to cover all current cellular and connectivity standards, as well as all their multi-mode, multi-band combinations.

Section

Conclusions

The main goal of this thesis was to answer the following question:

"What are the fundamental limits for the power dissipation of telecommunication front ends, and what design procedures can be followed that approach these limits and, at the same time, result in practical circuits?"

In Section 3.6, it was demonstrated that the fundamental limits are set by physics, but are typically not reached because they are dominated by limits in technology and circuit design.
To approach the limits imposed by technology, a design method was developed based on complexity reduction. A two-step approach was developed based on structure-independent transforms. In a first step, the optimum distribution of gain, noise, and linearity across the subcircuits in a front end have been derived. In a second step, the transforms are used to modify circuits from a library to match this optimum distribution.

On the basis of this design method, a design procedure was developed that allows:

- the selection of subcircuits that are the best basis for a low power front end design;
- the derivation of front-end specifications from system specifications;
- the derivation of subcircuit specifications from front-end specifications and a selection of subcircuits;
- the transformation of the selected subcircuits into optimized subcircuits for a specific front end.

This design method is described in Chapter 5 and was then applied to the DECT front end that was used as a case study in Chapter 4. In this case, a significant power savings of a factor of 2.7 is to be expected. The design procedure can also be used to demonstrate the additional reduction in power dissipation that can be achieved by relaxing the specifications until the system requirements are exactly met. The predicted power dissipation reduction exceeds two orders of magnitude!

This design method offers the best approximation of the technology-imposed low-power limits, within the boundary conditions posed by system specifications, circuit design and implementation technologies. Further savings, and therefore a closer approximation of the fundamental limits, can be achieved by changing these boundary conditions. For example, by reducing the margin between front-end specification and system requirements, the power dissipation of the DECT front end could be reduced by up to 2 orders of magnitude. But also the impact of other changes in boundary conditions, such as the addition of advanced antenna diversity, and the implementation in IC technologies that allow circuit scaling to high impedance levels, can be predicted using this design method. The impact of such changes can be very significant, as is shown in Chapter 6.

As a spin-off of the design method, a new class of figures of merit has been identified, called equivalent figures of merit. They are special, in that they not only project the performance of all subcircuits in a hyperplane in design space onto a point

in a one-dimensional performance ranking, but also generate a complete class of circuits from one parent via structure independent transforms.This allows such equivalent figures of merit to be used as the basis for circuit design.

A combination of many of the insights and results from this thesis has been used in the development of a very low power 2.5GHz transceiver in silicon-on-anything technology. This design, and the results, are described in Chapter 7. The simple receiver chain consumes only 3.5mW, or 7mW in angle-scanning antenna diversity mode. This is about one order of magnitude below other WLAN receivers.

Finally, a method to further improve the effectiveness and efficiency of RF front end design is proposed: RF platforms (Chapter 8). This approach fits well with the design method developed in this thesis, and has the potential of offering a very fast development track from specification through prototype to production.

Recommendations for further research

More research is required in the area of class-A/B front-ends. The building blocks for such front-ends are there, but the feasibility of a full front-end using such blocks still needs to be demonstrated. Also, the design method needs to be extended to cope with class-A/B circuits. Currently, the power-linearity parameter κ is assumed to be a constant for any circuit. This is correct for class-A circuits, but for class-A/B, κ becomes a function of the input signal.

Also, the proposed class-A/B differential pair is still relatively complex. A simpler circuit with similar functionality would be cheaper, and would probably work up to higher frequencies because of the reduced number of time constants. Such a new topology would greatly help in the implementation of a class-A/B front end.

Feedback requires also further study. Although it cannot be applied as a structure-independent transform in the strict sense, it might be possible to introduce it after all. Mixers could be split in a part of the circuit before the actual frequency conversion, and a part after the actual frequency conversion. In both parts, feedback could be implemente, e.g. through degeneration of the input stage. Therefore, a feedback SIT, and, if possible, an OSIT based on feedback needs to be developed.

A

Communication

Communication is the process of exchanging information between two or more entities. Three types of communication (figure 113) can be distinguished:

- Point-to-point communication, in which one entity communicates with one other entity (e.g. two people talking to each other)
- Point-to-multi-point communication, in which one entity communicates with several (many) other entities simultaneously (e.g. a formal presentation)
- Multi-point-to-multi-point communication, in which several (many) entities communicate with several (many) entities simultaneously (e.g. informal conversation at a party).

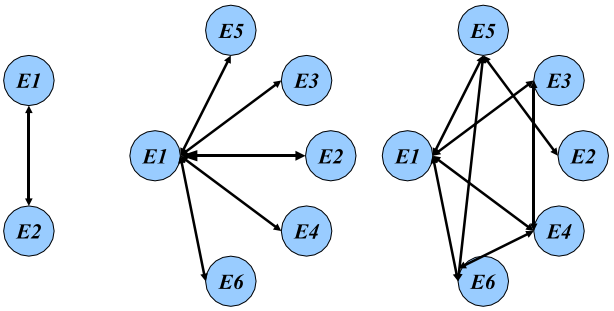


Figure 113 *Point-to-point (a), point-to-multi-point (b), and multi-point-to-multi-point communication*

Conceptually, multi-point-to-multi-point and point-to-multi-point communication can be thought of as several (many) point-to-point communication processes in parallel. Please note that one entity exchanging information with itself is not considered communication in the context of this thesis.

Figure 114 *Simple model of a uni-directional information stream*

Point-to-point communication consists of up to two uni-directional information streams, one from entity 1 to entity 2, and the other going in the opposite direction. In some cases, only one stream is present. During the exchange, the information in a uni-directional stream is embodied in a series of messages (figure 114).

A message consists of special arrangements of matter and/or energy. An example of a message consisting of a special arrangement of matter is the printed version of this thesis, in which colored ink particles are deposited on white paper in such a way that they form letters and pictures. An example of a message consisting of special arrangements of energy are the sound waves of a voice.

The information represented by a message is not uniquely defined. Instead, it depends on an agreement between the entities involved in the communication. The same message can represent different information to different entities. For example, the sound of the voice of a soprano in an opera might carry information about the underlying story and related emotions to an experienced listener. At the same time, the same sound of the same voice might carry the information that a lady is in distress (and thus in need of rescuing) to a less experienced listener. Obviously, both entities involved in an information exchange need to agree on a method for encoding information into messages to enable effective communication.

In telecommunication systems, messages are transmitted as electrical signals, and they need the same agreement on information encoding as other types of messages. Without such an agreement, the same electrical signal can represent, among others:

- the water level in a tank, using a predefined translation from average voltage to water

level, when measured with a voltmeter

- the sign of the difference between an ideal and an actual position of the position of a robot over time, when measured with an oscilloscope (figure 115)
- the level of interference at various frequencies in a radio system, when measured with a spectrum analyzer (figure 116)
- a picture from a webcam, when received by the port of a computer running suitable software (figure 117)

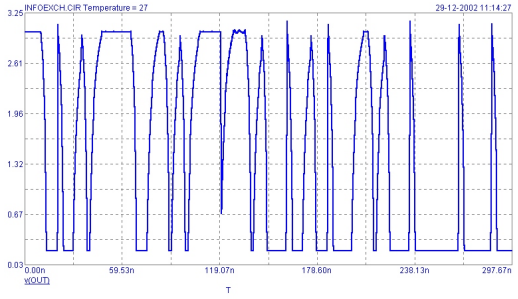


Figure 115 *Voltage versus time of a signal*

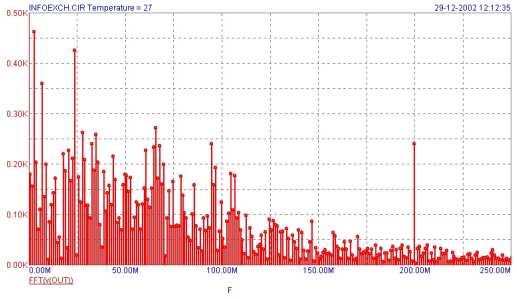


Figure 116 *Frequency spectrum of a signal*

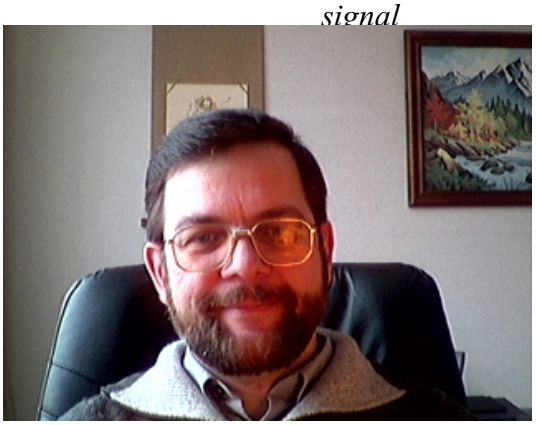


Figure 117 *Picture from a webcam*

Once information is encoded in a message, it can be translated into different types of messages in order to facilitate transmission, as long as it is translated back into the original message type before decoding (figure 118). This translating back and forth occurs frequently in communication processes, typically allowing the use of message types that can more easily be transmitted across large distances. For example, sound waves from a voice message are converted into messages constructed of electrical signals, and back to sound waves again, in a voice telephony system, allowing communication across distances much greater than would be feasible by sound waves.

The device translating an original message into another type is commonly called

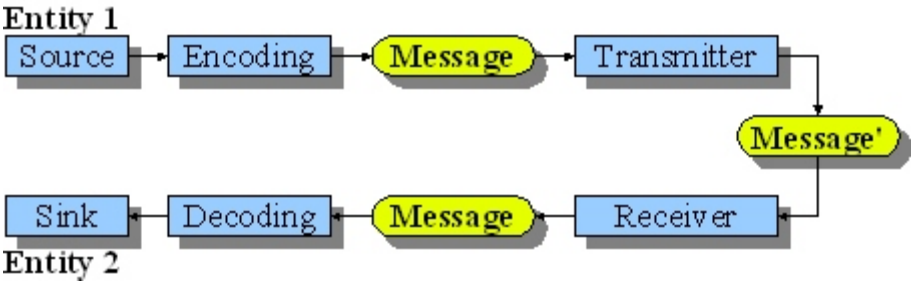


Figure 118 *Information transfer with translation of messages*

"transmitter", whereas the device translating the other message type back to the original is commonly called "receiver". This thesis will concentrate on the message conversion and transmission in transmitters and receivers (often called "transceivers" when combined into a single device). The construction and interpretation of the original messages (coding and decoding) is beyond the scope of this thesis.

A.1 Radio Communication

In most telecommunication systems, the signal as received by the receiver differs from the signal transmitted by the transmitter, because the transmission across a significant physical distance affects the signal in several ways. The most common impairments of radio signals are:

- *Path loss*, the attenuation of the signal between transmitter and receiver. Typically, only a small part of the energy propagates to the receiver, and other parts propagate in different directions. Also, part of the signal gets converted into other types of energy, most commonly thermal energy, by materials along the propagation path of the signal.
- *Multipath interference,* when different parts of the signal propagate over multiple paths from transmitter to receiver, and arrive at the receiver with different attenuation and phase (because of path length differences). Depending on the relative attenuation and phase of these signal parts, they can cancel each other partially or completely when combined, thereby causing extra attenuation. If the transmitter, the receiver, or objects that affect the propagation paths of the signal move around, the attenuation due to multipath interference becomes time variant.
- *Additive noise*, the addition of thermal noise and/or noise from natural processes (e.g. lightning), and manmade noise (e.g. spark plugs in a car engine).
- *Interference*, caused by signals from other transmitters.

These impairments are often modeled as a channel between the transmitter and receiver, as shown in figure 119. Depending on the properties of the channel and the properties of the translation in the transmitter and receiver, the receiver might or might not be able to recover the original message from the impaired signal received through the channel.

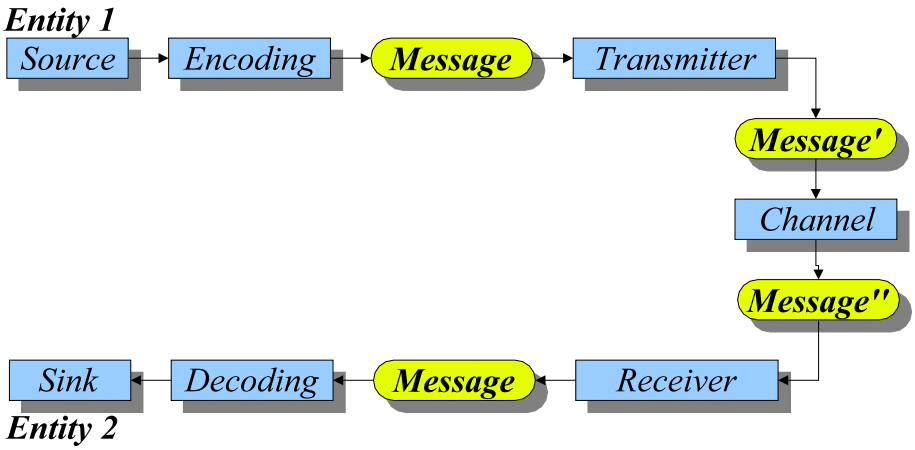


Figure 119 *Information transfer with translation and impairments of messages*

An important aspect in the design of radio transmitters and receivers is the robustness against channel impairments. This impacts directly the quality of the signal transmission, and therefore the perceived quality (and often value) of the radio. Important transmitter and receiver parameters with respect to this quality are:

- Transmitter power: transmitting a signal with a higher power level will result in received signals that are stronger compared to the noise and interference from other sources, and therefore in a more accurate received message, because the radio channel is highly linear. At the same time, these stronger signals can cause more interference for signals from other transmitters, so transmission power in a system is a compromise between the quality of the wanted signal at the receiver, and interference to other signals. In addition, higher transmitter power will result in higher battery power and therefore lower battery lifetime and/or larger (and more expensive) batteries.
- Transmitter linearity: signals that are distorted in the transmitter can result in more impaired messages at the receiver. Furthermore, they can cause unwanted signal components that interfere with signals from other transmitters.
- Transmitter noise: noise that is added to the signal in the transmitter can impair the messages at the receiver. In addition, it may impair signals from other transmitters.
- Transmitter spurious: any additional signals that are generated by the transmitter, but are not part of the intended message. These additional signals can cause impairments of the messages at the receiver, as well as interference for signals from other transmitters.
- Receiver sensitivity: a more sensitive receiver can better reconstruct a message from a smaller received signal, and is therefore capable of receiving better quality messages across larger distances.
- Receiver linearity: a non-linear receiver will distort the received signal, thus causing impairments to the message. In addition, other signals might combine with the wanted signals through the non-linearity of the receiver, making the receiver more sensitive to interference.
- Receiver selectivity: a less selective receiver will a larger part of an unwanted signal to combine with the wanted signal. This increases sensitivity of the receiver to

interference.

In addition to these parameters, transmitters and receivers can be designed to reduce the impact of channel impairments. For example, antenna diversity and channel equalization can reduce the impact of multipath propagation.

A.2 Translating Messages into Radio Signals

The translation of messages into radio signals in the transmitter typically involves converting lower frequency (analog or digital) signals representing the message into appropriately parameterized high frequency signals. This is often accomplished in two steps:

1. The message is converted into an appropriately parameterized low or intermediate frequency signal. This step is typically referred to as modulation.
2. This low frequency is then filtered, converted to a higher frequency and amplified. (though not necessarily in that order).

In the receiver, typically the complementary steps are carried out:

1. The high frequency signal is converted to a low or intermediate frequency, filtered and amplified (though not necessarily in that order).
2. The parameters of the low or intermediate signal are converted back into a signal comparable to the original message. This step is referred to as demodulation.

The part of the transmitter in which the signals are filtered, converted to higher frequencies, and amplified, and the part of the receiver in which the signals are converted to low or intermediate frequencies, filtered, and amplified, is referred to as the RF front end. This is shown in the simplified block diagram of figure 120. In this diagram, the box labeled “IF” (for intermediate frequency) contains the further signal processing that converts the amplified, filtered and frequency converted signals to and from the messages that need to be transmitted and received. The box labeled “Front end” contains the filtering, frequency shifting and amplification elements required to implement step 2 of the transmitter, and step 1 of the receiver, as defined above.

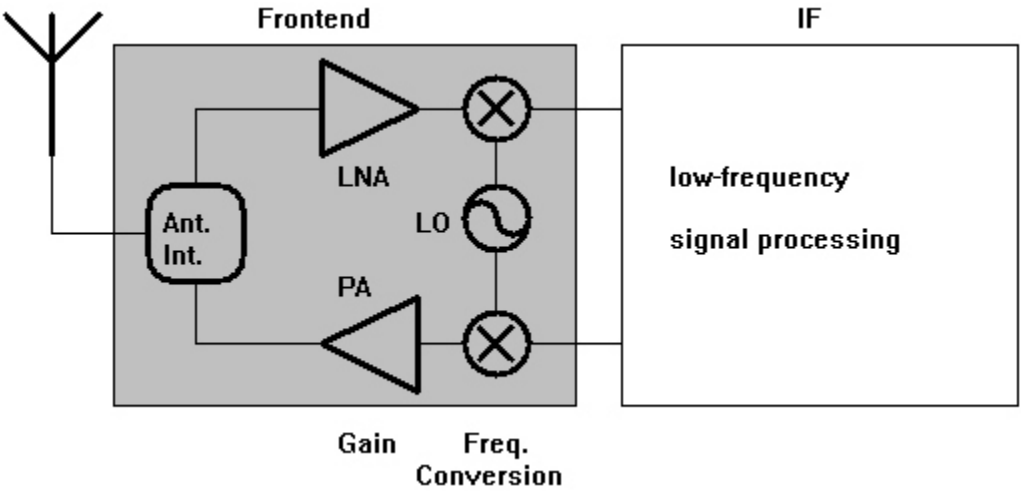


Figure 120 *Simplified block diagram of a radio transceiver*

In this thesis, the focus will be on the front end part of the radio, since this part is responsible for a significant part of the cost, performance and power consumption of the radio.

In principle, any combination of radio signal parameters can be used to represent the message, but only a limited number of methods have become popular:

- Frequency modulation (FM): the frequency of the radio signal contains the information representing the original message. For messages represented through digital signals, several forms of FM have become popular, such as frequency shift keying (FSK), in which a logical "1" is represented by a frequency shift Δf, and a logical "0" by a frequency shift $-\Delta f$ with respect to the original signal frequency. FM and FSK have the advantage of being simple to implement, especially since they allow non-linear distortion of the signal without significant effects on the message. They also allow for good quality reception with limited signal to noise and interference ratio of the modulated signal. The constant envelope of FM and FSK signals allows for the use of non-linear power amplifiers in the transmitter, and the robustness against non-linear distortion allows for the use of limiters in the receiver, adding to the simplicity of implementation. The main disadvantage is that this modulation method is rather spectrum inefficient. Derivatives of FSK, such as Gaussian frequency shift keying (GFSK) and minimum shift keying (MSK) have been developed to improve the spectrum efficiency of FSK systems to some degree.
- Amplitude modulation (AM): this modulation method represents the message in the amplitude of the RF signal, typically by frequency shifting the original signal to a higher frequency. It is spectrum efficient and fairly straightforward to implement, although it does require linear processing of the signal in the transmitter and receiver. To improve spectrum efficiency, single sideband modulation (SSB) has been developed. In this approach, only one of the two sidebands of a conventional AM signal is transmitted, resulting in a doubling of the spectrum efficiency at the cost of a higher complexity receiver and transmitter.
- Phase modulation (PM): this modulation method uses the phase of the RF signal to represent the information of the message. PM and its derivatives for digital signals

such as phase shift keying (PSK), quadrature phase shift keying (QPSK), and their many variations, allow for high spectrum efficiencies at the cost of fairly complex implementations and linear signal processing.

- Pulse modulation (PM): with the increasing popularity of UWB, several kinds of pulse modulation are becoming popular for radio transmission. Pulse position modulation (PPM) uses the time between pulses to encode the information of the message, but there are also systems using other properties of the pulse (e.g. its polarity) to encode the information. In principle, the encoding and decoding of PM can be very simple and straightforward. In practice, interferers, forbidden bands, multipath echos etc. make the system more complex than initially expected.

These modulation methods are outside the scope of this thesis, except where their properties and requirements influence the requirements of the front end part of the transceiver.

A.3 Sharing the radio channel

An important difference between wired and radio (wireless) communication is that a wire can be dedicated to a single signal, whereas the radio channel is, in principle, shared between all transmitters and receivers. Therefore, in the design of radio communication systems, it is essential to include some means of sharing the radio channel in such a way that there is a minimum of interference between the different transmitters and receivers. There are several popular methods for doing this:

- Frequency Division Multiple Access (FDMA): transmitters use different frequencies. These frequencies are separated sufficiently to allow negligible overlap of the transmitted spectrums. It is one of the oldest and most well-known methods: most people will be familiar with the tuning system of a broadcast radio, where signals from different stations can be selected by tuning to the frequency at which they are transmitting.
- Time Division Multiple Access (TDMA): transmitters use the same frequency, but transmit at different moments in time, in such way that there is negligible overlap in their transmissions over time. There are examples in which the transmission times are relatively long, e.g. when the same radio channel is used for one radio broadcast station during the day, and another at night. An example in which the time spans are much shorter, is the private mobile radio systems used by taxi drivers. They wait until they hear that nobody is using the channel, and then transmit their message. Currently, the most popular forms of TDMA use much shorter transmission times, often in the order of milliseconds. By appropriate buffering and processing of the signals, it is possible to emulate a transmission system with continuous transmission. In this case, the allocation of the time at which different handsets are allowed to use the radio channel is regulated through a strict protocol. From the user point of view, this time allocation is fully automatic: most people are not even aware that their GSM phone is timesharing the radio channel with other GSM handsets.
- Space Division Multiple Access (SDMA): transmitters can use the same frequency at the same time if they are sufficiently far apart. The reduction of signal strength with distance results in negligible signal strengths from other transmitters if they are

far enough away. SDMA is used for large geographical distances in many broadcast systems, but also across much shorter distances for cellular phones. Another form of SDMA which is currently recently receiving a lot of interest is the transmission of radio frequencies across longer distances but only in specific directions, using high gain antennas that transmit narrow beams of radio signals.

- Code Division Multiple Access (CDMA): transmitters can use the same frequency at the same time, even when they are close, if they code their signals in such a way that they can be separated from signals from other transmitters by cross correlation with specific codes.

Often, several access methods are combined to achieve the best trade-off for a specific system. For example, the next-generation cellular phones, based on the UMTS standard, are capable of using elements of all four access methods described above.

B

Telecommunication systems history

The next sections give an overview of current telecommunication systems and their recent history. The focus will be on cellular, cordless, and wireless data systems.

B.1 Cellular phone systems

The first attempts at mobile phones started in the 1920s. They slowly developed into the first commercial cellular phone systems in the 1980s. Before the year 1990, the prevalent cellular mobile phone systems around the world were analog systems, using mostly narrow band frequency modulation (FM) transmission. In the USA, the *Advanced Mobile Phone System* (AMPS) was used. Various other systems, often based on *Nordic Mobile Telephone* (NMT) systems, were put into operation in different countries of Europe. Initially, the mobile sets were limited to car phones. Already in the 1980s briefcase-like phones and later even handsets appeared, although they were huge by today's standards. Due to the cost of the subscription, air time, and equipment, use was limited and mostly restricted to businesses. These cellular systems are commonly called the first generation.

Figure 121*GSM phone*

In Europe, these analog systems were succeeded by a digital mobile phone system: the *Global System for Mobile Communications* (GSM, fig. 121). It operates in the 900MHz band, although later extensions (*Digital Communication System* or DCS, and *Personal Communication System* or PCS) allow for operation in the 1800MHz and 1900MHz band as well. GSM

uses the TDMA scheme. Together with technological advances, this allows for smaller and cheaper handsets. GSM became very popular, and was adopted by many countries outside of Europe as well.

In the USA, several competing digital cellular systems were developed: both TDMA (IS-54 and IS-136) and CDMA (IS-95). For a number of years, there was a lot of discussion about the relative merits of TDMA and CDMA, with the wildest claims going back and forth regarding system capacity, voice quality and power efficiency. These discussions often seemed to include at least as many political considerations as technical arguments. CDMA systems have been adopted by several countries outside the USA. Conversely, several operators in the USA are adopting GSM 1900 systems, and the market share of GSM in the USA is currently exceeding that of CDMA.

This first generation of digital systems is called the second-generation cellular systems. They all use circuit-switched connections, meaning that a connection is built up and kept open during the information exchange, even if for some time no data is being exchanged at all. For voice calls this is fine, since most of the time there will be information exchanged in at least one direction. However, with the increasing popularity of data connections (SMS, E-mail, web browsing, WAP, I-mode, etc.), circuit-switched connections have become less attractive, since they block scarce bandwidth and base station resources during the connection. This happens even if no data is transferred for some time (e.g. while the user is reading a web page or writing an E-mail). Using circuit-switched connections for data transfer results in high costs for the user. Packet-switched data connections do not suffer from this problem. Hence, they are quickly becoming more popular and will be discussed in more detail in Section B.2.1.

Extensions to these systems have been developed to improve spectrum efficiency and user data rates. This has resulted in the *high speed circuit switched data* (HSCSD), *general packet radio services* (GPRS) and EDGE extensions to GSM. HSCSD provides higher data rates by allowing the use of multiple time slots. GPRS allows multi-slot packet-based transmissions, and EDGE provides $3\pi/8$-8PSK based modulation instead of GFSK for higher data rates and bandwidth efficiency. These extensions are generally referred to as 2.5G (for generation 2.5, halfway between the first-generation digital cellular systems and the third-generation systems).

The third-generation cellular systems, commonly referred to as 3G, will offer much higher data rates (up to 2Mbps). The standard was fixed before the turn of the 21st century, and bandwidth allocations were settled through auctions in most countries around the same time. Originally, three standards were developed independently from each other: CDMA2000 in the USA, UMTS in Europe, and W-CDMA in Japan. The ITU was expected to select one of these as the basis for a world-wide 3G standard called IMT2000. To the surprise of many, the ITU decided eventually that all three standards would be covered by IMT2000, and that all three standards would coexist. W-CDMA and UMTS were already very similar from the start, and merged into a single standard. UMTS/W-CDMA uses a GSM-based infrastructure, and has an air interface that uses elements both from GSM (including a TDD mode) and CDMA-based systems. CDMA2000 is a further development of earlier CDMA-based systems such as IS-95. After the ITU decision, the CDMA2000 and UMTS standards have been further aligned. Products for both of these systems are currently being developed, but due to the economic downturn and high costs involved in the licenses and network infrastructure, the 3G systems have not yet been introduced on a wide scale.

The generation beyond 3G, often called 4G or NextG, is still in the early definition phase. It will to provide higher data rates (up to 1Gbps) for the user, and better quality of service. 4G is expected to achieve this through a combination and convergence of wireless data and cellular systems. In addition, more bandwidth efficient modulation methods will be employed, and the adaptivity towards applications and the environment is expected to increase [7].

B.2 Cordless phone systems

Cordless phones initially used analog narrowband FM transmissions. In the USA, the CT0 system was widespread and relatively cheap. CT0 systems operated in the 46MHz and 49MHz bands in the USA (figure 122), and features such as multi-channel operation, automatic channel selection, and limited forms of security were added over the years. The moderate frequencies and rather relaxed CT0 standard allowed for common and cheap technologies and components to be used, which probably contributed significantly to the success of this standard.

Figure 122 *CT0 phone*

In Europe, most countries initially allowed only CT1 systems, which operated using analog narrowband FM transmission in the 900MHz band. The tough specifications for such systems, in combination with the higher carrier frequency, made CT1 phones about one order of magnitude more expensive to the consumer than CT0 phones, without offering a clear advantage. In addition, the mandatory regular communication between handset and base station caused very short but quite noticeable (and to many people irritating) interruptions in the voice connection.

Figure 123 *DECT*

Another perceived limitation of both CT0 and CT1 at the time was that they could not cope with the amounts of voice traffic that can be expected in large office buildings, which was assumed to become one of the main applications for cordless phones. Most European countries were planning to solve this unsatisfactory situation by the introduction of a new digital cordless telephone standard *(Digital Enhanced Cordless Telecommunications System*, DECT) through the European Telecommunications Standardization Institute (ETSI), in parallel with the introduction of GSM. The specifications for the radio part of the DECT system were rather relaxed, and a DECT phone that would be competitive with a CT1 phone in terms of price and performance was easy to conceive.

When the DECT standard was nearing completion, CT0 phones became more and more popular in European countries, even those where such phones were illegal. At some point, the number of illegal CT0 phones in the Netherlands was estimated to be one order of magnitude larger than the number of legal CT1 phones. CT0 phones were legalized in most European countries shortly afterwards, around the same time that the first DECT

phones became available in the market (figure 123). Although DECT phones were significantly cheaper than CT1 phones while offering more features, in most countries they were initially around 5 times more expensive than CT0 phones, without offering obvious advantages to consumers.

Also, the main market for cordless phones turned out to be in consumer homes and small offices, rather than the large office environment originally envisioned. Hence, the larger capacity of DECT systems, and its related features such as handover, were not relevant for the main market, resulting in a slow growth of DECT market share. In the USA, digital cordless phones were also developed, initially as spread-spectrum phones in the 900MHz range with many proprietary air interfaces, later standardized as WCPE in the 2.4GHz band.

An intermediate digital cordless phone standard called CT2 [3] was developed to fill the perceived time gap between CT0/CT1 on one hand, and DECT on the other. CT2 was popular for a short period in the UK, where it could be used both as a cordless phone at home, and as a "poor man's" cellular phone through access points in the city and near freeways. However, it had several drawbacks compared to a real cellular phone:

- Coverage was limited to the vicinity of an access point. With only a few access points per a city, people had to go to a specific location to make a phone call. Most often, these locations also happened to have public phone booths nearby.
- Handover between access points was not provided for in the CT2 standard. Even if two access points were close enough to have their coverage areas overlap, phone calls would still be disconnected when moving from one access point to the next.
- Only outgoing phone calls were possible, since there was no provision in the standard, or in the implemented systems, for the location registers or broadcast functions required for incoming phone calls.
- Both the phones themselves and phone calls through public access points were fairly expensive.

Shortly after the introduction of CT2 telepoint services, the GSM system took off in the market. As a consequence, the CT2 system failed miserably in the UK, with large financial losses for the companies involved. This didn't keep the Dutch PTT from introducing the same system in the Netherlands shortly afterwards, with similar results.

B.3 Wireless data

Wireless data connections can be divided into three types:

1. Wide Area Networks (WAN), with coverage "everywhere" (typically spanning multiple countries).
2. Wireless Local Area Networks (WLAN), with a range sufficient to cover a small building, or a significant part of a larger one.
3. Wireless Personal Area Networks (WPAN), with a range of a few meters.

GSM already provided for wireless data services in the first generation of GSM products, allowing it to be used for WAN applications. Typically, the GSM phone needed to be connected to a computer through a special (and very expensive) cable. After this

connection was established, it basically operated like a 9.6kbps (later 14kbps) modem, also circuit-switched. At connection rates initially in excess of €1 per minute, it wasn't popular at all. With the introduction of HSCSD and GPRS, data rates went up (up to 38kbps) and costs went down, now based on data volume rather than connection time (currently around €0.50 per megabyte). The expensive data cable was replaced by an infrared and later a Bluetooth (BT) connection, creating a low threshold for the use of this service. Also, handsets with integrated E-mail clients, WAP, I-mode and web browsers, and even digital still picture cameras, have started to appear and are likely to make the use of WAN data connections even more popular.

For local connections, DECT already provided for wireless data rates of 64kbps using a single time slot. It also offered the possibility to link multiple time slots to increase the data rate. Given the typical range of a DECT system, this could have become the basis for a WLAN system. However, it lacked support for easy ad-hoc configuration, and for integration into the network layers of popular operating systems. DECT was also circuit-switched rather than packet-based. Some wireless data products based on DECT appeared in the market, but were not very popular. They typically combined the functionality of a small cordless phone system (typically up to 6 handsets) with cordless access to the POTS or ISDN telephone network from a computer.

With the introduction of pure WLAN systems, based on the IEEE 802.11b standard, cheap and convenient implementations of WLAN transceivers became widely available. They are quickly becoming very popular, especially with prices of transceivers dropping to around €30 or less (figure 124).

Figure 124 *WLAN transceiver for laptop*

IEEE 802.11b provides for raw data rates of up to 11Mbps, and allows handover between different access points to extend the coverage to larger areas. The system is being used both for home/office applications with private access points, as well as through so-called "hot spots" with public access points at e.g. airports, coffee stores, etc. However, IEEE 802.11b works at frequencies in the 2.4GHz industrial, scientific & medicinal (ISM) band, which is also used for other purposes, including microwave ovens, BT, WCPE, etc. This reduces the throughput, and currently solutions are being developed to work around these limitations.

The IEEE 802.11a standard is likely to offer significant improvement in these areas. It operates at less crowded frequencies between 5GHz and 6GHz, and offers data rates up to 54Mbps. IEEE 802.11g offers the same 54Mbps at the 2.4GHz band as a compromise between the high data rate of IEEE 802.11a and the low cost of 802.11b. Other, similar, standards have been proposed as well, most notably the European Hiperlan. However, at the moment, it seems that the IEEE 802.11 family of standards will become the dominant standards for WLAN applications for some time to come.

The next generation of WLANs will probably target even higher data rates. There are three likely approaches to achieve this:

1. Moving to even higher frequencies, such as the proposed Wind-Flex standard of the Wind-Flex consortium, that could offer in excess of 100Mbps data rates per channel in the 17GHz band. One obvious limitation of this approach is the very poor propagation of radio signals through walls at these frequencies, limiting the application to basically a single room for every access point.
2. Using even higher bandwidth efficiency at the same frequency bands that are already in use. The limitation here is the need for better signal-to-noise ratios, which translate in less range and lower resistance against interferers. Products using this approach are based on IEEE 802.11b (doubling the data rate to 22Mbps) and on 802.11a.
3. Re-using bandwidth that is already allocated to other transmitters, which allows much larger bandwidths at lower frequencies than would otherwise be possible. UWB is a likely candidate for this approach, but there is still relatively little experience of real-world behavior of UWB systems in a heavy-traffic environment.

In the WPAN area, several standards exist. Most notably, the Bluetooth (BT) standard was introduced in the middle of the 1990s, with the intention of providing a cheap replacement for the many wires between cellular phones, headsets, infrared connections etc. The radio specifications for BT were originally quite relaxed in order to allow cheap implementations. As happened with DECT, intense competition, especially on sensitivity, resulted in the need for much better performance than in the original standard. BT offers data rates of up to 720kbps in multi-slot mode over a range of a few meters for the low power classes. The high power class should cover a range of 100m. Extensions of the BT standard that allow higher data rates are currently in development.

Zigbee is a standard that offers even lower power and lower cost than BT, but also lower data rates. The main applications are expected to be in the area of remote control, and first products are currently under development.

B.4 Telecommunication system parameters

The main system parameters of the systems in this thesis are shown in the table at the end of this section (fig. 125). The indications for receive (Rx) and transmit (Tx) are from the perspective of the handset. The frequency indicated is the lower edge of the receive band. In TDMA systems, the receive band is often (but not always) identical to the transmit band. In FDMA systems, the transmit band is below the receive band. The bandwidth indicated in the table is the channel bandwidth of the receiver, but almost always identical to the channel bandwidth of the transmitter. The modulation type is the modulation as received from the base station, and often (but not always) identical to the transmit modulation.

The "Min SNR Rx" indicates the minimum signal-to-noise ratio required at the IF output of the receiver to achieve the bit-error (or frame-error) rate specified in the standard. The "Min Rx signal" indicates the required receiver sensitivity. The "Max Rx

signal" indicates the maximum desired signal strength at the receiver. Although not mentioned in this table, most standards also specify maximum levels for interferers that are at the same channel, or at adjacent or alternate channels. The "Min Tx signal" and "Max Tx signal" columns specify the required transmitter signal range as defined in the standard for a specific class of handsets. In several standards, multiple classes of handsets are specified, each with their own transmit power range. The GPS system has only a receiver in the handset. Therefore, there are no values for the transmitter power.

This table shows the wide range of parameters for these systems, with RF frequencies ranging from below 1GHz to over 5GHz, channel bandwidths from 10kHz to over 20MHz, required signal-to-noise ratios from 0 to 18dB, and sensitivity levels from -130dBm to -70dBm. This wide range is caused, among others, by the differences between long-range and short-range systems.

System Name	Frequency MHz	BW kHz	Modulation	Min SNR Rx dB	Min Signal Rx dBm	Max Signal Rx dBm	Min Signal Tx dBm	Max Signal Tx dBm
Table Conten;IEEE 802.11a	5150	16600	OFDM B/QPSK/QAM16/64	5	-82	-30	0	20
IEEE 802.11b Japan	2484	22000	DSSS 8ch CCK DB/DQPSK	6	-76	-10	0	23
IEEE 802.11b USA/EU	2412	22000	DSSS 8ch CCK DB/DQPSK	6	-76	-10	0	23
BlueTooth class 2	2402	1000	FHSS GFSK BT=0.5	18	-70	-20	-6	4
DCS 1800	1805	200	GMSK BT=0.3 TDMA	9	-103	-23	0	30
DECT	1880	1152	GMSK BT=0.5	12	-83	-4	26	26
DECT PAP	1880	1152	GMSK BT=0.5 TDMA	12	-86	-4	26	26
E-GSM 900	925	200	GMSK BT=0.3 TDMA	9	-103	-23	5	33
EDGE 1800	1805	200	8-PSK TDMA	16	-98	-23	5	30
EDGE 1900	1930	200	8-PSK TDMA	16	-98	-23	5	30
EDGE 900	921	200	8-PSK TDMA	16	-98	-23	5	33
EGSM-900	925	200	GMSK BT=0.3 TDMA	9	-98	-23	5	33
GPRS 1800	1805	200	GMSK BT=0.3 TDMA	16	-98	-23	5	30
GPRS 1900	1930	200	GMSK BT=0.3 TDMA	16	-98	-23	5	30
GPRS 900	921	200	GMSK BT=0.3 TDMA	16	-98	-23	5	33
GPS	1227.6	2046	BPSK	0	-130	-123	0	0
GSM-450	460.4	200	GMSK BT=0.3 TDMA	9	-98	-23	5	33
GSM-480	488.8	200	GMSK BT=0.3 TDMA	9	-98	-23	5	33
GSM-850	869	200	GMSK BT=0.3 TDMA	9	-98	-23	5	33
GSM-900	935	200	GMSK BT=0.3 TDMA	9	-98	-23	5	33
Hiperlan2	5180	16600	OFDM B/QPSK/QAM16/64	10	-85	-20	-15	20
IS-91 (AMPS,CAPCS)	869	30	FM dev=2.9kHz(FM)	8	-113	-25	8	36
IS-91 (NAMPS,CAPS)	869	10	FM dev=1.5kHz(NFM)	8	-116	-25	8	36
IS-95	869	1250	CDMA 1.2288Mcps	2.7	-104	-25	-50	36
IS-98/95 1900	1930	1250	CDMA 1.2288Mcps	2.7	-104	-25	-50	28
IS-98/95 900	869	1250	CDMA 1.2288Mcps	2.7	-104	-25	-50	36
P-GSM 900	935	200	GMSK BT=0.3 TDMA	9	-103	-23	5	33
PCS 1900	1930	200	GMSK BT=0.3 TDMA	9	-102		0	30
PHS	1880.15	288	FDMA/TDMA pi/4 QPSK					19
R-GSM 900	921	200	GMSK BT=0.3 TDMA	9	-103	-23	5	33

Figure 125 *Telecommunication system parameter table*

C

Technology impact on RF performance

In this appendix, the impact of technology on RF performance is discussed in more detail. The scaling of IC technologies, as well as the impact of limitations in modeling and interaction between circuit blocks will be discussed.

C.1 IC Technology scaling

RF performance is typically limited by IC, packaging, assembly, discrete component, PCB, and many other implementation technologies. All of these have an impact on the power dissipation of an RF front end, but IC technologies have the most direct impact, since most of the power flows from the power supply into ICs.

To keep this subject manageable, the focus will be on the impact of IC technologies on the following (main) RF performance parameters:

- gain (and loss), caused by losses in non-ideal passive components, interconnect, substrate, and parasitic elements of active devices. Typically, gain and loss are functions of frequency, and are defined at a specific frequency or within a frequency band.
- linearity (and distortion), caused by non-linear behavior of active and passive components. Linearity is also a function of frequency, since part of the distortion is caused by non-linear frequency-dependent components, e.g. junction capacitors.
- noise, caused by losses, substrate, passive and active devices, etc. Again, this is typically a function of frequency, especially in the case of active devices.
- power dissipation, which is typically dealt with as a dependent variable determined by the requirements for the other performance parameters.

When developing a product, many other parameters are important as well, such as cost, stability, isolation, ruggedness, and reliability. However, these parameters tend to have

a lower impact on power dissipation than the ones mentioned above (although there are some notable exceptions, such as the impact of ruggedness requirements on the *PAE* of an RF power amplifier). The same holds for technologies: although there is definitely an impact of packaging, assembly, discrete component and other technologies on power dissipation, the main impact comes generally from the IC technology, with again the RF power amplifier as a significant exception. Since this power amplifier is so different, in many aspects, from the small-signal circuits of a transceiver, it has become a separate specialism within the RF area, with its own problems, solutions, and sometimes even different technologies [22]. Figure 126 shows the power flows in a typical power amplifier from battery to antenna, illustrating the impact of different technologies on the power dissipation in this specific part of a transceiver.

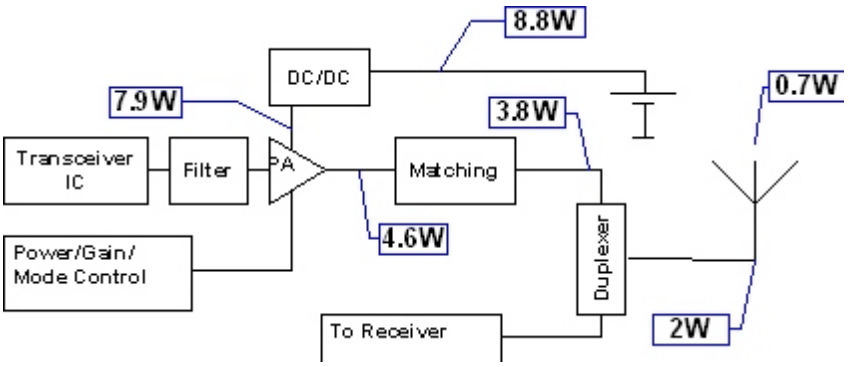


Figure 126 *Power flows from battery to antenna in a typical RF PA*

C.1.1 IC Device Types

Devices in IC technologies are built by depositing, implanting, etching, diffusing, or otherwise fabricating stacks of patterned horizontal layers in (or on) a substrate, often a semiconductor material such as silicon. The dimensions of these devices are of special importance for the power dissipation of RF circuits, since larger devices generally require more current to operate. The relation between technology scaling and active device performance will be investigated in more detail in the next two subsections. At this point, some basic concepts will be introduced that are needed for the discussion of the relation between gain and device scaling in the next section.

Two families of devices can be built using IC technologies. The distinction is based on the direction of the main current flow with respect to the substrate:

1. in lateral devices, the main current flows parallel to the substrate surface;
2. in vertical devices, the main current flows perpendicular to the substrate surface.

Although this distinction is most commonly made for active devices (e.g. “lateral PNP transistor” or “vertical NPN transistor”), it can be equally well applied to passive devices such as capacitors or resistors.

C.1.2 Lateral devices

In a lateral device, the current flow tends to scale proportionally to the width of the device, and reciprocally to the length of the device. For example, the current in a (lateral) CMOS device is in first order proportional to the W/L ratio, and the resistance of a lateral

resistor is in first order proportional to L/W ratio.

A simplified layout of a lateral device is shown in figure 127. For practical reasons, such as the need to add contacts to a device, the actual device is often larger than the part of the device that carries the main current, and across which the main voltage drops occur. This can be approximated by adding compensation factors (ΔL and ΔW) to the length (L) and width (W), resulting in a total device size of ($\Delta L + L$) by ($\Delta W + W$).

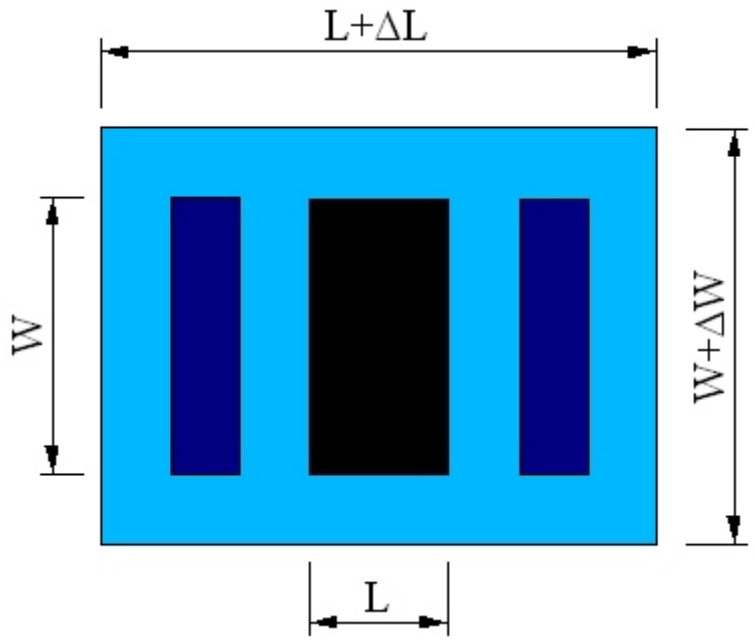


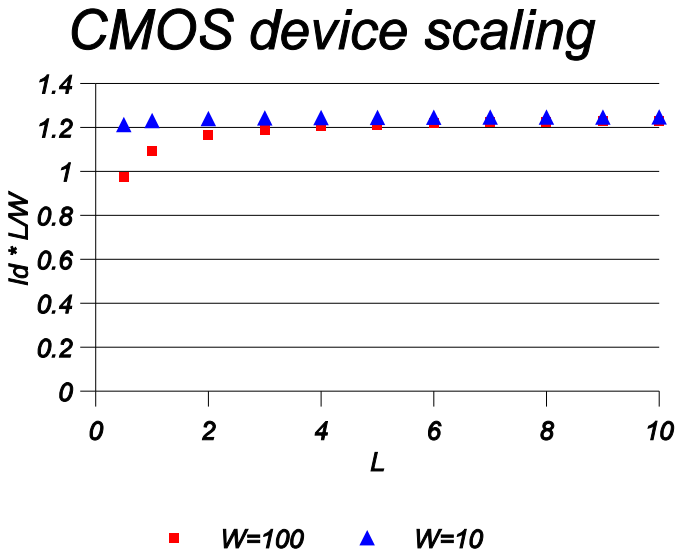
Figure 127 *Simplified lateral device layout*

A simple model for the current through a lateral device is based on the assumption that the current density J is proportional to the width/length (W/L) ratio when the voltage across the device is kept constant. For passive and homogeneous devices, such as resistors, this is an accurate model, since such devices have a constant electrical field strength:

$$J = \frac{E}{\rho} = \frac{V}{L\,\rho} = \frac{I}{W\,t}$$

$$\rho = R\,\frac{W\,t}{L} \qquad \textbf{(93)}$$

For active devices, this model is less accurate, especially for short channels and in devices in which the substrate or a common well acts as a common terminal to all (parts of) active devices. However, it still is accurate to within ±15% for e.g. 0.5µm CMOS devices when changing the gate width and length over one order of magnitude, as shown by the simulation results below:



The current I through a lateral device can therefore be modeled as:

$$I = \alpha_1 \frac{W}{L} \tag{94}$$

with α_1 a device-dependent scaling constant. Parasitic effects of lateral devices can be modeled as additional resistive and capacitive elements. The capacitive elements scale with bottom area and perimeter length:

$$C_{par} = \alpha_2 (W+\Delta W+L+\Delta L) + \alpha_3 (W+\Delta W)(L+\Delta L) \tag{95}$$

with α_2 and α_3 device-dependent scaling constants. Please note that the parasitics are mostly caused by the total device area and perimeter, not just the part that carries the main current. The additional parasitic resistive elements end up as series elements that scale with the length extensions:

$$R_{par} = \alpha_4 \frac{\Delta L}{W+\Delta W} \tag{96}$$

with α_4 a device-dependent constant. This results in the following simplified schematic for a lateral device with its related parasitics:

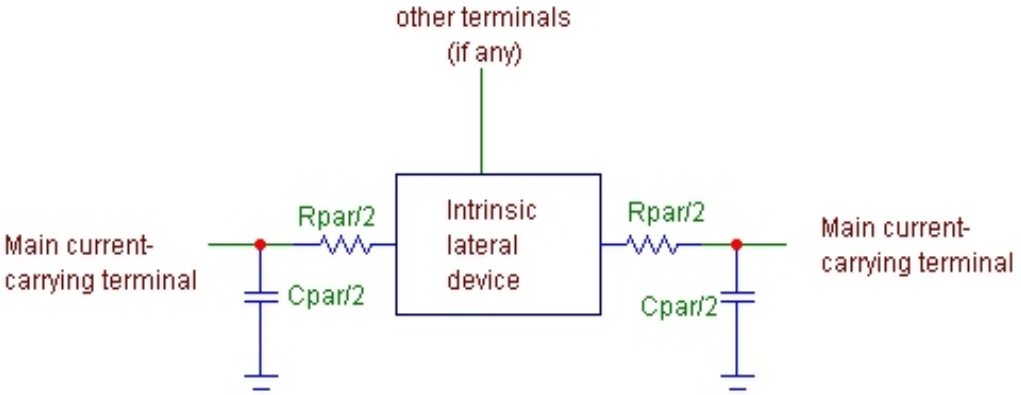


Figure 129 *Simplified schematic for a lateral device with related parasitic elements*

The parasitic will result in an additional, frequency-dependent current J_{cpar} that can be expressed as:

$$I_{cpar} = \alpha_5 f C_{par} \tag{97}$$

The ratio between intended current and parasitic current can now be expressed as:

$$\begin{aligned} \frac{I}{I_{cpar}} &= \frac{\alpha_1 \frac{W}{L}}{\alpha_5 f C_{par}} \\ &= \frac{\alpha_1 \frac{W}{L}}{\alpha_5 f\left(\alpha_2(W+\Delta W+L+\Delta L)+\alpha_3(W+\Delta W)(L+\Delta L)\right)} \\ &= \frac{\alpha_1}{\alpha_5 f L\left(\alpha_2(1+\frac{\Delta W}{W}+\frac{L}{W}+\frac{\Delta L}{W})+\alpha_3(1+\frac{\Delta W}{W})(L+\Delta L)\right)} \end{aligned} \tag{98}$$

The partial derivative of this ratio with respect to the width W is:

$$\frac{\partial\left(\frac{I}{I_{cpar}}\right)}{\partial W} = \frac{\alpha_1\alpha_2(\Delta L+\Delta W+L)+\alpha_3\Delta W(\Delta L+L)}{\alpha_5 f L(\alpha_2(\Delta L+\Delta W+L+W)+\alpha_3(\Delta L+L)(\Delta W+W))^2} \quad \textbf{(99)}$$

which is positive definite for all W and L. Similarly, the partial derivative of the I/I_{cpar} ratio is:

$$\frac{\partial\left(\frac{I}{I_{cpar}}\right)}{\partial L} = -\frac{\alpha_1 W(\alpha_2(\Delta L+\Delta W+2L+W)+\alpha_3(\Delta L+2L)(\Delta W+W))}{(\alpha_5 f L^2(\alpha_2(\Delta L+\Delta W+L+W)+\alpha_3(\Delta L+L)(\Delta W+W))^2)} \quad \textbf{(100)}$$

which is negative definite for all W and L. This shows that, for any lateral device, the ratio between intended current and parasitic current inevitably decreases with L and increases with W. It also decreases at higher frequencies. Since the device current is proportional to W/L, any attempt to decrease the device current at higher frequencies by reducing device width and/or increasing device length will result in a relatively larger impact of currents through the parasitic capacitance, bypassing the device.

Please note that this effect even occurs when ΔW and ΔL are zero. In that case, equation (6, 98, 101, 102, 102) simplifies to:

$$\frac{I}{I_{cpar}} = \frac{\alpha_1}{\alpha_5 f L\left(\alpha_2(1+\frac{L}{W})+\alpha_3 L\right)} \quad \textbf{(101)}$$

which has the same properties of the partial derivatives with respect to L and W. The gain bandwidth product $G\,BW$ can be derived by finding the frequency f at which the I/I_{cpar} ratio reaches 2, that is, half the output current ends up in the parasitic capacitance of the device, and therefore the gain is decreased by 3dB.

$$\begin{aligned} G\,BW &= \frac{\alpha_1\frac{W}{L}}{2\alpha_5\left(\alpha_2(W+\Delta W+L+\Delta L)+\alpha_3(W+\Delta W)(L+\Delta L)\right)} \\ &= \frac{\alpha_1}{2\alpha_5 L\left(\alpha_2(1+\frac{\Delta W}{W}+\frac{L}{W}+\frac{\Delta L}{W})+\alpha_3(1+\frac{\Delta W}{W})(L+\Delta L)\right)} \end{aligned} \quad \textbf{(102)}$$

C.1.3 Vertical devices

In a vertical device, the current is in first order proportional to both the length L and the width W of the device when the voltage across the device is kept constant. For example: in a vertical NPN transistor, the current is proportional to the $L*W$ product, and in a parallel-plate capacitor the capacitance (and therefore the current by constant voltage) is proportional to the $L*W$ product as well.

A simplified layout of a vertical device is shown in figure 130. Like lateral devices, most vertical devices need additional space for contacts, spacing between patterns, etc., which results in additional width and length (ΔL and ΔW). Added to the intrinsic device, this results in a total device size of ($\Delta L + L$) by ($\Delta W + W$), as shown in figure 130.

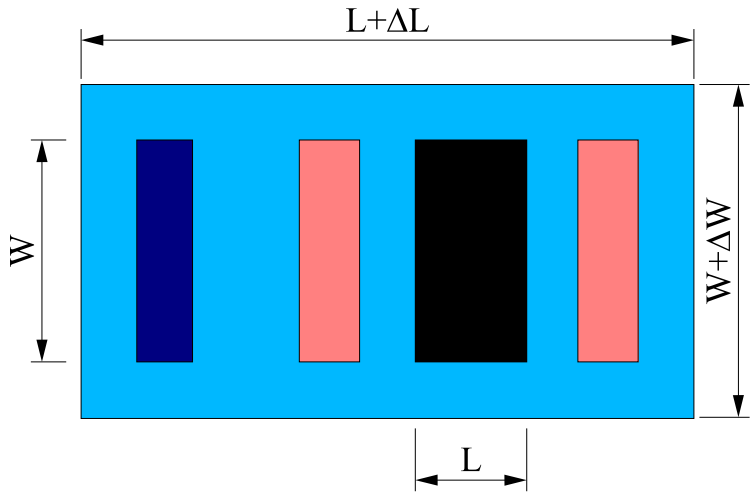


Figure 130 *Simplified layout of a vertical device*

In vertical devices, both passive and active, the current scales accurately with W*L. This is due to the fact that the electrical field distribution in such devices is almost independent of the scaling, since it is orthogonal to both scaling dimensions. As with lateral devices, expressions for the scaling of intended and parasitic currents can be derived. The scaling of the intended current is modeled as:

$$I = \alpha_6 L W \qquad \textbf{(103)}$$

Again, the parasitic capacitance of a vertical device scales with bottom area and perimeter length:

$$C_{par} = \alpha_7 (W+\Delta W+L+\Delta L) + \alpha_8 (W+\Delta W)(L+\Delta L) \qquad \textbf{(104)}$$

The current caused by the parasitic capacitance is:

$$I_{cpar} = \alpha_9 f C_{par} \tag{105}$$

This results in a ratio of intended and parasitic current of:

$$\begin{aligned} \frac{I}{I_{cpar}} &= \frac{\alpha_6 W L}{\alpha_9 f C_{par}} \\ &= \frac{\alpha_6 W L}{\alpha_9 f\left(\alpha_7(W+\Delta W+L+\Delta L)+\alpha_8(W+\Delta W)(L+\Delta L)\right)} \\ &= \frac{\alpha_6}{\alpha_9 f\left(\alpha_2(\frac{1}{L}+\frac{\Delta W+\Delta L}{WL}+\frac{1}{W})+\alpha_8(1+\frac{\Delta W}{W})(1+\Delta\frac{L}{L})\right)} \end{aligned} \tag{106}$$

From this, the partial derivatives for L and W respectively, are derived as follows:

$$\frac{\partial\left(\frac{I}{I_{cpar}}\right)}{\partial L} = \frac{\alpha_6 W(\alpha_7(\Delta L+\Delta W+W)+\alpha_8\Delta L(\Delta W+W))}{\alpha_9 f(\alpha_7(\Delta L+\Delta W+L+W)+\alpha_8(\Delta L+L)(\Delta W+W))^2} \tag{107}$$

$$\frac{\partial\left(\frac{I}{I_{cpar}}\right)}{\partial W} = \frac{\alpha_6 L(\alpha_7(\Delta L+\Delta W+L)+\alpha_8\Delta W(\Delta L+L))}{\alpha_9 f(\alpha_7(\Delta L+\Delta W+L+W)+\alpha_8(\Delta L+L)(\Delta W+W))^2} \tag{108}$$

With both partial derivatives being definite positive, the ratio between intended and parasitic current will be smaller (and thus worse) for smaller values of either L or W. In other words: any vertical device will unavoidably show a worse ratio between intended and parasitic current for smaller values of W and L, that is, for smaller intended currents. Also, the ratio decreases at higher frequencies. In a similar way, it can be shown that the same conclusion holds even if ΔL and ΔW are zero.

The gain bandwidth product $G\,BW$ can be derived in the same way as for lateral devices:

$$GBW = \frac{\alpha_6 WL}{2\alpha_9\left(\alpha_7(W+\Delta W+L+\Delta L)+\alpha_8(W+\Delta W)(L+\Delta L)\right)}$$

$$= \frac{\alpha_6}{2\alpha_9\left(\alpha_2(\frac{1}{L}+\frac{\Delta W+\Delta L}{WL}+\frac{1}{W})+\alpha_8(1+\frac{\Delta W}{W})(1+\Delta\frac{L}{L})\right)} \quad (109)$$

Since in IC processes, devices are built either as lateral or vertical devices, and since for both types of devices the ratio between intended and parasitic current becomes smaller (and thus worse) for smaller currents through the device and for higher frequencies, there is a fundamental problem with scaling devices in any IC technology to achieve low power at high frequencies. These problems can be reduced by special technologies, such as vertical scaling, trenches, SOI, and SOA, but the fundamental problem remains.

C.2 Other technology-induced performance limitations

In addition to the scaling of devices, there are other performance limitations that are related to, but not directly part of, IC technology. The two most relevant ones are discussed below.

C.2.1 Modeling problems

One complicating factor in developing RF front-end circuits, and especially when low power is an important design target, is the accurate modeling of the circuit. As discussed in the previous section, the impact of device extensions ΔW and ΔL, and of parasitics related to the perimeter of devices, increases with decreasing device sizes and currents. Therefore, low-power circuits operate closer to the high-frequency limits of an IC process than circuits with higher currents at the same frequency. In other words, the design margin between requirements and process performance is smaller for low-power RF circuits. In addition, the impact of additional parasitics, such as perimeter capacitance, becomes more important.

Accurate modeling of all devices, but also of interconnect, substrate, packaging etc., is required in order to achieve a low-power RF front-end design within a reasonable amount of time. For some devices, very accurate models are available, for example for bipolar transistors. Models such as Mextram [36] provide sufficient accuracy for most circuits. As a result, bipolar transistor model inaccuracies no longer contribute significantly to overall simulation inaccuracies. For other effects, only very limited models are available, for example for substrate electrical coupling in IC processes that use intermediate substrate impedances (around 10Ωcm).

It is not likely that accurate models for all relevant effects will be available in the near future. And even if they would be available, along with the appropriate parameters

for each of the models and for the relevant IC process, the number of potentially relevant effects is so high, that it would be impractical to simulate circuits with models for each of these effects. The simulation time would be prohibitive, even with the computer capacity increases that can be expected over the next five years. Moreover, many of these models require information that is simply not available in the circuit-design phase of a top-down design approach. For example, wire lengths and shapes, interconnect levels, relative positions of components and circuits on the IC substrate, etc. will not be known until the layout phase is (partially) completed, and layout parameter extraction and back-annotation are carried out.

In practice, designers make choices - often based on experience - about what to model, and use estimates for parameters that are not exactly known at this stage in the design process. This is where experienced designers distinguish themselves from less experienced designers: making the right choices for the effects that need to be modeled, and the right estimates for unknown parameters, has a very big impact on the performance of the IC, and determines to a high degree the number of redesigns needed to arrive at the required level of performance.

The small margin between required and achievable performance, and as a consequence the large number of effects that need to be taken into account, are the main drivers for improved modeling. One way to reduce or even avoid this problem is to use technologies that have a much higher performance than is required for a specific application. This is of course not very cost efficient, and therefore not always possible.

Another way to reduce the modeling problem is to integrate the circuit-design and layout phases in the design process. This is already done for very high frequency circuits, but is limited to circuits of low complexity. Making changes in a layout is often much more time consuming than making the same changes in a schematic representation of the circuit. For example, changing a resistor or capacitor value can have dramatic consequences in a layout, requiring major changes to the position of many other devices and their interconnect, whereas at the schematic level, only a single number changes in the schematic. For larger circuits and bigger changes, the amount of work can cause long delays in the development project.

A third way to reduce the modeling problem is to build circuits from larger building blocks that have been previously designed and characterized, similar to the design of digital circuits using building blocks at higher abstraction levels, such as gates, registers, memories, and cores. However, this does not solve all modeling problems, since there can also be significant unintended interactions between blocks.

C.2.2 Inter-block interaction

Unintended, or parasitic, interaction between circuit blocks can be caused by several effects, including:

- Non-linear input and output impedances
- Thermal and/or electrical coupling through the substrate
- Magnetic coupling of internal inductors in each of the blocks
- Magnetic coupling of internal inductors with bond wires
- Injection of spurious signals that are present at inputs and/or outputs
- Coupling through power supply and ground lines

Modeling and characterizing all these effects is cumbersome at best, and almost impossible in those cases where the relative positions of the circuit blocks involved, and

of other circuit blocks, play a significant role. This situation can be improved by designing circuit blocks in such a way that they are relatively insensitive to most of these effects, and that they cause as little thermal, electrical, and magnetic disturbance as possible. This is of course not always possible because of the limited margins between available and required performance, and because of cost.

One technology that helps to reduce inter-block interaction, and at the same time alleviates modeling problems, is RF-module technology. Using this technology, circuit blocks can be implemented on individual physical ICs, and packaged and connected in an RF module. By modeling and characterizing the building blocks based on measurements of the physical ICs, their behavior can be accurately predicted. The interaction between circuit blocks is minimized by the larger physical distance in a module, the elimination of a common semiconductor substrate, and decoupling of the power and ground signals between the blocks.

Appendix

D

Current design methods & results

In the following sections, short descriptions of the different categories of design methods are presented.

D.1 Automatic versus manual design methods

Automatic design methods are based on tools that automatically carry out the steps prescribed in the design method. From a methodology point of view, there is little difference between automatic and manual design methods: automatic design methods can always be carried out manually, but this might in some cases not be practical because of the amount of work involved. Conversely, manual design methods can often be implemented in an automatic way, although this is less obvious. Especially design methods based on pattern recognition and experience are difficult to automate.

An example of an automatic design method is circuit sizing using the tool Adapt [65]. This design method covers part of the schematic design abstraction level, in which the circuit topology, implementation technology, and target specifications have already been defined. Using this method, the parameter values of the various devices can be optimized according to a specified optimization criterion.

An example of a manual design method will be discussed in the next chapter.

D.2 Iterative versus single-shot methods

Iterative design methods contain a loop in their design process. At the beginning of an iteration through this loop, a version of the design is available that does not meet all performance requirements. In the loop, some operations are carried out on this design that modify its performance. At the end of the loop, the performance is compared to the performance requirements. If the requirements are not met, the design is used as a starting point for the next iteration through the loop. Iterative circuit design methods, in which the evaluation of the achieved performance is determined through a circuit simulator, are also called "simulation-based methods" [66, 66, 66]. Many, and maybe even most, current RF products are designed based on iterative methods as described in Section 1.4.

Single-shot design methods achieve their results without the need for iterations. These design methods are often based on equations or rules, and are often associated with properties such as "correct by construction" and "first time right". Single-shot methods have become popular in the digital area, but are currently also being proposed and used for RF design.

As with all design methods, it is important to realize that results achieved through single-shot design methods are only as valid and accurate as the models and assumptions on which such methods are based. Therefore, methods such as single-shot, exact, and algorithmic are not necessarily better than heuristic, iterative, and approximate design methods. An approximate solution based on accurate models can be much more useful than an exact solution based on unrealistic models. This also explains the differences in popular design methods between digital and RF design. In digital design, the models more accurately predict the performance of a design, making single-shot methods more likely to be successful than in the RF area.

D.3 Algorithmic versus heuristic methods

Heuristic methods are based on practical and empirical rules rather than theory or scientific principles. This category includes expert systems, but also the manual design methods employed by many experienced designers. The main disadvantage of heuristic methods is that it cannot be guaranteed that an achieved solution is optimal, or even functional. However, this method often results in solutions that would be very difficult or impossible to achieve otherwise, especially for very complex design problems.

D.4 Model versus reality based methods

Most current RF IC design methods use models of the system, circuit, and components that are used in the design. This has several advantages:

- Models allow access to all signals, parameters, and internal states of the components, circuits and system, without disturbing the operation of the model. This can be an enormous help in understanding the behavior of a component, circuit or system.
- Models do not break, and give 100% reproducible results.
- Models can be used for designs at a time that the actual components, circuits, or systems are not yet available, or cannot yet be fabricated.

There are also disadvantages to the use of models. Especially the accuracy of the model is usually limited, either because the model does not accurately describe the behavior of some effects, or because some effects are not included in the model at all.

As an alternative, a design method can directly use the components, circuits and system that will be part of the final design. An example of this approach is the traditional iterative "trial and error" method of RF design by soldering components on a PCB, carrying out measurements, and adjusting the design until the required performance is achieved. The outcome of such a design method is simultaneously the design and the final implementation of this design.

Sometimes it is not obvious which method is being followed. For example, a prototype of a volume product could be developed using a reality based method. However, a prototype that meets all requirements does not guarantee a design that meets all requirements: because of parameter spread in volume production, a significant percentage of products could still fail to meet the requirements. Therefore, a prototype has to be considered a model in the context of design methods.

D.5 Exact versus approximate methods

An exact design method results in a design that exactly meets the requirements, and therefore is the optimum design for the given target specifications. An approximate design method results in a design that approximately meets the requirements, and therefore is close to the optimum design for the given target specifications. Approximate methods are popular, because the limited accuracy of the models, and assumptions used in RF design methods, reduce the relevance of an exact solution.

Whether an exact method is more appropriate than an approximate one depends also on the application for which a design is being developed. In some applications, there are no hard requirements on performance, although performance of course plays an important role in the value of a design. In other applications, hard performance requirements do exist, for example as type approval requirements that need to be met before any product can be sold.

D.6 Explicit and implicit design methods

In explicit design methods, the goal of the design and the method to be used are formally described. In implicit methods, the design method or the requirements are not explicitly stated. Also in this case, the method being used might not be obvious. Even if a complete formal specification and design method is available, several aspects might not be formally specified but are still very much expected. For example, the specifications for a low noise amplifier usually provide gain, noise figure, linearity, power dissipation, input and output impedance, power supply voltage and temperature range. Often, they do not explicitly include the requirement to:

- not oscillate
- not generate significant spurious signals on the power supply
- not explode
- have limited switch-on currents

etc. Still, if one of these requirements would not be met, the customers for such a design would probably not be satisfied by the result. In practice, it is very difficult or even impossible to provide a complete, formal, specification for anything but the most trivial product.

D.7 Bottom-up versus top-down design methods

In bottom-up design methods, the design starts with the investigation of the possibilities and performance of the individual parts of a design, and subsequently these parts are combined into a complete design that meets the specifications. Top-down starts from the specifications for the complete design, and derives from these the architecture and specifications for subcircuits, from which in turn topologies and subcircuit parameters are derived [13].

Both design methods have disadvantages:

- by starting from individual parts of a design, a global optimization of e.g. gain, noise and linearity distribution is difficult to achieve.
- by starting from the specifications, assumptions have to be made about the feasibility of subcircuit performance that can only be verified in a later stage of the design process.

For that reason, several meet-in-the-middle approaches have been defined, such as bottom-to-top creation [8], that combine partial top-down with partial bottom-up methods.

D.8 Custom versus reuse design methods

One way to speed up the design process is by reusing parts of a circuit that has been designed previously. Reuse can occur at any abstraction level of the design, such as system, schematic, and layout. The main disadvantage of a reuse method is that it only is practical if a circuit or part of a circuit with acceptable specification and implemented in an appropriate technology is available. This is often not the case. Custom design from scratch is one way to solve this, but in some cases it can be more efficient to design a circuit by modifying an existing circuit, to convert it to another technology, or to modify its performance. Modifying circuits for reuse is called recycling. An example of recycling is the widespread use of double-balanced mixers based on the Gilbert cell. Even though all such mixers use the same circuit topology, they can have widely varying performance depending on bias conditions, device sizes, resistor values etc. Also, they can be implemented in many different technologies.

D.9 Abstraction levels covered by design methods

Finally, different design methods address different abstraction levels of a design process. In the context of this thesis, we will concentrate on system, schematic, and layout levels. Design methods often only cover one or two of these abstraction levels, requiring additional design methods to cover the other levels, and yet another design method for using the combination of two or more design methods.

E

Limiting in zero-IF receivers

In a superhet receiver for frequency modulated systems (fig. 131, 133, 133), it is possible to realize the IF gain by limiters without significantly affecting the demodulation accuracy. This can easily be seen from a square-wave baseband signal "data(t)", representing a digital data sequence "0101010101" (Fig. 132).

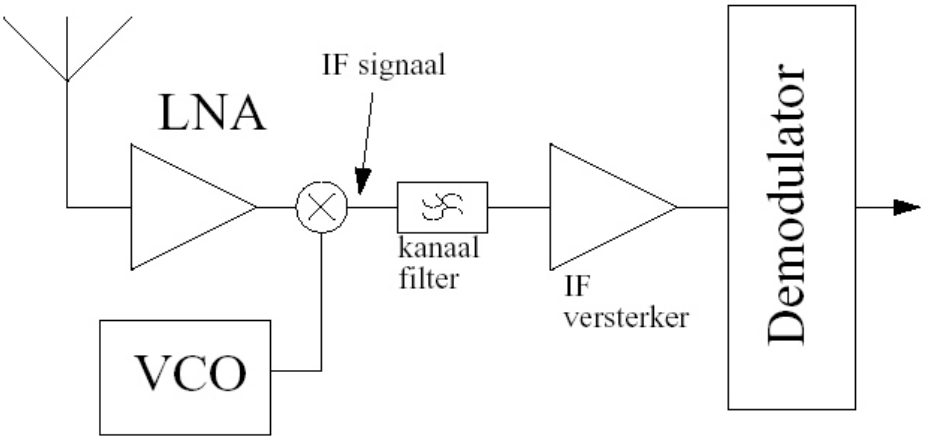


Figure 131 *Superhet architecture*

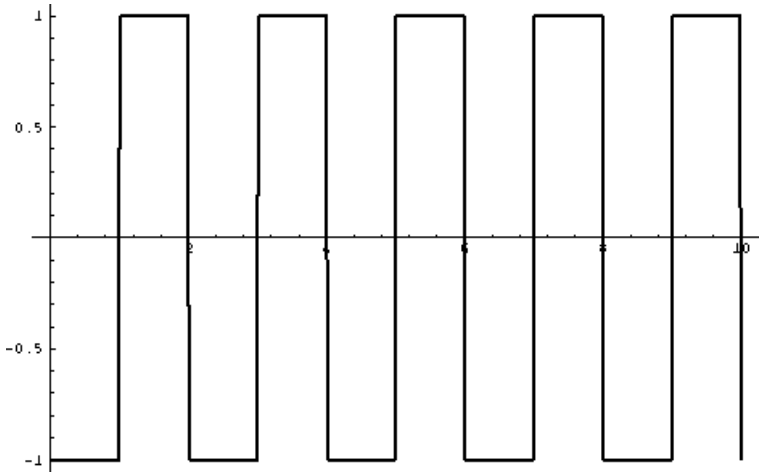


Figure 132 *Digital baseband signal*

If this baseband signal is applied to the input of a frequency shift keying (FSK) modulator, which has a center frequency f_c and a swing f_δ, this will result in an IF signal with the following phase $\phi(t)$ versus time (fig. 133).

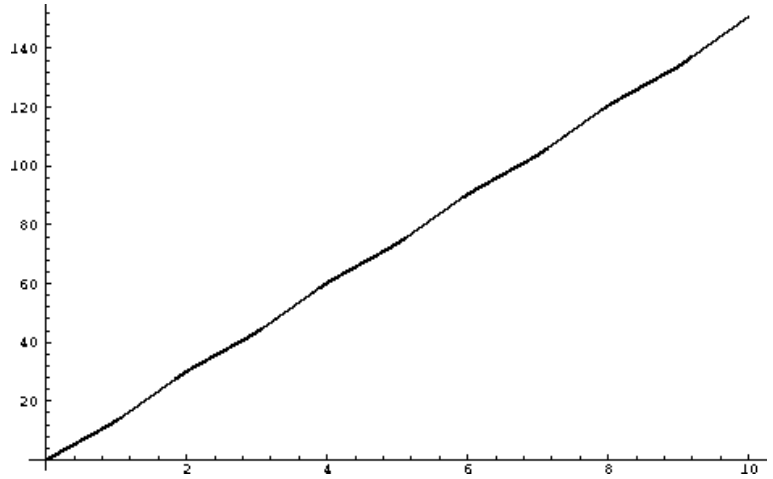


Figure 133 *Phase of the IF signal as a function of time*

With $\phi(t)$ equal to:

$$\Phi(t) = \int_0^t 2\pi f_c + 2\pi f_\delta data(\tau)\ d\tau \qquad \textbf{(110)}$$

A more conventional representation of the IF signal is obtained by plotting the instantaneous signal value *I(t)* rather than the phase value against time, as shown in fig. 134. The information in the two representations, however, is the same and the structures that will be discussed in this text are simpler to understand with reference to the phase curve rather than the instantaneous signal value curve of the IF signals.

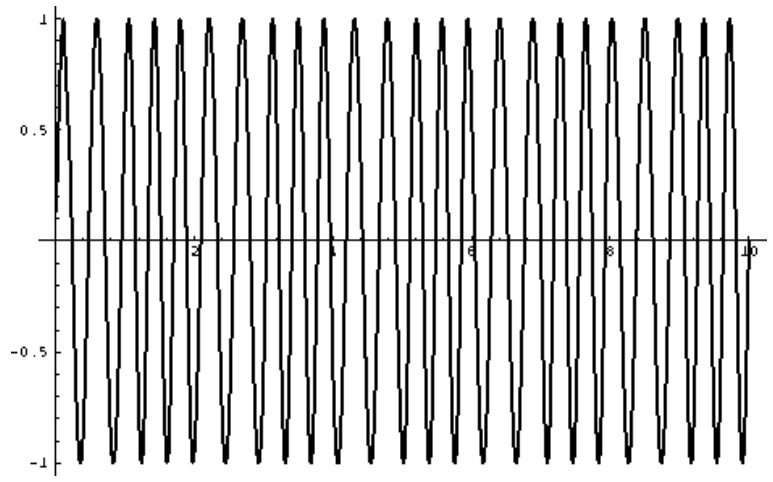


Figure 134 *Instantaneous signal value of the IF signal as a function of time*

With *I(t)* equal to:

$$I(t) = \sin\left(\int_0^t 2\pi f_c + 2\pi f_\delta \, data(\tau) \, d\tau\right) \qquad \textbf{(111)}$$

After limiting this IF signal, a signal *LI(t)* is obtained as shown in fig. 135.

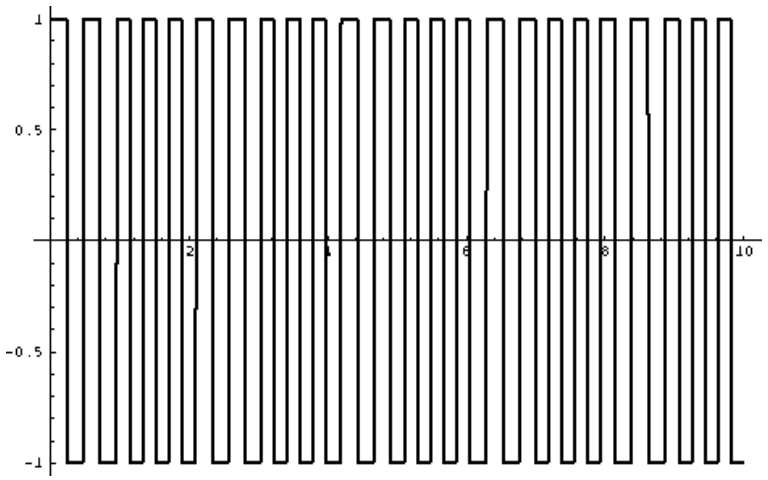


Figure 135 *Instantaneous value of the IF signal after limiting*

Due to the lack of amplitude information, the phase of this signal is often determined on the basis of the last zero crossing (as is the case, for example, in a count detector or pulse count demodulator). The reconstructed phase variation *ImpRec(t)* is shown in fig. 136, again plotted as a function of time.

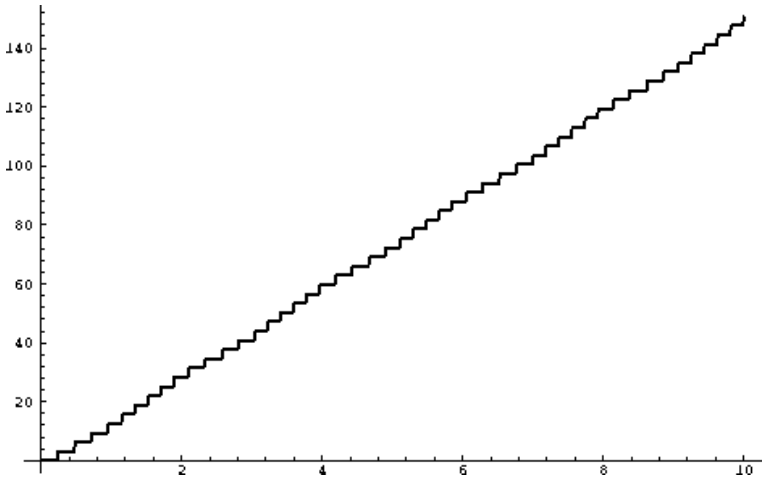


Figure 136 *Phase of the IF signal after limiting.*

The value of the function *ImpRec* equals π times the number of zero crossings of the

instantaneous IF signal after limiting, in the interval [0,t). Reconstruction by other methods, in which more information than only the last zero crossing is used (for example, PLL demodulators), will result in a different (and sometimes more accurate) phase

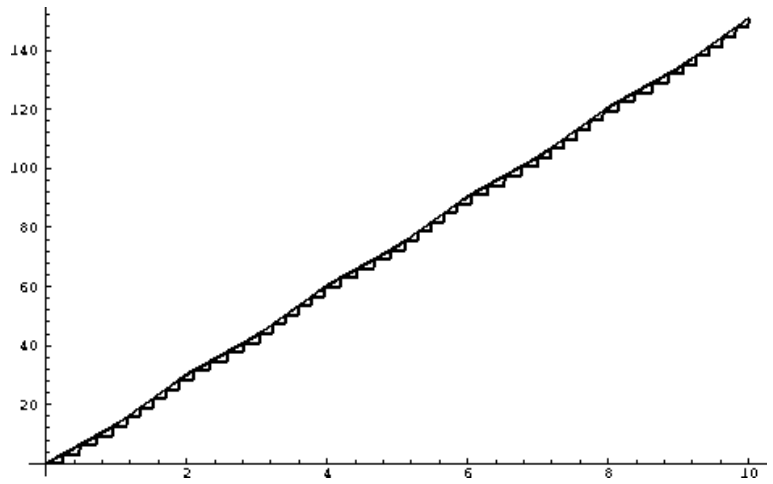


Figure 137 *Original and reconstructed phase of the IF signal*

variation. A pulse counter demodulator is relatively simple to implement, and the reconstructed phase rather accurately approaches the original phase, as long as the IF frequency and modulation index are both high. The output of this type of demodulator is shown in fig. 137:

The error introduced after demodulation can be modeled in the same way as quantization noise in an A/D converter, because the operation in the phase domain corresponds to quantizing the phase. A major difference is that the phase is a cyclic phenomenon, so that the "range" of the phase quantizer is not limited, unlike A/D converters for voltages or currents. This property makes it possible to enlarge the apparent range per period of the baseband signal by increasing the carrier frequency. This corresponds to adding a linearly increasing voltage to the input signal of an A/D converter, and is similar to the effect of dithering in such a converter: in either case the number of transitions is enhanced while the input signal remains the same, so that the input signal can be approximated more accurately.

The quantization effect increases when reducing the IF frequency. Therefore it also becomes more obvious in the phase domain (fig. 138).

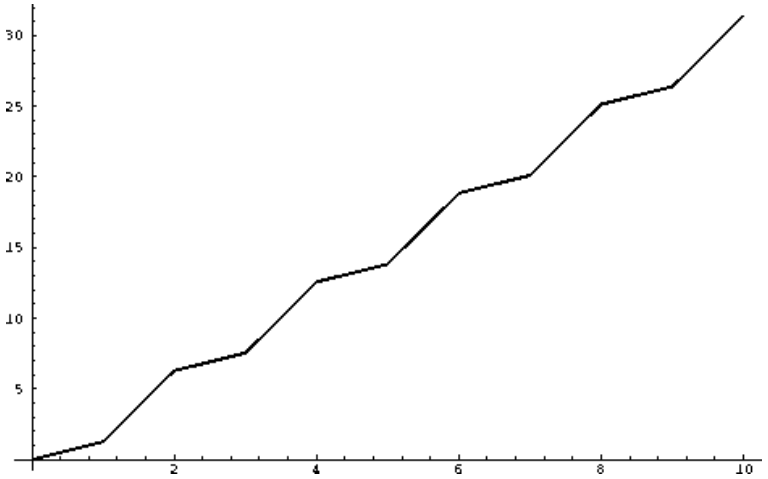


Figure 138 *Phase of an IF signal at a lower IF frequency*

The instantaneous value of the IF signal is shown in fig. 139:

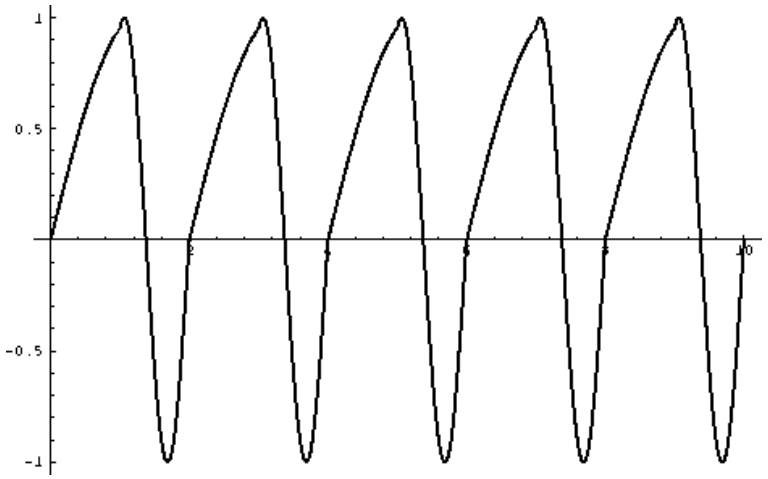


Figure 139 *Instantaneous value of the IF signal at a lower IF frequency*

After limiting, this signal is transformed into (fig. 140):

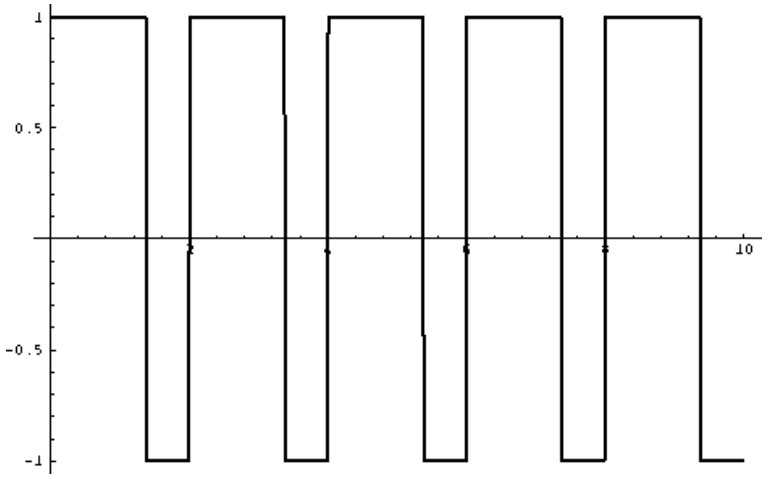


Figure 140 *Instantaneous value of the IF signal with a lower IF frequency after limiting*

The reconstructed phase of the limited IF signal now becomes (fig. 141):

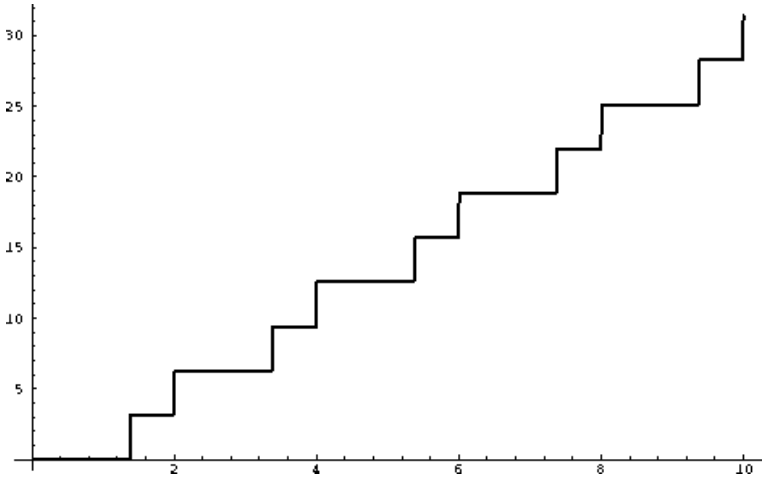


Figure 141 *Reconstructed phase of the limited IF signal at a lower IF frequency*

By comparing the reconstructed phase to the original phase, it is obvious that the differences have increased compared to the differences at a higher IF frequency (fig. 142):

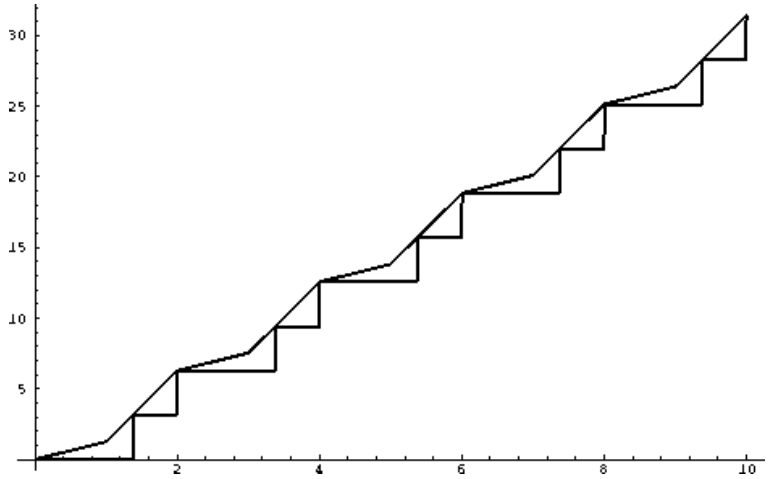


Figure 142 *Reconstructed and original phase of the IF signal at low-IF frequency*

This increased difference corresponds to the increased quantization noise. Since the phase change in one bit period has decreased at a lower IF frequency, the resolution of the quantizer, in terms of "least significant bit transitions", has decreased proportionally.

E.1 Problems with limiting in zero-IF receivers

In a zero-IF receiver, the quantization error becomes very big since the IF frequency is zero. The only zero crossings are caused by the frequency deviation. If the frequency deviation is small compared to the bit frequency, this can result in extreme quantization errors in the phase domain. Fig. 143 shows the phase variation of the IF signal in a zero-IF receiver, using the same data signal and frequency deviation as in the previous section:

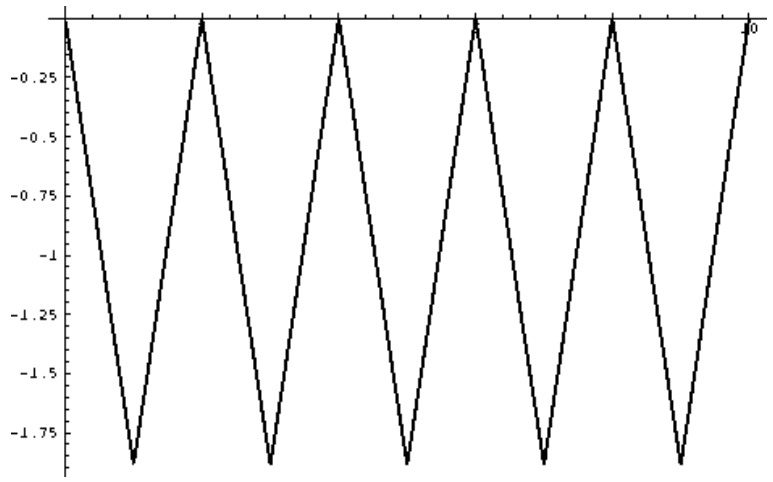


Figure 143 *Phase variation of the IF signal in a zero-IF receiver*

The instantaneous IF signal value for this situation is shown in fig. 144:

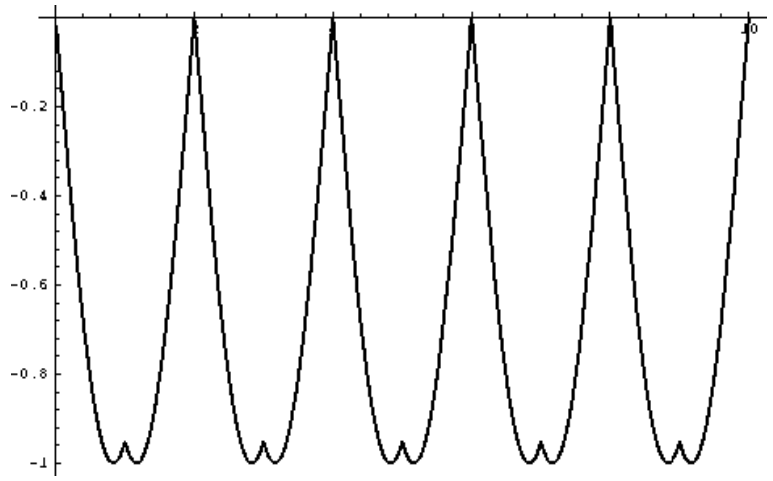


Figure 144 *Instantaneous signal value of the IF signal in a zero-IF receiver*

The result after limiting is shown in fig. 145:

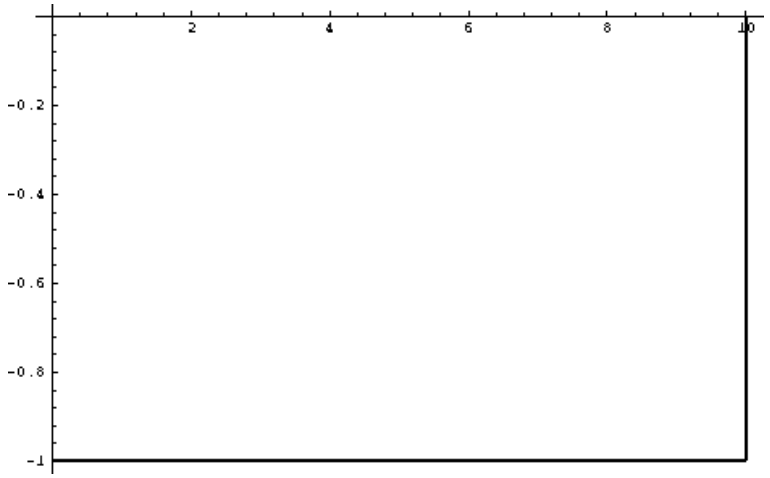


Figure 145 *Instantaneous value of the IF signal after limiting in a zero-IF receiver*

Since it is impossible to distinguish positive and negative frequencies on the basis of this single IF signal, phase reconstruction is very limited (fig. 146):

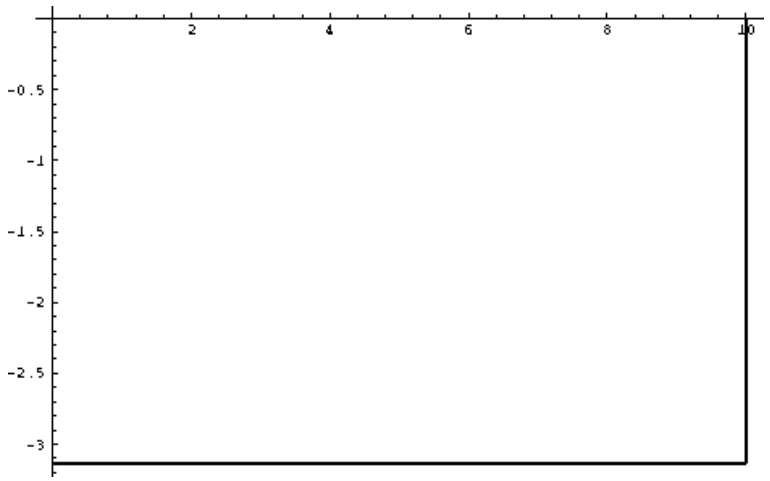


Figure 146 *Reconstructed phase of the limited IF signal in a zero-IF receiver*

When comparing this reconstructed phase with the original phase, the differences are very obvious (fig. 147):

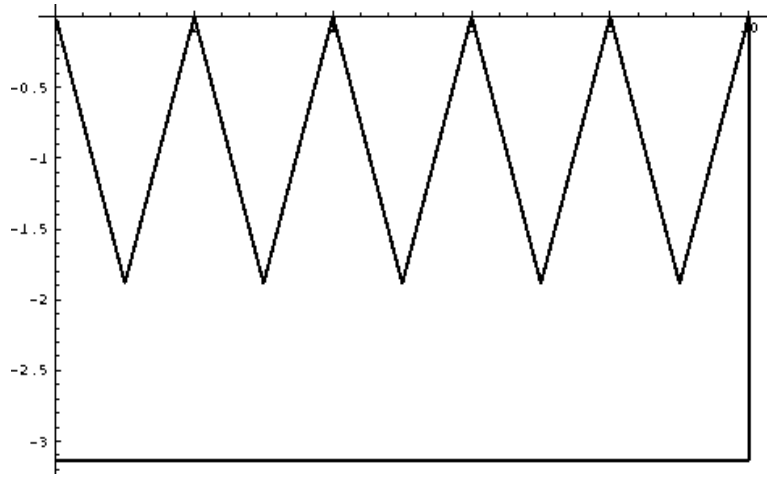


Figure 147 *Reconstructed and original phase of the limited IF signal in a zero-IF receiver*

In a zero-IF receiver, at least to IF signals are necessary to distinguish between positive and negative IF frequencies. These signals are usually referred to as I and Q. The phase variation of these signals is represented in Fig. 148:

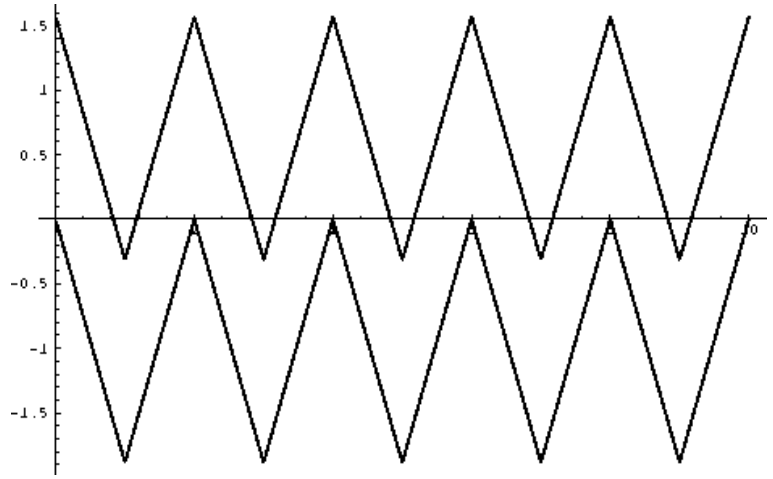


Figure 148 *Phase variation of the I and Q signals in a zero-IF receiver*

The associated instantaneous signal values are represented in fig. 149:

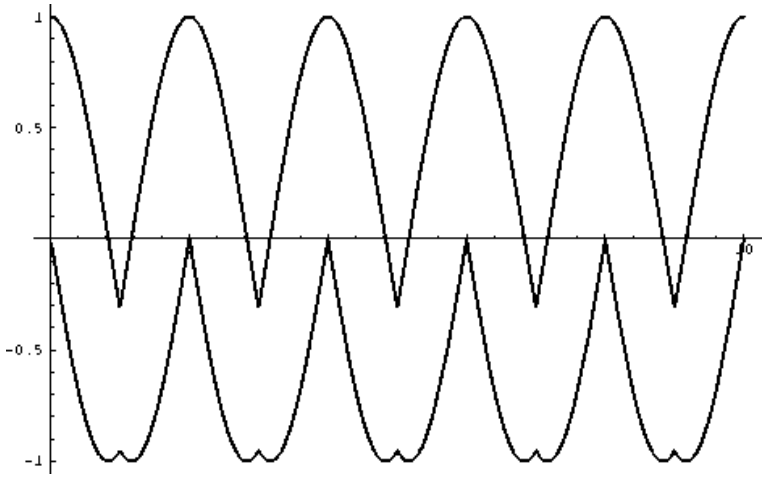


Figure 149 *Instantaneous values of the I and Q IF signals in a zero-IF receiver*

When sending these signals through a limiter, the following results will be obtained (fig. 150):

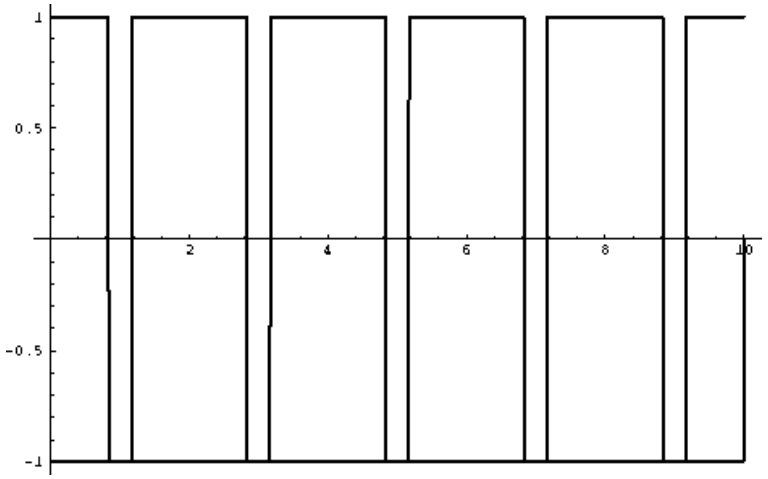


Figure 150 *I and Q signals after limiting*

The combination of the I and Q signs indicates in which quadrant the phase is at that moment. Since there is assumed that the phase variation is a continuous function of time, the phase variation may be reconstructed from the successive quadrants, except for one constant value (the initial phase). The frequency of the IF signal can now also be determined, both for positive and negative frequencies. The reconstruction of the phase variation established in this manner shows the following pattern (fig. 151):

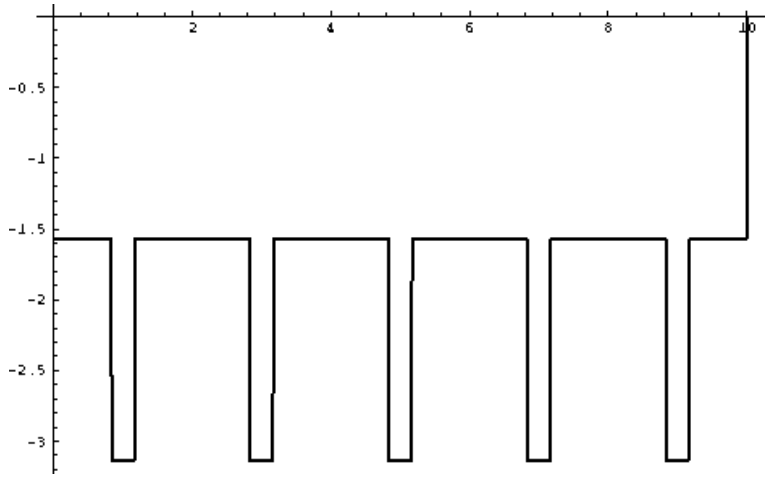


Figure 151 *Reconstructed phase of the I and Q signals*

This phase variation follows the IF phase more accurately than the reconstruction based upon a single IF signal (Fig. 22), but it is still very limited (Fig. 152):

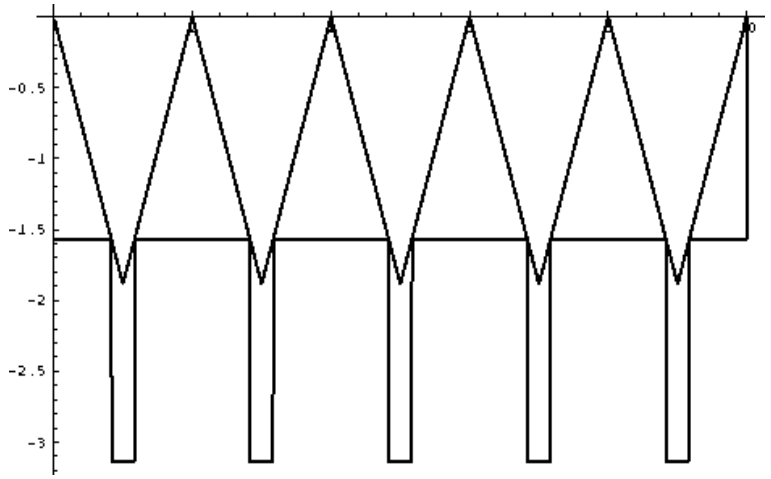


Figure 152 *Reconstructed and original phase of the I and Q signals*

The previous examples have been based on a modulation index of 0.3. When reducing the modulation index to less than 0.25, no reconstruction of the limited signal will be possible any more (fig. 153).

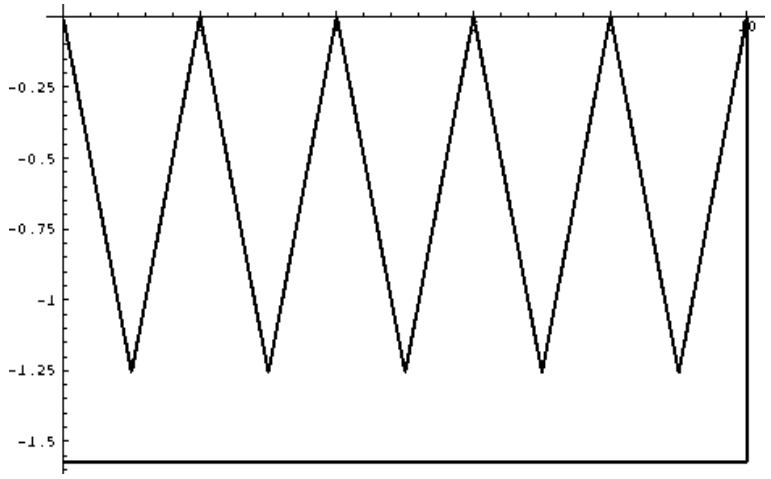


Figure 153 *Reconstructed and original phase of a limited IF signal with a modulation index of 0.2*

This demonstrates the problem of reconstructing the phase variation in FSK-modulated systems with a low modulation index, when using a zero IF architecture that has limiters in the IF signal path.

E.2 Existing solutions

There are at least two different approaches to solving this problem:

- the quantization of the phase can be carried out using a smaller phase difference between quantization levels;
- the reconstruction of the signal in between two quantization levels can be carried out more accurately.

Reducing the step size can be achieved by using additional IF signals [94] (fig. 154).

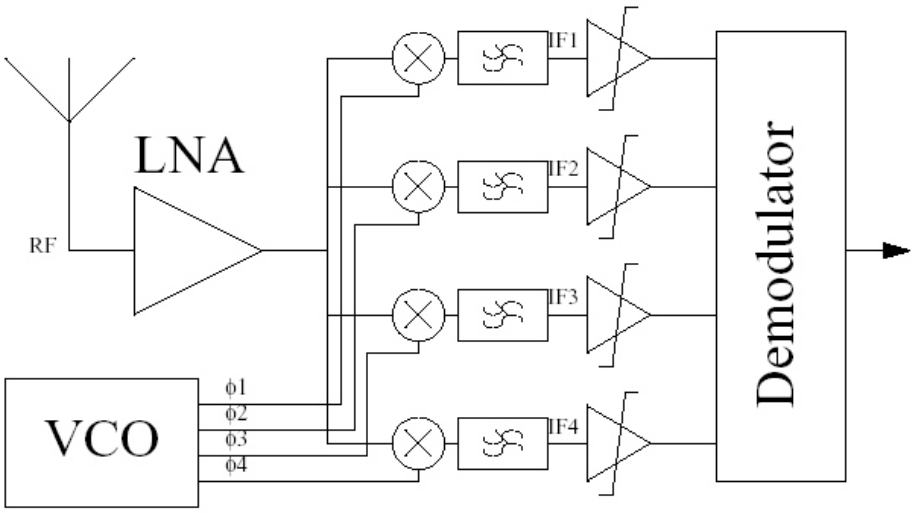


Figure 154 *Multi-branch receiver*

In a multi-branch receiver, the VCO needs to generate multiple (in this example 4) local oscillator signals (in this case with phases ϕ1, ϕ2, ϕ3 and ϕ4). Each LO signal is connected to a different mixer, and the output of these mixers provide the multiple IF signals (in this case "IF1" through "IF4"). This architecture is rarely used, however, since the complexity of the receiver increases significantly: each additional IF signal calls for a new mixer, a new filter, and an additional VCO output.

Accurate reconstruction in between quantization levels is achieved, for instance, by PLL demodulators. The use of such a demodulator in a zero-IF receiver, however, requires an oscillator that can generate both positive and negative frequencies. This implies that such an oscillator needs to cover an infinite number of octaves of tuning range. This is difficult to realize with practical boundary conditions.

E.3 New solutions

Both approaches to solve the limited zero-IF receiver problem have drawbacks: high circuit complexity for the multi-branch receiver, and a difficult VCO for the PLL demodulator. In this section, new solutions that avoid these drawbacks will be introduced. Finally, an implementation that combines both new solutions will be presented.

E.3.1 Low-complexity generation of multiple IF signals

The multiple IF signals in a multi-branch receiver are not independent. Two IF signals with different phase are can be converted into an orthogonal pair by linear transformation. In fact, most two-branch receivers already use such an orthogonal pair. Other IF signals can then be constructed from these IF signals by forming linear, weighted, combinations of the orthogonal pair.

Let $IF(\alpha)$ represent the IF signal with phase α. The orthogonal pair of IF signals I and Q now correspond to:

$$I = IF(0)$$
$$Q = IF(\frac{\pi}{2}) \qquad \textbf{(112)}$$

Any new IF signal with phase β can now be generated by a linear combination of I and Q:

$$IF(\beta) = I\cos(\beta) + Q\sin(\beta) \qquad \textbf{(113)}$$

Therefore the creation of additional IF signals can be postponed to a later stage in the signal processing chain. This reduces the increase in circuit complexity, since additional signal processing stages only need to be implemented from the point where the additional IF signals are created. Postponing this to the last possible stage result in the implementation with the lowest complexity.

The last possible stage at which the linear combination of I and Q, as described in (113), can be carried out, is the stage before the first non-linear processing of the IF signal, i.e., before the limiters. The linear combination from (113) can be accomplished through a simple resistive interpolation network, similar to the interpolation circuits used in an folding & interpolation ADC [95]. A block diagram of a receiver using this interpolation approach is shown in fig. 155.

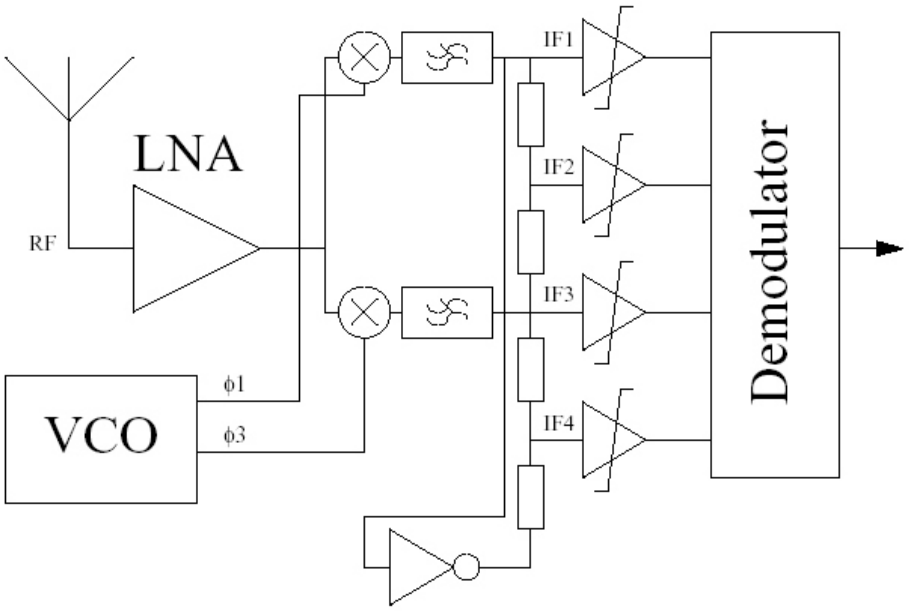


Figure 155 *Block diagram of a zero-IF interpolation receiver*

In a zero-IF interpolation receiver, the additional VCO phase generators, mixers and filters can be replaced by a small number of resistors and an inverter. The inverter in a balanced system can be replaced by a crossing of the signal wires. Further resistors can be added between the resistor network and the inputs of the limiters in order to maintain the same signal amplitude and source impedance, and therefore the same phase shift at high frequencies. This is comparable to the delay compensation in interpolation networks of folding A/D converters [95].

An interpolation zero-IF receiver can achieve a arbitrarily small error when demodulating frequency modulated signals by increasing the number of IF signals. Since each halving of the quantization step calls for a doubling of the number of limiters and resistors in the interpolation network, this approach is limited both because of the exponentially increasing cost and because of the exponentially increasing accuracy requirements of these elements. Therefore, a reconstruction method to complement the interpolation approach would be very useful.

E.3.2 Accurate reconstruction in superhet receivers

The reconstruction method used in the previous examples is a hold function of the last known phase value. Although this is an obvious and easy to implement method, it is known from sampling theory that better reconstruction possibilities exist for equidistantly sampled signals.

The continuous-time phase quantization process may also be considered a sampling process. In a quantization process a signal is described as a function of time while only the interval in which the signal occurs is known at any instant. This usually results in a statistical description of the signal and its quantization error. In fact, the quantization error is often modeled as a noise source in quantization process models.

A continuous-time quantization process may also be considered as a sampling process, because there are instants in time at which the value of the signal is exactly known, i.e. at the moment that a transition between two quantization levels occurs (fig. 156).

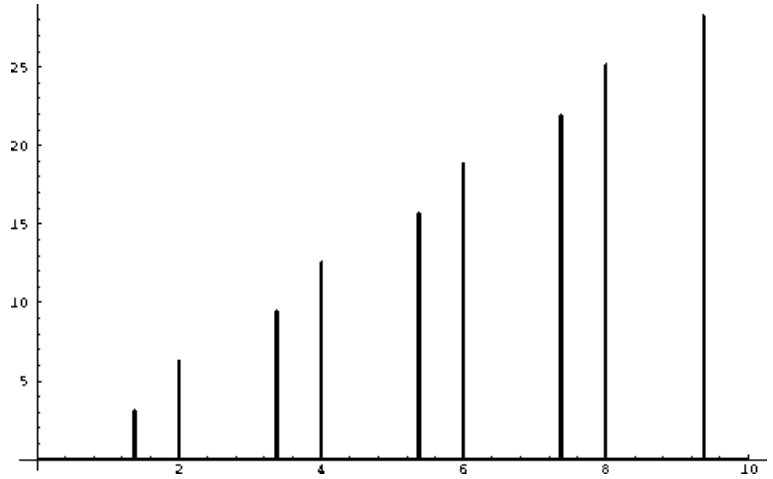


Figure 156 *Sampled value of an IF signal with a low frequency*

When comparing this signal with the original signal (fig. 157), it is obvious that the samples represent the exact value, without any quantization error.

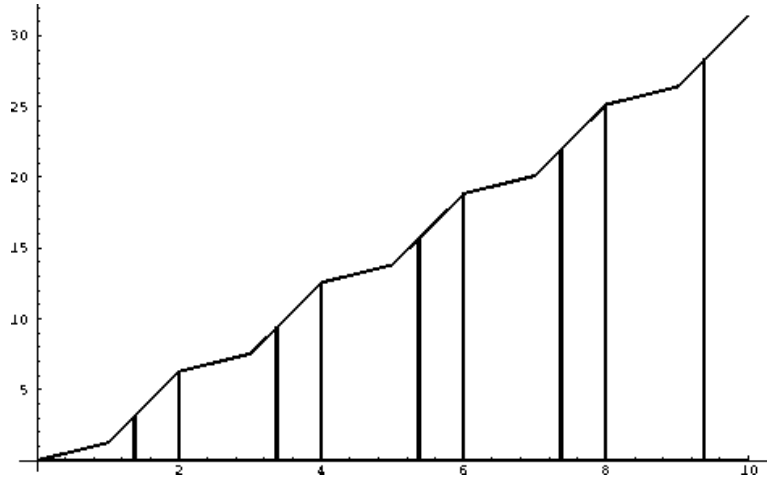


Figure 157 *Sampled and original values of the low frequency IF signal*

Since traditional sampling theory is based on equidistant samples, and requires Nyquist bandwidth limits for reconstruction, it does not apply directly to the sampling as shown in fig. 157. The sampling is obviously non-equidistant, and the bandwidth of the phase trajectory is seldom, if ever, specified or controlled in telecommunication transceiver systems.

In a superhet architecture, this can be solved by considering the inverse function of the IF signal, namely, the time as function of the phase of the signal. This is possible because the IF frequency is significantly higher than the frequency deviation, making the phase trajectory as a function of time strictly monotonic. Hence, an inverse function is guaranteed to exist. Fig. 158 shows a curve of time plotted against phase for the same signal:

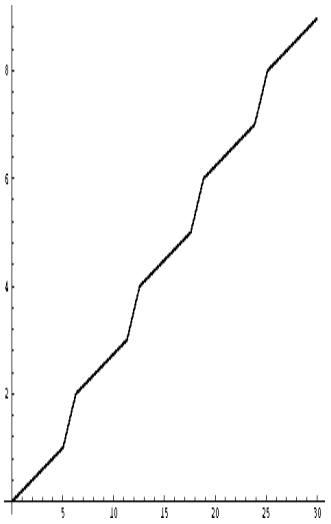


Figure 158 *Time as a function of IF signal phase*

Fig. 159 shows that the quantized version of this IF signal can be interpreted as an equidistantly sampled signal:

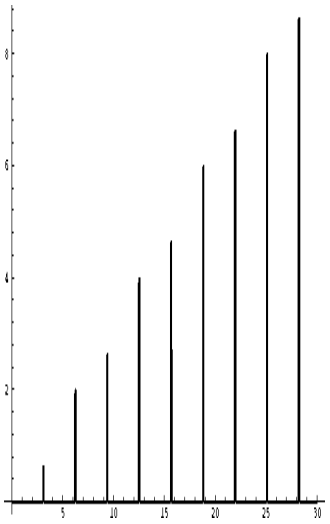


Figure 159 *Quantized IF signal plotted as time samples versus phase*

If the time as a function of phase is band-limited, a simple and perfect reconstruction based upon the standard sampling theory is possible. This does call for a sufficiently high sampling rate with respect to the bandwidth of the time function, but this requirement can always be satisfied by making the sampling period (thus the distance in the phase domain between two quantization steps) sufficiently small. Even if, as will be often the case, a system does not guarantee that the time as function of phase is bandlimited, the errors introduced by aliasing might still be smaller than those introduced by a simple hold function.

This method cannot be applied to receivers in which the phase variation as a function of time can not be inverted, i.e. if this function is not strictly monotonic. Therefore, low-IF receivers and superhet receivers can use this method, but zero-IF receivers cannot.

E.3.3 Accurate reconstruction in zero-IF receivers

In the special case of binary FSK modulation, a solution exists in which the phase function need not be inverted. This solution uses the symmetry of the phase function around a transition of the baseband signal. When the phase quantization levels are close enough, the transition of the baseband signal is halfway between the nearest transitions of the phase signal through the phase quantization levels. This is shown in fig. 160.

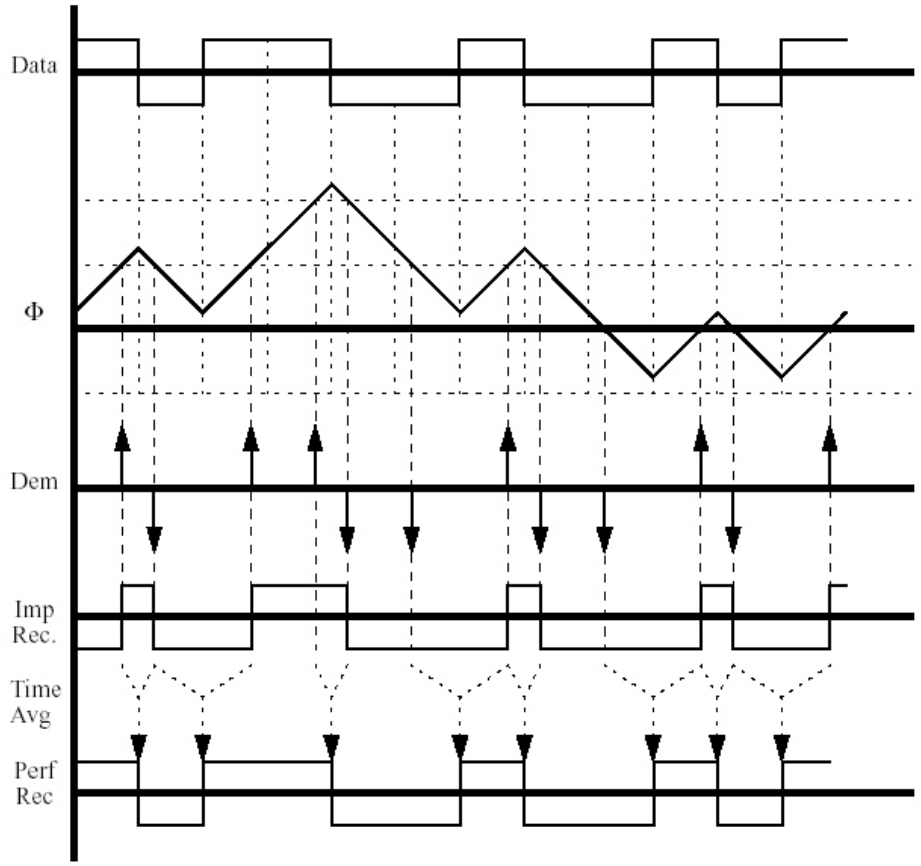


Figure 160 *Accurate reconstruction of the baseband signal*

The top curve of fig. 160 shows the baseband signal, i.e. the data to be transmitted. In this example, a random data sequence "101100100101" is used. The associated phase variation of the IF signal in a zero-IF receiver is shown in the second curve ("ϕ"). The horizontal dashed lines represent the quantization levels, with a quantization step of π/2. This IF signal is then converted into time samples as a function of phase using a differentiate-and-cross-multiply (*DACM*) demodulator, as shown in fig. 161.

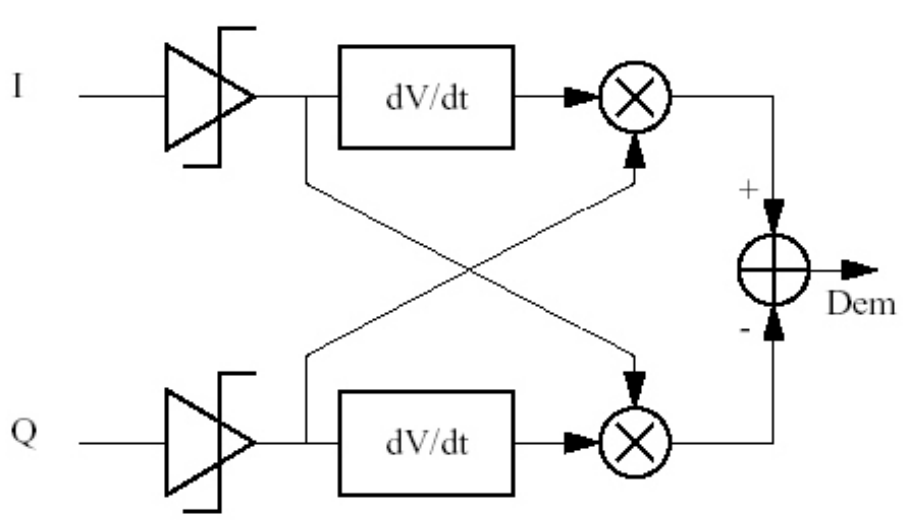


Figure 161 *Diagram of a DACM demodulator for zero-IF receivers*

This DACM demodulator produces positive pulses at a positive transition of the phase of the IF signal through a phase quantization level, and a negative pulse at a negative transition. These pulses are shown in the third curve ("Dem") of fig. 160. A simple, but inaccurate reconstruction can be obtained by integrating the pulses, or by holding the sign of the last pulse, as is shown in the fourth graph ("Imp.Rec") of fig. 160.

A perfectly accurate reconstruction ("Perf. Rec." in fig. 160) can be obtained by delaying the transitions of the demodulator output to the moment halfway between two successive pulses with opposite signs ("Time Avg" in fig. 160). This method requires that sufficient pulses per data symbol are obtained, so that any possible bit sequence can be decoded. This limits the quantization step size. In addition, a fixed time delay needs to be incorporated in the reconstruction circuit, to guarantee causality. This delay is at least half the maximum time between two opposite pulses.

The following variables will be used for a more precise formulation of these conditions:

T: symbol period
T_d: fixed delay in the reconstruction algorithm
R: symbol rate
X: maximum phase excursion in time T
$\Phi(t)$: phase at instant t
f_{IF}: frequency of the IF signal
n: number of phase quantization levels as a result of which the quantization distance is π/n.

The following relations exist between these variables:

$$T = \frac{1}{R} \qquad \textbf{(114)}$$

and:

$$X = 2\pi f_{IF} T \qquad \textbf{(115)}$$

A perfect reconstruction can be achieved the following condition is satisfied:

$$X < \frac{\pi}{n} < 2X \qquad \textbf{(116)}$$

using the following algorithm:

1. Look for the first pulse in the "Dem" signal;
2. Set the reconstructed data to a value that corresponds to the sign of the pulse;
3. Repeat the next 3 steps until the end of the signal;
4. If a positive pulse is followed by a negative pulse, set the reconstructed data to "0" at an instant that corresponds to the fixed delay T_d plus the average of the instants of the positive pulse and the negative pulse;
5. If a negative pulse is followed by a positive pulse, set the reconstructed data to "1" at an instant that corresponds to the fixed delay T_d plus the average of the instants of the positive pulse and the negative pulse;
6. If the period of time between two pulses exceeds 2T, change the value of the signal every period of time T after the symbol that follows the symbol in which the last pulse occurred, and before there is another period of time T in which a pulse occurs.

The major disadvantage of this algorithm and the corresponding boundary conditions is that the period of time between two successive opposite pulses may be arbitrarily long. As a consequence, T_d needs to be infinitely large as well, so that the memory capacity for the implementation will also have to be infinitely large.

This may be avoided if more information about the signal is known, for example, that it will be partitioned in packets of predefined maximum length, and/or that the symbol sequence "10101010101010 ..." has a predefined maximum length. There are modulation systems in which this can be guaranteed, for example, (d,k,N)-constrained sequences ($d \geq 1$) for super density CD, and the PHY and MAC layers for the DECT protocol.

If no further information about the signal is known, a practical implementation of this algorithm can still be obtained by using a stronger boundary condition for the quantization level:

$$\frac{X}{2} < \frac{\pi}{n} < X \qquad \textbf{(117)}$$

With this stricter boundary condition, the maximum period of time between two pulses is 2T, so that the minimum delay in the reconstruction becomes T. This means that a memory with a capacity of only one symbol is sufficient, and that step 6 from the algorithm may be omitted.

The smaller quantization step can be obtained by using a multi-branch receiver. This does require an extension of the DACM demodulator, as shown in fig. 162:

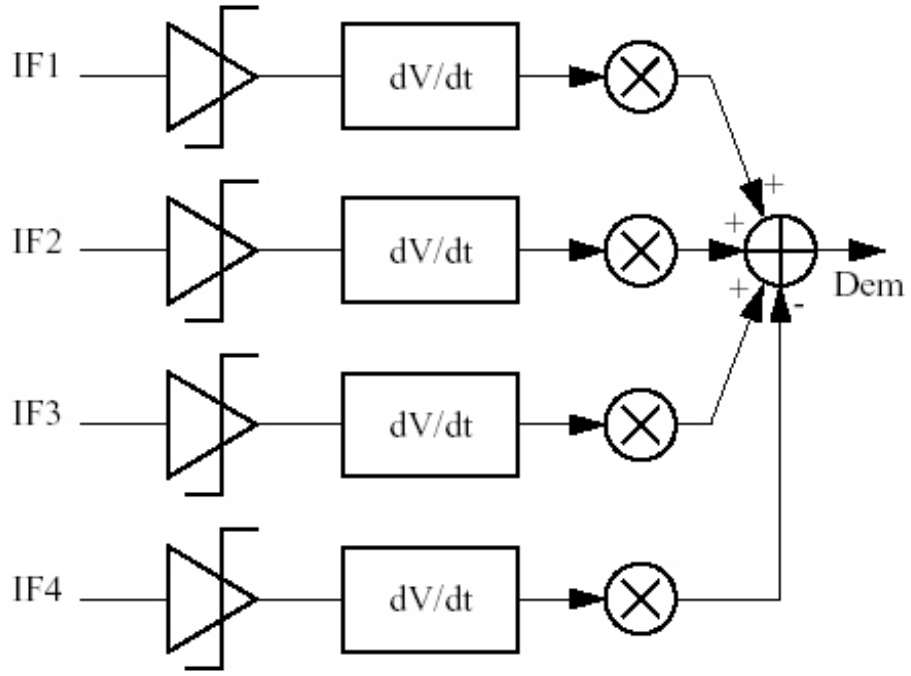


Figure 162 *Extended DACM demodulator for interpolation receiver*

F

Solution of the general problem

The mathematical formulation of the general problem of the optimum distribution of gain, linearity and noise distribution for an arbitrary number of cascaded stages, as formulated in this thesis, was solved by dr. ir. A.J.E.M. (Guido) Janssen at the Philips Research Laboratories, Eindhoven. He reformulated the problem in a different way using the following definitions:

$$\begin{aligned} x_i &= G_i\, IP3_i && i = 1,\ldots,n \\ y_i &= \prod_{j=1}^{i} G_j && i = 0,\ldots,n \end{aligned} \qquad \textbf{(118)}$$

and, by convention:

$$y_0 = \prod_{j=1}^{0} G_j = 1 \qquad \textbf{(119)}$$

Note that x_i can be interpreted as the output *IP3* of block i, whereas y_i can be interpreted as the partial gain of blocks 1 through i.

Also, let

$$\begin{aligned} a &= F_{tot} - 1 \\ a_i &= F_{i+1} - 1 && i = 0,\ldots,n-1 \\ b &= \frac{1}{IP3_{tot}} \end{aligned} \qquad \textbf{(120)}$$

Note that a can be interpreted as the part of the noise factor that exceeds the noise factor of an ideal circuit block (F=1).

Then the x_i, y_i are positive but otherwise unrestricted by the definitions in equation (118). Also:

$$\frac{\prod_{j=1}^{i-1} G_j}{IP3_i} = \frac{\prod_{j=1}^{i} G_j}{G_i IP3_i} = \frac{y_i}{x_i} \qquad i = 1,\ldots,n \qquad \textbf{(121)}$$

The problem now takes the following form:

Given

$$\begin{aligned} \kappa_i &> 0 \qquad i=1,\ldots,n \\ a_i &> 0 \qquad i=0,\ldots,n-1 \\ a &> 0 \\ b &> 0 \end{aligned} \tag{122}$$

find the minimum of:

$$P_{tot} = \sum_{i=1}^{n} \kappa_i x_i \tag{123}$$

under the conditions that:

$$\sum_{i=0}^{n-1} \frac{a_i}{y_i} = a, \qquad y_n = G_{tot},\ y_0 = 1 \tag{124}$$

$$\sum_{i=1}^{n} \frac{y_i}{x_i} = b \tag{125}$$

Condition (124) is the equivalent of Friis formula, whereas condition (125) is the equivalent of the $IP3_{tot}$ spec of a front end. Please note that at this point, y_i is a function of x_i, b, a_i and a.

This problem was solved as follows:

Take $y_0, y_1, \ldots, y_{n-1}, y_n > 0$ such that the constraints (124) and (125) are satisfied. Then determine the minimum of P in (123) as a function of $x_1, \ldots, x_n$ under the constraint (125). This yields a minimum value $P(\underline{y} = (y_0, y_1, \ldots, y_{n-1}, y_n))$. Now determine $\underline{y} = (y_0, y_1, \ldots, y_{n-1}, y_n)$ such that the constraints (124) are satisfied while $P(\underline{y})$ is minimal.

Thus there are two steps:

Step 1:
Given $y_0, y_1, \ldots, y_{n-1}, y_n > 0$ such that (125) holds, minimize (i.e. find x_i):

$$P_{tot} = \sum_{i=1}^{n} \kappa_i x_i \quad \text{under the condition that} \quad \sum_{i=1}^{n} \frac{y_i}{x_i} = b \tag{126}$$

which yields a minimum value $P(\underline{y})$ and $x_i(\underline{y})$.

Step 2:
Minimize $P(\underline{y})$ under the condition that (124) holds.

Details of step 1:
Fix $y_0, y_1, \ldots, y_{n-1}, y_n > 0$
Based on Lagrange's multiplier rule [40], there is a point where the extrema of P occur under the constraint that there is a multiplier λ such that:

$$\nabla_{\underline{x}} P_{tot} = \lambda \, \nabla_{\underline{x}} \left(\sum_{i=1}^{n} \frac{y_i}{x_i} - b \right) \tag{127}$$

The extrema represent the minimum power for $OIP3_i$ while meeting the IP3 specs. This gives:

$$\kappa_i = -\lambda \frac{y_i}{x_i^2} \qquad i = 1,\ldots,n \tag{128}$$

The constraint in (126) needs to be satisfied, hence λ needs to be determined such that:

$$\sum_{i=1}^{n} \frac{y_i}{x_i} = b \qquad x_i^2 = -\lambda \frac{y_i}{\kappa_i} \tag{129}$$

The first part of this equation ensures that the IP3 spec is met, while the second part represents the solution of x_i for minimum power.

Evidently $-\lambda > 0$. This yields

$$b = \sqrt{-\frac{1}{\lambda}} \sum_{i=1}^{n} \left(y_i \sqrt{\frac{\kappa_i}{y_i}} \right) = \sqrt{-\frac{1}{\lambda}} \sum_{i=1}^{n} \sqrt{\kappa_i y_i} \tag{130}$$

This can be used to express λ as a function of y_i, resulting in a new function for $\underline{x}$:

$\underline{x} = \mathrm{f}(y_i, k_i, b)$.

At the extremum point this evaluates to:

$$\begin{aligned} P(\underline{y}) &= \sum_{i=1}^{n} \kappa_i x_i \\ &= \sum_{i=1}^{n} \kappa_i \sqrt{-\lambda \frac{y_i}{\kappa_i}} \\ &= \sqrt{-\lambda} \sum_{i=1}^{n} \sqrt{\kappa_i y_i} \\ &= \frac{1}{b} \left(\sum_{i=1}^{n} \sqrt{\kappa_i y_i} \right)^2 \end{aligned} \qquad \textbf{(131)}$$

In this expression, the value for $\sqrt{-\lambda}$ that follows from (130) was substituted.

Details of step 2:
It is also valid to minimize:

$$\sum_{i=1}^{n} \sqrt{\kappa_i} z_i \qquad z_i = \sqrt{y_i} \qquad i = 1..n \qquad \textbf{(133)}$$

where the $z_i > 0$ are constrained by the equivalent of Friis formula, see (124):

$$\sum_{i=0}^{n-1} \frac{a_i}{z_i^2} = a \qquad z_n = \sqrt{G_{tot}},\ z_0 = 1 \qquad \textbf{(134)}$$

The constraints $z_n = \sqrt{G_{tot}},\ z_0 = 1$ are dealt with as follows:

$$\begin{aligned} \sum_{i=1}^{n} \sqrt{\kappa_i} z_i &= \sqrt{\kappa_n G_{tot}} + Q(z) \\ Q(z) &= \sum_{i=1}^{n-1} \sqrt{\kappa_i} z_i \end{aligned} \qquad \textbf{(136)}$$

which takes out z_n, and

$$\sum_{i=1}^{n-1} \frac{a_i}{z_i^2} = c$$
$$c = a - a_0 \qquad \textbf{(137)}$$

which takes out z_0.

Thus Q(z) in (136) must be minimized under the constraint (137). By again applying Lagrange's multiplier rule, there has to be a multiplier μ at the minimum point such that:

$$\nabla_{\mathbf{z}} Q = \mu \nabla_{\mathbf{z}} \left(\sum_{i=1}^{n-1} \frac{a_i}{z_i^2} - c \right) \qquad \textbf{(138)}$$

This gives:

$$\sqrt{\kappa_i} = \frac{-2\mu a_i}{z_i^3} \qquad i=1,...,n-1 \qquad \textbf{(139)}$$

The constraint (137) needs to be satisfied, hence μ should be determined such that:

$$\sum_{i=1}^{n-1} \frac{a_i}{z_i^2} = c \qquad z_i^3 = \frac{-2\mu a_i}{\sqrt{\kappa_i}} \qquad \textbf{(140)}$$

This gives:

$$\sqrt[3]{-2\mu} = \sqrt{\frac{1}{c} \sum_{i=1}^{n-1} \sqrt[3]{\kappa_i a_i}} \qquad \textbf{(141)}$$

Then, at the minimum point, Q has the value:

$$
\begin{aligned}
Q(z) &= \sum_{i=1}^{n-1} \sqrt{\kappa_i} z_i \\
&= \sum_{i=1}^{n-1} \sqrt{\kappa_i} \sqrt[3]{\frac{-2\mu a_i}{\sqrt{\kappa_i}}} \\
&= \sqrt[3]{-2\mu} \sum_{i=1}^{n-1} \sqrt[3]{\kappa_i a_i} \\
&= \frac{1}{\sqrt{c}} \left(\sum_{i=1}^{n-1} \sqrt[3]{\kappa_i a_i} \right)^{\frac{3}{2}}
\end{aligned}
\qquad \textbf{(142)}
$$

Therefore, the minimum value of P equals, see (131):

$$
P_{\min} = \frac{1}{b} \left(\sqrt{\kappa_n G_{tot}} + \frac{1}{\sqrt{c}} \left(\sum_{i=1}^{n-1} \sqrt[3]{\kappa_i a_i} \right)^{\frac{3}{2}} \right)^2 \qquad \textbf{(143)}
$$

and is reached for, see (133):

$$
\begin{aligned}
& y_i = z_i^2 = \left(\frac{a_i}{\sqrt{\kappa_i}} \right)^{\frac{2}{3}} (-2\mu)^{\frac{2}{3}} \qquad (-2\mu)^{\frac{2}{3}} = \frac{1}{c} \sum_{i=1}^{n-1} \sqrt[3]{\kappa_i a_i} \qquad i = 1,\ldots,n-1 \\
& y_n = G_{tot}
\end{aligned}
\qquad \textbf{(144)}
$$

and, see (129):

$$
x_i = \sqrt{-\lambda} \sqrt{\frac{y_i}{\kappa_i}} \qquad i = 1,\ldots,n-1 \qquad \textbf{(145)}
$$

where, see (130):

$$
\sqrt{-\lambda} = \frac{1}{b} \sum_{i=1}^{n} \sqrt{\kappa_i y_i} \qquad \textbf{(146)}
$$

Thus, using (141), x_i simplifies to:

$$x_i = \frac{1}{b}\sqrt{\frac{y_i}{\kappa_i}}\sum_{j=1}^{n}\sqrt{\kappa_j y_j} \qquad \textbf{(147)}$$

G

FAT transforms

Many of the specification parameters of a transceiver correspond one-to-one to the system specification. Table 14 shows a list of such parameters.

Field	*Description*	*Unit*	*Fmt*
RxMinFreq	Minimum receive frequency	MHz	N
RxMaxFreq	Maximum receive frequency	MHz	N
TxMinFreq	Minimum transmit frequency	MHz	N
TxMaxFreq	Maximum transmit frequency	MHz	N
RXChannelPitch	Grid of the RX channel raster	kHz	N
RXChannelBW	3dB bandwidth of one RX radio channel	kHz	N
TXChannelPitch	Grid of the TX channel raster	kHz	N
TXChannelBW	3dB bandwidth of one TX radio channel	kHz	N
BasebandBW	3dB bandwidth of one baseband channel	kHz	N
NrOfChannels	Number of receive/transmit channel pairs		I
RXFreqError	Maximum error in RX RF frequency	kHz	N
RXModulation	Description of RX modulation type		A60
RXDeviation	Frequency deviation of RX FM type signals	kHz	N
TXFreqError	Maximum error in TX RF frequency	kHz	N
TXModulation	Description of TX modulation type		A60
TXDeviation	Frequency deviation of TX FM type signals	kHz	N
TXMinSignal	Minimum transmit power level	W	N
TXMaxSignal	Maximum transmit power level	W	N
TxAvgSignal	Average transmit power level	W	N
TXStepSignal	Transmit power level step size	dB	N

Field	*Description*	*Unit*	*Fmt*
RXMinRssi	Minimum RSSI indication required	dBm	N
RXMaxRssi	Maximum RSSI indication required	dBm	N
RxStepRssi	Maximum step size of RSSI	dB	N
DataRate	Raw data rate across the radio channel	kbps	N
TimeSlot	Duration of a single timeslot	ms	N

Transceiver specification parameters that are identical to system spec parameters

Other receiver specifications are derived from the system specification through formulas. In these formulas, the field names of the database table are used as variables to indicate the appropriate values from the database tables. The table from which these database fields are taken is indicated as a subscript. The following parameters are derived (table 14).

Field	*Description*	*Unit*	*Fmt*
RxNF	Receiver Noise Figure	dB	N
Cochannel	Cochannel suppression	dB	N
Rx1dBCompr	Receiver 1dB compression point	dBm	N
RxAgcRange	Receiver optimum AGC range	dB	N
RxDutyCycleAct	Rx Percentage "on" time in active mode	%	N
TxDutyCycleAct	Tx Percentage "on" time in active mode	%	N

Transceiver parameters derived from system parameters

Also, the following parameters are derived from detail tables of the System.db table, and translated to detail tables of the Trx.db table, to show offset-frequency dependence (table 14).

Parameter	*Description*	*Detail Table*
Selectivity	Selectivity	TrxSel.db
IP3	Receiver IP3 spec	TrxIP3.db
RxVcoNoise	Receiver VCO phase noise	RxVCO.db

Transceiver parameters derived from other system tables

The noise figure is derived from the thermal noise floor in the channel bandwidth, the minimum signal power and the minimum required signal-to-noise ratio. It is assumed that the noise figure consumes the margin between minimum signal level, thermal noise in the channel bandwidth, and minimum signal to noise ratio of the system:

$$RxNF_{Trx.db} = (RxMinSignal_{System.db} - 30) - RxMinSNR_{System.db} - 10\log(kT\ 1000\ RxChannelBW_{System.db}) \quad \textbf{(148)}$$

Please note that the minimum channel is in dBm and needs to be adjusted to dBW since the calculation of the noise power results in dBW as well (this explains the number 30). Also, the channel bandwidth is in kHz and needs to be adjusted to the noise density of

kT in 1Hz (this explains the factor 1000). This noise figure calculation gives the highest noise figure for which the system specifications can still be met. It assumes that all other parts of the receiver, and especially the demodulator, are ideal.

The allowed co-channel interference level can be found from the single-tone interferer at 0Hz offset in the SysI1.db table:

$$Cochannel_{Trx.db} = SignalPower_{SysI1.db} - InterfererPower_{SysI1.db}$$
$$\text{all values at: } offset_{SysI1.db} = 0 \quad \textbf{(149)}$$

The third order distortion input intercept point is computed from:

$$P_{dist} = SignalPower_{SysI2.db} - cochannel \quad \textbf{(150)}$$

$$IP3_{TrxIp3.db} = InterfererPower_{SysI2.db} + \frac{InterfererPower_{SysI2.db} - Pdist}{2} \quad \textbf{(151)}$$

Please note that this formula will give different values for IP3 based on the $offset_{SysI2.db}$ at which the $SignalPower_{SysI2.db}$ and the $InterfererPower_{SysI2.db}$ are selected. This allows for specifications in which larger interferers are allowed at larger offsets from the desired channel. This can be accommodated in implementations with selectivity distributed through the receiver chain. Therefore, IP3 is included in the transceiver specifications as a separate table: *TrxIp3.db*.

The 1dB compression point of the front end needs to be chosen in such way that the largest interfering signal can still be handled without any compression that would reduce the sensitivity of the receiver for the desired signal. The following dynamic ranges are introduced to simplify this discussion:

$$\begin{aligned} DR_0 &: \text{Dynamic range of the desired signal} \\ DR_1 &: \text{Dynamic range of the desired signal and one interferer} \\ DR_2 &: \text{Dynamic range of the desired signal and two interferers} \end{aligned} \quad \textbf{(152)}$$

The 1dB compression point does not need to accommodate the maximum desired signal. If both DR_1 and DR_2 are smaller than DR_0, then the requirements for the transceiver dynamic range can be reduced by introducing AGC in or before the first active stage of the transceiver. The range ΔG_{AGC} of this AGC for a system with fixed dynamic range should ideally be chosen according to the following formula:

$$\Delta G_{AGC} = DR_0 - \text{Max}(DR_1, DR_2) \quad \textbf{(153)}$$

A somewhat larger ΔG_{AGC} might be chosen to allow for process spread, temperature etc. If ΔG_{AGC} is negative, AGC only makes sense if the dynamic range of the receiver can be adjusted. On the other hand, adjustable dynamic range makes sense even if ΔG_{AGC} is positive, since there might be situations in practice where DR_1 and DR_2 are small compared to DR_0. ΔG_{AGC} is stored in the Trx.db in field RxAgcRange.

The value for the 1dB compression point of the transceiver can now be derived from the values of DR_1, DR_2 and $RxMinSignal_{Trx.db}$:

$$Rx1dBCompr_{Trx.db} = RxMinSignal_{Trx.db} + Max(DR_1, DR_2) + Margin \quad \textbf{(154)}$$

The reasoning behind this equation is that with ideal AGC, the receiver should be able to cope with a dynamic range Max(DR_1, DR_2), and therefore the maximum signal after AGC should be sensitivity plus maximum dynamic range. Obviously, the circuit in front of the AGC should be able to cope with maximum signals as defined in $RxMaxSignal_{System.db}$, but ideally there should be no active circuits in front of the AGC. Margin is a margin that should be taken into account to allow the receiver to operate at the desired BER in the presence of an interferer. Many telecom standards are defined in such a way that this margin is 3dB, resulting in the same interferer signal and noise power.

Selectivity offers two benefits:

1. The noise bandwidth is reduced to the channel bandwidth, thereby increasing the sensitivity of the receiver;
2. Interferers are attenuated until their level is low enough to allow reception of a weak desired signal.

The first benefit can already be achieved with rather moderate filter requirements. Therefore, the selectivity requirements are typically determined by the second point. It is possible to increase the selectivity of a receiver beyond the requirements posed by the second point. This might reduce the IP2 and IP3 requirements of receiver stages. If this selectivity can be implemented with passive components and with low signal loss, then the total power dissipation of the receiver might be reduced. Since the amount of selectivity that can be introduced this way is theoretically unlimited, it is not possible to take this aspect of selectivity or even an upper limit for it into account. Instead, the minimum amount of selectivity needed to meet the specifications will be calculated.

The minimum selectivity is determined by the relative increase in out-of-channel interferers relative to the co-channel interferer. This increase can only be accommodated by selectivity:

$$\begin{aligned} &Attenuation_{TrxSel.db}(offset_{TrxSel.db}) = \\ &\quad \mathrm{Max}\Big(CoChannel_{Trx.db} - \\ &\qquad (InterfererPower_{SysI1.db}(offset_{SysI1.db}) - SignalPower_{SysI1.db}(offset_{sysI1.db})), \\ &\qquad CoChannel_{Trx.db} - \\ &\qquad (InterfererPower_{SysI2.db}(offset_{SysI2.db}) - SignalPower_{SysI2.db}(offset_{sysI2.db}))\Big) \end{aligned} \quad \textbf{(155)}$$

Receiver VCO noise is determined on the basis of single-tone interferers in combination with reciprocal mixing. The requirement is that the product of the VCO sideband noise integrated over the bandwidth of a channel is less than the noise level defined by the $RxMinSNR_{system.db}$ at the desired input signal levels defined for the appropriate single tone interferers:

$$\begin{aligned} CNR_{RxVco.db}(Offset_{RxVco.db}) = {} & SignalPower_{SysI1.db}(Offset_{SysI1.db}) - RxMinSNR_{System.db} - \\ & InterfererPower_{SysI1.db}(Offset_{SysI1.db}) - \\ & 10 * \log(RxChannelBW_{System.db}) - Margin \end{aligned} \quad \textbf{(156)}$$

Again, *Margin* is a margin that should be taken into account to allow the receiver to operate at the desired BER in the presence of an interferer. Many telecom standards are defined in such a way that this margin is 3dB, resulting in the same interferer signal and noise power.

The duty cycle of the transmitter and receiver is assumed to be the reciprocal of the number of time slots:

$$\begin{aligned} RxDutyCycleAct_{Trx.db} &= \frac{1}{TimeSlots_{System.db}} \\ TxDutyCycleAct_{Trx.db} &= \frac{1}{TimeSlots_{System.db}} \end{aligned} \quad \textbf{(157)}$$

More complex systems might need manual adjustments of these fields. Finally, the name of the transceiver and the *comment* and *source* fields can be chosen freely.

G.1 Deriving front end specifications

The front end specification is derived from the transceiver specification, taking into account the properties of the selected antenna and IF/baseband circuits. Again, many fields can be copied directly.

Field	*Description*	*Unit*	*Fmt*
Cochannel	Cochannel suppression	dB	N
RxMinFreq	Minimum receive frequency	MHz	N
RxMaxFreq	Maximum receive frequency	MHz	N
TxMinFreq	Minimum transmit frequency	MHz	N
TxMaxFreq	Maximum transmit frequency	MHz	N
RxChannelPitch	Grid of the rx channel raster	kHz	N
RxChannelBW	3dB bandwidth of one rx radio channel	kHz	N
TxChannelPitch	Grid of the tx channel raster	kHz	N
TxChannelBW	3dB bandwidth of one tx radio channel	kHz	N
BasebandBW	3dB bandwidth of one baseband channel	kHz	N
NrOfChannels	Number of receive/transmit channel pairs		I
RxFreqError	Maximum error in RX RF frequency	kHz	N
RxModulation	Description of RX modulation type		A60
RxDeviation	Frequency deviation of RX FM type signals	kHz	N
TxFreqError	Maximum error in TX RF frequency	kHz	N
TxModulation	Description of TX modulation type		A60
TxDeviation	Frequency deviaton of TX FM type signals	kHz	N
TxStepSignal	Transmit power level step size	dB	N
RxStepRssi	Maximum step size of RSSI	dB	N
DataRate	Raw data rate across the radio channel	kbps	N
TimeSlot	Duration of a single timeslot	ms	N
RxDutyCycleAct	Rx Percentage “on” time in active mode	%	N
TxDutyCycleAct	Tx Percentage “on” time in active mode	%	N

The following parameters are derived by simple tranformations, that will be described later in this section:

Field	*Description*	*Unit*	*Fmt*
RxNF	Receiver noise figure	dB	N
RxGain	Gain of the receiver frontend	dB	N
TxMinSignal	Minimum transmit power level	W	N
TxMaxSignal	Maximum transmit power level	W	N
TxAvgSignal	Average transmit power level	W	N
Rx1dBCompr	Receiver 1dB compression point	dBm	N
RxAgcRange	Receiver optimum AGC range	dB	N
RxMinRssi	Minimum RSSI indication required	dBm	N
RxMaxRssi	Maximum RSSI indication required	dBm	N

Finally, some specifications can be chosen freely:

Field	*Description*	*Unit*	*Fmt*
FrontEndName	Name of the front end		A80
Antenna	Antenna used to derive this frontend spec		A80
IFBB	IF/baseband chip used to derive this frontend		A80
Vsupply	Supply voltage	V	N
Manufacturer	Manufacturer name, if known		A80
Process	IC Process		A80
Technology	IC Technology		A80
Comment	Comment on transceiver specification		M
Source	Description of the source of the data		M

There is one degree of freedom in deriving the specifications from the front end: the IF IC only has specifications for NF and IP3, whereas the front end has specs for NF, IP3 and gain. Because of the trade-off between these parameters, a common-sense default value for the gain is calculated first:

$$RxGain_{FE.db} = \frac{10\log\left(10^{\frac{NF_{IFBB.db}}{10}} - 1\right) - NF_{Trx.db} + IP3_{IFBB.db} - IP3_{Trx.db}}{2} \qquad \textbf{(158)}$$

Basically, this formula selects the geometrical means for the gain of the front end between the two limits posed by meeting the receiver noise figure spec for low front end gains and meeting the receiver IP3 specs for high front end gain. The FAT tool also provides a simple mechanism to explore this trade-off, showing that typically there is a rather flat, and therefore non-critical, optimum.

Based on this first estimate for the gain, the noise figure is calculated based on Friis' equation:

$$NF_{Fe.db} = 10\log\left(10^{\left(\frac{RxNF_{Trx.db} + Gain_{Antenna.db}}{10}\right)} - \frac{10^{\left(\frac{NF_{IFBB.db}}{10}\right)} - 1}{10^{\left(\frac{RxGain_{Fe.db}}{10}\right)}}\right) \qquad \textbf{(159)}$$

The IP3 is calculated based on the worst case equation for IP3 of cascaded stages:

$$IP3_{Fe.db} = 10\log\left(\frac{1}{\frac{1}{10^{\left(\frac{IP3_{Trx.db}}{10}\right)}} - \frac{10^{\left(\frac{RxGain_{Fe.db}}{10}\right)}}{10^{\left(\frac{IP3_{IFBB.db}}{10}\right)}}}\right) + Gain_{Antenna.db} \quad \textbf{(160)}$$

H

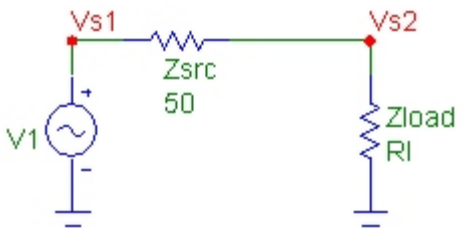
Definitions & derivations

In this section, several definitions and derivations of RF front end related issues will be discussed, such as s-parameter definitions for complex source impedances, and definitions and derivations for third order intercept points.

H.1 S-parameters for a complex source impedance

This section will discuss the inconsistency between the traditional expression for the reflection coefficient and the reflected power in case of a complex source impedance. Also, a new reflection coefficient will be introduced that is consistent both with the original reflection coefficient for real source impedances, and with the reflected power for complex source impedances.

The following picture shows a general situation of a source with source impedance Zsrc and load impedance Zload:



The load impedance Zload causes an output voltage Vs2 that can be described by the following equation:

$$Vs2 = Vs1 \frac{Zload}{Zload + Zsrc}$$

The current Iload flowing through Zload can be expressed as:

$$Iload = \frac{Vs1}{Zsrc + Zload}$$

The power Pload dissipated in the load is therefore:

$$Pload(Zsrc,Zload) = Vs2\ Iload^* = Vs1 \frac{Zload}{Zsrc + Zload}\ \frac{Vs1^*}{(Zsrc + Zload)^*} = |Vs1|^2\ \frac{Zload}{|Zrsc + Zload|^2}$$

The maximum available power from the source, Pmax, is:

$$Pmax(Zsrc) = Pload(Zsrc, Zsrc^*)$$

The reflected power from the load, Prefl, is:

$$Prefl(Zsrc,Zload) = Pmax(Zsrc) - Pload(Zsrc,Zload)$$

The reflected power coefficient Γ' is the ratio between reflected power and maximum available power:

$$\Gamma`(Zsrc,Zload) = \sqrt{\frac{Pref(Zsrc,Zload)}{Pmax(Zsrc)}}$$

Traditionally [41], the reflection coefficient Γ is defined as:

$$\Gamma(Zsrc,Zload) = \frac{Zload - Zsrc}{Zload + Zsrc}$$

For real values of Zsrc, these expressions are consistent:

$$|\Gamma`(Zsrc,Zload)| = |\Gamma(Zsrc,Zload)| \qquad Zsrc \in \Re$$

This can be proven by simple expansion of the left hand side and right hand side, splitting the value of Zsrc and Zload in their respective real and imaginary parts:

$$\Gamma(Zsrc,Zload)\,\Gamma(Zsrc,Zload)^* = \frac{Zload_i^2 + Zload_r^2 - 2\,Zload_i Zsrc_i + Zsrc_i^2 - 2\,Zload_r Zsrc_r + Zsrc_r^2}{Zload_i^2 + Zload_r^2 + 2\,Zload_i Zsrc_i + Zsrc_i^2 + 2\,Zload_r Zsrc_r + Zsrc_r^2}$$

$$\Gamma`(Zsrc,Zload)\,\Gamma`(Zsrc,Zload)^* = \frac{Zload_i^2 + Zload_r^2 + 2\,Zload_i Zsrc_i + Zsrc_i^2 - 2\,Zload_r Zsrc_r + Zsrc_r^2}{Zload_i^2 + Zload_r^2 + 2\,Zload_i Zsrc_i + Zsrc_i^2 + 2\,Zload_r Zsrc_r + Zsrc_r^2}$$

For real Zsrc, the term in which these expressions differ becomes zero. This also happens to be the case when the load impedance is real but the source impedance is complex.

Γ` can be rewritten in a form that is only slightly different from the traditional definition for Γ, and is consistent with Γ for real source impedances:

$$\Gamma`(Zsrc,Zload) = \frac{Zload - Zsrc^*}{Zload + Zsrc}$$

Therefore, the reflected power coefficient Γ` is a more universally applicable parameter than the reflection coefficient Γ, and compatible within the domain Zsrc $\in \Re$. By measuring the voltage ratio Vs2 / Vs1, we can find the value of Rload. The derivation of this formula is shown below:

$$\frac{Vs2}{Vs1} = \frac{Rload}{Rload + Rsrc} = 1 - \frac{Rsrc}{Rload + Rsrc}$$

$$1 - \frac{Vs2}{Vs1} = \frac{Rsrc}{Rload + Rsrc}$$

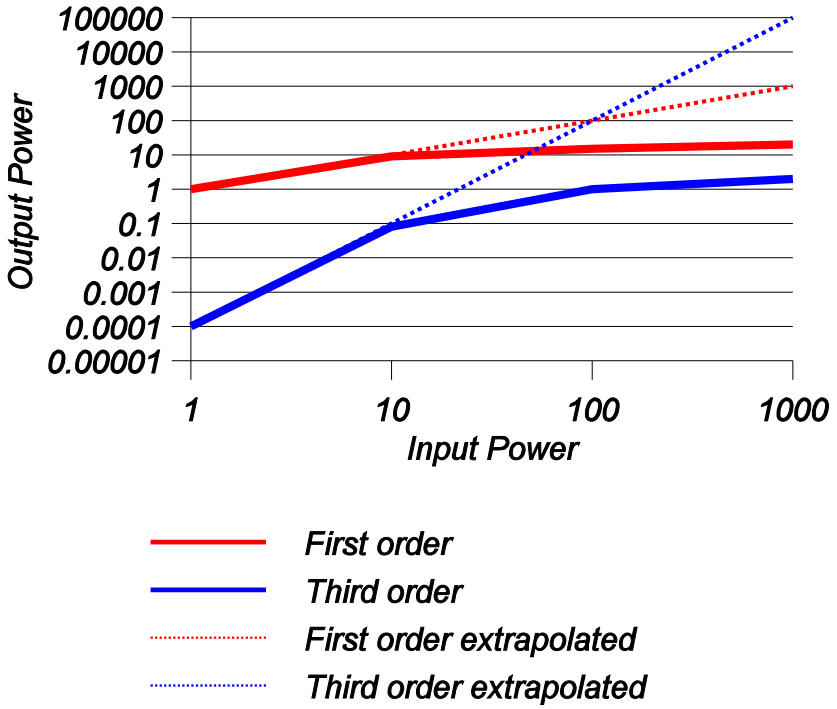
$$Rload = \frac{Rsrc}{1 - \frac{Vs2}{Vs1}} - Rsrc = Rsrc\left(\frac{1}{1 - \frac{Vs2}{Vs1}} - 1\right) = Rsrc\ \frac{\frac{Vs2}{Vs1}}{1 - \frac{Vs2}{Vs1}}$$

H.2 Distortion

In this section, some background issues with respect to third order distortion will be discussed.

H.2.1 IP3

The third order intercept point *IP3* is typically defined as the input power level at which the extrapolated third order intermodulation component of a processed signal would have equaled the extrapolated first order output component. The figure below shows the relations between first order output component, third order input component, and input power level:



IP3 is a convenient linearity parameter for radio circuits, since it directly relates to the common situation where a wanted signal is affected by the distortion products from adjacent and alternate channels. An adjacent channel is a channel that is one channel spacing above or below the wanted channel. An alternate channel is a channel that is two channel spacings above or below the wanted channel. Assume a wanted channel at frequency f_0, and a channel spacing of Δf, If both interferers are at frequencies above f_0, then the interferer at the adjacent channel is at frequency $f_1 = f_0 + \Delta f$, and the interferer at the alternate channel is at frequency $f_2 = f_0 + 2\ \Delta f$. The third order distortion products are now at $3\ f_1$, $3\ f_2$, $2\ f_1 - f_2$, and $2\ f_2 - f_1$, as shown in fig. 165. The third order product at $2\ f_1 - f_2$ coincides with the wanted channel, and effectively acts as a co-channel interferer.

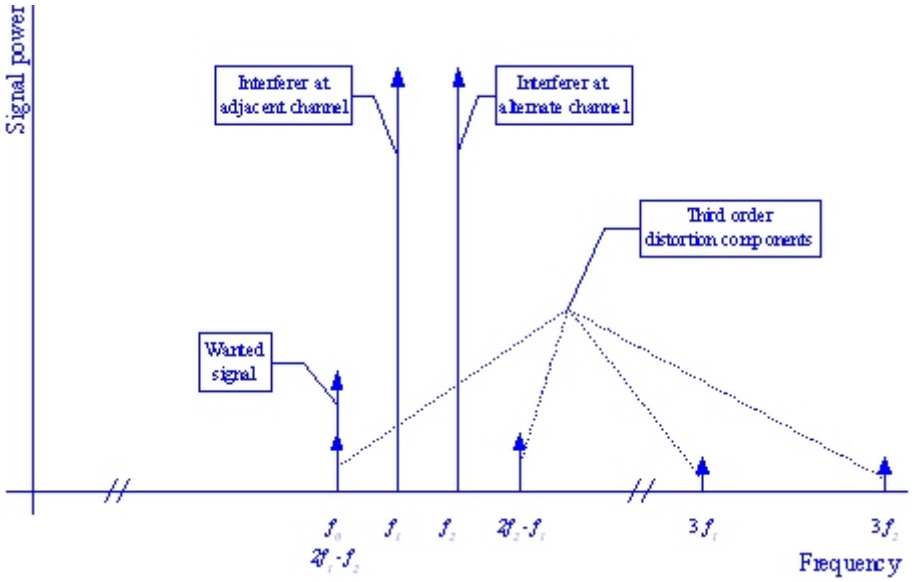


Figure 165 *Effect of third order distortion in radio systems with interferers*

Assume the general case of a non-linear voltage two-port. Such a two-port can be modeled through a Taylor Series expansion around its biasing point V_0:

$$V_{out} = a + b(V_{in} - V_0) + c(V_{in} - V_0)^2 + d(V_{in} - V_0)^3 + \ldots \quad \textbf{(173)}$$

$VIP3_0$ is defined as the level of the DC input signal V_{in} for which the extrapolated distortion components are equal to the extrapolated amplified input signal. In case of DC signals, this results in a value of:

$$VIP3_0 = \sqrt{\frac{b}{d}} \quad \textbf{(174)}$$

The *IP3* that corresponds to this *VIP3* can be easily found when taking into account the input impedance Z_{in} of the circuit. In this section, Z_{in} is assumed to be a real (i.e. resistive) impedance:

$$IP3_0 = \frac{VIP3_0^2}{Z_{in}} = \frac{b}{dZ_{in}}$$

For single-tone input signals, an input signal is defined as:

$$V_{in} = \alpha\sqrt{2}\sin(\omega t) \quad \textbf{(176)}$$

The factor $\sqrt{2}$ was introduced to get the same signal power level as DC for the same amplitude α. This will be expanded around V_0=0. The amplified input signal around this biasing point is:

$$b V_{in} = b \alpha \sqrt{2} \sin(\omega t) \tag{177}$$

The third order distortion component is now:

$$d V_{in}^3 = d \alpha^3 \sqrt{2} \frac{(3 \sin(\omega t) - \sin(3 \omega t))}{2} \tag{178}$$

Please note that the third order distortion component also generates signals at the same frequency as the input signal. These components are responsible for the non-constat gain of the signals at the input frequency, such as compression effects. Since the third order signal component is distributed across two frequency components, but single-tone $VIP3_1$ is traditionally based on the power at three times the input frequency, the expression for single-tone $VIP3_1$ is different from the formula for DC $VIP3_0$:

$$VIP3_1 = \sqrt{\frac{2b}{d}}$$

The $IP3_1$ that corresponds to this $VIP3_1$ can be found easily, because different frequency components add independently to the signal power. Therefore, equal voltages of different frequency components imply equal power, hence:

$$IP3_1 = \frac{VIP3_1^2}{Z_{in}} = \frac{2b}{dZ_{in}}$$

For a two-tone input signal, which is the usual measurement method for the narrowband systems typically found in telecom applications, the situation is again different. An input signal can now be defined as:

$$V_{in} = \alpha \left(\sin(\omega_1 t) + \sin(\omega_2 t) \right) \tag{181}$$

The amplified signal is now:

$$b V_{in} = b \alpha \left(\sin(\omega_1 t) + \sin(\omega_1 t) \right) \tag{182}$$

The third order distortion component is now:

$$dV_{in}^3 = \frac{1}{4}\alpha^3 d\Big(9\sin(\omega_1 t) + 9\sin(\omega_2 t) - \sin(3\omega_1 t) - \sin(3\omega_2 t) - 3\sin((\omega_1 - 2\omega_2)t) + 3\sin((2\omega_1 - \omega_2)t) - 3\sin((2\omega_1 + \omega_2)t) + 3\sin((\omega_1 + 2\omega_2)t)\Big) \quad \textbf{(183)}$$

In this case, there are even more frequency components in the distorted signal, and only part of the distortion power ends up at the beat frequencies ($2\omega_1$-ω_2) and ($2\omega_2$-ω_1) . Only these distortion components are typically used in determining the dual-tone IP3. Therefore, the expression will be different again. The two-tone $VIP3_2$ can be determined by solving for equal amplitude of the beat frequencies and the desired signal. This results in:

$$VIP3_2 = 2\sqrt{\frac{b}{3d}} \quad \textbf{(184)}$$

The two-tone *IP3* can be found from the *VIP3* using the same approach:

$$IP3_2 = \frac{VIP3_2^2}{Z_{in}} = \frac{4b}{3dZ_{in}} \quad \textbf{(185)}$$

When this is translated from voltages to power levels, another inconsistency is often introduced: the amplitude of the two-tone input signal is typically taken as the amplitude of the individual tones. This inconsistent $IP3_2$ value will be indicated by *IP3′*. It is defined by the following equation:

$$IP3' = \frac{VIP3_2^2}{2\,Z_{in}} = \frac{2b}{3dZ_{in}} \quad \textbf{(186)}$$

Please note that the *IP3′* inconsistency cannot be applied to the single-tone or DC case, since there is no second input signal component that can be neglected. Summarizing, the following expressions represent the different measurement methods of *IP3*:

DC IP3 ($IP3_0$)	$\frac{b}{dZ_{in}}$

Single Tone IP3 ($IP3_1$)	$\frac{2b}{dZ_{in}}$
Dual Tone IP3 ($IP3_2$)	$\frac{4b}{3dZ_{in}}$
Historical Dual Tone IP3 ($IP3'$)	$\frac{2b}{3dZ_{in}}$

From this table, it becomes clear that the different methods that exist to derive *IP3* from a circuit result in different *IP3* figures. This is caused by neglecting various input- and distortion output signal components. Unfortunately, the differences are large enough to cause confusion, but not large enough to make it implicitly obvious which method has been used. Therefore, it is preferable to indicate the measurement method used to arrive at a quoted *IP3*. In this thesis, the conventions listed in the table above will be used. If *IP3* without further indication is used, *IP3'* is meant.

H.2.2 IP3 of cascaded stages

When cascading several stages, all stages contribute directly and indirectly to the distortion at the output. Direct contributions consist of intermodulation products at the output of a stage that will be amplified by subsequent stages until they reach the output. Indirect contributions consist of amplification of the input signal for subsequent stages that causes increased distortion in these stages.

Predicting the IP3 of cascaded stages from the gains and IP3s of the individual stages is very desirable. Unfortunately, it is not possible to provide an exact solution, since the IP3 only specifies the magnitude of the distortion of a stage, and not the phase relationships If only gain and IP3 are known, the distortion components of different stages have undefined phases. They could all add in phase (worst case), but they could also cancel each other (best case), or one might argue that their phases are uncorrelated and therefore they will add as uncorrelated signals (i.e. power summing).

Usually, power summing is used to get an impression of typical specifications, and voltage summing is used for worst case specifications. In this section, expressions for both summing methods will be derived, based on the cascaded stages shown in fig. 166.

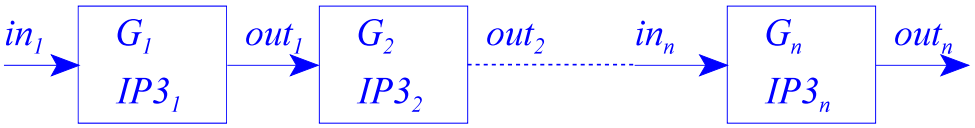


Figure 166 *Cascaded stages*

For a single stage, IP3 is defined as:

$$IP3 = P_{in} \sqrt{\frac{P_{out}}{P_{Dist}}} \tag{191}$$

P_{in} and P_{out} are related through:

$$P_{out} = GP_{in} \tag{192}$$

Substituting (192) in (191) gives:

$$IP3 = P_{in}^{\frac{3}{2}} \sqrt{\frac{G}{P_{dist}}} \tag{193}$$

Solving for P_{dist} results in:

$$P_{dist} = \frac{GP_{in}^3}{IP3^2} \tag{194}$$

The IP3 of multiple cascaded stages can now be determined for the worst case situation: in-phase addition of distortion components. The distortion at the output of the cascaded stages is:

$$P_{dist} = \left(\sqrt{G_2 G_3 \ldots G_n P_{dist_1}} + \sqrt{G_3 \ldots G_n P_{dist_2}} + \ldots + \sqrt{P_{dist_n}} \right)^2 \tag{195}$$

or:

$$P_{dist} = \left(\sum_{i=1}^{n} \sqrt{P_{dist_i} \prod_{j=i+1}^{n} G_j} \right)^2 \tag{196}$$

Substituting (194) in (196) gives:

$$P_{dist} = \left(\sum_{i=1}^{n} \sqrt{\frac{G_i {P_{in_i}}^3}{IP3_i^2} \prod_{j=i+1}^{n} G_j} \right)^2 \tag{197}$$

The input power of stage i (denoted as $P_{in,i}$) can be expressed in terms of the input power P_{in}, as is shown in the next formula:

$$P_{dist} = \left(\sum_{i=1}^{n} \sqrt{\frac{G_i P_{in}^{\,3} \prod_{j=1}^{i-1} G_j^3}{IP3_i^2} \prod_{j=i+1}^{n} G_j} \right)^2 = \left(\sum_{i=1}^{n} \sqrt{\frac{P_{in}^{\,3} \prod_{j=1}^{i-1} G_j^2 \prod_{j=1}^{n} G_j}{IP3_i^2}} \right)^2 \quad \textbf{(198)}$$

Now P_{in} can be extracted from the sum, and the products of gain can be combined:

$$P_{dist} = P_{in}^{\,3} \left(\prod_{i=1}^{n} G_i \right) \left(\sum_{i=1}^{n} \frac{\sqrt{\prod_{j=1}^{i-1} G_j^2}}{IP3_i} \right)^2 \quad \textbf{(199)}$$

The products of stage gains G_i can be combined into the total gain G:

$$G = \prod_{i=1}^{n} G_i \quad \textbf{(200)}$$

This results in the following expression for the output distortion power:

$$P_{dist} = P_{in}^{\,3} G \left(\sum_{i=1}^{n} \frac{\prod_{j=1}^{i-1} G_j}{IP3_i} \right)^2 \quad \textbf{(201)}$$

Substituting (201) in (191) results in the expression for worst case *IP3* of cascaded stages:

$$\frac{1}{IP3} = \sum_{i=1}^{n} \frac{\prod_{j=1}^{i-1} G_j}{IP3_i} \quad \textbf{(202)}$$

or:

$$IP3 = \frac{1}{\sum_{i=1}^{n} \frac{\prod_{j=1}^{i-1} G_j}{IP3_i}} \quad \textbf{(203)}$$

In a similar way, the typical *IP3* can be derived for uncorrelated phases of distortion components:

$$P_{dist} = P_{dist_1} + P_{dist_2} + \ldots + P_{dist_n} = \sum_{i=1}^{n} P_{dist_i} \tag{204}$$

Substituting 252 in this equation yields:

$$P_{dist} = \sum_{i=1}^{n} \frac{G_i P_{in_i}^3}{IP3_i^2} \prod_{j=i+1}^{n} G_j \tag{205}$$

Again, the input power of stage *i* can be expressed in terms of P_{in}:

$$P_{dist} = \sum_{i=1}^{n} \frac{G_i P_{in}^3 \prod_{j=1}^{i-1} G_i^3}{IP3_i^2} \prod_{j=i+1}^{n} G_j \tag{206}$$

Now P_{in} can be extracted:

$$P_{dist} = P_{in}^3 \sum_{i=1}^{n} \left(\frac{G_i \prod_{j=1}^{i-1} G_j^3}{IP3_i^2} \prod_{j=i+1}^{n} G_j \right) = P_{in}^3 \sum_{i=1}^{n} \left(\frac{\prod_{j=1}^{i-1} G_j^2 \prod_{j=1}^{n} G_j}{IP3_i^2} \right) \tag{207}$$

And the gains of the separate stages can be combined into a total gain *G*:

$$P_{dist} = P_{in}^3 G \sum_{i=1}^{n} \left(\frac{\prod_{j=1}^{i-1} G_j^2}{IP3_i^2} \right) \tag{208}$$

Using P_{dist} to calculate the overal typical (instead of worst-case) *IP3* results in:

$$\frac{1}{IP3} = \sqrt{\sum_{i=1}^{n} \left(\frac{\prod_{j=1}^{i-1} G_j^2}{IP3_i^2} \right)} \tag{209}$$

The total typical *IP3* now becomes:

$$IP3 = \frac{1}{\sqrt{\sum_{i=1}^{n}\left(\frac{\prod_{j=1}^{i-1} G_j^2}{IP3_i^2}\right)}} \tag{210}$$

Section

References

[1] Article “Telecommunication”, in “Compton’s Interactive Encyclopedia 1995", CD-ROM, Compton’s Newmedia Inc., 1995

[2] Article “Telecommunication”, in “Britannica 2001 Standard Edition CD-ROM”, Britannica.com Inc., Bristol, U.K., 2001

[3] "Wireless Communication, The Interactive Multimedia CD-ROM", Edition 2001, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2001

[4] A. Bruce Carlson, “Communication Systems”, second edition, McGraw-Hill Kogakusha Ltd., Tokyo, Japan, 1975

[5] The story on wired car phones in 1910 can be found at http://www.privateline.com/TelephoneHistory2A/ericsson.htm

[6] Robert G. Gallager, “Information Theory and Reliable Communication”, John Wiley & Sons, Inc., New York, U.S.A., 1968

[7] Robert Berezdivin e.a., “Next-Generation Wireless Communications Concepts and Technologies”, IEEE Communications Magazine, March 2002, pg. 108-116

[8] C. Takahashi et al., “A 1.9GHz Si Direct Conversion Receiver IC for QPSK modulation Systems”, Proc. ISSCC, San Francisco, February 1995, pp. 138-139

[9] M. McDonald,"A 2.5GHz Silicon Bipolar Image-Reject Front End", Proc. ISSCC, San Francisco, February 1993

[10] J. Sevenhans e.a., "An Integrated SI bipolar RF transceiver for a Zero-IF 900MHz GSM Digital Mobile Radio Front End of a Hand Portable Phone.", IEEE CICC Conference, May 12-15 1991, San Diego

[11] J. Rudell et al., "A 1.9GHz Wide-Band IF Double Conversion CMOS Integrated Receiver For Cordless Telephone Applications", proc. ISSCC, San Francisco, 1997

[12] S. Heinen et al., "A 2.7V 2.5GHz Bipolar Chipset for Digital Wireless Communication", proc. ISSCC, San Francisco, 1997

[13] A. Abidi et al., "The Future of CMOS Wireless Transceivers", proc. ISSCC, San Francisco, 1997

[14] G. Dawe et al., "A 2.7V DECT RF Transceiver with Integrated VCO", proc. ISSCC, San Francisco, 1997

[15] J. Crols, "A 1.5GHz Highly Linear CMOS Down conversion Mixer", IEEE JSSC, Vol. 30, No. 7, 1995

[16] R. Meyer et al., "A 2.5GHz BiCMOS Transceiver for Wireless LAN", proc. ISSCC, San Francisco, 1997

[17] A. Shahani et al., "A 12mW Wide Dynamic Range CMOS Front-End for a Portable GPS Receiver", proc. ISSCC, San Francisco, 1997

[18] Isao Yoshida et. al., “A 3.6V 4W 0.2cc Si Power-MOS-Amplifier Module for GSM Handset Phones”, ISSCC Digest of Technical Papers, San Francisco, February 1998, pg. 3.4-1 .. 3.4-8

[19] S. Wong, S. Luo and L. Hadley, “A 2.7-5.5V 0.2-1W BiCMOS RF Driver Amplifier IC with Closed-loop Power Control and Biasing”, ISSCC Digest of

Technical Papers, San Francisco, February 1998, pg. 3.5-1 .. 3.5-10

[20] P. Asbeck et al., “Device and Circuit Approaches for next Generation Wireless Communications”, Technical Feature, Microwave Journal, February 1999, fig. 8

[21] H. Hashemi, “The Indoor Radio Propagation Channel”, Proceedings of the IEEE, Vol. 81, No. 7, July 1993

[22] Peter Baltus et al., "Design Considerations for RF Power Amplifiers demonstrated through a GSM/EDGE Power Amplifier Module", AACD 2001, Noordwijk, The Netherlands

[23] J. Meindl, A. Davis, “The Fundamental Limit on Binary Switching Energy for Terascale Integration (TSI)”, JSSC Vol.35 Nr.10, October 2000, pg. 1515 .. 1516

[24] T. Liu and E. Westerwick, "5-GHz CMOS radio transceiver front-end chipset," *IEEE Journal of Solid-State Circuits,* vol. 35, pp. 1927 - 1933, December 2000.

[25] M. Harada, T. Tsukahara, J. Kodate, A. Yamagishi, and J. Yamada, "2-GHz RF front-end circuits in CMOS/SIMOX operating at an extremely low voltage of 0.5 V," *IEEE Journal of Solid-State Circuits,* vol. 35, pp. 2000 - 2004, December 2000

[26] G. Grau, U. Langmann, W. Winkler, D. Knoll, J. Osten, and K. Pressel, "A current-folded up-conversion mixer and VCO with center-tapped inductor in a SiGe-HBT technology for 5-GHz wireless LAN applications," *IEEE Journal of Solid-State Circuits,* vol. 35, pp. 1345 - 1352, September 2000

[27] J. D. van der Tang and D. Kasperkovitz, "A low-phase-noise reference oscillator with integrated pMOS varactors for digital satellite receivers," *IEEE Journal of Solid-State Circuits,* vol. 35, pp. 1238 - 1243, August 2000

[28] B. De Muer, M. Borremans, M. Steyaert, and G. Li Puma, "A 2-GHz low-phase-noise integrated LC-VCO set with flicker-noise upconversion minimization," *IEEE Journal of Solid-State Circuits,* vol. 35, pp. 1034 - 1038, July 2000

[29] H. R. Rategh, H. Samavati, and T. H. Lee, "A CMOS frequency synthesizer with an injection-locked frequency divider for a 5-GHz wireless LAN receiver," *IEEE Journal of Solid-State Circuits,* vol. 35, pp. 780 - 787, May 2000

[30] F. Svelto, S. Deantoni, and R. Castello, "A 1.3 GHz low-phase noise fully tunable CMOS LC VCO," *IEEE Journal of Solid-State Circuits,* vol. 35, pp. 356 - 361, March 2000

[31] C. Hung and K. K. O, "A packaged 1.1-GHz CMOS VCO with phase noise of -126 dBc/Hz at a 600-kHz offset," *IEEE Journal of Solid-State Circuits,* vol. 35, pp. 100 - 103, January 2000

[32] K. W. Kobayashi, A. K. Oki, L. T. Tran, J. C. Cowles, A. Gutierrez-Aitken, F. Yamada, T. R. Block, and D. C. Streit, "A 108-GHz InP-HBT monolithic push-push VCO with low phase noise and wide tuning bandwidth," *IEEE Journal of Solid-State Circuits,* vol. 34, pp. 1225 - 1232, September 1999

[33] F. Piazza and Q. Huang, "A 1.57-GHz RF front-end for triple conversion GPS receiver," *IEEE Journal of Solid-State Circuits,* vol. 33, pp. 202 - 209, February 1998

[34] S. Wu and B. Razavi, "A 900-MHz/1.8-GHz CMOS receiver for dual-band applications," *IEEE Journal of Solid-State Circuits,* vol. 33, pp. 2178 - 2185, December 1998

[35] H. Voorman and H. Veenstra, "Tunable high-frequency Gm-C filters," *IEEE*

Journal of Solid-State Circuits, vol. 35, pp. 1097 - 1108, August 2000

[36] H.C. de Graaff, W.J. Kloosterman, and T.N. Jansen, "Compact Bipolar Transistor Model for CACD, with Accurate Description of Collector Behaviour", Proc. Conf. on Solid-State Dev. and Mat. (Tokyo), pp. 287-290, 1986

[37] M. Madihian, T. Drenski, L. Desclos, H. Yoshida, H. Hirabayashi, and T. Yamazaki, "A 5-GHz-band multifunctional BiCMOS transceiver chip for GMSK modulation wireless systems," *IEEE Journal of Solid-State Circuits,* vol. 34, pp. 25 - 32, January 1999

[38] B. Razavi, "CMOS technology characterization for analog and RF design," *IEEE Journal of Solid-State Circuits,* vol. 34, pp. 268 - 276, March 1999

[39] J. L. Tham, M. A. Margarit, B. Pregardier, C. D. Hull, R. Magoon, and F. Carr, "A 2.7-V 900-MHz/1.9-GHz dual-band transceiver IC for digital wireless communication," *IEEE Journal of Solid-State Circuits,* vol. 34, pp. 286 - 291, March 1999

[40] Q. Huang, P. Orsatti, and F. Piazza, "GSM transceiver front-end circuits in 0.25-µm CMOS," *IEEE Journal of Solid-State Circuits,* vol. 34, pp. 292 - 303, March 1999

[41] K. Yamamoto, T. Moriwaki, T. Fujii, J. Otsuji, M. Miyashita, Y. Miyazaki, and K. Nishitani, "A 2.2-V operation, 2.4-GHz single-chip GaAs MMIC transceiver for wireless applications," *IEEE Journal of Solid-State Circuits,* vol. 34, pp. 502 - 512, April 1999

[42] B. Razavi, "A 900-MHz/1.8-GHz CMOS transmitter for dual-band applications," *IEEE Journal of Solid-State Circuits,* vol. 34, pp. 573 - 579, May 1999

[43] M. T. Terrovitis and R. G. Meyer, "Noise in current-commutating CMOS mixers," *IEEE Journal of Solid-State Circuits,* vol. 34, pp. 772 - 783, June 1999

[44] P. J. Sullivan, B. A. Xavier, and W. H. Ku, "Doubly balanced dual-gate CMOS mixer," *IEEE Journal of Solid-State Circuits,* vol. 34, pp. 878 - 881, June 1999

[45] K. Tsai and P. R. Gray, "A 1.9-GHz, 1-W CMOS class-E power amplifier for wireless communications," *IEEE Journal of Solid-State Circuits,* vol. 34, pp. 962 - 970, July 1999

[46] J. T. Colvin, S. S. Bhatia, and K. K. O, "Effects of substrate resistances on LNA performance and a bondpad structure for reducing the effects in a silicon bipolar technology," *IEEE Journal of Solid-State Circuits,* vol. 34, pp. 1339 - 1344, September 1999

[47] K. S. Kundert, "Introduction to RF simulation and its application," *IEEE Journal of Solid-State Circuits,* vol. 34, pp. 1298 - 1319, September 1999

[48] B. Razavi, "A 2.4-GHz CMOS receiver for IEEE 802.11 wireless LANs," *IEEE Journal of Solid-State Circuits,* vol. 34, pp. 1382 - 1385, October 1999

[49] S. S. Mohan, M. del Mar Hershenson, S. P. Boyd, and T. H. Lee, "Simple accurate expressions for planar spiral inductances," *IEEE Journal of Solid-State Circuits,* vol. 34, pp. 1419 - 1424, October 1999

[50] P. Orsatti, F. Piazza, and Q. Huang, "A 20-mA-receive, 55-mA-transmit, single-chip GSM transceiver in 0.25-µm CMOS," *IEEE Journal of Solid-State Circuits,* vol. 34, pp. 1869 - 1880, December 1999

[51] W. Simbürger, H. Wohlmuth, P. Weger, and A. Heinz, "A monolithic transformer coupled 5-W silicon power amplifier with 59% PAE at 0.9 GHz," *IEEE Journal of Solid-State Circuits,* vol. 34, pp. 1881 - 1892, December 1999

[52] A. Pärssinen, J. Jussila, J. Ryynänen, L. Sumanen, and K. A. I. Halonen, "A 2-GHz wide-band direct conversion receiver for WCDMA applications," *IEEE Journal of Solid-State Circuits,* vol. 34, pp. 1893 - 1903, December 1999

[53] H. Darabi and A. A. Abidi, "Noise in RF-CMOS mixers: A simple physical model," *IEEE Journal of Solid-State Circuits,* vol. 35, pp. 15 - 25, January 2000

[54] C. Samori, A. L. Lacaita, A. Zanchi, S. Levantino, and G. Calì, "Phase noise degradation at high oscillation amplitudes in LC-tuned VCO's," *IEEE Journal of Solid-State Circuits,* vol. 35, pp. 96 - 99, January 2000

[55] C. Hung and K. K. O, "A packaged 1.1-GHz CMOS VCO with phase noise of -126 dBc/Hz at a 600-kHz offset," *IEEE Journal of Solid-State Circuits,* vol. 35, pp. 100 - 103, January 2000

[56] C. C. Enz and Y. Cheng, "MOS transistor modeling for RF IC design," *IEEE Journal of Solid-State Circuits,* vol. 35, pp. 186 - 201, February 2000

[57] T. H. Lee and A. Hajimiri, "Oscillator phase noise: A tutorial," *IEEE Journal of Solid-State Circuits,* vol. 35, pp. 326 - 336, March 2000

[58] S. S. Mohan, M. d. M. Hershenson, S. P. Boyd, and T. H. Lee, "Bandwidth extension in CMOS with optimized on-chip inductors," *IEEE Journal of Solid-State Circuits,* vol. 35, pp. 346 - 355, March 2000

[59] M. S. Kappes, "A 3-V CMOS low-distortion class AB line driver suitable for HDSL applications," *IEEE Journal of Solid-State Circuits,* vol. 35, pp. 371 - 376, March 2000

[60] J. Duan, L. Ko, and B. Daneshrad, "A highly versatile beamforming ASIC for application in broad-band fixed wireless access systems," *IEEE Journal of Solid-State Circuits,* vol. 35, pp. 391 - 400, March 2000

[61] F. Behbahani, W. Tan, A. Karimi-Sanjaani, A. Roithmeier, and A. A. Abidi, "A broad-band tunable CMOS channel-select filter for a low-IF wireless receiver," *IEEE Journal of Solid-State Circuits,* vol. 35, pp. 476 - 489, April 2000

[62] S. F. Tin, A. A. Osman, K. Mayaram, and C. Hu, "A simple subcircuit extension of the BSIM3v3 model for CMOS RF design," *IEEE Journal of Solid-State Circuits,* vol. 35, pp. 612 - 624, April 2000

[63] H. Samavati, H. R. Rategh, and T. H. Lee, "A 5-GHz CMOS wireless LAN receiver front end," *IEEE Journal of Solid-State Circuits,* vol. 35, pp. 765 - 772, May 2000

[64] A. Deiss, D. Pfaff, and Q. Huang, "A 200-MHz sub-mA RF front end for wireless hearing aid applications," *IEEE Journal of Solid-State Circuits,* vol. 35, pp. 977 - 986, July 2000

[65] M. Kole, T. Heijmen, T. Kevenaar, H.-J. Pranger, and M. Sevat, "Adapt, an interactive tool for analog synthesis", Proc. SAME-2001 Conf., Sophia Antipolis, France, pp. 15-19, 2001

[66] T. Heijmen, “Optimization methods in integrated circuit design”, Workshop on Optimization and Coupled Problems in Electromagnetism, Napoli, 2003

[67] Jung-Suk Goo, Hee-Tae Ahn, Donald J. Ladwig, Zhiping Yu, Thomas H. Lee, and Robert W. Dutton, "Design Methodology for Power-Constrained Low Noise RF Circuits," Proceedings of the Workshop on Synthesis and System Integration of Mixed Technologies (SASIMI), Nara, Japan, pp. 394-401, October 2001

[68] http://www.fme.fujitsu.com/product/rfic/methodology.html

[69] R. Naiknaware and T. S. Fiez, "Automated hierarchical CMOS analog circuit

stack generation with intramodule connectivity and matching considerations," *IEEE Journal of Solid-State Circuits,* vol. 34, pp. 304 - 317, March 1999

[70] Jacob Rael, Ahmadreza Rofougaran, Asad Abidi, "Design Methodology Used in a Single-Chip CMOS 900 MHz Spread-Spectrum Wireless Transceiver", DAC, San Francisco, CA, USA, pg. 44, 1998

[71] http://www.barcelonadesign.com/

[72] D. Bayer, "Methodology for reuse of high speed mixed signal and rf integrated circuits", thesis, Ohio State University, 2003

[73] Klumperink, E.A.M., Bruccoleri, F., Nauta, B., "Finding All Elementary Circuit Exploiting Transconductance", IEEE transactions on circuits and systems-II, 48(11), 1039-1053, November 2001

[74] F. Bruccoleri, E. A. M. Klumperink, B. Nauta, "Generating All Two-MOS-Transistor Amplifier Leads to New Wide-Band LNAs", IEEE Journal of Solid-State Circuits, Vol. 36, No. 7, pp. 1032-1040, July 2001

[75] Eric A. M. Klumperink, "Transconductance based cmos circuits: Circuit generation, classification and analysis", ISBN 90-3650921-1, PhD Thesis, University of Twente, Enschede, The Netherlands, 1997

[76] K. Fong, "Design and Optimization Techniques forMonolithic RF Downconversion Mixers", thesis, University of California, Berkeley, 1997

[77] H. Duric, "Method for optimisation of mobile and wireless transceiver design", M.Sc. Thesis, TUE, Eindhoven, 10 December 2002.

[78] C. Hull, "Analysis and Optimization of Monolithic RF downconversion receivers", thesis, University of California, Berkeley, 1992

[79] Reza Mahmoudi, "A multidisciplinary design method for second and third generation mobile communication system microwave components", thesis, University Twente, The Netherlands, 2001, ISBN 90-365-1661-7

[80] D. K. Shaeffer, A. R. Shahani, S. S. Mohan, H. Samavati, H. R. Rategh, M. del Mar Hershenson, M. Xu, C. P. Yue, D. J. Eddleman, and T. H. Lee, "A 115-mW, 0.5-µm CMOS GPS receiver with wide dynamic-range active filters," *IEEE Journal of Solid-State Circuits,* vol. 33, pp. 2219 - 2231, December 1998

[81] J. P. Maligeorgos and J. R. Long, "A low-voltage 5.1-5.8-GHz image-reject receiver with wide dynamic range," *IEEE Journal of Solid-State Circuits,* vol. 35, pp. 1917 - 1926, December 2000

[82] T. Liu and E. Westerwick, "5-GHz CMOS radio transceiver front-end chipset," *IEEE Journal of Solid-State Circuits,* vol. 35, pp. 1927 - 1933, December 2000.

[83] P. G. M. Baltus, A. G. Wagemans, R. Dekker, A. Hoogstraate, H. Maas, A. Tombeur, and J. van Sinderen, "A 3.5-mW, 2.5-GHz diversity receiver and a 1.2-mW, 3.6-GHz VCO in silicon on anything," *IEEE Journal of Solid-State Circuits,* vol. 33, pp. 2074 - 2079, December 1998

[84] Q. Huang, F. Piazza, P. Orsatti, and T. Ohguro, "The impact of scaling down to deep submicron on CMOS RF circuits," *IEEE Journal of Solid-State Circuits,* vol. 33, pp. 1023 - 1036, July 1998

[85] Lawrence E. Larson, "Integrated Circuit Technology Options for RFIC's—Present Status and Future Directions", IEEE Journal of Solid-State Circuits, vol. 33, no. 3, pp. 387, March 1998

[86] Derek K. Shaeffer, Arvin R. Shahani, S. S. Mohan, Hirad Samavati, Hamid R. Rategh,Maria del Mar Hershenson, Min Xu, C. Patrick Yue, Daniel J. Eddleman,

and Thomas H. Lee, “A 115-mW, 0.5- m CMOS GPS Receiver with Wide Dynamic-Range Active Filters”, IEEE Journal of Solid-State Circuits, vol. 33, no. 12, Pg. 2219, December 1998

[87] Eyad Abou-Allam, John Nisbet, Michael Maliepaard, “Low-Voltage 1.9-GHz Front-End Receiver in 0.5um CMOS Technology”, IEEE Journal of Solid States Circuits, vol. 36, no. 10, pg. 1434, October 2001

[88] Francesco Gatta, Enrico Sacchi, Francesco Svelto, Paolo Vilmercati, Rinaldo Castello, “A 2-dB Noise Figure 900-MHz Differential CMOS LNA”, IEEE Journal of Solid State Circuits, vol. 36, no. 10, pg. 1444, October 2001

[89] Guiseppe Gramegna, Mario Paparo, Pietro Erratico, Placido De Vita, “A Sub-1-dB NF ±2.3-kV ESD-Protected 900-MHz CMOS LNA”, IEEE Journal of Solid-State Circuits, vol. 36, no. 7, pg. 1010, July 2001

[90] ETSI, “Radio Equipment and Systems. Digital European Cordless Telecommunications Common Interface. Part 2: Physical Layer”, DE/RES 3001-2, 1992

[91] Kazuaki Murota, Kenkichi Hirade, “GMSK Modulation for Digital Mobile Radio Telephony”, IEEE Transactions on Communications, Vol. COM-29, No. 7, July 1981, pg. 1044..1050

[92] M. Steyaert, R. Roovers, “A 1GHz Single Chip Quadrature Modulator without External Trimming”, proceedings of the 17th European Solid State Circuits Conference, 11..13 September 1991, pg. 261..264

[93] This graph was proposed by Maja Vidojkovic during a private conversation in the summer of 2003 in Eindhoven, the Netherlands.

[94] M. Pardoen, “A Fully Integrated 1V/100uA high bitrate CP-FSK receiver, AACD 1994

[95] Rudy J. van de Plassche, Peter Baltus, “An 8-bit 100-MHz Full-Nyquist Analog-to-Digital Converter”, IEEE Journal of Solid-State Circuits, Vol. 23, No. 6, December 1988

[96] David M. Pozar, “Microwave and RF Design of Wireless Systems”, 1st Edition, December 2000, John Wiley & Sons, ISBN 0-471-32282-2

[97] G. Dolmans, L. Leyten, “Performance study of an adaptive dual antenna handset for indoor communications”’ IEE Proc.-Microw. Antennas Propag., Vol. 146, No. 2, April 1999

[98] B. Gilbert, “The MICROMIXER: A Highly Linear Variant of the Gilbert Mixer Using a Bisymmetric Class-AB Input Stage”, IEEE JSSC, Vol. 32, No. 9, September 1997, pg. 1412-1423

[99] G. Hurkx et. al., “An Efficient Simulation Method for Linking Bipolar Process and Device Optimization to Circuit Performance”, IEDM, Washington, 1993, pg. 105..108

[100] ETSI DTR-RES 3001, “DECT Reference Document Version 2.1”

[101] P. Baltus, A. Tombeur, “DECT Zero IF Receiver Front End”, Proc. AACD, Leuven, 1993

[102] J.L. de Jong et al., BCTM Tech. Dig. 1989, pp. 121-124

[103] R. Dutton, “The Role of TCAD in Parasitic Analysis of ICs”, Proc. of the 23rd ESSDERC ‘93, Grenoble, France, pg. 75

[104] Wen Fang, “Accurate Analytical Delay Expression for ECL and CML Circuits and Their Applications to Optimizing High-Speed Bipolar Circuits”, IEEE

Journal of Solid State Circuits, Vol. 25, No. 2, April 1990, pg. 572..583
[105] H.C. de Graaff e.a., "A Figure of Merit for the high-frequency noise behaviour of bipolar transistors", Bipolar Circuit & Technology Meeting, Minneapolis, 1993, pg. 118..1210
[106] Ronald Dekker, "An Ultra Low-Power RF Bipolar Technology on Glass", IEDM 1997
[107] G. Hurkx, "The Relevance of fT and fmax for the Speed of a Bipolar CE Amplifier Stage", proc. Of the BCTM, Minneapolis, U.S.A., September 1996
[108] H.J. Butterweck, "Elektrische Netwerken", tweeded druk, Uitgeverij Het Spectrum, Utrecht, The Netherlands, 1979, ISBN 90 274 7555 5
[109] This picture was found at http://www.bell-labs.com/history/75/gallery.html, and is used according to the terms of use set forth in http://www.lucent.com/copyright.html
[110] P. Baltus, "Influence of process- and device parameters on the performance of portable RF communication circuits", proceedings of the 24th ESSDERC, Edinburgh, Scotland, 1994, pg. 3-10
[111] P. Baltus,"Differential amplifier, an integrated circuit, and a telephone", US patent 055999050, 12/07/1999 (class-AB differential pair)
[112] P. Baltus, L. Leyten, H. Visser, A. Tombeur, A. Wagemans, J. van Sinderen, "Radio transmission system and a radio apparatus for use therein", US patent 05887247 03/23/1999 (Angle scanning diversity principle)
[113] P. Baltus, L. Leyten, H. Visser, A. Tombeur, A. Wagemans, J. van Sinderen, "Radio transmission system and a radio apparatus for use in such a system", US patent 05751249 05/12/1998 (Angle scanning diversity implementation)
[114] P. Baltus, A. Janssen, "Receiver and demodulator for phase or frequency modulated signals", US patent 05808509 09/15/1998 (reconstruction of data in limited ZIF FSK receiver)
[115] F. Behbahani, J. C. Leete, Y. Kishigami, A. Roithmeier, K. Hoshino, and A. A. Abidi, "A 2.4-GHz low-IF receiver for wideband WLAN in 0.6-μm CMOS architecture and front-end", IEEE Journal of Solid-State Circuits, vol. 35, pp. 1908 - 1916, December 2000
[116] B. Razavi, "A 2.4-GHz CMOS receiver for IEEE 802.11 wireless LANs", IEEE Journal of Solid-State Circuits, vol. 34, pp. 1382 - 1385, October 1999
[117] M. Madihian, E. Bak, H. Yoshida, H. Hirabayashi, K. Imai, Y. Kinoshita, T. Yamazaki, and L. Desclos, "A 2-V, 1-10 GHz BiCMOS transceiver chip for multimode wireless communications networks", IEEE Journal of Solid-State Circuits, vol. 32, pp. 521 - 525, April 1997
[118] V. Thomas, J. Fenk, and S. Beyer, "A one-chip 2GHz single superhet receiver for 2Mb/s FSK radio communication"" IEEE International Solid-State Circuits Conference, vol. XXXVII, pp. 42 - 43, February 1994
[119] This figure is from an internal Philips Semiconductors report (ERA/IR02019.1) by Arno Neelen.

List of Publications and patents

Publications

1. R. van de Plassche, P. Baltus, "The Design of an 8-bit Folding Analog-to-Digital Converter", Philips Journal of Research, Vol. 42, No. 5/6, 198, pg. 482-510
2. R. van de Plassche, P. Baltus, "An 8b 100MHz Folding ADC", IEEE International Solid State Circuits Conference Digest of Technical Papers, Vol. 31, San Francisco, U.S.A., 1988, pg. 222
3. R. van de Plassche, P. Baltus, "An 8-bit 100-MHz Full-Nyquist Analog-to-Digital Converter", IEEE Journal of Solid State Circuits, Vol. 23, No.6, Dec. 1988, pg. 1334
4. M. Ligthart, P. Baltus, M. Freeman: “RPST : A ROM Based Pseudo-Exhaustive Self Test Approach”, International Test Conference (ITC87). Washington, D.C., USA, 1987, pg. 915-922
5. P. S. van der Meulen, D. Huang, U. Bar-Gadda, E. Lee, P. Baltus, "EXIST: an Interactive VLSI Architectural Environment", 1988 IEEE International Conference on Computer Design: VLSI in Computers & Processors, Rye Brook, NY, U.S.A, 1988, pg. 312-319
6. P. Baltus, P. van der Meulen, R. Morley: “An Efficient Multi-Level Multi-Wire Differential Interface”. Proceedings of the 20th International Symposium on Multiple-Valued Logic (ISMVL90), University of Illinoins, Urbana-Champaign, 1990, pg. 181-188
7. G. Hurkx, P. Baltus, E. Bladt, M. Knuvers, "An Efficient Simulation Method for Linking Bipolar Process and Device Optimization To CIrcuit Performance", Technical Digest of the 39th IEEE International Electron Device Meeting, Washington, 1993, pg. 105-108
8. P. Baltus, A. Tombeur, "DECT Zero IF Receiver Front End", Proceedings of the workshop Advances in Analog Circuit Design, Leuven, Belgium, April 6-8, 1993
9. P. Baltus, A. Tombeur, "DECT Zero IF Receiver Front End", section in “Analog Circuit Design: Mixed A/D Circuit design, Sensor Interface Circuits and Communication Circuits” by W. Sansen et al. (ed.), pg. 295-318
10. W. van der Wel, R. Koster, A. Jansen, E. Bladt, G. Hurkx, P. Baltus, "Poly-Ridge Emitter Transistor (PRET): Simple Low-Power Option to a Bipolar Process", Technical Digest of the 39th IEEE International Electron Device Meeting, Washington, 1993, pg. 453-356
11. P. Baltus, "Influence of Process and Device Parameters on the Performance of Portable RF Communication Circuits", Proc. ESSDERC '94, Edinburgh, Scotland, September 11-15, 1994, pg. 3-11 (invited plenary paper)
12. P. Baltus, "Design Issues for Low Power Mobile Transceiver Frontends", VLSI TSA Workshop, Taiwan, 1997 (invited plenary paper)
13. A. Wagemans, P. Baltus, "A 3.5mW 2.5GHz diversity receiver and a 1.2 mW 3.6 GHz VCO in Silicon-On-Anything", ISSCC, San Francisco, February 1998

14. P. Baltus, "Put your power into SOA LNAs", Proceedings of the workshop Advances in Analog Circuit Design, Kopenhagen, Denmark, 1998
15. P. Baltus, "Put your power into SOA LNAs", section in “Analog Circuit Design: Volt electronics, Mixed Mode Systems, Low-Noise and RF Power Amplifiers for Telecommunication” by W. Sansen et al. (ed.), pg. 337-358
16. R. Dekker et al., "An Ultra Low-Power RF Bipolar Technology on Glass", IEDM '97
17. P. Baltus et. al., "A 3.5mW 2.5GHz diversity receiver in Silicon-On-Anything", JSSCC, December 1998
18. P. Baltus et al., “Optimizing RF Front Ends for Low Power”, special issue on low power of the Proceedings of the IEEE (around December 2000)
19. Peter Baltus et al., “Design Considerations for RF Power Amplifiers demonstrated through a GSM/EDGE Power Amplifier Module”, AACD 2001, Noordwijk, The Netherlands
20. P. Baltus et al., “RF Power Amplifier Fundamentals”, Short course at CICC 2001, San Diego, U.S.A.
21. P. Baltus et al., “Efficient Design Methods for 5GHz and beyond”, ESSCIRC 2002, Florence, Italy (invited plenary paper)

US Patents

1. US06282413 28/8/2000 Multi-staged frequency conversion with single local oscillator (6-mixer single LO dual image cancellation receiver) Peter Baltus
2. US06192229 20/02/2001 Diode Mixer circuit (Schottky Ring Mixer) Eduard Stikvoort, Peter Baltus, Anton Tombeur
3. US06060968 05/09/2000 Device with circuit element and transmission line formed by a dielectric between facing conductor strips (flip chip transmission line) Peter Baltus, Ronald Dekker, Lukas Leyten
4. US055999050 12/07/1999 Differential amplifier, an integrated circuit, and a telephone (class A/B differential pair) Peter Baltus
5. US05887247 03/23/1999 Radio transmission system and a radio apparatus for use therein (Angle scanning diversity principle) Peter Baltus, Lukas Leyten, Henk Visser, Anton Tombeur, Ton Wagemans, Jan van Sinderen
6. US05808509 09/15/1998 Receiver and demodulator for phase or frequency modulated signals (reconstruction of data in limited ZIF FSK receiver) Peter Baltus, Guido Janssen
7. US05751249 05/12/1998 Radio transmission system and a radio apparatus for use in such a system (Angle scanning diversity implementation) Peter Baltus, Lukas Leyten, Henk Visser, Anton Tombeur, Ton Wagemans, Jan van Sinderen
8. US05694071 12/02/1997 Electronic device comprising means for compensating an undesired capacitance (Electronic inductor device) Fred Hurkx, Peter Baltus, Marinus Knuvers, Cees Hart
9. US05224103 06/29/1993 Processing device and method of programming such

a processing device (microcontroller built-in self-test) Michiel Ligthart, Peter Baltus

10. US05189422 02/23/1993 Analog-to-digital converter with delay correction; Rudy van de Plassche, Peter Baltus
11. US05166956 11/24/1992 Data transmission system and apparatus providing multi-level differential signal transmission (multi-level multi-wire interface) Peter Baltus, Pieter van der Meulen
12. US05124588 06/23/1992 Programmable combinational logic circuit (PLD in BiCMOS) Peter baltus, Michiel Ligthart
13. US04939517 07/03/1990 Circuit for synchronizing transitions of bits in a digital code (Bit sync for folding ADC) Peter Baltus, Rudy van de Plassche
14. US04897656 01/30/1990 Complementary voltage interpolation circuit with transmission delay compensation (compensation resistors & interconnect for folding ADC) Rudy van de Plassche, Peter Baltus
15. US04870417 09/26/1989 Error correction circuit suitable for thermometer or circular code (basic error corrector: rounding instead of truncation for folding & flash ADC) Rudy van de Plassche, Peter Baltus
16. US06697606 02/24/2004 Transceiver and a telecommunication system having a transceiver (IS95 LO generation), Ton Wagemans, Peter Baltus

Japanese Patent

1. JP01189227 28/07/1989 Analog/Digital Converter (corresponds to US4897656)

Section

Summary

This thesis describes an investigation into the design of RF front ends with minimum power dissipation. The central question is:

"**What are the fundamental limits for the power dissipation of telecommunication front ends, and what design procedures can be followed that approach these limits and, at the same time, result in practical circuits?**"

After a discussion of the state of the art in this area, the elementary operations of a front end are identified. For each of these elementary operations, the fundamental limits for the power dissipation are discussed, divided into technology imposed limits and physics imposed limits. A traditional DECT front end design is used to demonstrate the large difference between the fundamental limits and the power dissipation of existing circuits.

To improve this situation, first the optimum distribution of specifications across individual subcircuits needs to be determined, such that the requirements for a specific system can be fulfilled. This is achieved through the introduction of formal transforms of the specifications of subcircuits, which correspond with transforms of the subcircuit itself. Using these transforms, the optimum distribution of gain, noise, linearity and power dissipation can be determined. As it turns out, this optimum distribution can even be represented by a simple, analytical expression. This expression predicts that the power dissipation of the DECT front end can be reduced by a factor of 2.7 through an optimum distribution of the specifications.

Using these optimum specifications of the subcircuits, the boundaries for further power dissipation reduction can be determined. This is investigated at the system, circuit and technology level. These insights are used in the design of a 2.5GHz wireless local area network, implemented in an optimized technology ("Silicon on Anything"). The power dissipation of the complete receiver is 3.5mW, more than an order of magnitude below other wireless LAN receivers in recent publications.

Finally, the combination of this minimum power design method with a platform based development strategy is discussed.

Section

Samenvatting

Dit proefschrift beschrijft onderzoek naar het ontwerp van RF front ends met minimale vermogensdissipatie. De centrale vraag hierin is:

"Wat zijn de fundamentele grenzen voor de vermogensdissipatie van RF front ends, en welke ontwerpprocedure kan worden gevolgd zodat deze limieten met praktische circuits kunnen worden benaderd?"

Na een bespreking van de huidige stand van zaken op dit gebied worden de elementaire bewerkingen van een RF front end geïdentificeerd. Voor ieder van deze elementaire bewerkingen worden de fundamentele grenzen voor de vermogensdissipatie besproken, opgesplitst naar grenzen bepaald door de fysica en grenzen bepaald door de technologie. Aan de hand van een traditioneel ontwerp van een DECT front end wordt vervolgens gedemonstreerd dat er nog een groot verschil is tussen de fundamentele grenzen en de vermogensdissipatie van bestaande circuits.

Om hier verbetering in aan te brengen is het allereerst nodig om te bepalen wat de optimale verdeling is van de specificaties van de individuele subcircuits waarmee aan de eisen van een bepaald systeem kan worden voldaan. Dit wordt mogelijk door de introductie van formele transformaties op de specificaties van subcircuits, die corresponderen met een transformatie op het circuit zelf. Aan de hand van deze transformaties kan een optimale verdeling van versterking, ruis, lineariteit en vermogensdissipatie worden bepaald. Deze optimale verdeling blijkt zelfs te kunnen worden weergegeven in een eenvoudige, gesloten uitdrukking. Deze uitdrukking voorspelt dat de vermogensdissipatie van het DECT front end door een optimale verdeling van de specificaties met een factor 2.7 kan worden gereduceerd.

Met deze optimale specificaties van de subcircuits wordt het mogelijk om te bepalen waar er ruimte is voor verdere reductie van de vermogensdissipatie. Dit is onderzocht op systeem-, circuit- en technologieniveau. Deze inzichten zijn vervolgens gebruikt bij het ontwerpen van een 2.5GHz front end voor een draadloos netwerk in een geoptimaliseerde IC-technologie ("Silicon on Anything"). De vermogensdissipatie van de complete ontvanger is 3.5mW, ruim een orde lager dan andere draadloze netwerk-ontvangers in recente publicaties.

Tenslotte wordt besproken hoe deze ontwerpmethode kan worden gecombineerd met een productontwikkelingsstrategie gebaseerd op platformen.

Section

Acknowledgments

The work described in this thesis was not, and could not have been, carried out in isolation. The environment that existed in Philips Research Eindhoven was essential, especially the open and stimulating interaction between people in different competence areas, as well as the interaction with people in the business organizations of Philips. It was never a problem to get help, support, and feedback from people. Therefore, I would like to thank everyone at Philips who has, in some way or another, been involved in my work.

Some people had a special impact on this work. Among them, Rudy van de Plassche, my first group leader at Philips, spent a lot of time teaching me things about analog electronics that cannot be found in text books, and in some cases I could only learn them by being allowed to make the required mistakes myself. Thank you very much, Rudy!

Dieter Kasperkovitz provided unlimited amounts of wisdom and reflection, as well as guidance in technical and organizational problems, especially during the first years of the RF design work at Philips Research, for which I am very grateful.

A lot of the work was done together with people in the RF front end modules cluster. Ton Wagemans combined a positive attitude and lots of optimism with a good insight in RF design. He was always ready for, and looking forward to, the next challenge. His passing away after an illness at such a young age is still difficult to accept, and I miss him a lot. Thanks for everything, Ton! Special thanks also go to Lukas Leyten and Guido Dolmans for their excellent research in the area of radio propagation and antenna diversity, to Luuk de Maaijer for the first RF CMOS circuit designs, to Henk Visser for the 'high power' part of low-power front ends, and of course to Anton Tombeur for working together on most projects, from the very beginning (DECT front end) through the SOA transceiver. Thank you very much! During a lot of this work, I shared an office with Johan van der Tang, who was always there to discuss any topic with me, and still keeps this up even after we both left that office. Thank you very much, Johan!

IC technology plays an important role in RF design, and a lot of this work would not have been possible without a lot of help from, and interaction with, people in IC process research. At Philips Research in Sunnyvale, Richard Lane was the first "IC technology" colleague with whom I discussed the impact of IC process properties on

analog circuit performance. Later, at Philips Research in Eindhoven, Ronald Dekker turned out to be the ideal “IC technology” colleague with both the patience to listen to my ‘complaints’ about IC technology performance, as well as the creativity to do something drastic about it. Ronald, thank you very much for all the great discussions and for all your help! Harrie Maas invented SOA with Ronald, and was always ready and willing to help us with many practical problems such as die photos, etching of plastic IC packages etc. Fred Hurkx and Dick Klaassen helped me understand the details of transistor models, as well as the various figures of merit related to these devices. If things became really complicated, Jan Slotboom always had the patience and insight to help analyze any problem. Carel van der Poel, first as group leader and later as director of the IC technology research sector, encouraged cooperation and provided inspiration. Thanks to you all!

I also received a lot of help from colleagues at Philips Semiconductors Hamburg, especially with SOA design and processing. Special thanks go to Steffen Hahn, Johann Fock, Andreas Wichern, and Knud Holtvoeth.

Guido Janssen solved the mathematics of the general problem of distribution of gain, linearity and noise across signal processing stages in just one day, after I’d been trying on and off for months to do this myself. This is another example of benefits of an environment such as Philips Research. It was very useful for my research, although not necessarily for my self-confidence. Guido also helped with the reconstruction of FSK data after limiting in a zero-IF receiver. Thank you very much, Guido!

The work on RF platforms is being carried out for a large part by staff and students that are/were in the Advanced Systems Lab (ASL) RF group: Martin Barnasconi, Tom Buss, Stefan Crijns, Oswald Moonen, Arno Neelen, Jeroen Snelten, and Marc Vlemmings, as well as Maja Vidojkovic and Haris Duric all worked on various subjects in this area. Theo Claasen anticipated the need for a platform approach in RF, and made sure that work in this area was started. Hans Brouwhuis provided a lot of the inspiration and motivation for this work, and Pieter Jan van Bommel helped develop a sound approach by asking the right questions. This work would not have been possible without the continuous and never failing support of Bob Payne. Bram van den Berge, my current manager, who made sure that finishing this thesis became a priority, and provided lots of practical support as well. Thank you all very much for your help and support with this challenging work!

Martin Schuurmans, at the time director of the IC design center of Philips Research, motivated me to start this line of research by agreeing that it was very ‘unscientific’ that the limits of low-power RF circuit design were unknown, and by subsequently challenging me to find them. Thank you for this and for your help, Martin!

The successive group leaders of the integrated transceiver group at Philips Research in Eindhoven allowed and often motivated me to continue this research: Hans Brandsma, Aad Sempel and Pieter Hooijmans. Thank you!

My promotor, Arthur van Roermund, convinced me to move my original target of finishing my thesis before my retirement to a much earlier date. In addition, he contributed a lot through extensive comments on the first drafts, and discussions about the contents, structure and form of this thesis. Thank you very much, Arthur! The members of the core committee, prof. Koonen, prof. Bergmans and prof. Nauta provided valuable feedback that is included in this thesis, and I appreciate their help as well as the help from the other committee members, prof. Otten, prof. Brombacher, prof. Long and prof. Blom.

Lots of thanks also go to my family and friends. Special thanks go to my parents, who never failed to encourage and support me and my brother, even when our hobby involved e.g. the drilling of holes in the kitchen table. I very much regret that my father is no longer with us.

Finally, thanks to Christel for sharing my life, and for spending so much time together working on this thesis.

Section

Biography

Petrus Gerardus Maria Baltus was born in Sittard, The Netherlands, on July 5th 1960. From a very early age he showed an interest in analyzing the workings of his toys by taking them apart at the very earliest opportunity. Soon afterwards, he became interested in building new things from parts obtained in this way, especially electrical and electronic circuits. After radio receivers and transmitters, his interests broadened to include measurement equipment and computer hardware and software.

Given these interests, studying Electrical Engineering at the University of Technology, Eindhoven, was an obvious choice. During this time, he worked on projects such as an HPIB bus tester, passive video filter synthesis and analysis software, accuracy improvements of industrial weighing equipment, and software models of industrial chemical processes for a real-time simulator. He also worked as an assistant in the organizational behavior research group. He received his M.Sc. Degree (cum laude) in February 1985. The subject of his Master's thesis, carried out at the Philips video lab in Eindhoven, was "visibility of errors generated in time discrete video filters".

After his graduation, he started working at the Philips Research Laboratories in Sunnyvale, California, initially designing fast folding analog-to-digital converters in the group of Rudy van de Plassche. This was a relatively small lab focusing on research in the areas of CAD tools, IC design, and IC technology, offering a very stimulating environment with lots of possibilities for interactions between the different research areas. As a result, he became interested in the relations between system design, circuit design and IC technology. Later, he worked on the design of RISC micro controllers, as well as system simulators and compilers for such micro controllers. In 1990, he moved to the Netherlands and started working in the radio and data transmission systems department of Philips Research in Eindhoven, where he worked for 10 years as research scientist and cluster leader on the design of RF front ends for telecommunication systems. Projects during this time included front ends for cordless phones, pagers, and wireless LANs. Although Philips Research in Eindhoven (the "Nat.Lab.") was a much larger organization than Philips Research Labs in Sunnyvale, it offered again lots of opportunities for interaction with other research groups, especially in the IC technology sector. Most of the work on which this thesis is based was carried out in this period. In 2000, he joined Philips Semiconductors to work as development lab manager at the RF-business development center in Tokyo, Japan, where W-CDMA and Bluetooth products were developed, and later as RF domain manager at the system labs in Eindhoven.

Currently he works as program manager and group leader RF at the Advanced Systems Lab in Eindhoven, with a special interest in RF platforms.

Section

Biografie

Petrus Gerardus Maria Baltus werd geboren op 5 juli 1960 in Sittard. Al vanaf een heel jonge leeftijd toonde hij interesse in het analyseren van de werking van zijn speelgoed door het bij de eerste de beste gelegenheid uit elkaar te halen. Kort daarna raakte hij geïnteresseerd in het opbouwen van nieuwe elektrische en elektronische schakelingen met behulp van de onderdelen die hij op deze manier had verkregen. Na radiozenders en ontvangers verbreedde zijn interesse zich tot meetinstrumenten en computerhardware en software.

Gezien deze interesses was het een voor de hand liggende keuze om elektrotechniek te gaan studeren aan de Technische Universiteit in Eindhoven. Tijdens deze studie werkte hij aan projecten zoals een HPIB bus-tester, passieve videofilter synthese- en analyse-software, verbetering van de nauwkeurigheid van industriële weegwerktuigen en softwaremodellen van industriële chemische processen voor een real-time simulator. Ook werkte hij als student-assistent in de groep organisatiepsychologie. Hij studeerde cum laude af in 1985. Het onderwerp van zijn afstudeerproject, dat werd uitgevoerd op het Philips videolab in Eindhoven, was “zichtbaarheid van beeldfouten gegenereerd in tijddiscrete videofilters”.

Na zijn studie ging hij werken bij de Philips Research Laboratories Sunnyvale in Californië, waar hij begon met het ontwerpen van snelle analoog-digitaal convertors in de groep van Rudy van de Plassche. Dit was een relatief klein laboratorium dat zich richtte op research op het gebied van CAD tools, IC-ontwerp en IC-technologie. Het bood een heel stimulerende omgeving met veel mogelijkheden voor interactie tussen de verschillende research-onderwerpen. Tengevolge hiervan raakte hij geïnteresseerd in de verbanden tussen systeemontwerp, circuitontwerp en IC-technologie. Later werkte hij aan het architectuur- en circuitontwerp van RISC-microcontrollers en de systeemsimulatoren en compilers hiervoor.

In 1990 verhuisde hij terug naar Nederland en begon te werken in de groep radio- en datatransmissiesystemen van Philips Research in Eindhoven, waar hij 10 jaar lang als wetenschappelijk medewerker en clusterleider onderzoek deed naar het ontwerp van RF front ends voor telecommunicatiesystemen. Projecten in deze periode omvatten onder andere front ends voor draadloze telefoons, semafoons, en draadloze netwerken. Ofschoon Philips Research in Eindhoven (het ‘Nat.Lab.’) een veel grotere organisatie was dan Philips Research Labs in Sunnyvale, waren er ook hier weer veel mogelijkheden voor interactie met andere research- groepen, in het bijzonder in de IC-technologiesector.

Veel van het werk waarop dit proefschrift is gebaseerd werd uitgevoerd in deze periode.

In 2000 ging hij werken voor Philips Semiconductors als development lab manager bij het RF-business development center in Tokyo, Japan, waar W-CDMA en Bluetooth producten werden ontwikkeld, en later als RF domein manager op het systeemlaboratorium in Eindhoven. Hij werkt nu als program manager en groepsleider RF op het Advanced Systems Lab in Eindhoven, met een speciale interesse in RF-platformen.