Circuit Design for RF Transceivers

Domine Leenaerts, Johan van der Tang
and Cicero Vaucher

CIRCUIT DESIGN FOR RF TRANSCEIVERS

CIRCUIT DESIGN
FOR
RF TRANSCEIVERS

By

Domine Leenaerts
Philips Research Laboratories Eindhoven

Johan van der Tang
Eindhoven University of Technology

and

Cicero S. Vaucher
Philips Research Laboratories Eindhoven

EEL 01|479

To Lisanne, Nienke, and Viviane

Contents

Preface

One of the key parts in a mobile telecommunication terminal is the transceiver. The term transceiver stems from the words transmitter and receiver. These words refer to the main task of a transceiver. In the context of a mobile telecommunication terminal, the receiver transforms the signals coming from the antenna into signals which can then be converted into the digital domain. The transmitter converts the analog version of the digital data stream at baseband into a signal at radio frequencies, and delivers this signal to the antenna with a certain amount of power.

Radio transceivers have been around since the 1900s, with the invention of AM and later FM radio broadcasting. In the 1920s, pioneers like Armstrong developed transceiver concepts which we still use today. The frequency bands ranged from several kHz up to a few 100 MHz. Transceivers for mobile telecommunication started to appear in the 1980s, with the development of DECT and GSM standards. These and other telecommunication standards for mobile telephony used instead radio frequencies between 800 MHz and 3 GHz. Another difference between these transceivers and those of previous generations is that they were integrated on silicon, instead of being made with discrete components. The first integrated transceivers were designed in bipolar processes, eventually combined with GaAs technology. The transceiver itself consisted of several ICs.

With the demand for higher data rates, attempts were made to develop wireless data standards using concepts similar to those used in the successful mobile telecommunication standards. To achieve high data rates, the radio frequency was increased, resulting in the 5 GHz carrier frequency for HiPerLAN/2 (or IEEE 802.11*a*) with data rates up to 54 Mb/s. Another wireless standard is Bluetooth (or IEEE 802.11*b*) where the primary goal was to obtain a standard which can be produced at very low cost.

Due to pressure from the market economy, the trend in radio frequency design is to integrate the complete transceiver (except, possibly, for the power amplifier) on a single substrate as a multi-chip module, or even on a single die. This integration is not simple, due to the complexity of the system, its technical specifications and the need for good components at radio frequencies. Although the active devices currently have RF capabilities, this is not necessarily true for the integrated passive components such as inductors, varactors, bond pads and
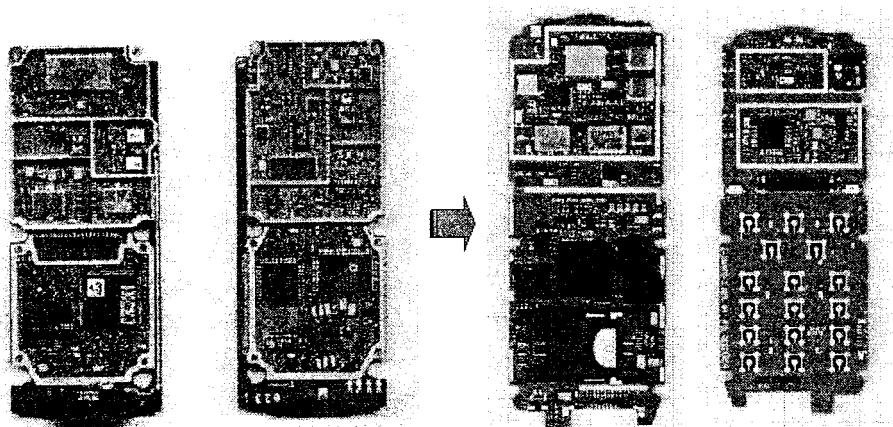
Figure 1. Mono band GSM in 1994 (left) and dual band GSM (right) in 1999.

electrostatic protection devices. Still, this integration trend achieved impressive results, as can be seen in Figure 1. The mobile terminal for mono band GSM operated at 880 to 910 MHz, and consisted of 270 components. Five years later, the dual band solution (900 and 1800 MHz bands) entered the market with only 130 components. A reduction of 50% in discrete components, 50% in PCB area, and more than 60% in RF PCB area had been achieved.

Modern transceivers are built up around a few basic building blocks, namely amplifiers, filters, mixers and oscillators. A frequency synthesizer, to generate the correct local oscillator frequency, completes the transceiver. With these building blocks several architectures can be realized.

In a single-conversion technique, a single local oscillator frequency is used for down-conversion of the RF signals. To circumvent the image rejection problem a dual-conversion architecture can be employed. Two local oscillators are then used; the first one to take care of the image rejection issue and the second one to ease the channel selection problem.

With the use of digital computing power on chip, complex (de-)modulation is possible, leading to quadrature up- and down-conversion architectures. A commonly used transceiver architecture is the direct-conversion architecture. The intermediate frequency is set to zero, implying that the desired signal is its own image. The image rejection problem is therefore eliminated, in first order. Also near zero-IF concepts are used, where the frequency difference between the local oscillator and the desired frequency is close to the channel bandwidth. The (near) zero-IF architecture can be found in, for example GSM, DECT and, wireless LAN front ends.

SCOPE OF THE BOOK

Chap. 4

Chap. 3

Chap. 2

LNA

I/Q filter

ADC

demodulator

duplexer

PA

filter

DAC

modulator

Chap. 5

Chap. 6

Chap. 7

VCO

PLL
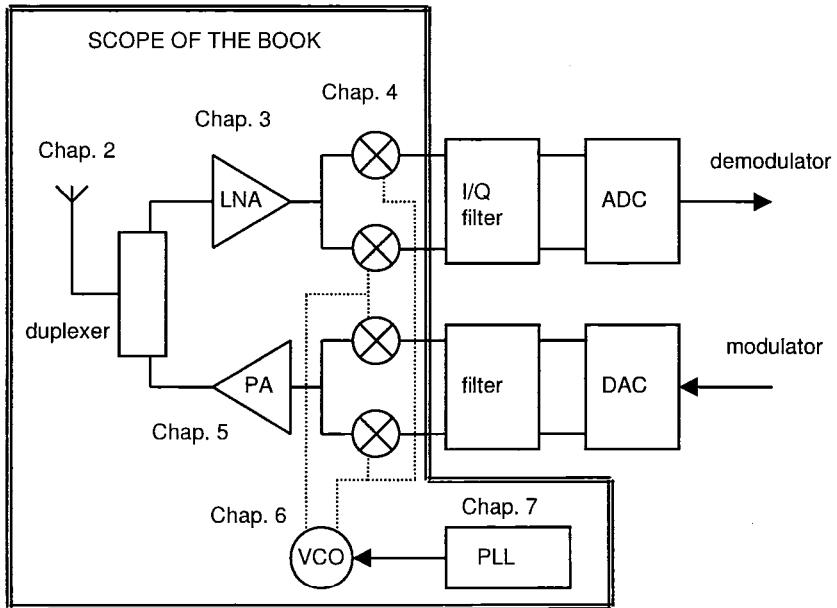
Figure 2. (Near) Zero-IF transceiver architecture. The architecture includes analog-to-digital and digital-to-analog converters.

An example of a (near) zero-IF transceiver is depicted in Figure 2. After the duplexer, a low noise amplifier (LNA) first amplifies the signals in the receiver path. Then quadrature mixing is performed to down-convert the RF signals to an IF frequency at (near) zero Hertz. Two quadrature (90° out-of-phase) signals from the local oscillator (VCO) are needed, for the realization of a (effective) mixing operation with a single positive frequency. Intermediate frequency filtering can be performed to attenuate adjacent and non-adjacent channels. The resulting complex base band signals are then digitized in a (quadrature) analog-to-digital converter (ADC) before demodulation can be performed. At the transmit side, the signals are transformed into the analogue domain by a digital-to-analog converter (DAC). The resulting baseband signals are still complex, and are by means of quadrature up-conversion converted to a real signal at the radio frequency. Then the power amplifier (PA) boosts the signals towards the antenna at the required power transmit level. A phase-locked loop (PLL) is needed to generate a stable and correct RF frequency from a reference oscillator, most often a crystal oscillator. This book will discuss the design of the circuits needed to built a RF transceiver like the one in Figure 2. Filters, data converters, and digital (de-)modulation are however outside the scope of this work.

The contents of this book are based on ongoing research activities in the Integrated Transceivers department at Philips Research Laboratories Eindhoven, The Netherlands. Our primary goal is to find solutions to problems which arise when a radio frequency transceiver is integrated on a single die. Close co-operation with researchers in the IC technology groups is required, since the current designs are continually pushing the limits of what is possible in a particular technology. Reflections of these co-operations will be found in this book. We have assumed that the reader has a basic knowledge of analog RF design. This book should be seen as a follow-up of the university text books on RF analog design. The chapters are summarized below.

Chapter 1 contains detailed discussions of the basic principles of RF design and commonly-used RF terminology. RF designers need to understand the limits of the used technology with respect to the active and passive components. Basic active devices terminology will be discussed, namely cut-off frequency and maximum oscillation frequency. We will then cover several technologies, ranging from standard bipolar, RF CMOS, to advanced silicon technologies such as Silicon-On-Anything. The chapter concludes with a detailed discussion of the RF performance of passive elements.

For RF applications, bringing signals onto or off the silicon is not trivial. The off-chip antenna plays an important role, as it can be considered the signal source for the receiver and the load for the transmitter. The bond wires and bond pads may affect the RF performance of the signals and circuitry on the chip. Electrostatic discharge protection devices also heavily influence the RF performance. These topics will be discussed in Chapter 2, together with a study of transmission lines.

Chapter 3 discusses the design of low noise amplifiers. With the help of several design steps, the reader experiences the problems arising when realizing this circuit in CMOS or bipolar technologies.

Mixers and, in particular, their noise behavior are currently receiving a great deal of attention in the literature. Active and passive mixers will be discussed in Chapter 4.

Integration of power amplifiers with 20 dBm or more delivered power in silicon technology is an ongoing research topic, and some working examples have recently been presented at conferences. Temperature stabilization and ruggedness are a few of the problems incurred when designing this type of circuit. These problems will be discussed in Chapter 5.

Chapter 6 will treat the design of voltage controlled oscillators for RF applications; do we use RC-oscillators or LC based oscillators and how do we generate quadrature signals? All of these topics, including the phase noise problem, will be highlighted.

Finally Chapter 7 will make use of the oscillators to realize frequency synthesizers for radio frequencies. Spectral purity performance will be addressed, in addition to the design of low-power frequency dividers and high speed phase-frequency detector/charge-pump combinations.

Many books, papers and internal technical notes from Philips Research contributed in making the presented material state-of-the-art. The authors do not pretend that this book is complete, however; each chapter could provide enough discussion for a book in itself. The authors would therefore want to apologize for any imperfections.

Acknowledgements

GLOSSARY

Symbol	Description	Unit
$\alpha_{lf}(f_m)$	relative magnitude of the phase noise due to loop filter elements	
a_{sp}	relative amplitude of a spurious signal with respect to the carrier	dBc
A	area	m^2
A_e	amplitude error	
A_p	power conversion gain	
A_v	voltage gain	
A_{LO}	amplitude of the carrier signal	V
A_{sp}	amplitude of a spurious signal	V
b	ratio of the time constants of the loop filter τ_2/τ_3	
β	small-signal common-emitter current gain	
$\beta(j\omega)$	transfer function in oscillator feedback model	
C	capacitance	F
C_1, C_2	capacitances of the loop filter	F
C_{db}	drain-bulk capacitance	F
C_{fixed}	fixed capacitance in resonator circuit	F
C_{gd}	gate-drain capacitance	F
C_{gg}	intrinsic gate capacitance	F
C_{gs}	gate-source capacitance	F
C_{js}	collector-substrate junction capacitance	F
C_{junc}	drain junction capacitance	F
$CNR(f_m)$	carrier to noise ratio at offset frequency f_m	dBc/Hz
CNR_{norm}	normalized carrier to noise ratio	dBc/Hz
C_μ	collector-base capacitance	F
C_p	capacitance in parallel resonator circuit	F
C_{par}	parasitic capacitance	F
C_π	emitter-base capacitance	F
C_s	capacitance with series resistance	F
C_{sb}	source-bulk capacitance	F
CP_{1dB}	1−dB compression point (output referred)	dB
δ_{cp}	duty-cycle of the output pulse of a charge-pump	
E_g	bandgap energy	eV
ϵ	dielectric constant	F/m
ϵ_{ox}	oxide permittivity ($= 3.45 \cdot 10^{-15}$ F/m)	F/m
ϵ_0	vacuum permittivity ($= 8.85 \cdot 10^{-16}$ F/m)	F/m
η_0	characteristic impedance in free space ($= 377\ \Omega$)	Ω

$\phi_d(f_m)$	rms phase noise power density of main divider	rad/$\sqrt{\text{Hz}}$
ϕ_e	phase error	rad
$\phi_{eq}^2(f_m)$	equivalent synthesizer phase noise floor at the input of the phase detector	rad^2/Hz
$\phi_{lf}^2(f_m)$	open-loop phase noise power density generated by the loop filter elements	rad^2/Hz
ϕ_m	phase margin (radians in equations, degrees in figures)	
$\phi_o^2(f_m)$	phase noise power density of the PLL output signal	rad^2/Hz
$\phi_{olp}^2(f_m)$	"low-pass" phase noise power component of $\phi_o^2(f_m)$	rad^2/Hz
$\phi_{ohp}^2(f_m)$	"high-pass" phase noise power component of $\phi_o^2(f_m)$	rad^2/Hz
$\phi_{pd}(f_m)$	rms phase noise power density of phase frequency detector	rad/$\sqrt{\text{Hz}}$
$\phi_{ref}(f_m)$	rms phase noise power density of reference divider	rad/$\sqrt{\text{Hz}}$
$\phi_{vco}(f_m)$	rms phase noise power density of free-running VCO	rad/$\sqrt{\text{Hz}}$
$\phi_x(f_m)$	rms phase noise power density of crystal oscillator	rad/$\sqrt{\text{Hz}}$
f	frequency	Hz
δf	residual FM	Hz
Δf	peak frequency deviation of a carrier	Hz
f_a	available bandwidth	Hz
f_c	open-loop bandwidth, 0 dB cross-over frequency	Hz
f_{center}	center frequency	Hz
f_{div}	frequency of the signal at the output of a frequency divider	Hz
f_m	fourier frequency (offset, modulation or baseband frequency)	Hz
f_{max}	maximum oscillation frequency	Hz
f_{min}	minimum oscillation frequency	Hz
f_{osc}	oscillation frequency	Hz
f_{out}	output limited frequency	Hz
f_{ref}	operation frequency of the PFD	Hz
f_v	input limited frequency	Hz
f_T	cut-off frequency	Hz
f_{xtal}	frequency of crystal oscillator	Hz
F	noise factor	
F_{out}	output frequency of a PLL	Hz
FOM_n	figure of merit related to parameter n	
Γ	reflection coefficient	
G	gain	

$G(s)$	open-loop transfer function of a PLL	
G_a	available power gain	
g_m	transconductance	A/V
$g_{m-startup}$	minimum transconductance for oscillator startup	
G_{max}	maximum power gain	
G_p	(delivered) power gain	
G_T	transducer gain	
$H(s)$	closed-loop transfer function of a PLL	
$H_f(j\omega)$	transfer function in oscillator feedback model	
$i_{carrier-rms}$	rms carrier current	A
I	current	A
I_{cp}	amplitude of the output current of a charge pump	A
I_{leak}	leakage current in the tuning line of the VCO	A
I_{out}	instantaneous output current of a charge pump	A
I_{tail}	tail current of a differential pair	A
I_{tune}	tuning current	A
IP_n	n-th order intercept point	dB
IIP_n	n-th order input referred intercept point	dB
IM_n	n-th order inter modulation	dB
$\overline{I_n^2}$	mean square noise current over a bandwidth Δf	A^2/Hz
$\overline{i_{n\,Bipolar}^2}$	bipolar output noise current in 1 Hz	A^2/Hz
$\overline{i_{n\,MOS}^2}$	MOS output noise current in 1 Hz	A^2/Hz
$\overline{i_{noise-out}^2}$	closed loop mean square noise output current in 1 Hz	A^2/Hz
k	Boltzmann's constant ($=1.38 \cdot 10^{-23}$ J/K)	J/K
k_p	propagation constant	
K_{cco}	CCO tuning constant	Hz/A
$K_{tail-rms}$	rms sensitivity of f_{osc} for tail current variations	Hz/A
K_{pd}	gain of PFD/CP combination	A/rad
K_{vco}	VCO tuning constant	Hz/V
$\mathcal{L}(f_m)$	SSB phase noise to carrier ratio at offset freq. f_m	dBc/Hz
L	inductance	H
L	length of transistor	m
λ	wavelength	m
L_{finger}	finger length of transistor	m
L_p	inductance in parallel resonator circuit	H
L_s	inductance with series resistance	H
n	an integer	
N	number of oscillator stages	
N	main divider division ratio, integer	

NBW	noise bandwidth	Hz
NF	noise figure	dB
OIP_n	n-th order output referred intercept point	dB
P_{av}	available power	W
$P_{carrier}$	carrier power	W
P_{DC}	DC input power	W
P_{del}	delivered power	W
P_{diss}	power dissipation	W
P_{in}	input power	W
P_{noise}	source resistance noise power	dBm/Hz
P_o	output power	W
P_{RF}	RF output power	W
P_t	transmitted power	W
θ	phase shift	rad
$\theta(t)$	stochastic phase variable	rad
θ_p	peak phase deviation of phase modulation	rad
q	charge of the electron ($=1.602 \cdot 10^{-19}$ C)	C
Q	quality factor	
Q_{cs}	capacitor quality factor	
Q_{LC}	quality factor of parallel resonator circuit	
Q_{ls}	inductor quality factor	
Q_{RC}	RC oscillator quality factor	
R	reference divider division ratio, integer	
R	resistance	Ω
R_1	resistance used in the PLL loop filter	Ω
$Ratio_{var}$	varactor capacitance ratio	
r_b	base resistance	Ω
R_{eff}	effective bulk resistance	Ω
R_g	gate resistance (due to poly)	Ω
R_i	input resistance	Ω
R_{in}	gate resistance due to non-quasi static behavior	Ω
R_L	load resistance	Ω
R_o	output resistance	Ω
R_p	loss resistance in parallel resonator circuit	Ω
R_r	radiation resistance	Ω
R_s	source resistance	Ω
R_{series}	series resistance	Ω
r_π	base-emitter resistance	Ω
$R_\square$	sheet resistance	$\Omega/\square$
ρ	resistivity of metal	$\Omega\cdot$cm

$S(f)$	spectral density over 1 Ohm resistor	V^2/Hz
s	Laplace transform complex variable ($s = \sigma + j\omega$)	
s-parameter	two-port scatter parameters ($s_{11}, s_{12}, s_{21}, s_{22}$)	
$SFDR$	spurious free dynamic range	dB
SNR	signal-to-noise ratio	dB
T	(absolute) temperature	K
$T_{hp}(s)$	high-pass transfer function	
τ	base transit time	s
τ_{delay}	large signal delay	s
τ_2, τ_3	time constants of the loop filter	s
$\tau_{3,sp}$	time constant determined from spectral purity considerations	s
v	normalized frequency deviation from ω_{osc}	
$v_{nf}(f_m)$	rms voltage noise density originated in the loop filter	$V/\sqrt{Hz}$
V	voltage	V
$V_{carrier}$	peak carrier amplitude	V
V_p	peak voltage	V
V_{pp}	peak-to-peak voltage	V
$V_{mismatch}$	magnitude of the ripple voltage due to mismatch in the CP current sources	V
$V_{reverse}$	reverse varactor voltage	V
V_{ripple}	ripple voltage at the VCO tuning line	V
V_{rms}	rms voltage	V
$\overline{V_n^2}$	mean square noise voltage over a bandwidth Δf	V^2/Hz
V_{tune}	voltage at the tuning input of a VCO	V
W	width of transistor	m
y-parameter	two-port admittance parameters ($y_{11}, y_{12}, y_{21}, y_{22}$)	
Z	impedance	Ω
$Z_f(s)$	transimpedance of the loop filter	Ω
Z_i	input impedance	Ω
Z_L	load impedance	Ω
Z_o	output impedance	Ω
Z_{opt}	optimal noise impedance	Ω
Z_s	source impedance	Ω
Z_0	characteristic impedance	Ω
z-parameter	two-port impedance parameters ($z_{11}, z_{12}, z_{21}, z_{22}$)	
ω	angular frequency	rad/s
ω_c	open-loop bandwidth $\omega_c = 2\pi f_c$	rad/s
ω_m	angular modulation frequency	rad/s
ω_{osc}	angular oscillation frequency	rad/s

Abbreviations

AC	Alternating Current
ACPR	Adjacent Channel Power Ratio
A/D	Analog/Digital
AGC	Automatic Gain Control
AM	Amplitude Modulation
CCO	Current Controlled oscillator
(C)MOS	(Complementary) Metal Oxide Semiconductor
CNR	Carrier to Noise Ratio
CP	Charge-Pump
CP	Compression Point
dB	Decibel
dBc	dB relative to the carrier
DC	Direct Current
DECT	Digital European Cordless Telephone
D-FF	D-type Flip-flop
DPO	Double Poly
DR	Dynamic Range
DRO	Dielectric Resonator Oscillator
DSB	Double-Sideband
EDGE	Enhanced Data Rates for GSM Evolution
EMC	Electromagnetic Compatibility
ESD	Electro Static Discharge
FDD	Frequency-Division Duplex
FM	Frequency Modulation
FOM	Figure of Merit
ggNMOST	Grounded Gate NMOS Transistor
GSM	Global System for Mobile communication
HBM	Human Body Model
IC	Integrated Circuit
IF	Intermediate Frequency
IP	Intercept Point
IIP	Input Referred Intercept Point
IM	Inter Modulation
IRR	Image Rejection Ratio
I/O	Input/Output
I/Q	In-phase/Quadrature
LNA	Low Noise Amplifier
LO	Local Oscillator
LPF	Low-Pass Filter

LTI	Linear Time Invariant
NQS	Non-Quasi Static
OIP	Output Referred Intercept Point
PA	Power Amplifier
PAE	Power Added Efficiency
PFD	Phase-Frequency Detector
PLL	Phase-Locked Loop
PM	Phase Modulation
QAM	Quadrature Amplitude Modulation
rms	Root-Mean-Squared
RF	Radio Frequency
RX-band	Receive Band
SAW	Surface Acoustic Wave
SFDR	Spurious-Free Dynamic Range
SNR	Signal-to-Noise Ratio
SOA	Silicon-on-Anything
SSB	Single-Sideband
TDD	Time-Division Duplex
TX-band	Transmit Band
UMTS	Universal Mobile Telecommunications System
VCO	Voltage Controlled Oscillator
VHF	Very High Frequency

Chapter 1

RF Design: Concepts and Technology

Radio Frequency (RF) circuitry very often operates at the boundaries of achievable performance within a particular process technology. This means that the designers not only need analog circuit design knowledge but must also have a deep understanding of the device technology. The performance of many RF designs is primarily determined by the quality of the passives used. High quality inductors are certainly difficult to realize in silicon. This chapter starts with a review of commonly used RF terminology and concepts. The second part of this chapter treats technology-related issues, with a particular emphasis on passives.

1.1 RF SPECIFICATIONS

As in many other design domains, RF and microwave designers[1] specify the behavior of their circuits using terminology which originates from signal theory. In contrast to low and intermediate frequency analog design, however, the specifications in RF design are usually related to powers instead of voltages. In addition, the small signal behavior is described using s-parameters rather than using z-parameters. The result is that analog designers and microwave designers often misunderstand each other, even though they are discussing the same behavior. Some commonly used terminology is discussed in this section.

---

[1] Strictly speaking, the microwave frequency band is defined from 1 GHz to 30 GHz. Designers working with these frequencies are therefore referred to as "microwave designers". However, many telecommunication systems (i.e. radios) operate in the 0.9 to 3 GHz range, and this frequency band is often mistakenly referred to as the radio frequency (RF) band, hence the name "RF designer". Throughout this book, we will often interchange the names microwave designer and RF designer.
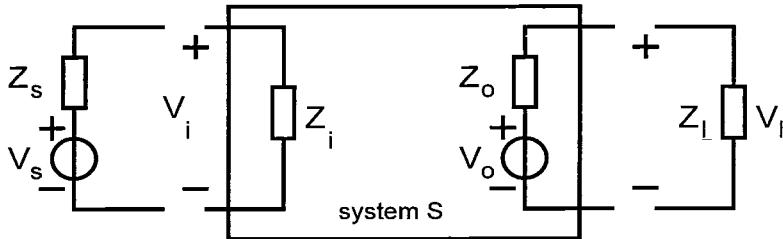
Figure 1.1. Time-invariant system S with load impedance Z_L. System S has an input impedance Z_i and an output impedance Z_o. The system is driven by a voltage source V_s with internal impedance Z_s.

1.1.1 Gain

One issue which often leads to misunderstandings between analog designers and microwave specialists is the specification of *gain*. Analog designers normally use the concept of *voltage gain*, while RF designers use a broad range of terminology, related to signal powers.

Consider a linear, memoryless, time-invariant system S with an input impedance Z_i, and an output impedance Z_o. This system is driven by a source with an impedance Z_s, and the system is loaded with an impedance Z_l. This situation is depicted in Figure 1.1 and will be used to explain several concepts of gain.

Assume that the input impedance is much higher than the source impedance and, in the extreme case, that $Z_i \to \infty$. This means that no current will flow into system S, and the voltage of the source, V_s, is equal to the voltage at the input of S, V_i. Similarly, assuming $Z_o \to 0$ will give a voltage V_l across the load equal to the output voltage of S, V_o. This latter voltage is equal to the amplified voltage at the input of S,

$$A_v = \frac{V_o}{V_i} \to \frac{V_l}{V_s} \tag{1.1}$$

Here A_v is called the (unloaded) *voltage gain*, and is normally used by analog designers.

Let us focus on the input of system S and assume that Z_i is far from infinity[2] and hence $V_s \neq V_i$. Then a current will also flow into system S, and we can define types of power. To achieve maximum *power transfer*, the load impedance must be the conjugate of the source impedance. Suppose that $Z_s \neq Z_i^*$, where

---

[2]For RF frequencies, the input impedance of bipolar and MOS transistors is definitely small compared to the value at DC or at low frequencies.

the asterix denotes conjugation. RF designers call this an input impedance mismatch. Consequently, the power *available* from the source ($P_{s,av}$) is not the same as the power *delivered* at the input ($P_{i,del}$). Let us consider this in more detail and assume that the impedances in Figure 1.1 can be treated as resistors. The input voltage V_i is derived as

$$V_i = \frac{R_i}{R_i + R_s} V_s \tag{1.2}$$

and consequently

$$P_{i,del} = \frac{R_i}{(R_i + R_s)^2} V_s^2 \tag{1.3}$$

The power $P_{s,av}$ is the power that the source V_s would deliver to a conjugate matched circuit, and can be expressed as

$$P_{s,av} = \frac{V_s^2}{4R_s} \tag{1.4}$$

We can also define the available power at the output $P_{o,av}$, i.e., the power that V_o would deliver to a conjugate matched circuit, as

$$P_{o,av} = \frac{V_o^2}{4R_o} = \left(\frac{R_i}{R_i + R_s}\right)^2 V_s^2 A_v^2 \frac{1}{4R_o} \tag{1.5}$$

The mismatch between R_i and R_s influences the *available gain* G_a,

$$G_a = \frac{P_{o,av}}{P_{s,av}} = \left(\frac{R_i}{R_i + R_s}\right)^2 A_v^2 \frac{R_s}{R_o} < \frac{P_{o,av}}{P_{i,del}} \tag{1.6}$$

Obviously, for a perfect input impedance match $P_{s,av} = P_{i,del}$ and the available gain increases. In a similar way, we can consider an impedance mismatch between Z_o and Z_L. The available power at the output of system S, $P_{o,av}$ is higher than the delivered power at the load, $P_{l,del}$ which is equal to

$$P_{l,del} = \frac{V_l^2}{R_L} = V_i^2 A_v^2 \frac{R_L}{(R_L + R_o)^2}, \tag{1.7}$$

and the power delivered at the input is equal to

$$P_{i,del} = \frac{V_i^2}{R_i} \tag{1.8}$$

The resulting gain, while assuming a perfect input match, is called (*delivered*) *power gain* and is defined as

$$G_p = \frac{P_{l,del}}{P_{i,del}} = A_v^2 \frac{R_L R_i}{(R_L + R_o)^2} < \frac{P_{o,av}}{P_{i,del}} \tag{1.9}$$

We obtain the *maximum gain* for system S when we have perfect input and output impedance matches:

$$G_{\max} = \frac{P_{o,av}}{P_{i,del}} = \frac{P_{l,del}}{P_{s,av}}$$ (1.10)

$$Z_s = Z_i, Z_o = Z_L$$ (1.11)

and therefore yields

$$G_{\max} = A_v^2 \frac{R_i}{4R_o}$$ (1.12)

which follows from (1.6) with $R_i = R_s$ and from (1.9) with $R_L = R_o$. Similarly, the lowest gain, called the *transducer gain* is obtained if we have a mismatch on both input and output,

$$G_t = \frac{P_{l,del}}{P_{s,av}} = \left(\frac{R_i}{R_i + R_s}\right)^2 A_v^2 \frac{4R_L R_s}{(R_L + R_o)^2}$$ (1.13)

However, this transducer gain is the actual gain for system S when driven by a source V_s and loaded by an impedance Z_L. Note that from (1.13) we indeed obtain G_a for $R_L = R_o$ and G_p for $R_i = R_s$.

We end up with five possible gain definitions and it is therefore important to specify the used gain.

There are two main reasons for power matching in RF applications. One reason is when reflections of signals occur. Signals are reflected when the distance between the source and load is larger than $\frac{1}{4}\lambda$ where λ is the wavelength, i.e. $\lambda = \frac{c}{f}$ where c is the speed of light and f is the frequency of interest. In free space, for 1 GHz, λ is approximately equal to 30 cm, but under realistic conditions it can reduce to a few centimeters. Hence $\frac{1}{4}\lambda$ can only be a few millimeters in integrated RF applications. The distance between the antenna output (i.e. the source) and the input of the low noise amplifier (LNA) (i.e. the load) is normally large and reflections can therefore take place. A general rule might be that reflections are only of interest for the "on-off" chip connections; connections between blocks which are both on-chip will not be deteriorated by reflections, due to the small size of the IC.

The second reason to focus on power matching in RF is when input impedance mismatch drastically influences the performance of the circuit. This happens at off-chip antenna filters, for instance. They normally are designed for 50 Ω input/output impedance. If the on-chip LNA has an input impedance other than 50 Ω, the filter characteristic will change and hence the performance of the

overall system will change. In such cases, it is important to have a 50 Ω input impedance match.

Therefore, even for RF applications it is not always necessary to focus on optimal power transfer.

Microwave designers normally use *scatter (s)-parameters* to define the small signal performance of their circuits [1]. These scatter parameters are based on: voltage waves entering system S, i.e. the incident waves and waves leaving S at the same port, the scattered waves. The ratios between the powers of these incident and scattered waves at the input and output are the s-parameters. Let us define E_{i1} as the incident voltage wave and E_{r1} as the scattered voltage wave for port 1 of two-port system S. In a similar way we define E_{i2} and E_{r2} for port 2. If we normalize these waves by the square root of Z_0, i.e. $a_j = E_{ij}/\sqrt{Z_0}$ and $b_j = E_{rj}/\sqrt{Z_0}$ then the two-port relations may be written as

$$b_1 = s_{11}a_1 + s_{12}a_2 \tag{1.14}$$
$$b_2 = s_{21}a_1 + s_{22}a_2 \tag{1.15}$$

The five gain definitions can all be formulated in terms of s-parameters and the reflection coefficient Γ, defined as

$$\Gamma = \frac{Z - Z_0}{Z + Z_0} \tag{1.16}$$

where Z_0 is the characteristic impedance. Equation (1.16) expresses the quality of matching and $20 \log |\Gamma|$ is defined as the "return loss"[3]. In fact, it means that the load is reflecting back some of the power delivered by the source. You can also view it as that the load not taking up all of the available power from the source. As an example, assume $Z = R + \Delta R$ where S is a 50 Ω system. For an impedance variation between 20% and 30%, (1.16) results in a return loss of -15 to -20 dB, i.e. at least 1% of the power is reflected back. There will be no reflection for a perfect match. The formulations for the gains, expressed

---

[3] Note that $s_{11} = \Gamma_1$ and $s_{22} = \Gamma_2$.

in s-parameters and reflection coefficients, are given by

$$A_v = \frac{s_{21}(1 + \Gamma_l)}{(1 - s_{22}\Gamma_l) + s_{11}(1 - s_{22}\Gamma_l) + s_{21}\Gamma_l s_{12}} \tag{1.17}$$

$$G_a = \frac{1 - |\Gamma_s|^2}{|1 - s_{11}\Gamma_s|^2}|s_{21}|^2 \frac{1}{1 - |\Gamma_{out}|^2} \tag{1.18}$$

$$G_p = \frac{1 - |\Gamma_l|^2}{|1 - s_{22}\Gamma_l|^2}|s_{21}|^2 \frac{1}{1 - |\Gamma_{in}|^2} \tag{1.19}$$

$$G_t = \frac{1 - |\Gamma_s|^2}{|1 - s_{11}\Gamma_s|^2}|s_{21}|^2 \frac{1 - |\Gamma_l|^2}{1 - |\Gamma_{out}\Gamma_l|^2} \tag{1.20}$$

1.1.2 Noise

One of the most important parameters in analog and RF circuit design is the ratio between the signal power to the total noise power, i.e. the *signal-to-noise ratio (SNR)*. The *noise factor (F)*, a measure normally used in RF design, is strongly related to this definition,

$$F = \frac{SNR_{in}}{SNR_{out}} \tag{1.21}$$

The noise factor[4] is a ratio between the SNR at the input and the SNR at the output of a system. Clearly, (1.21) can be rewritten as

$$F = \frac{P_{in}N_{out}}{G_{av}P_{in}N_{in}} = \frac{N_{out}}{G_{av}N_{in}}, \tag{1.22}$$

showing that the noise factor is the ratio between the total noise at the output of the system N_{out} to the total noise at the output of the system due to the noise at the input of the system N_{in}. The noise factor is a measure of how much the SNR degrades as the signal passes through a system.

The noise factor can be calculated by referring the noise sources of the system to the input of the system. If the system can be considered as a two-port system, the noise of the system can be modeled by two input noise generators, a parallel current source $\overline{I_n^2}$ and a series voltage source $\overline{V_n^2}$. Here we define $\overline{X^2}$ as the mean square of X over a certain bandwidth Δf. This stems from the formal definition of the power spectral density (PSD) of X, $S_X(f)$

$$S_X(f) = \lim_{T \to \infty} \frac{\overline{|X_T(f)|^2}}{T}$$

---

[4]The noise figure NF is normally defined as $10\log(\text{noise factor})$.
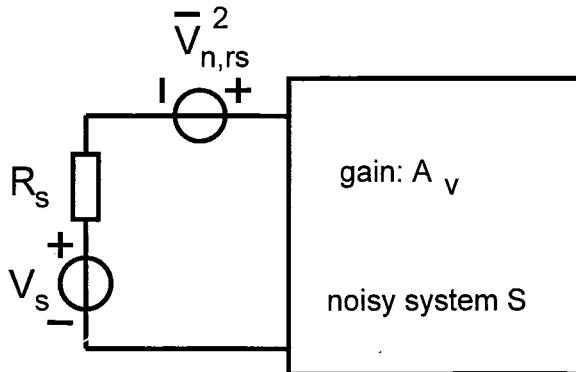
Figure 1.2. Noisy system S driven by voltage source V_s. The internal resistance R_s of V_s produces noise, represented by $\overline{v_{n,rs}}$. The voltage gain of S is equal to A_v.

where $X_T(f)$ is the one-sided Fourier transformation of signal $x(t)$. Normally, the one-sided PSD is used in circuit design, so that the thermal noise power spectral density $S_X(f) = 4kTR$ for a resistor, where k is the Boltzmann constant and T is the absolute temperature. The dimension of mean square voltage per unit bandwidth stems from the fact that the power spectral density refers to the (rms) voltage across a resistor with resistance R to generate a power of $4kTR$ in a 1 Hz bandwidth. The mean square noise voltage of this resistor is then given by $\overline{V_n^2} = 4kTR \cdot \Delta f$.

Again we can see the difference between analog and RF design. Assuming an almost infinite input impedance in analog design results in only an input referred voltage noise source. This assumption no longer holds at radio frequencies and, consequently, the input referred current noise source also has to be taken into account. It is important to keep in mind that these two noise sources are correlated. The noise influence of these sources to the total system can only be found after adding their influences.

We can now use these input referred noise sources to calculate the noise factor. Assume a voltage source V_s with a resistance R_s driving a noisy system with voltage gain A_v, as depicted in Figure 1.2. This source resistance produces noise, modeled as $\overline{V_{n,rs}^2}$. The noise of the system can be referred to the input, making the system itself noiseless, as depicted in Figure 1.3. The SNR_{in} can now be found as

$$SNR_{in} = \frac{V_s^2}{V_{n,rs}^2}, \tag{1.23}$$
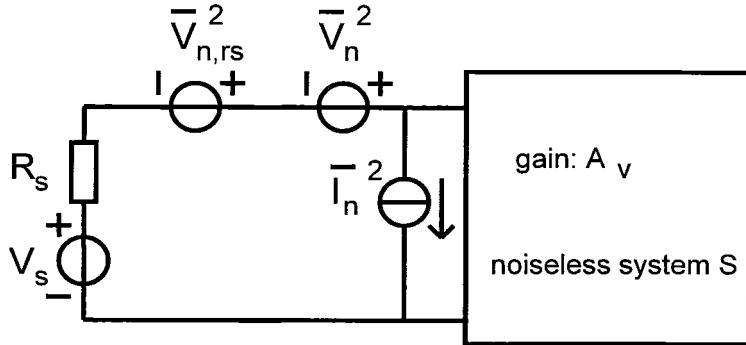
Figure 1.3.   The noise of S is now represented by the input referred noise sources. The system still has a voltage gain of A_v.

where we have assumed that all power available at the source is delivered to the input of the system. The SNR_{out} can be measured as

$$SNR_{out} = \frac{A_v^2 V_s^2}{\left(\overline{V_{n,rs}^2} + \overline{(V_n + R_s I_n)^2}\right) A_v^2} \tag{1.24}$$

Note that $\overline{V_n^2}$ and $\overline{I_n^2}$ are added before squaring, to take their correlation into account. The noise factor is now obtained by dividing (1.23) by (1.24), yielding

$$F = \frac{SNR_{in}}{SNR_{out}} = \frac{\overline{V_{n,rs}^2} + \overline{(V_n + R_s I_n)^2}}{\overline{V_{n,rs}^2}} = 1 + \frac{\overline{(V_n + R_s I_n)^2}}{\overline{V_{n,rs}^2}} \tag{1.25}$$

The noise factor is usually specified for a 1 Hz bandwidth at a particular frequency. The specification can be obtained from (1.25) by using the expression $\overline{V_{n,rs}^2} = 4kTR_s$, and is called the spot noise factor to emphasize the 1 Hz bandwidth.

Designers are not only interested in the noise factor of a single circuit, but also in the noise behavior of the total system (e.g. frontend), which is normally a cascade of circuits. Assuming we know the noise factor F_i and the available gain $G_{a,i}$ for each separate circuit i, Friis' formula gives the expression of the overall F_{total} for m circuits as a function of F_i

$$F_{total} = 1 + (F_1 - 1) + \frac{F_2 - 1}{G_{a,1}} + \frac{F_3 - 1}{G_{a,1} G_{a,2}} + \cdots + \frac{F_m - 1}{G_{a,1} G_{a,2} \cdots G_{a,m-1}} \tag{1.26}$$

We can conclude several properties from Friis' formula:

- The F_{total} is referred to the input of the first circuit.

- The F of each circuit is calculated with respect to the source impedance driving that circuit. This stems from the use of the available gain in the definition.

- The F of the first stage is dominant when the gain of each stage is reasonable. However, if the first circuit exhibits a low gain or attenuation (i.e. loss), for instance, then the noise figure of the second circuit is amplified when referred to the input of the first circuit.

As F is dependent on the source impedance, we can minimize F for a given two-port system by choosing a proper value for the source impedance. We therefore simultaneously maximize the SNR. Assume a voltage source V_s with an impedance $Z_s = R_s + jX_s$ driving a noisy system similar to Figure 1.3. The noise produced by the source impedance is modeled as $\overline{V_{n,rs}^2}$. If we define the input referred noise voltage as $\overline{V_n^2} = 4kTR_{nv} \cdot \Delta f$, and the input referred noise current as $\overline{I_n^2} = 4kTG_{ni} \cdot \Delta f$, the system can be treated as noiseless. According to (1.25) we have

$$F = \frac{\overline{V_{n,rs}^2} + \overline{(V_n + Z_s I_n)^2}}{\overline{V_{n,rs}^2}} = 1 + \frac{R_{nv}}{R_s} + \frac{|Z_s|^2}{R_s R_{ni}} \tag{1.27}$$

for the noise factor, where $R_{ni} = 1/G_{ni}$ and assuming that the two input referred noise sources are not correlated. Differentiating F to R_s yields

$$R_s = \left(R_{nv} R_{ni} + X_s^2\right)^{\frac{1}{2}},$$

and if we assume that we only have a resistive source impedance, the optimal source impedance for noise matching is found to be

$$R_{opt} = \sqrt{R_{nv} R_{ni}} \tag{1.28}$$

with

$$F_{\min} = 1 + 2\sqrt{\frac{R_{nv}}{R_{ni}}} \tag{1.29}$$

as the minimum noise factor.

Note that the optimal source resistance for minimum noise does not coincide with that for maximum power transfer.

The derivation above holds only when the noise sources are not correlated. If they are completely correlated, $\overline{I_n^2}$ is completely determined by $\overline{V_n^2}$. For each

spectral component within $\overline{I_n^2}$, $I_n = Y_c V_n$ where Y_c describes the correlation. If the correlation is only partial, we have $I_n = Y_c V_n + Y_n V_n$, where Y_n describes the uncorrelation. Using these expressions we can again derive a relationship for the optimal source impedance. For the general case, the optimal noise match is achieved for $Z_s = Z_{opt}$, while the optimal power transfer is obtained for $Z_s = Z_{in}^*$. The optimal noise impedance for a given circuit is not generally easy to calculate.

Noise matching and power matching are in principle not obtained at the same time. Any deviation of the source impedance from this optimal value will result in an increase of the noise factor of the overall system. In the RF community, this deviation is normally expressed in the reflection parameters, yielding

$$F = F_{\min} + \frac{4 r_n \left| \Gamma_s - \Gamma_{s,opt} \right|^2}{\left| 1 + \Gamma_{s,opt} \right|^2 \left(1 - \left| \Gamma_s \right|^2 \right)} \tag{1.30}$$

where $r_n = R_s / Z_o$ is the normalized source resistance.

One last design issue regarding noise is the following rule of thumb used by analog designers. Assume that the input referred noise of system S can be represented by only a noise voltage source. Then (1.27) reduces to

$$F = 1 + \frac{R_{nv}}{R_s} \tag{1.31}$$

Thus, for equal resistance values, the noise figure is 3 dB. The circuit produces the same amount of noise as the source resistance. To let the source resistance be dominant, the input referred noise resistance should be smaller than the source resistance.

1.1.3   Non-linearity

Radio frontend specifications are usually related to the non-linear behavior of the circuits from which the system is built. We can think of specifications such as gain compression, inter-modulation and blocking. We will briefly discuss these effects assuming a memoryless, time-variant system S with input $x(t)$ and output $y(t)$, with a non-linearity up to the order 3^5,

$$y(t) = \alpha x(t) + \beta x^2(t) + \gamma x^3(t) \tag{1.32}$$

---

[5]Note that effects of memory and higher order non-linearities may have a significant influence on the overall behavior of the system.

Suppose a sinusoidal signal $x(t) = A\cos(\omega t)$ is applied to system (1.32), then the amplitude of the fundamental frequency is given as

$$\alpha A + \frac{3\gamma A^3}{4} \tag{1.33}$$

from which it follows that the small signal gain α is varied with the input amplitude due to the third order non-linearity. For $\gamma < 0$, the gain $\alpha + \frac{3\gamma A^2}{4}$ is then decreasing as a function of A. The input level at which the small signal gain has dropped by 1 dB is called the *"1 dB compression point"* (CP_{1dB}) and can be derived as

$$CP_{1dB} = \sqrt{0.145 \left| \frac{\alpha}{\gamma} \right|} \tag{1.34}$$

Obviously, the non-linearities in (1.32) will also cause higher order terms of the fundamental frequency. In fully differential systems, the even-order harmonics will be attenuated, but they will never vanish due to mismatches which corrupt the symmetry.

Many interesting phenomena can be explained by assuming a two-tone input signal, i.e. $x(t) = A\cos(\omega_1 t) + B\cos(\omega_2 t)$, where the first tone represents the wanted signal and the second tone the interferer. Assuming the desired signal to be related to ω_1, then according to (1.32), the amplitude of the fundamental tone is found to be

$$\alpha A + \frac{3\gamma A^3}{4} + \frac{3\gamma A B^2}{2} \tag{1.35}$$

Assume that the desired signal is weak compared to the interferer, $A \ll B$, then with $\gamma < 0$ the small signal gain is a decreasing function of the amplitude of the interferer[6]. Consequently, the gain drops to zero for sufficiently large B, an effect referred to as *blocking*. We may even take into account a modulation of the amplitude of the interferer, for example $B(1 + m\cos(\omega_m t))\cos(\omega_2 t)$ (with $m < 1$) which will change (1.35) into

$$\alpha A + \frac{3\gamma A^3}{4} + \frac{3\gamma A B^2}{2}\left(1 + \frac{m^2}{2} + \frac{m^2}{2}\cos(2\omega_m t) + 2m\cos(\omega_m t)\right) \tag{1.36}$$

The formula tells us that the small signal gain of the weak desired signal is modulated with the amplitude modulation of the strong interferer at ω_m and $2\omega_m$. This effect is called *cross modulation*.

So far we have only investigated the influence of non linearities to the fundamental frequency. But the non linearities give also rise to mixing terms of the

---

[6]This effect is called desensitization.

frequencies, called *inter modulation (I M)*. Assuming for the two-tone input signal both components equal or almost equal in strength then the following inter modulation products can be obtained:

- second order: $\beta A B \cos((\omega_1 + \omega_2)t) + \beta A B \cos((\omega_1 - \omega_2)t)$

- third order: $\frac{3\gamma A^2 B}{4} \cos((2\omega_1 + \omega_2)t) + \frac{3\gamma A^2 B}{4} \cos((2\omega_1 - \omega_2)t)$

   $+ \frac{3\gamma A B^2}{4} \cos((\omega_1 + 2\omega_2)t) + \frac{3\gamma A B^2}{4} \cos((\omega_1 - 2\omega_2)t)$

Obviously, when $\omega_1 \approx \omega_2$ the third order inter modulation terms $(I M_3)$ again appear in the band of interest, that is, they cannot be removed by means of a filter. Therefore, signals may therefore be corrupted. The metric *third order intercept point* $(I P_3)$ has been defined as a measure of this effect. It is defined as the point where the amplitude of the fundamental (1.35) is equal to the amplitude of the third order inter modulation (assuming $A = B$),

$$|\alpha| A + \frac{9 |\gamma| A^3}{4} = \frac{3 |\gamma| A^3}{4}$$

yielding (in (peak) voltage, not in power),

$$A_{I P3} = \sqrt{\frac{4 |\alpha|}{3 |\gamma|}} \tag{1.37}$$

for the $I P_3$, assuming $\alpha \gg \frac{9|\gamma|A^2}{4}$. The output $I P_3$ is then obtained as $\alpha A_{I P3}$. For practical circuits, the assumption $\alpha \gg \frac{9|\gamma|A^2}{4}$ no longer holds, and the intercept point is often beyond the allowable input range. It is therefore found by linear extrapolation of the measured behavior for small input amplitudes. The input $I P_3$ $(I I P_3)$ can also be found from a single tone measurement. Using the $I M_3$ products and a given input power P_{in} we obtain

$$I I P_3 = \frac{\Delta P}{2} + P_{in} \tag{1.38}$$

where ΔP is the difference between the magnitudes of the fundamentals and the $I M_3$ products at the *output*. All parameters in expression (1.38) are in decibels. The output intercept point $O I P_3$ can be found to be the difference between the $I I P_3$ and the power gain G_p, i.e. $O I P_3 = G_p P_{in} + \frac{\Delta P}{2}$. Figure 1.4 shows a geometric interpretation of this expression[7].

---

[7]In general, the n-th order intercept point is related to the n-th order intermodulation term and can be found as follows. Define the angle of the fundamental amplitude line as θ_1 and the angle of the amplitude of the nth-order harmonic as θ_n. For the intercept point, we obtain $\tan(\theta_n) = n \tan(\theta_1)$ or $Q + \Delta P = nQ$ and $\Delta P = (P_1 - P_n)$ where P_1 and P_n are the powers at the output. Here, $Q = O I P_n - \Delta P$. Consequently, $O I P_n = G_p P_{in} + \frac{\Delta P}{n-1}$, $I I P_n = P_{in} + \frac{\Delta P}{n-1}$.
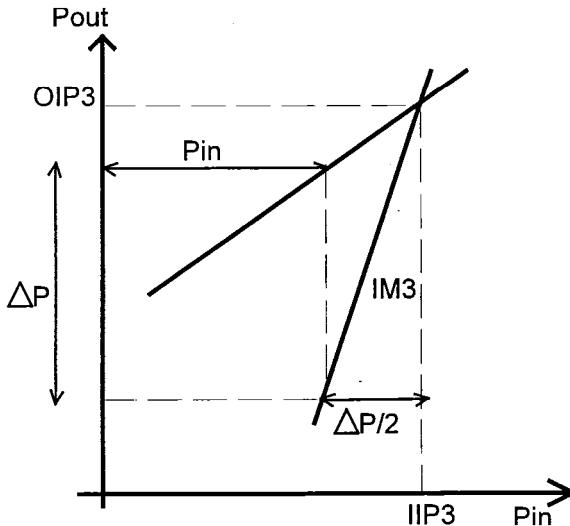
Figure 1.4.   Input IP_3 as a function of the IM_3 and input power P_{in}. Both axes are in decibels.

From (1.37) and (1.34) we can see that the difference between the 1 dB compression point and the input IP_3 is roughly -9.6 dB.

If stage (1.32) is cascaded with a second non-linear stage,

$$z(t) = \kappa y(t) + \lambda y^2(t) + \mu y^3(t)$$

the overall A_{IP3} results in

$$A_{IP3,all} = \sqrt{\frac{4}{3} \left| \frac{\alpha\kappa}{\gamma\kappa + 2\alpha\beta\lambda + \alpha^3\mu} \right|} \tag{1.39}$$

which can be approximated[8] to

$$\frac{1}{A_{IP3,all}^2} = \frac{1}{A_{IP3,1}^2} + \frac{\alpha^2}{A_{IP3,2}^2} + \frac{3\beta\lambda}{2\kappa} \tag{1.40}$$

where $A_{IP3,1}$ and $A_{IP3,2}$ represent the input IP_3 points of the first and second stages, respectively. Increasing the gain of the first stage results in a decrease of the overall IP_3; the second stage senses a large input signal thereby producing much greater IM_3 products. Equation (1.39) can be extended for more cascaded

---

[8]We have used the relationship $|x + y| \leq |x| + |y|$

stages into the general expression

$$\frac{1}{A_{IP3,all}^2} = \frac{1}{A_{IP3,1}^2} + \frac{\alpha^2}{A_{IP3,2}^2} + \frac{\alpha^2\kappa^2}{A_{IP3,3}^2} + \cdots \qquad (1.41)$$

where we should emphasize that the expression is still an approximation.

1.1.4   Sensitivity

The minimum signal level that a system can detect with an acceptable SNR is called the *sensitivity* of that system. This minimum input signal $P_{in,\min}$ is found to be the integral over the bandwidth B of the product of noise factor F, the acceptable SNR at the output, and the source resistance noise power P_{noise}, yielding

$$P_{in,\min} = B \cdot SNR \cdot P_{noise} \cdot F \qquad (1.42)$$

This expression can be simplified by assuming power matching at the input. Under this condition, P_{noise} equals kT, which for room temperature results in $P_{noise} = -174$ dBm/Hz. The *noise floor* is total integrated noise of the source available at the input of the system and can be obtained from $B \cdot P_{noise}$. It represents the minimum amount of noise that a system must be able to cope with, assuming the system itself is noiseless. The maximum amount of noise that the system should cope with is found to be the difference between the sensitivity level and the minimum accepted SNR. This noise level is known as the total *in-band* noise . The difference between the in-band noise and the noise floor is the amount of noise the system may contribute, that is, the noise factor .

The noise floor is also important in the definition of the *dynamic range (DR)* of the system. In RF design, the definition is based on the sensitivity and the inter modulation, and is called the *spurious-free dynamic range (SFDR)*[9].

1.2   RF DEVICE TECHNOLOGY

Drawing an analogy with digital integrated circuit technology, we would expect the optimum technology choice for radio frequency IC applications to follow the same path that digital IC implementations followed, that is, towards CMOS. However, technical requirements for a transceiver function are considerably more complex than that of a digital integrated circuit. Issues such as noise (near the carrier), linearity and gain are performance specifications that

---

[9]In general, the DR is defined as the ratio of the maximum tolerable input signal to the minimum input level (often the noise floor).

RF transceivers have to deal with. As a result, the optimum integrated circuit technology choices for RF transceivers are still evolving. Designers currently have a wide range of possibilities: CMOS, BiCMOS, GaAs, bipolar technologies, etc. In addition, improvements of existing technologies have resulted in technologies such as double poly devices, SiGe and Silicon-on-Anything (SOA).

The remainder of this chapter includes a short discussion on technology performance factors and device modeling for RF, followed by an explanation of a few technologies suitable for RF applications.

1.2.1 Characterization and Modeling

It is often desirable to consider the performance of a transistor in terms of some simple, and preferable easily measurable parameters. Some commonly used Figures Of Merits (FOMs) are discussed here, enabling us to choose the right technology for a particular application. However, we have to bear in mind that non-technical factors very often dictate the technology choice as well. These factors include cost, production volume, yield, time to market, etc. After a brief discussion on modeling, we will focus on the technical parameters in the remainder of this section.

Modeling

Modeling of devices is as important as characterization. Without good models, it becomes difficult to analyse and predict circuit behaviour on transistor level.

Bipolar devices have a long history in modeling, even for radio frequencies. Models such as those in Figure 1.7 are usually used, and they give very accurate simulation results when compared to measurements. We will therefore not focus on bipolar device modeling, but will concentrate on MOS device modeling, where this accuracy is still lacking.

For the past few years, intensive research has been going on in the field of RF MOS modeling. Studies have shown that the "standard" (digital oriented) MOS models do not allow for RF modeling. Three major effects play an important role at high frequencies: the terminal resistors, bulk effects, and non-quasi static effects. We will briefly discuss each of these effects below.

- The gate resistance is an important parameter and can be approximated by

$$R_g = \frac{1}{3} \frac{W}{L} R_{\square, poly} \tag{1.43}$$

where $R_{\square, poly}$ is the sheet resistance of polysilicon, used in the gate of the device. The value of the gate resistance can be reduced below a few Ohms if

there is a proper MOST layout. Resistances in series with the drain (R_d) and source (R_s) degrade the drain current and transconductance of the MOST. The source resistance will degrade the noise performance and the maximum oscillation frequency. The drain resistance will reduce the output impedance of the device.

- Device operations are influenced by signals on the bulk node as a result of the body effect. A proper bulk resistance network is therefore mandatory in the intrinsic device model. The values of these resistors are highly dependent on the substrate used, i.e. a high-ohmic or low-ohmic substrate beneath an epi layer.

- When signal frequencies are in the same order of the maximum operating frequencies as the MOS device, non-quasi static effects can occur. To explain the effect of non-quasi static behavior, consider a MOS device with a sinusoidal signal on its gate terminal. At low frequencies, the thickness of the channel will change due to the gate modulation, but the change will be the same everywhere in the channel. In other words, there will be no phase differences in the channel. The dynamics of the device can be modeled by applying the charge equations to the static DC equations. This is no longer true at high frequencies, as the channel thickness will be modulated. This effect is strongest for long devices and can only be measured for minimum device lengths near the cut-off frequencies. As a consequence, the device equation should be solved in a non-quasi static approach. One of the important results of non-quasi static effects is the value of the gate input resistance. At low frequencies, the gate is assumed to be purely capacitive, i.e. Re (Z_i) = 0. This is no longer true at high frequencies, as there will also be a resistive part (see Figure 1.5). In fact, we "see" the channel[10]. These effects result in an extra real contribution to the input impedance, which can be taken into account by adding an additional resistance of approximately (in the case of saturation)

$$R_{in} = \frac{1}{n g_m} \tag{1.44}$$

to the gate resistance, where $n \approx 5 - 7$. Note that this effect cannot be modeled by including external gate and source resistances as can be seen in Figure 1.5.

---

[10]If the input resistance at RF were zero, the device would have an infinite power gain, because no dissipation takes place at the input. This is impossible of course.
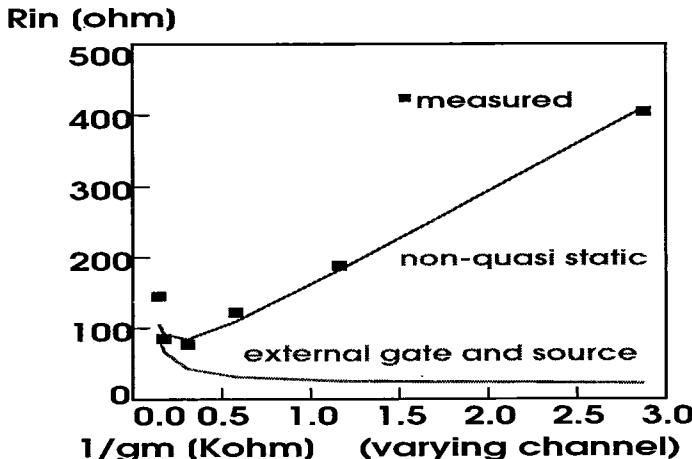
Figure 1.5. The measured input resistance of a MOST is inversely proportional to the transconductance.

The effects mentioned above can be incorporated into a low-frequency MOST model such as the one depicted in Figure 1.6. The resulting model is accurate for RF analysis, although it only partly includes the non-quasi static effects. If a complete non-quasi static behavior has to be included, the low-frequency MOST should be divided into a chain of smaller MOSTs in a very particular way [2].

Cut-off Frequency

Probably the most used FOM for a bipolar device is the *cut-off frequency* $f_{T,bipo}$, sometimes called the transition frequency or unity-current-gain frequency. The latter name refers to the definition of the cut-off frequency: the f_T is defined as the transition frequency at which the common-emitter small signal gain drops to unity for a short-circuit load.

We can use the equivalent circuit from Figure 1.7 to calculate the small-signal common-emitter current gain β with

$$r_\pi = \frac{\beta_0}{g_m}, \; C_\mu = C_{jc}, \; C_\pi = \tau_F g_m + C_{je} \tag{1.45}$$

and τ_F the forward transit time. The junction capacitances C_{je} and C_{jc} are functions of the voltage v across the junction,

$$C_j(v) = \frac{CJ}{\left(1 + \frac{v}{VJ}\right)^{MJ}} \tag{1.46}$$
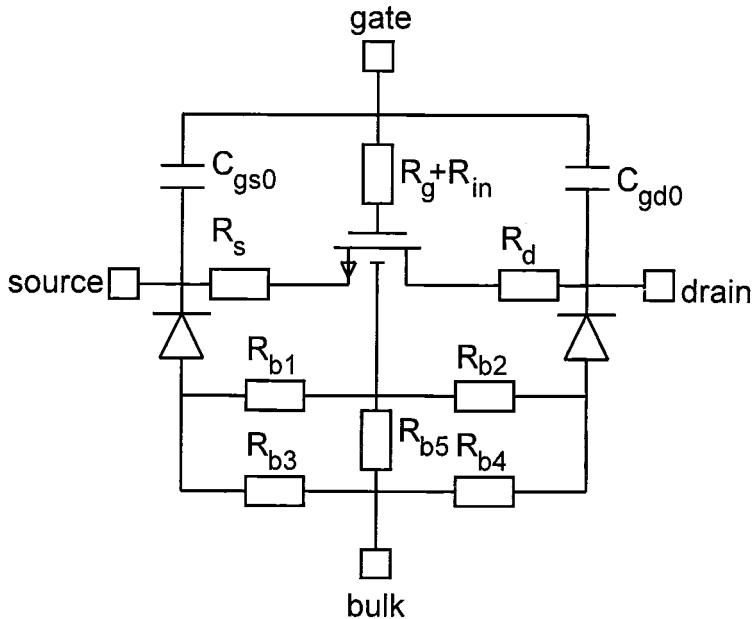
Figure 1.6.   MOST model including RF effects.

where CJ, MJ, and, VJ are constants, defined by the technology.

For a bipolar device, the small-signal common-emitter current gain β is now given by

$$\beta\,(\omega) = \beta_0 \frac{1 + s/z_1}{1 + s/p_1} \qquad (1.47)$$

with the pole and zero defined as

$$z_1 = -\frac{g_m}{C_\mu} \quad p_1 = \frac{1}{r_\pi (C_\pi + C_\mu)} \qquad (1.48)$$

and with the static common-emitter gain β_0 defined as

$$\beta_0 = g_m r_\pi \qquad (1.49)$$
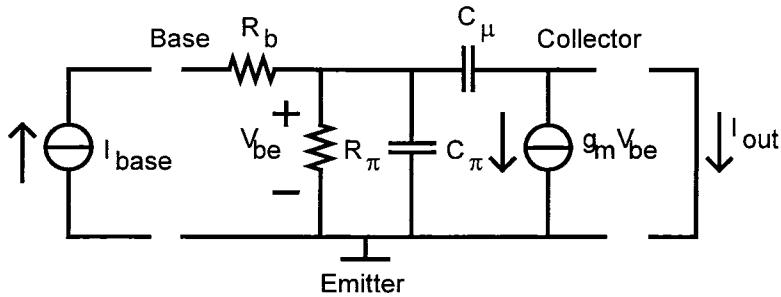
Figure 1.7. Equivalent small-signal circuit for common-emitter configuration with a short-circuit load.

For high frequencies and normal operation, $\omega C_\mu << g_m$. The current gain β drops to unity at a frequency $f_{T,bipo}$[11]

$$f_{T,bipo} = \frac{g_m}{2\pi \left(C_\mu + C_\pi\right)} \tag{1.50}$$

In a similar way, we can define the cut-off frequency $f_{T,most}$ for a MOS device,

$$f_{T,most} = \frac{g_m}{2\pi \left(C_{gg} + C_{par}\right)} \tag{1.51}$$

where we have used the small-signal equivalent circuit from Figure 1.8 in combination with the RF extentions mentioned in Figure 1.6. Here, C_{gg} represents the intrinsic gate capacitance and hence reflects the total sum of C_{gs} and C_{gd}. Capacitance C_{par} includes overlap capacitance and interconnect capacitance to the substrate. For a MOS device, the gate length has a major impact on $f_{T,most}$; a small gate length is needed for a high cut-off frequency.

---

[11]The cut-off frequency can also be obtained from the chain matrix k, defined for a two-port as

$$\begin{pmatrix} u_1 \\ i_1 \end{pmatrix} = \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix} \begin{pmatrix} u_2 \\ i_2 \end{pmatrix}$$

The f_T is now measured for $u_2 = 0$, while the input port is only loaded by a current source. According to the matrix definition, we obtain

$$f_T \rightarrow \frac{i_2}{i_1} = \frac{1}{k_{22}} = 1$$

and for low frequencies for a bipolar device, $k_{22} = -1/\beta_0$.
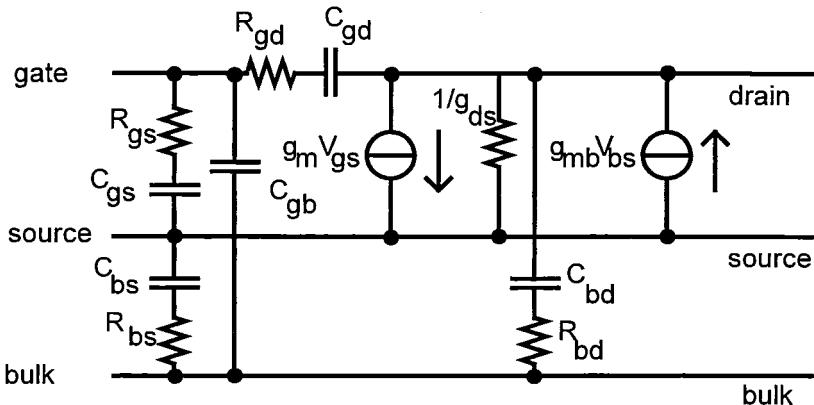
Figure 1.8.   Small equivalent circuit of a MOS device.
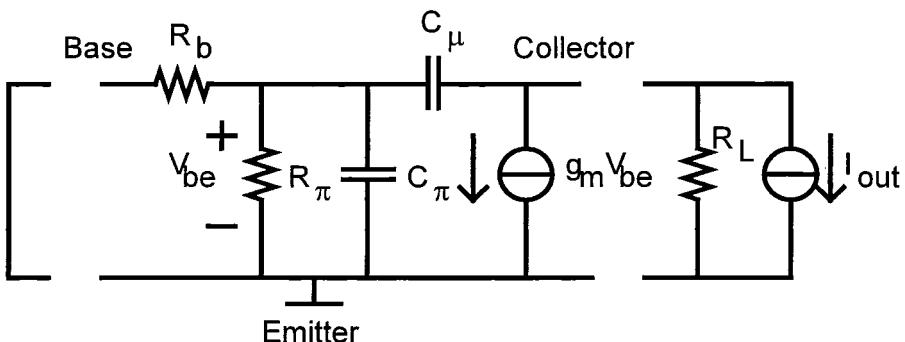
Figure 1.9.   Small-signal equivalent circuit of a bipolar device to calculate the voltage gain of the device.

Maximum Oscillation Frequency

The *maximum oscillation frequency* f_{max} has been defined to include the effect of the base resistance for bipolar transistors and similarly the gate, source and channel resistance for MOS devices. At this frequency, the maximum available power gain, i.e. G_{max}, becomes unity. This implies that both the voltage gain and the current gain are important.

For a bipolar device the voltage gain can be obtained using the small signal equivalent circuit of Figure 1.9. The input is short-circuited while the output is loaded by a current source. The frequency dependent voltage gain a_v is derived

as

$$a_v(\omega) = a_v \frac{1 + s/z_1}{1 + s/p_1} \tag{1.52}$$

with

$$z_1 = \frac{g_m}{C_\mu} \quad p_1 = -\frac{g_m}{|a_v| C_\mu} \quad a_v = -g_m R_L \tag{1.53}$$

From this expression together with the current gain expression, $f_{max,bipo}$ can be found to be

$$f_{max,bipo} = \left(\frac{f_{T,bipo}}{8\pi r_b C_\mu} \right)^{1/2} \tag{1.54}$$

where C_μ is the depletion layer capacitance of the collector-base junction. In a similar way, the $f_{max,mos}$ can be found to be [12]

$$f_{max,mos} = \left(\frac{f_{T,most}}{4\pi C_{junc}\sqrt{(R_g + R_{in} + R_s) R_{eff}}} \right) \tag{1.55}$$

where C_{junc} is the drain junction capacitance and R_{eff} the effective bulk resistance. Clearly, $f_{max,mos}$ is more influenced by parasitic elements than $f_{T,most}$, and the device layout is therefore extremely important.

Input Limited Frequency

The *input limited frequency* f_v is related to the frequency dependency of the transconductance of the device. If we use the small signal equivalent circuit of Figure 1.10, then the transconductance is derived as

$$g_m(\omega) = g_m \frac{1 + s/z_1}{1 + s/p_1} \tag{1.56}$$

---

[12] The maximum oscillation frequency can also be obtained using the chain matrices. The G_{max} is defined as $G_{max} = a_v \cdot a_i$ under the condition that the two port is perfectly terminated, $Z_L = Z_{out}^*$ and $Z_s = Z_{in}^*$. The voltage gain is defined as

$$a_v = \frac{Z_L}{k_{12} + Z_L k_{11}}$$

and the current gain is defined as

$$a_i = \frac{1}{k_{22} + Z_L k_{21}}$$

where Z_L is a function of Z_{in}. An analytical expression of G_{max} is difficult to obtain because of the matching property.
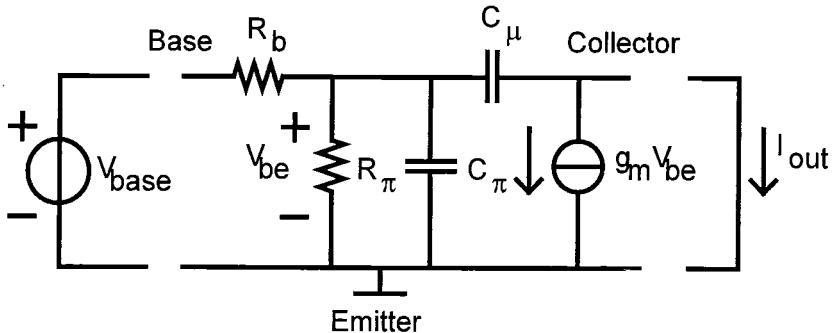
Figure 1.10.   Small signal equivalent circuit of a bipolar device to calculate the transconductance as function of the frequency.

with

$$z_1 = \frac{g_m}{C_\mu} \quad p_1 = -\frac{1}{r_b(C_\mu + C_\pi)} \qquad (1.57)$$

The parameter f_v is defined as the frequency at which $g_m(\omega)$ is dropped by 3 dB. For a bipolar device, $f_{v,bipo}$ can then be derived as

$$f_{v,bipo} \approx \frac{1}{2\pi r_b \left(C_\mu + C_\pi\right)} \qquad (1.58)$$

In a similar way, we can obtain the input limited frequency for a MOS device,

$$f_{v,mos} \approx \frac{1}{2\pi r_g \left(C_{gg} + C_{par}\right)} \qquad (1.59)$$

Output Limited Frequency

If the device is loaded with a resistor, we can define the voltage gain, as we have done to calculate the maximum oscillation frequency. The frequency at which the voltage gain drops with 3 dB is know as the *output limited frequency*, f_{out}. For a bipolar device this yields

$$f_{out,bipo} = \frac{1}{2\pi R_L C_\mu} \qquad (1.60)$$

where R_L is the load resistance of the device[13].

A similar expression can be derived for a MOS device, yielding

$$f_{out,mos} = \frac{R_{out} + R_L}{2\pi R_L R_{out} C_{out}}$$ (1.61)

where R_{out} is the output resistance for the device and, in the first order can be approximated as $1/g_{ds}$.

Maximum Available Bandwidth

From a circuit point of view, it is more convenient to know the device *available bandwidth* f_a. This parameter takes the input limited frequency into account as well as the output limited frequency f_{out}. The available bandwidth is then defined as

$$f_a = \frac{1}{\frac{1}{f_v} + \frac{1}{f_{out}}}$$ (1.62)

The available bandwidth parameter is a function of the device current. If we consider a bipolar device, then for low currents, the f_a is determined by f_{out} and for high frequencies it is mainly determined by f_v. For a bipolar device, we can derive that the $f_{a,bipo}$ is equal to

$$f_{a,bipo} = \frac{1}{2\pi} \cdot \frac{1}{r_b C_\pi + (r_b + R_L) C_\mu}$$ (1.63)

For a MOS device, the input limited frequency is normally large and the $f_{a,mos}$ is therefore determined by the output limited frequency.

1.2.2   Technology Choice

We will discuss several technology options a designer can choose from when designing a circuit. As already mentioned, in an RF design both active and passive devices are important. We will discuss both type of devices for several technologies.

---

[13]The voltage gain for a two port is obtained under the condition that the output current is zero. Using the chain matrix parameters, the voltage gain a_v is then found to be $1/k_{11}$. To obtain f_{out} the two-port needs to be loaded with a load R_L. This load can be treated as a two-port in itself, expressed by chain matrix k^2, while we assume that the loaded two-port is expressed by k^1. Then the voltage gain of the overall two port, i.e. a cascade of two blocks, is found from the overall chain matrix $k = k^1 \cdot k^2$,

$$a_v = \frac{1}{k_{11}} = \frac{1}{k_{11}^1 k_{11}^2 + k_{12}^1 k_{21}^2}$$
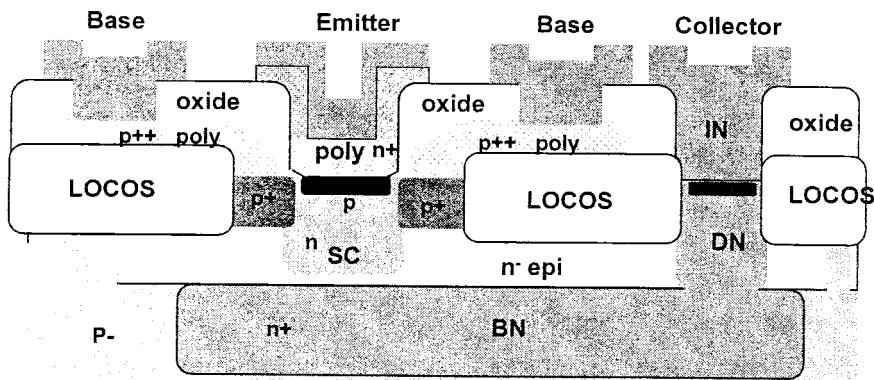
Figure 1.11.   Schematic structure of a double poly bipolar device. The substrate has a resistivity of 20 Ω·cm.

Double Poly Devices

Practically all modern high performance bipolar devices with base widths of 100 nm or smaller employ a polysilicon emitter. In this case, the emitter is formed by heavily doping a polysilicon layer and then activating the doped polysilicon layer just sufficiently to obtain reproducible base currents and low emitter resistance.

The base region can be divided into two regions: the intrinsic base and the extrinsic base. The intrinsic base is the region directly underneath the emitter and determines the collector current. The extrinsic base connects the intrinsic base with the base terminal, and can be considered to be a parasitic component. A self-aligned process is used to decouple the intrinsic and extrinsic base. Instead of making contact with the base directly by metal, contact is made indirectly with the extrinsic base via a p-type polysilicon layer, hence the name double poly process (DPO). A metal contact to the poly layer is then made on top of a field-oxide region. A small base area is obtained due to the self-alignment process. Consequently, small extrinsic base collector junction capacitances are obtained, as well as low contact resistances. A schematic diagram illustrating the structure of a double poly bipolar device is given in Figure 1.11.

Typical values for a minimum device of 1 μm x 0.8 μm are: $f_T = 23.4$GHz at 194 μA, $f_{\max} = 10.6$ GHz and $f_{out} = 1.9$ GHz.

Polysilicon filled deep trench isolation can be used instead of p-type diffusion regions to isolate sub collectors of adjacent transistors. As the trenches cut through the sub collector layer, the collector-substrate capacitance is reduced by a factor of six. Although deep-trench isolation gives far better RF performance,
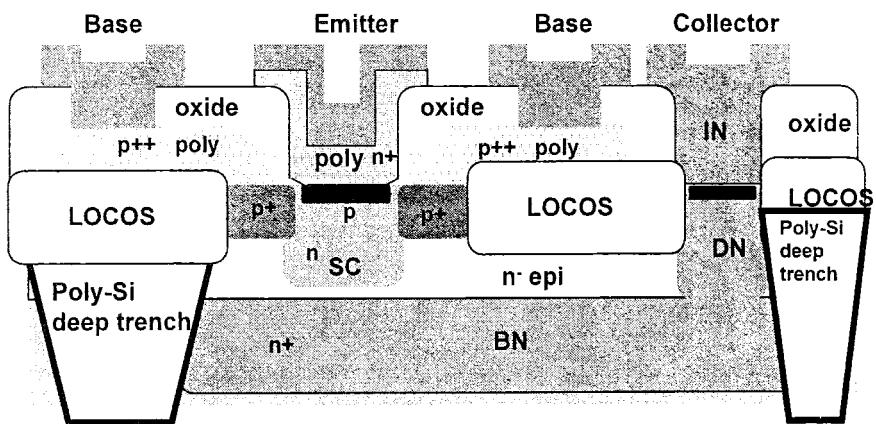
Figure 1.12.   Schematic illustration of deep-trench isolation. The trench consists of 30 nm Nitride and 30 nm Oxide around the polysilicon.
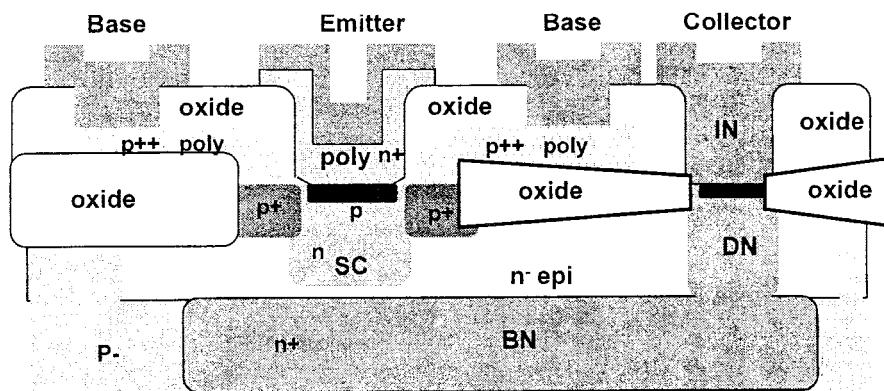
Figure 1.13.   Schematic of shallow-trench isolation. The oxide layer cuts into the p^+ layer of the base, instead of the LOCOS layer of Figure 1.12.

diffusion isolation is still used in many bipolar processes for financial reasons. The typical performance value for f_{out} due to deep-trench isolation is increased to $f_{out} = 3.9$ GHz. An illustration is given in Figure 1.12.

In a similar way as deep trenches cut through the sub-collector (SC) layer to minimize parasitic capacitance, a trench can be added to minimize the collector base capacitor. This shallow trench cuts through the p^+ layer of the base. An example is provided in Figure 1.13. The improvement of performance due to shallow trenches can be found in a reduction of the collector-base capacitor by approximately 30%.
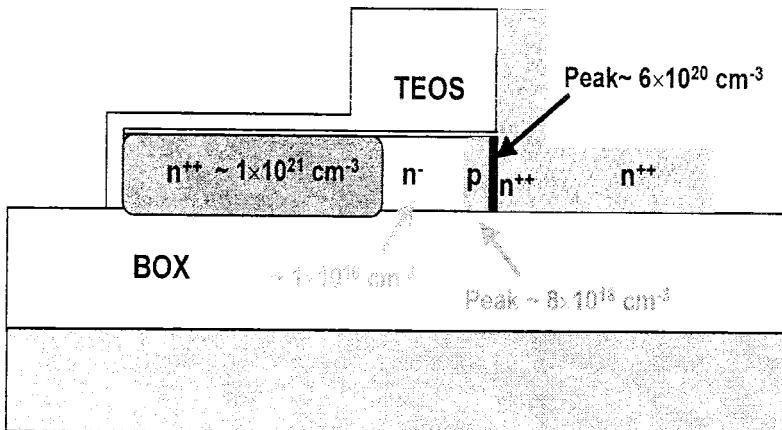
Figure 1.14.   Lateral bipolar device in SOA.

In general, the advantages of double poly techniques are: very high cut-off frequencies, reduced contact and base resistances, and a sub-lithographic emitter. The process is also compatible with SiGe epitaxial base. The drawback is the increased complexity and therefore cost, and the difficulty of merging with CMOS processing. A BiCMOS process based on double poly techniques is therefore difficult.

Silicon-on-Anything

One drawback of bipolar technology is the existence of a substrate, causing parasitic elements which negatively influence the performance. To increase performance, one option is to remove the substrate. A second option is to use a lateral transistor geometry to have a peak f_T at much lower current densities. These two issues are the basic ideas behind a new RF technology, called Silicon-on-Anything (SOA) [3]. A geometrical structure of a lateral bipolar device in SOA is depicted in Figure 1.14. A buried oxide (BOX) layer, 0.4 μm thick, is grown on a 20 Ω·cm p-type substrate. The n^{++} implant for the collector region takes place on top of this (right side of the picture). The collector width is realized by a self-alignment process. The base is then realized in a similar way by a p implant. Finally, the emitter region is defined. A microphotograph of the device is shown in Figure 1.15. The next step in SOA is to remove the substrate. This is realized by flipping the die (steps 1 to 2) and glueing it onto a glass substrate (step 2). This means that the active and passive devices are now on top of a glass substrate, as indicated in Figure 1.16. Finally, the substrate is removed by an etching process (step 3). All these process steps are backend
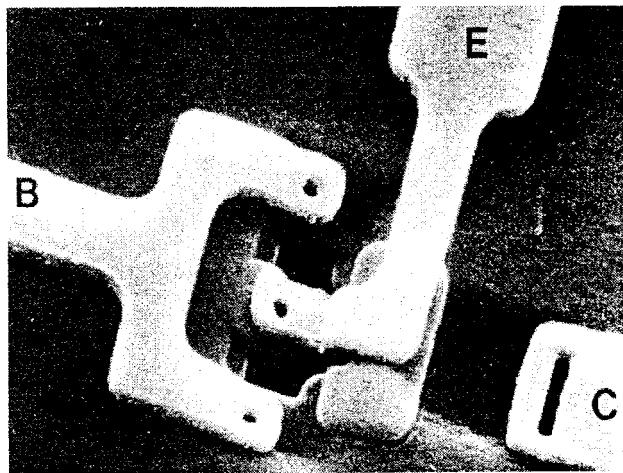
Figure 1.15. Die microphotograph of an SOA device.
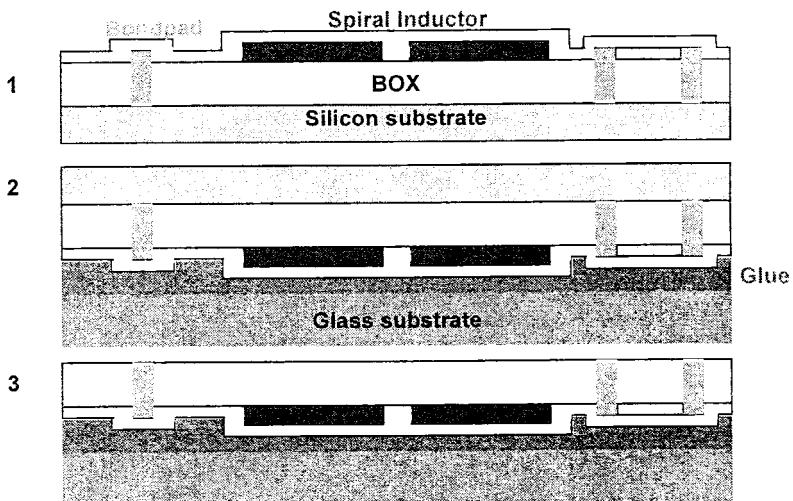
Figure 1.16. SOA transfer to glass in three post processing steps. The technique is valid for any type of device, not only active devices.

steps, that is, they can be performed after processing and passivation. These additional steps are cheap to realize, and do not need complicated equipment. The advantages of SOA are not only in the reduced parasitics for the active devices. The passives such as spiral inductors will also improve significantly
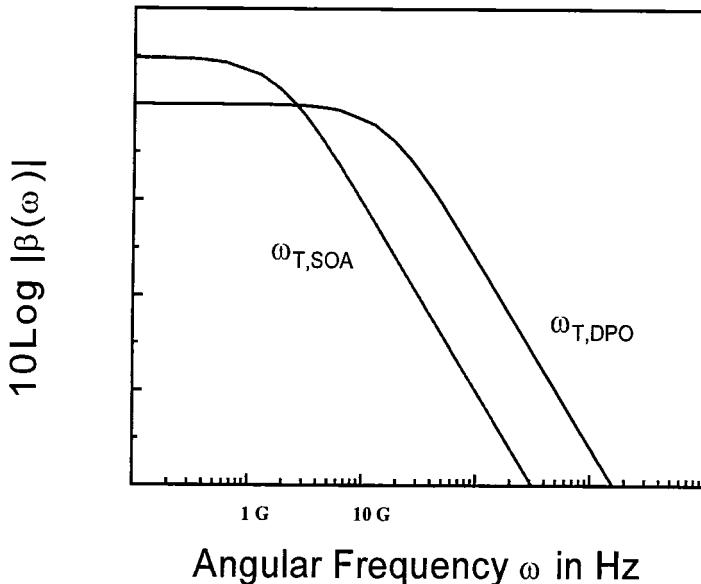
Figure 1.17.   Comparison of cut-off frequency for SOA and DPO devices.

due the absence of any parasitics towards the substrate. We will return to this issue in the next section.

Comparison

As a technology benchmark, we will now compare an existing DPO process with an existing SOA process, using the minimum device geometries. In Figure 1.17 the cut-off frequencies are compared, where $f_{T,DPO} \approx 25$ GHz and $f_{T,SOA} \approx 8$ GHz.

   In a similar manner, a comparison of the maximum oscillation frequency is given in Figure 1.18. For SOA we have $f_{max,DPO} \approx 5.6$ GHz and $f_{max,SOA} \approx 11$ GHz. The influence of the reduced base resistance is now clearly visible. From a circuit point of view, designers are normally also interested in the relationship between current consumption and the cut-off, maximum oscillation and available frequencies. These relationships are shown in Figure 1.19. The available bandwidth for an SOA device is the same as for a DPO device, although obtained by a current reduced by a factor of 30. Hence, very low power design is possible in SOA, as we will see later in other chapters.
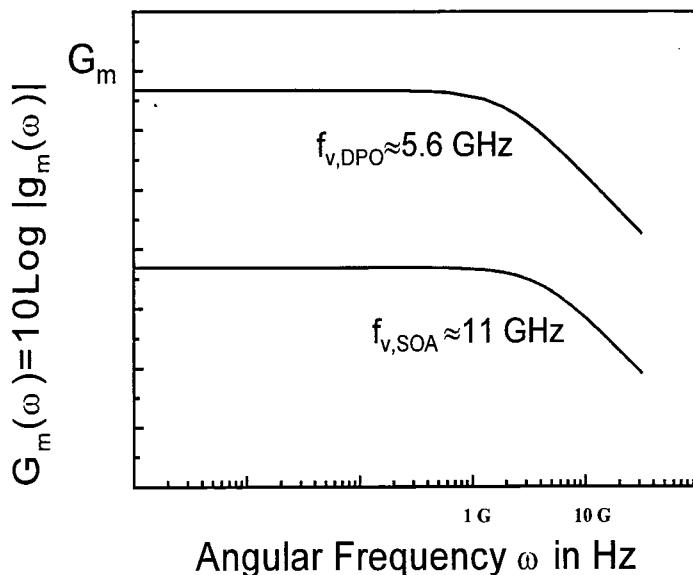
Figure 1.18.   Maximum oscillation frequency for an SOA and DPO device.
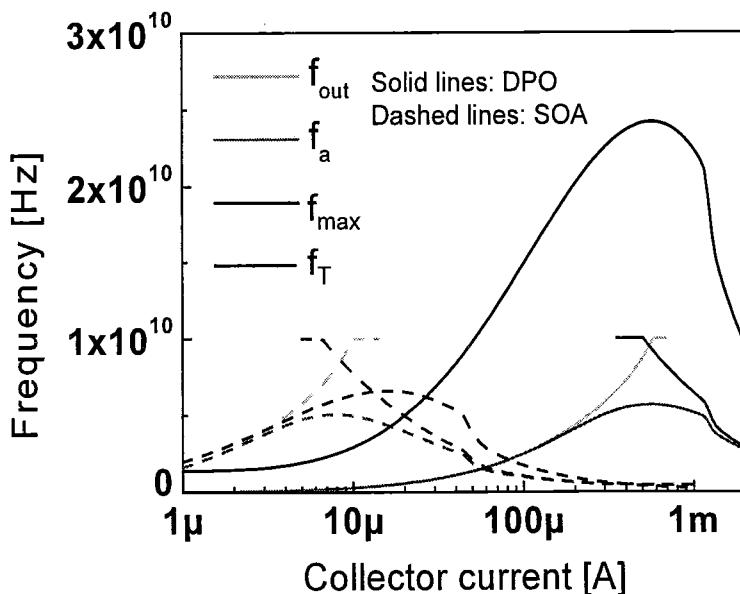
Figure 1.19.   Current versus FOMs for SOA and DPO technology. The FOMs used are f_T, f_a, f_{out}, and f_{max} respectively.

SiGe Bipolar Technology

In order to get more current gain out of a bipolar device, the drift field for the electrons in the base should be maximized. This can be achieved by lowering the bandgap in the base region. Silicon (Si) has a bandgap E_g of approximately 1.2 eV, but germanium (Ge), for instance, has a bandgap of almost half that value, namely 0.66 eV. If there is germanium in the base region, this will allow the device characteristic to be modified. Optimal performance is obtained by a gradual profile of Ge over the base region. The base bandgap should be small near the base-collector junction and large near the emitter-base junction [4].

Since the emitter of a SiGe device is the same as that of a Si-base device, the base currents should be the same. Only the collector current is affected by the SiGe dope profile in the base. The current gain β using Ge can be derived as

$$\frac{\beta(Si\,Ge)}{\beta(Si)} = \frac{\Delta E_{g,Si\,Ge}/kT}{1 - \exp(-\Delta E_{g,Si\,Ge}/kT)} \tag{1.64}$$

where k is the Boltzmann constant and T is absolute temperature. The ratio value is typically about four to six. In a similar way, the low-current base transit time ratio can be defined as

$$\frac{\tau(Si\,Ge)}{\tau(Si)} = \frac{2kT}{\Delta E_{g,Si\,Ge}} \left[1 - \frac{kT}{\Delta E_{g,Si\,Ge}} \cdot \left(1 - \exp(\frac{-\Delta E_{g,Si\,Ge}/kT}{kT}) \right) \right] \tag{1.65}$$

and is normally about 2.5. The cut-off frequency of the bipolar device is limited by the base transit time and the emitter delay time, which is inversely proportional to the collector gain. SiGe therefore allows substantially higher cut-off frequencies than Si-base transistors.

RF CMOS

In contrast to bipolar technologies, CMOS technology is driven by digital circuitry and hence cost. More functionality per unit area is required, which in turn forces down the scale of the technology. Sub-micron technology down to 0.12 μm is currently standard and deep sub-micron technology such as 100 nm technology will shortly become available in production lines. Although 0.25 μm CMOS is sufficient for RF applications upto 5 GHz, the demand for high integration densities is also forcing RF circuitry to be designed in these new processes.

One advantage of CMOS compared to bipolar processes is the degree of freedom with respect to the dimensions of the transistor. This degree of freedom is important, since the cut-off frequency, for instance, depends on the gate
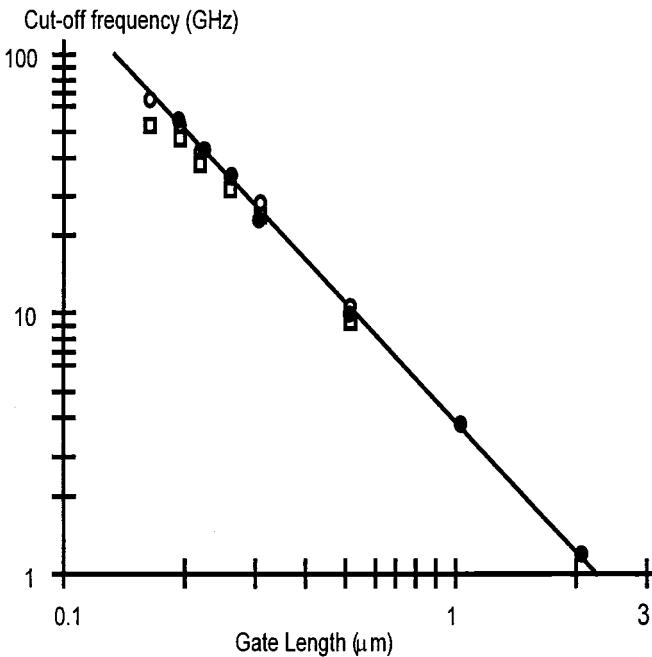
Figure 1.20. Cut-off frequency vs. gate lengths in a 0.25μm CMOS technology for different device geometries: (•) single, (o) parallel and (□) finger.

length as can be seen in Figure 1.20 for a 0.25 μm CMOS process. A similar relationship exists for the maximum oscillation frequency, as shown in Figure 1.21. For a chosen process technology, it is therefore important to remember that the dimensions of the device are largely for determining its operational frequency.

Both figures also show another phenomenon. The performance of the device is also largely determined by the geometry of the device, resulting from its layout. In principle, there are three possible layout structures. Considerer a device of width W and length L.

- The first one is a single geometry, which actually means that the device is laid out as one rectangle, with aspect ratio W for its width and L for the length. This structure is used most in digital circuitry.

- The second option is a parallel structure. To reduce the gate resistance, the devices can be laid out as n devices in parallel with each device with a width of W/n. A schematic drawing of this layout is shown in Figure 1.22. The basic layout is the same as that for the single device. The gate has contacts
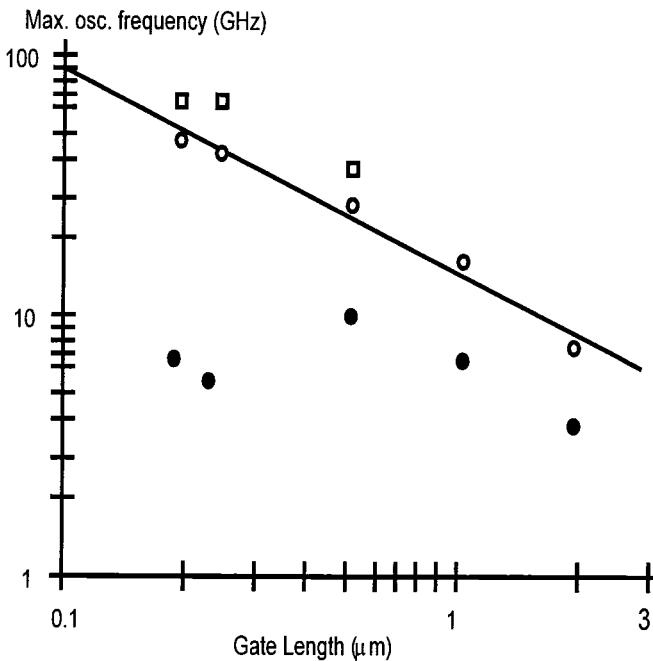
Figure 1.21.   Maximum oscillation frequency vs. gate lengths in a 0.25μm CMOS technology for different device geometries: (●) single, (○) parallel and (□) finger.
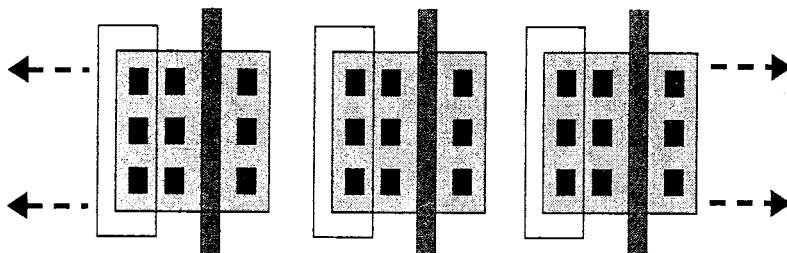
Figure 1.22.   Parallel geometry for a MOS device.

at both sides. This device has appreciably lower gate resistance than the single device because of the reduced width. The parallel gate capacitance is increased, however, and the drain junction capacitance is not optimal. This device will therefore give a better f_{max} at the cost of a reduced f_T. This effect can also be observed from Figure 1.20 and Figure 1.21.

■ The last option is a finger structure, as depicted in Figure 1.23. The gate is connected from two sides using metal interconnect, as only then can a very
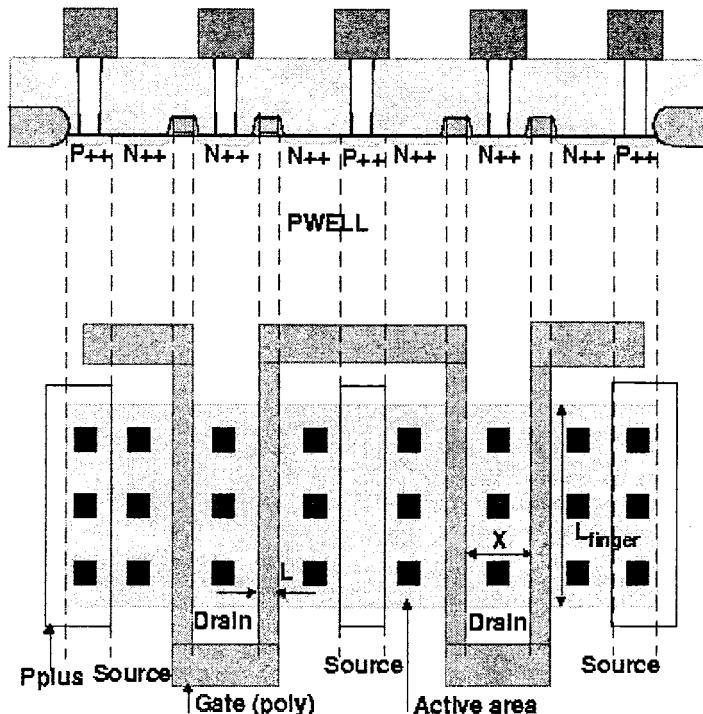
Figure 1.23. Finger structure geometry for a MOS device to reduce the gate resistance.

low total gate resistance be achieved. To reduce the contribution of the poly resistance to the gate resistance to less than a few ohms, the device must have as many fingers as possible. The drain area is shared by two transistors, almost halving the drain junction capacitance per transistor. The substrate contact is merged with the source contact, and is shunted by a silicide layer. This structure has the lowest gate resistance and lowest junction capacitance of all structures. It is a very interesting structure for power, high f_{max} and low noise figures. The relatively high parasitic capacitance at the gate will degrade the f_T as seen in Figure 1.20 and Figure 1.21.

1.3 PASSIVES

Most RF circuits do not only consist of active devices, but also use passive components such as resistors, capacitors and, particularly at high frequencies, planar inductors. In many situations, the overall performance of the circuit is determined by the (parasitic) behavior of the passives. This means that the choice of technology should also be based upon the performance of the passive
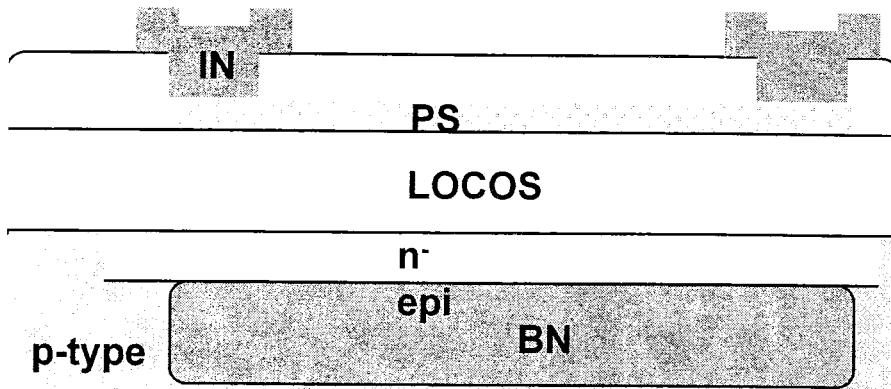
Figure 1.24. Polysilicon resistor in double poly technology. The resistance value is determined by the $R_\square$ of the PS region.

components. In the following sections, we will discuss some issues related to the design of passives in a few technologies.

1.3.1   Resistors

To have large resistance values, the resistor should be made from a material with a high resistance per unit area (i.e. sheet resistance $R_\square$). The parasitic capacitances should also be as small as possible. In a double poly technology, the resistor is laid out in a polysilicon layer on top of a LOCOS layer, as shown in Figure 1.24. The first order approximation of the resistance value is given as $R = R_\square \frac{L}{W}$, where L and W ($L > W$) are the length and width of the resistor geometry, respectively. From Figure 1.24 it is clear that parasitic capacitances will also contribute to the performance of the resistor. A fair model of the parasitics is given in Figure 1.25 and should be used on both terminals of the resistor. First of all, we have the plate capacitance (C_{LOCOS}) as a result of the LOCOS. Electric edge, or fringe, fields will also contribute to the parasitics and are modeled by C_{Fringe}. The buried n-layer (BN) will reduce the parasitic resistance of these two capacitances, hence the quality factor will increase. The capacitance C_{LOCOS} can be modeled as $C_{LOCOS} = W \cdot L \cdot \frac{\varepsilon}{t_{ox}}$, where $\varepsilon = \varepsilon_0 \varepsilon_r$ are the dielectric parameters and t_{ox} is the thickness of the LOCOS layer. In a similar way, $C_{Fringe} = 2\alpha (W + L + 2\alpha t_{ox}) \varepsilon$ where α is a process dependent parameter. The substrate also contributes, but normally $C_{sub} \ll C_{LOCOS} + C_{Fringe}$ for low frequencies; this capacitance becomes dominant at RF (see Figure 1.26). Removing the substrate will therefore significantly improve the quality of the resistor.

Polyresistance
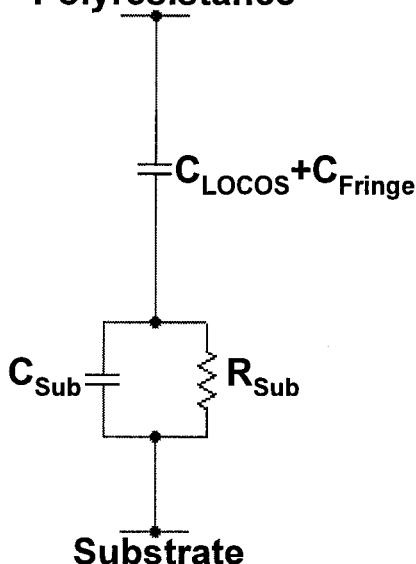
Figure 1.25.   Parasitic capacitances for a poly-resistor.

1.3.2 Capacitors

Decreasing the oxide thickness implies an increase of the capacitance value per unit area. In a double poly process, a capacitor is therefore realized by adding a thin oxide layer on top of the polysilicon layer in Figure 1.24, as can be seen in Figure 1.27.

The capacitance is given by $C = \varepsilon_{ox}/t_{ox}$ where $\varepsilon_{ox} = 3.9 \cdot \varepsilon_0$ and $t_{ox} = 0.3$ μm. The parasitic components can be modeled using the model in Figure 1.25.

In standard technologies, metal-to-metal capacitors can be designed to have capacitance values up to a few hundred femto Farad. These metal-to-metal capacitors are designed by using metal layer one as the bottom plate, in combination with the top metal layer as the top plate of the capacitor. There is only oxide between the two plates. However, in RF designs larger capacitance values are very often needed. These large values can be realized by fringe capacitors . When designing fringe capacitors, the device is laid out in such a manner that a maximum number of parasitic capacitances are created to increase the value of the capacitor. As an example, consider Figure 1.28 where five metal layers have been used to realize the capacitor. Capacitance C_1 represents the parasitic capacitance between two metal layers, whilst C_2 reflects the parasitic capacitance between two metal lines on the same layer. A finger structure has been
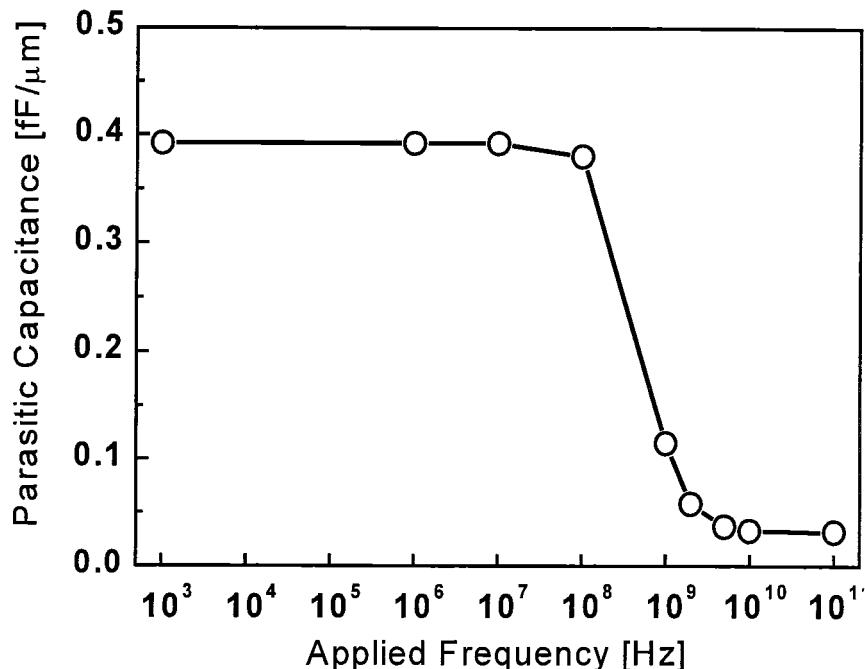
Figure 1.26.   At high frequencies, the C_{sub} becomes dominant.
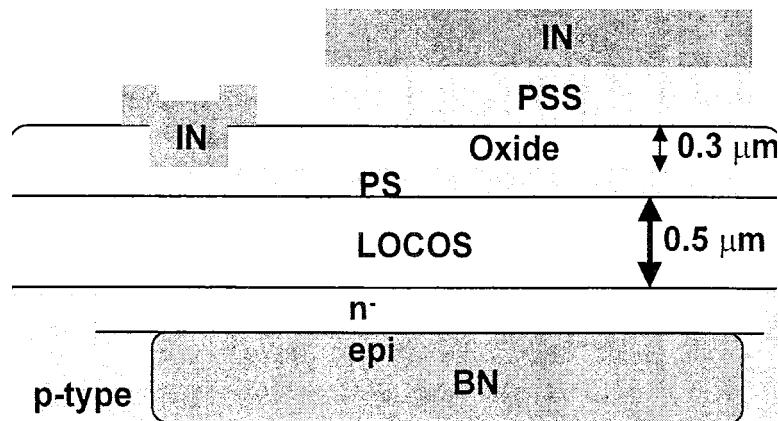
Figure 1.27.   Capacitor in double poly technology. The capacitance value is determined by the thickness of the oxide layer, in this case 0.3 μm.
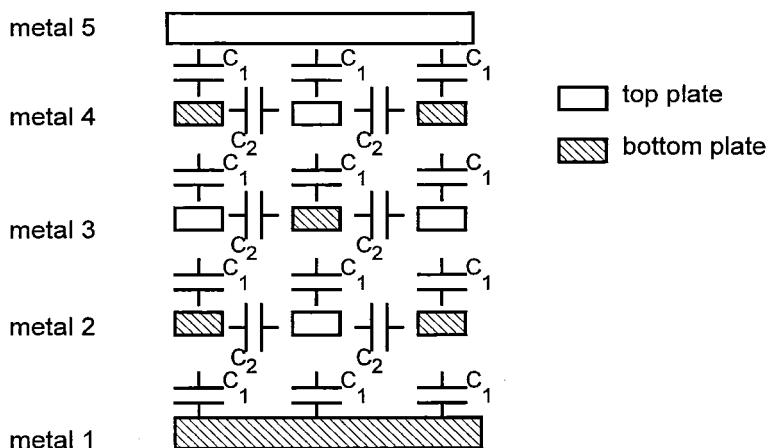
Figure 1.28. Fringe capacitor using five metal layers, in cross-section. A finger structure has been laid out per layer.

laid out per layer. The overall capacitance is measured between metal 5 and metal 1. For a 0.25 μm CMOS process, this geometry yields a capacitance-area ratio of 0.42 fF/μm^2, with a Q of 50 to 100 at 2 GHz. A simpler geometry for a fringe capacitor is shown in Figure 1.29, where metal layers are connected in a column structure Now, capacitance C_1 represents the parasitic capacitance between two vias. For the same process, this capacitor has a ratio of 0.3 fF/μm^2 and similar Q values.

Another option to realize large capacitance values is to make use of a special metal-insulator-metal (MiM) structure. The advantage of such a capacitor is the capacitance-area ratio of $1 - 5$ fF/μm^2. The drawback, however, is the special process option needed for this structure, making the technology more expensive.

In conclusion, when working in MOS technology, the designer can also use the MOS device as a capacitor . By connecting the source to the drain terminal, the gate capacitance reflects the capacitor value. We should bear in mind that the device should then be operational, otherwise no channel will exist and hence no charge. The MOST capacitor should therefore be properly biased.

1.3.3   Planar Monolithic Inductors

One of the major research topics in RF design is the realization of inductors on-chip. Although many rules of thumb have been derived from experiments, synthesis of planar inductors is still in its infancy. For a particular technology, a given inductance value and a given quality factor at a certain operating fre-
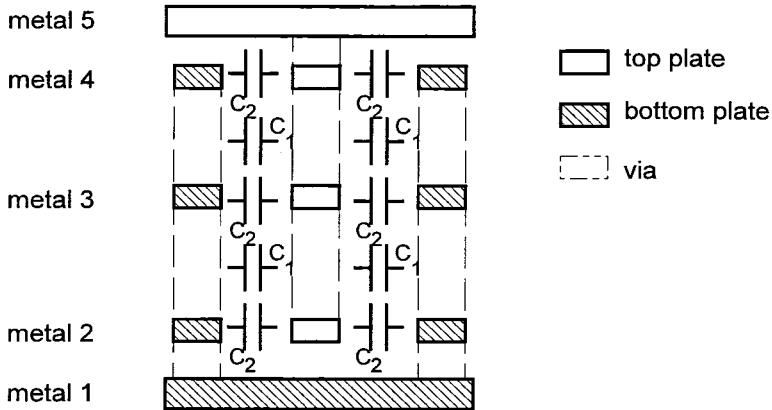
Figure 1.29.   Cross-section of a simple multilayer fringe capacitor with column structure.

quency, the designer has many parameters to tune. The most important ones, as shown in Figure 1.30, are the diameter d, inner radius r, track width w and number of turns. An approximation of the inductance value L and series resistance R_{seri} are given by the following model equations [5]:

$$L = K_L \frac{(\pi d^2/4)^{3/2}}{w^2} \left(\frac{A_{metal}}{\pi d^2/4} \right)^{5/3} \left(\frac{w}{w+s} \right)^{1/7} \tag{1.66}$$

$$R_{seri} = \frac{\rho}{t_{metal}} \frac{\pi d^2/4}{w} \left(\frac{1}{w} + \left(\frac{t_{metal}\omega_0}{K_r \rho} \right)^2 \right) \frac{A_{metal}}{\pi d^2/4}$$

where A_{metal} is the metal area, ρ the resistivity of metal, K_L an inductor constant, K_r a resistor constant, s the spacing between two lines, and ω_0 the resonance frequency in rad/s. This model takes skin effects into account by means of an equivalent reduced line width $(K_r\rho/t_{metal}\omega_0)^2$, depending on geometry and frequency. Although it is not very accurate, this expression gives reasonable first order predictions.

The equivalent lumped model for an inductor also includes the parasitic effects due to substrate and coupling between turns. These parasitic effects are represented by capacitor C_p and series resistor R_p,

$$C_p = \frac{\epsilon_{0x}}{t_{ox}} A_{metal} \tag{1.67}$$

$$R_p = R_{\square,sub} \left(\frac{t_{sub}^2}{t_{sub}^2 + \pi d^2/4} \right)$$

where $R_{\square,sub}$ is the sheet resistance of the substrate.
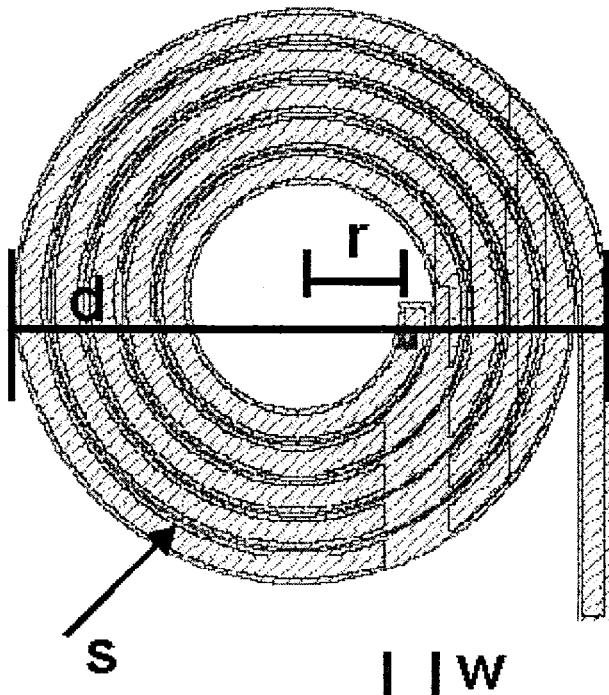
Figure 1.30. Geometry of a planar inductor. Diameter d, inner radius r, line spacing s, and track width w for 5 turns are shown.

Some parameters not taken into account in these model equations are the layer configuration and bridging, for example. Normally, the top metal layer is thicker than the other metal layers. This layer should, therefore, definitely be used to realize the inductor. In general, it is not true to say that the more connected metal layers used the better the inductor becomes; this largely depends upon the process. Bridging is normally done in the metal layer beneath the ones used to realize the inductor. Often only a single metal layer is used, because otherwise the bridge is too close to the substrate. Figure 1.30 shows a circular inductor geometry, but square or hexagonal geometries are also used. Square geometries do have the advantage that they do not spoil area and are easier to lay out. The spiral geometry should comply with the lay out rules, but in the end they give a better inductor performance.

It is obvious that the substrate plays an important role in the quality factor of the inductor. The higher the resistance of the substrate the less losses due to parasitics, and hence the better inductors obtained. This effect can be measured and is shown for three types of substrate in Figure 1.31.
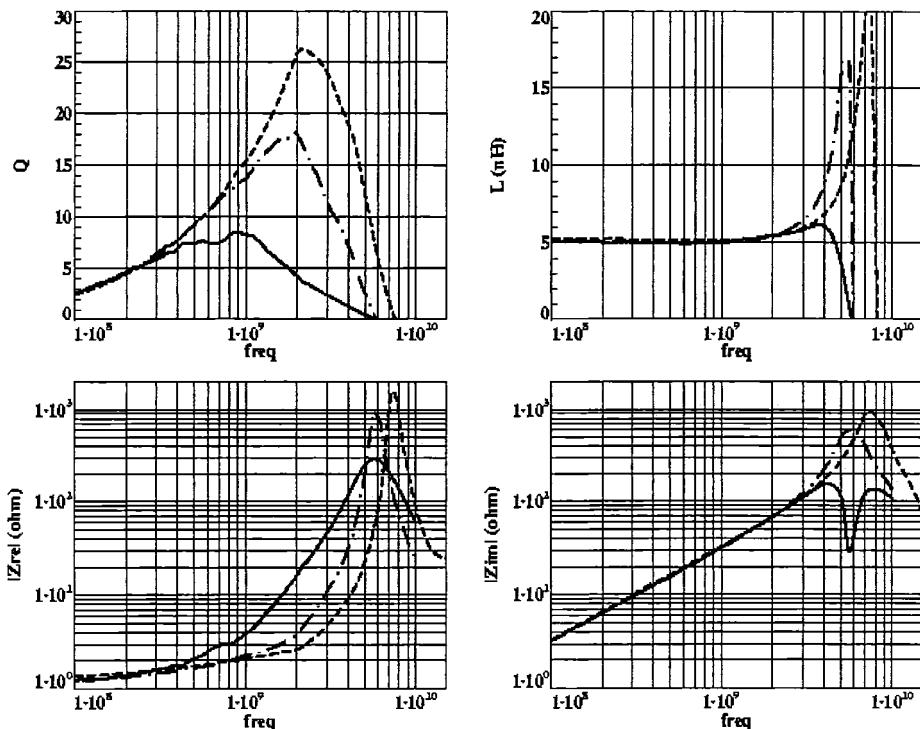
Figure 1.31.   Relationship between inductor performance and substrate. The solid line represents 20Ω·cm substrate, the dot-dashed line a 1000Ω·cm substrate and the dashed line the SOA technology. The 20Ω·cm substrate is from a DPO technology.

Shielding plates can be used beneath the inductor to shield it from the substrate. These shields can be made of metal or polysilicon and can have several geometries in order to reduce the amount of circulating currents through the shielding plate. As an example, consider the spiral inductor photograph in Figure 1.32. A poly shield has been used to reduce influences to and from the substrate. Bridging is performed in metal 3 and 2, while metal layers 6 up to 4 are used for the spiral itself. The square plane underneath is the poly shield. The performance of this 4 nH inductor is shown in Figure 1.33. As high-ohmic substrate has been used, the quality factor is good enough for RF applications in the 2 GHz band.

Finally, we will give a table with some measurement results for several inductors in a 0.18 μm CMOS technology with six metal layers (Table 1.1). The substrate had a resistivity of 7 Ω·cm. All the inductors are realized in
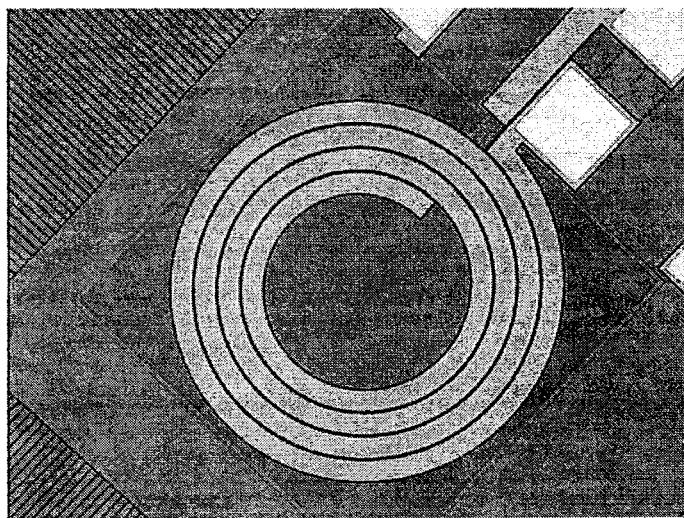
Figure 1.32.   Spiral inductor with a square poly shield beneath it.
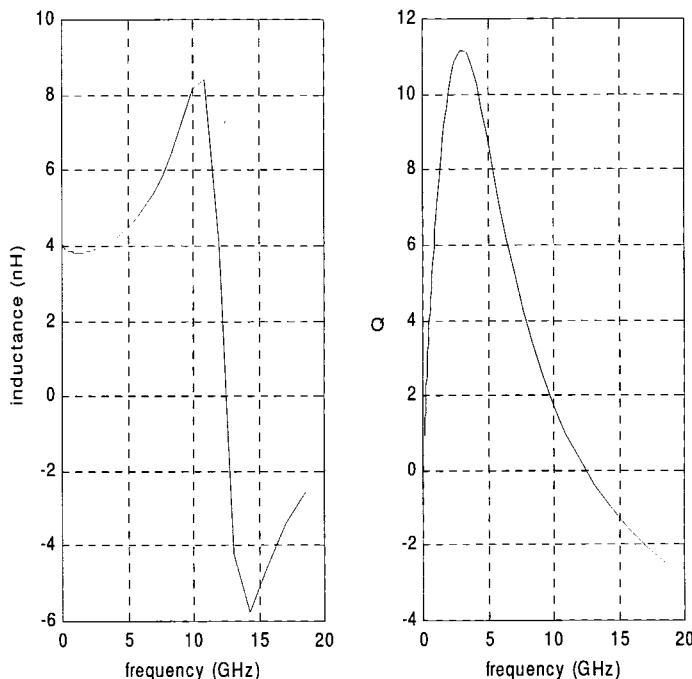
Figure 1.33.   Measured quality factor and inductance value as a function of the frequency for the spiral inductor of Figure 1.32.

metal 6 up to metal 4, and bridging is performed in metal 3 and 2. The chosen geometry is a spiral structure. The inductance and quality factor have been measured at 2.45 GHz. From these results, it becomes clear that shielding, for instance, will not affect the performance of the inductor apart from lowering the resonance frequency. Measurements on the same inductor structures, but laid out on a low-ohmic substrate (10 mΩ·cm), show that the quality factor drops by a factor 2 or more.

Table 1.1.   Measured inductors with geometry according to Figure 1.30

Inductor	d (μm)	turns	w (μm)	s (μm)	L (nH)	Q	f_{res} (GHz)
L1+shield	400	4	25	3	4.2	9	6.5
L2+shield	300	4	15	3	3.8	11	9
L3+shield	200	4	8	3	3.6	12	9.2
L4	400	5	25	3	4.2	10	9.9
L5	300	4	15	3	3.9	11	12
L6	400	4	20	3	5.3	10	8.3
L7	500	4	20	3	8.2	9	5.8

REFERENCES

[1] G. Gonzales, *Microwave Transistor Amplifiers: analysis and design*, Prentice-Hall, Inc, New Jersey, 2nd edition, 1997.

[2] L.F. Tiemeijer *et al.*, "MOS Model 9 based non-quasi-static small-signal model for RF circuit design," in *European Solid-State Device Research Conf. (ESSDERC)*, 1999, pp. 652–655.

[3] A. Wagemans *et al.*, "A 3.5mW 2.5GHz diversity receiver and a 1.2mW 3.6 GHz VCO in Silicon-On-Anything," in *IEEE International Solid-State Circuits Conf. (ISSCC)*, 1998, pp. 250–251.

[4] Y. Taur and T.H. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge University Press, New York, 1998.

[5] J. Crols, *Full Integration of Wireless Transceiver Systems*, Katholieke Universiteit Leuven, Leuven, 1997.

Chapter 2

Antennas, Interface and Substrate

Analog designers only concentrate on the circuits on the silicon. However, the connection with the outside world, i.e. going on/off chip, is equally important for RF designers. Bond pads and their electrostatic discharge protection circuits may heavily influence the RF performance of the actual circuit. The same holds for the packaging and choice of substrate, for example, a high-ohmic bulk or low-ohmic substrate with a high- ohmic epitaxial layer on top of that. Because wavelengths can be in the same range as the sizes of the circuit, interconnect wire lengths also play an important role. The choice of antenna also influences the RF performance nowadays. This chapter highlights a few of the most important problems and their solutions.

2.1 ANTENNAS

Basically, a power amplifier can be considered as an AC source, that is connected to an antenna. The resulting current delivered by the power amplifier has a component in-phase with the applied voltage, hence the impedance of the antenna must have a real part. The antenna therefore draws power from the amplifier. An efficient antenna will let most of this power flow in the form of electromagnetic waves. Only a small fraction of the power will be dissipated by the antenna due to ohmic heating. The impedance of the antenna will also have a reactive part. If the antenna is resonant, this reactance is zero at the operating frequency.

In RF design, the antenna for the low noise amplifier in the receive path is mainly treated as a source impedance in series with a voltage source. Alternatively, the antenna acts as a load impedance for the power amplifier in the transmit path. Depending on the realization of the antenna, the impedance value may fluctuate significantly as a function of the environment. Small an-

tenna sizes have low resistive impedance levels and are very sensitive to the environment. For instance, moving your hand from a position close to the antenna to far away from it, may let the resistive part of the impedance fluctuate between 30% and 40% from its nominal value. For a 50 Ω antenna, this means that the actual impedance varies from 30 Ω to 70 Ω.

The gain of a transmitting antenna is defined as the ratio of the power density to the maximum power density. The gain G for an isotropic antenna is given as

$$G = \frac{4\pi r^2 S}{P_t} \tag{2.1}$$

where S is the power density of the antenna, r the distance from the antenna and P_t the transmitted power. It is common to assume that an antenna is lossless; the losses must be small to have a reasonable approximation.

One antenna type that is very often used in the radio community is the dipole antenna. The dipole antenna consists of two equal arms of a certain length, configured in the same plane. Assume that the receive antenna is such a dipole of total length l, see Figure 2.1. Let θ be the angle from the dipole axis for the incident electric field E. Then the effective length is defined as

$$h = \frac{l}{2} \sin \theta \tag{2.2}$$

and the peak voltage at the terminals of the two wires is derived as follows

$$V_p = hE \tag{2.3}$$

For receive antennas, we talk about effective area instead of gain, to define the ratio of the available power P_r to the maximum power. The available power can be derived as follows

$$P_{av} = \frac{|V_p|^2}{8R} = \frac{|hE|^2}{8R} \tag{2.4}$$

where R is the antenna resistance. The effective area can now be derived as

$$A = \frac{|hE|^2}{8RS} = \frac{|h|^2 S\eta_0}{4RS} = \frac{|h|^2 \eta_0}{4R} \tag{2.5}$$

where η_0 is the characteristic impedance of a transmission line in free space, $\eta_0 = \sqrt{\mu_0/\epsilon_0} = 377 \ \Omega$.

Assuming the gain and effective area of a transmit antenna defined by G_t and A_t is similar to the receive antenna, i.e. G_r and A_r, then the following relationship holds

$$G_t A_r = G_r A_t \tag{2.6}$$
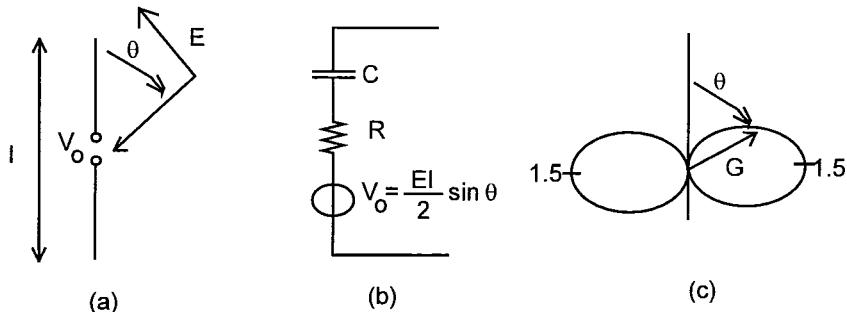
Figure 2.1. Dipole antenna with its effective length (a), its Thevenin equivalent circuit (b) and the polar plot of the gain pattern (c).

independently of the position of the antenna in space or their type. This means that the ratio of gain to effective area is a universal constant, derived as follows

$$G/A = 4\pi/\lambda^2 \tag{2.7}$$

where λ is the wavelength.

The radiation resistance R_r of the dipole can be derived as

$$R_r = 20\pi^2 \, (l/\lambda)^2 \tag{2.8}$$

where l is the total length of the antenna. Using (2.5), the effective area is found to be

$$A = \frac{3\lambda^2}{8\pi} \sin^2 \theta \tag{2.9}$$

and therefore, using (2.7),

$$G = 1.5 \sin^2 \theta \tag{2.10}$$

as is depicted in Figure (2.1). If we consider the dipole as an open-out transmission line (see later in this chapter), then the dipole has a resonance when the total length of the dipole is half a wavelength, or when each arm is a quarter wavelength. The resonant length for practical dipoles is actually a little shorter, at 0.48λ. From (2.8) for the resonant resistance, we obtain 49 Ω, which makes it convenient for connecting to a 50 Ω input impedance circuit.

Obviously, the dipole antenna is not the only antenna configuration. Removing one arm of the dipole and placing the remaining arm vertically above a ground plane results in a monopole antenna. The effective length doubles compared to the dipole due to reflections off the ground plane which double the

component of the electric field along the wire. The resistance would increase by a factor four according to (2.8). However, no power is received for angles below the ground plane, which halves the angle. The resistance is therefore then doubled compared to the dipole. Adding wires at the top of the monopole allows the potential of the antenna to reach the potential at the top, rather than half way up. This so-called "loading" results in a doubling of the effective length and a quadrupling at the radiation resistance.

2.2   BOND WIRES

The bond wire is the interface from the circuits on the silicon to the outside world. The bond wire connects the bond pad to the package pin which, in turn, is connected to the PCB. For RF applications, the bond wire cannot be treated as a simple wire, but must be considered as an inductor for which we should calculate the inductance value. This value is determined by the package used, die size and other parameters. Some basic information on bond wire shapes and dimensions as well as inductance calculations are given below. The anticipated Q of the bond wire is also calculated, and an estimate of the expected inductance variation is given.

The general picture shown in Figure 2.2 is valid for a bond wire that is connected from the die to the lead frame. In the middle of the figure is the die. The die thickness is in the range of 300 to 400 μm. The bond pads are placed at a distance z from the die edge. The die is placed in the center of the die pad or cavity. For a general package like the LQFP64 package, the cavity size is 6600 μm (not drawn to scale in the figure). The distance from the edge of the die to the lead frame is x. The minimum value of x is 4% of the cavity size, which is 4% of y in the figure, with a minimum value of 125 μm. As y is 6000 μm in our case, x is derived from:

$$2x = 6600 \mu m - y \rightarrow x = 300 \mu m$$

As 4% of 6600 μm is 264 μm, x is indeed more than 4% of the cavity size. As a rule of thumb, the die placement accuracy is 100 μm. This means that x can have any value between 200 and 400 μm. From the die to the lead frame, the bond wire is first raised by 200 μm. The shape is as shown in the Figure 2.2, where the bond wire arrives on the lead finger at a 30° angle. The distance from the lead frame edge to the position where the bond wire is attached is l_{lead}. This distance depends on the pin number in the package.

The projected bond wire length can now be derived from $l_{lead} + x + z$, as indicated in Figure 2.2. For the real bond wire length, the shape of the bond

$l_{projected} = l_{lead} + x + z$

Bondwire

200 μm 200 μm

30°

Lead finger

Lead frame

Lead frame

$l_3 = 380\mu m/sin(30°)$
$= 760\mu m$

$l_2 = 380\mu m$

$l_4 = 200\mu m$

30°

$l_{bondwire} = l_{projected} - l_1 + l_3 + l_4 = l_{projected} - 658.2\mu m + 760\mu m + 200\mu m = l_{projected} + 301.8\mu m$

$l_1 = 380\mu m/tan(30°)$
$= 658.2\mu m$
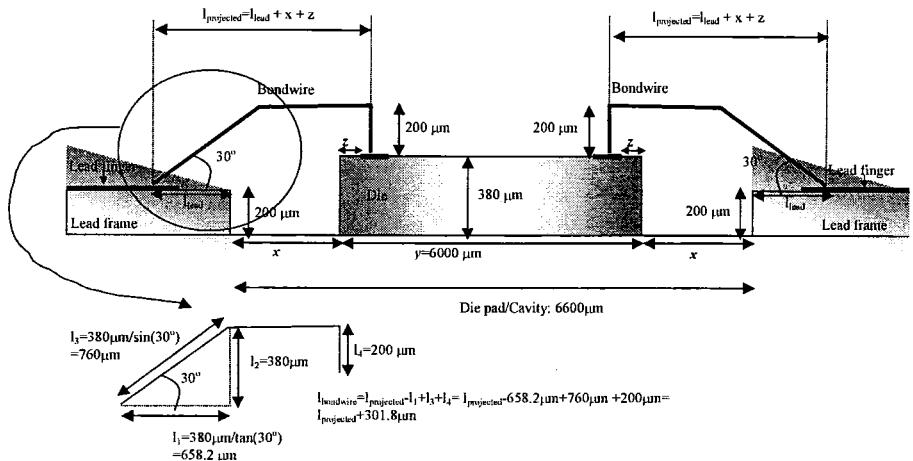
Figure 2.2. Shape and dimensions for bond wires from a die to a lead frame.

wire has to be taken into account. We can see that the real bond wire length $l_{bondwire}$ can be derived from $l_{projected} + 300$ μm.

For a bond wire that is bonded from die to die, we anticipate the same shape as depicted in Figure 2.2. This means that the bond wire is raised by 200 μm, then runs in parallel to the die surface, and finally lands on the die again at a $30°$ angle.

Two bond wires are shown schematically in Figure 2.3, which can be considered to be a general configuration in a standard package. The following first-order equations can be used to calculate the inductance of the bond wire:

$$L_1 = L_2 = \frac{l}{5}\left(\ln\left(\frac{2l}{5}\right) - 0.75 + \frac{r}{l}\right) \tag{2.11}$$

$$M_{12} = M_{21} = \frac{l}{5}\left(\ln\left(\frac{l}{d} + \sqrt{1 + \left(\frac{l}{d}\right)^2}\right) - \sqrt{1 + \left(\frac{d}{l}\right)^2} + \frac{d}{l}\right) \tag{2.12}$$

where L_i is the inductance of bond wire i in nH, M_{ij} is the mutual inductance between bond wires i and j in nH, l is the bond wire length in mm, r is the radius of the bond wire cross-section in mm, and d is the distance between the two conductors in mm. As an example, a standard bond wire has a radius of 16 μm and the material is gold.
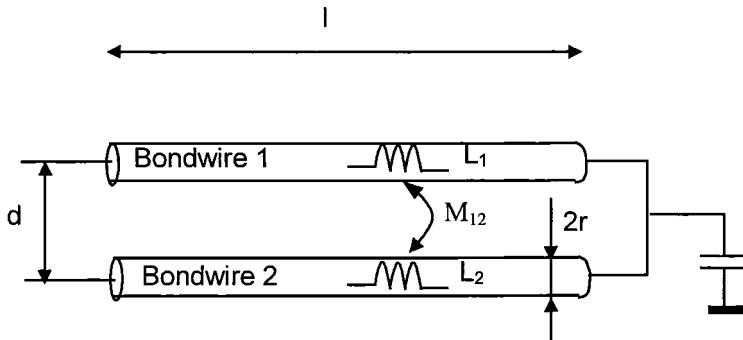
Figure 2.3.   Two bond wires and their mutual inductance.

Using equation (2.12), the inductance of a bond wire can be calculated. The quality factor Q of the bond wire can be calculated from:

$$Q = \frac{\omega L}{R_{series}} \qquad (2.13)$$

where R_{series} is the series resistance of the bond wire. This resistance can be calculated from the following well-known equation

$$R_{series} = \frac{\rho l}{A}, \ A = \pi r^2 \qquad (2.14)$$

where ρ is the conductance of gold, which equals 2.35 $\mu\Omega{\cdot}$cm, l is the wire length and A is the area of the wire cross-section. An ohmic series resistance of 0.03 Ω/mm wire is obtained for a radius of 16 μm.

At higher frequencies, the ohmic series resistance in the wires will increase due to the skin effect. This effect causes the current to flow through the outside part of the inductor only. The skin depth δ is defined as the thickness of a hollow inductor with the same high-frequency resistance

$$\delta = \sqrt{\frac{2\rho}{\omega\mu}} \qquad (2.15)$$

where μ is the permeability of gold, which equals $4\pi \cdot 10^{-7}$ F/m. This means that at 2.4 GHz, the skin depth is 1.57 μm. The area of the cross-section through which current flows can now be estimated by $2\pi r\delta$, where r is the wire radius of 16 μm. This area is $1.58 \cdot 10^{-4}$ mm^2. The ohmic series resistance now becomes 0.15 Ω/mm. It should be noted that the inductance of the bond wire is also influenced by the skin effect. However, the increase in R_{series} is found to be a lot larger than the decrease in L.

Many designers use the well-known "1 mm equals 1 nH" bond wire rule-of-thumb. Indeed, it turns out that this rule is pretty accurate if compared to the values obtained using the above derived equations.

Some estimations can be made with regard to the spread. Due to bond wire bending (both in a horizontal and vertical direction), as well as variations in bond wire height and wire diameter and also due to imperfect modelling, the anticipated inductance variation is 6%. On top of this, we have to take the die placement inaccuracy into account, as mentioned above. For each coil structure, the relative influence of the die placement accuracy of 100 μm can be calculated from the change in bond wire length due to the variation in x. The total coil inaccuracy can then be derived as

$$\Delta L(\%) = \sqrt{6^2 + \Delta L^2_{dieplacement}} \qquad (2.16)$$

where $\Delta L_{dieplacement}$ is the relative change in the inductance value due to the 100 μm variation in x.

2.3 TRANSMISSION LINES

In this section we will highlight a few general properties of transmission lines. Furthermore, we will discuss how transmission lines can be used to realize impedance matching. Finally, we will say some words on how transmission lines can be realized on silicon, i.e. microstrip lines and coplanar lines.

2.3.1 General Theory

Analog designers consider their circuits to be built up with lumped components connected by lines with a zero-length. This assumption holds as long as the circuit has dimensions that are large compared to the wavelength. This means that the frequencies used may not be too high. For radio frequencies, this assumption no longer holds, the length of the wires, i.e. the interconnects, begin to play an important role. The wires should then be treated as transmission lines, possibly modeled by lumped elements with length-dependent component values.

Consider Figure 2.4 where a piece of transmission line is drawn. Every increment δx of a transmission line contributes a series inductance $L\delta x$ and shunt capacitance $C\delta x$ (assuming a lossless transmission line[1]). The increment

---

[1] In the case of a lossy transmission line, the inductor in Figure 2.4 will have a series resistance $R\delta x$ and the capacitor will have a parallel conductance $G\delta x$. The resistance represents the ohmic losses of the metal conductors, while G represents the dielectric loss. The characteristic impedance of the cable in this lossy case is defined as $Z_0 = \sqrt{(R + j\omega L)/(G + j\omega C)}$.
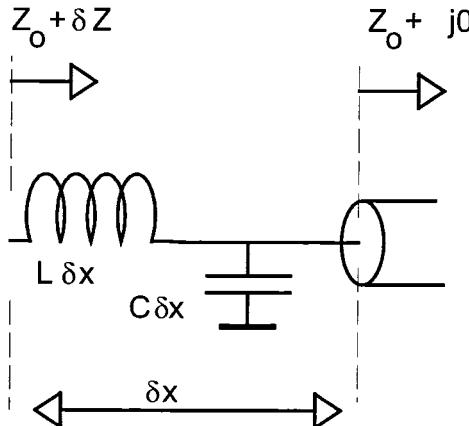
Figure 2.4.   Transmission line with another infinitesimal section added to it.

δx contributes to the characteristic impedance Z_0 as

$$Z_0 + \delta Z = j\omega L \delta x + \frac{Z_0}{j\omega C \delta x Z_0 + 1} \tag{2.17}$$

In the extreme case that δx goes to zero, this equation results in

$$Z_0 + \delta Z = Z_0 + j\omega \delta x (L - Z_0^2 C) \tag{2.18}$$

The increment of Z_0 equals zero when $Z_0 = \sqrt{L/C}$, i.e. the impedance remains constant as the transmission line is lengthened [1] [2].

Assume now that an input $\exp(j\omega t)$ is applied at the input of the line. A voltage drop δV across the incremental length δx of the line will occur, resulting in the differential equation

$$\frac{dV}{dx} = -j\omega \frac{L}{Z_0} V = -j\omega \sqrt{LC} V \tag{2.19}$$

with solution

$$V = V_o \exp(-j\omega \sqrt{LC} x) = V_o \exp(-jk_p x) \tag{2.20}$$

where k_p is known as the *propagation constant* and k_p can also be defined as $2\pi/\lambda$ where λ is the wavelength. Including the time dependency of the input signal leads to the final expression of the voltage at point x

$$V = V_o \exp j(\omega t - k_p x) \tag{2.21}$$
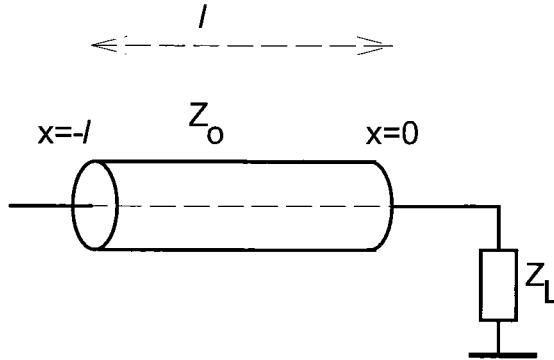
Figure 2.5. Transmission line of length l terminated by a load Z_L.

It is important to notice that we may write $\exp(-jk_px)$ as $\exp(-\alpha x - j\beta x)$ where α is the *attenuation factor* and $\omega/\beta = v$ is the velocity of the wave. Consequently, the velocity is dependent on the frequency, a property called *dispersion*[2].

2.3.2 Impedance Matching using Transmission Lines

To start the discussion on impedance matching by means of transmission lines, let us recall the definition of the reflection coefficient Γ

$$\Gamma = \frac{Z - Z_0}{Z + Z_0}$$

(see Section 1.1 for more details).

Let us assume that some source at the starting position of a transmission line produces an incident wave to the right and that the load at the other end of the line causes a reflection wave to the left. This situation is depicted in Figure 2.5. At any point x somewhere along the line with length l, the voltage on the line and the current through the line are derived as

$$V(x) = \exp(-jk_px) + \Gamma \exp(jk_px) \tag{2.22}$$

$$I(x) = \frac{1}{Z_0 \left(\exp(-jk_px) + \Gamma \exp(jk_px)\right)}$$

---

[2] A nice historical example of dispersion is the first transatlantic telegraph cable. The cable had the following element values: $L = 460$ nH/m, $C = 75$ pF/m and $R = 7m\Omega$/m. At 12 Hz we have $\alpha = \sqrt{\omega RC/2} = 4.4 \cdot 10^{-3}$ nepers/km and $v = \sqrt{2\omega/(RC)} = 17.000$ km/s. The loss of the entire line was therefore $\alpha l = 140$ dB and the delay was $l/v = 210$ ms. At 3 Hz a change of a factor 2 in the loss and delay is achieved, making 70 dB and 240 ms respectively. The 12 Hz component arives 210 ms ahead of the 3 Hz component and attenuates by an extra 70 dB.

The input impedance at $x = -l$ is now defined as

$$Z_i = \frac{V(-l)}{I(-l)} = \frac{\exp(-jk_p(-l)) + \Gamma \exp(jk_p(-l))}{\exp(-jk_p(-l))/Z_0 + \Gamma \exp(jk_p(-l))/Z_0} \qquad (2.23)$$

yielding

$$Z_i = Z_0 \frac{Z_L + jZ_0 \tan(k_p l)}{Z_0 + jZ_L \tan(k_p l)} \qquad (2.24)$$

after some manipulations. Several observations can be made after looking at this expression:

- If the line is terminated with the characteristic impedance, that is $Z_L = Z_0$, (2.24) results in $Z_i = Z_0$.

- If the line is short-circuited at $x = 0$, then $Z_i = jZ_0 \tan(k_p l)$. Depending on the length of the line, the input impedance is inductive or capacitive. For instance, for $k_p l < \pi/2$, the line is purely inductive. Because $\tan(k_p l) = \tan(\omega l/v)$ is *not* proportional to ω, the term "inductive" or "capacitive" does not refer to a lumped inductor or lumped capacitor, as these elements are proportional to ω. Only for small pieces of line length l, does $\tan(k_p l) \approx \omega l/v$, and the line input impedance becomes a good approximation of a lumped element.

- If the line is open-circuited, then $Z_i = Z_0/(j \tan(k_p l))$. Again, depending on the length of the transmission line, the input impedance is capacitive or inductive. If $k_p l < \pi/2$, then the input impedance is capacitive.

- Suppose that the length is a quarter wavelength, $l = \lambda/4$. Consequently, $k_p l = \pi/2$ and (2.24) therefore becomes $Z_i = Z_0^2/Z_L$ or

$$\frac{Z_i}{Z_0} = \frac{Z_0}{Z_L} \qquad (2.25)$$

If the impedances are normalized to the characteristic impedance Z_o, then (2.25) tells us that a quarter wavelength transmission line inverts the load impedance. Another interesting aspect of this line length is the fact that $Z_0 = \sqrt{Z_i Z_L}$. Consequently, if the source and load impedance are known, the characteristic impedance can be chosen such that all of the available power can be delivered to the load.

The Smith chart is often used to design the impedance matching network. In the Smith chart, the reflection coefficient Γ is drawn as a function of the
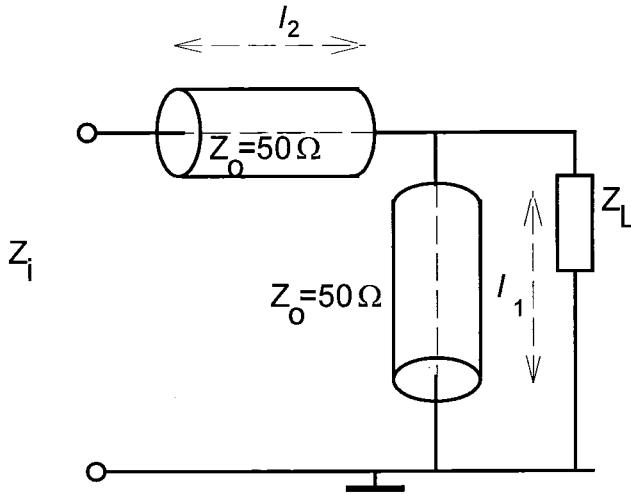
Figure 2.6. A matching network using short-circuited stub of length l_1 in parallel with a load followed by a series transmission line of length l_2. The characteristic of both lines is 50 Ω.

frequency[3]. We can also obtain the length l of the transmission line as function of λ from the Smith chart. As an example, consider a matching network to a load $Z_L = 50$ Ω for a given input impedance Z_i. Such a matching can be realized with a configuration in which a short-circuited stub is connected in parallel with the load, followed by a series transmission line as shown in Figure 2.6. Due to the parallel setting, we can best use the y-Smith chart (series of y-parameters instead of parallel z-parameters). The normalized load admittance is $y_L = 1/z_L = Z_0/Z_L = 1$. According to (2.25) the short circuited transmission line gives an impedance

$$Z = jZ_0 \tan(kl) \qquad (2.26)$$

Normalizing this to Z_0 yields $Z = j\tan(k_p l) = \pm j/b$, and thus the impedance is pure capacitive or inductive, depending on the length. For the equivalent admittance this yields $y = \pm jb$, hence the replacement admittance of the parallel network is derived as $y_p = y_L + y = 1 \pm jb$. The design of the matching network follows a few steps:

---

[3]As the reflection parameter Γ is a bilinear operator on Z, the loci of constant resistance are circles in the Γ-plane with the centres at the real axis. Similarly, the loci of constant inductance (capacitance) are circles with their centre along a vertical line above (below) the real axis.
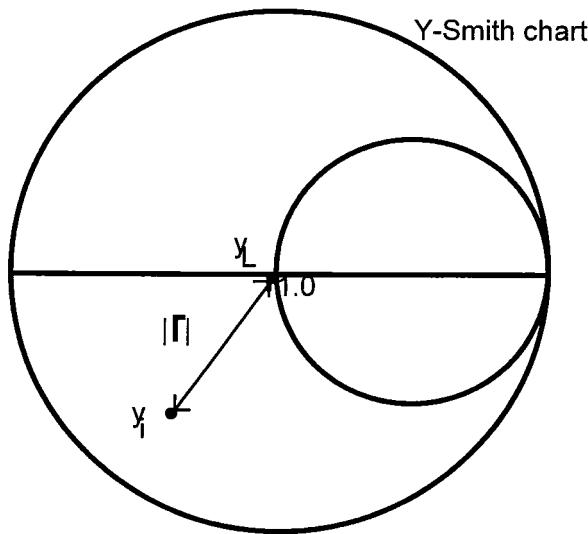
Figure 2.7.   Location of the admittances in the y-Smith chart.

1. Mark the admittances y_i, and y_L in the normalized y-Smith chart (see Figure 2.7).

2. The series transmission line converts the admittance y_p into y_i. Consequently, both admittances must have the same reflection coefficient $|\Gamma|$, and therefore the value b must be selected so that y_p and y_i are on a constant $|\Gamma|$-circle. This is depicted in Figure 2.8.

3. Length l_1 produces the conversion from y_L into y_p, and l_2 is needed to convert y_p into y_i. Both lengths can be found as a function of the wavelength λ, as arcs around the Smith chart, as shown in Figure 2.9.

Similar techniques can be applied to match a given input impedance to a certain output impedance. One can also fix the length of the transmission line and use the characteristic impedance as parameter according to (2.25).

2.3.3   Microstrip Lines and Coplanar Lines

In RF design, the transmission lines discussed in the previous section are realized using microstrip lines or coplanar strip lines.

A microstrip line is a transmission line consisting of a metal strip and a ground plane with the dielectric medium in between, as can be seen in Figure 2.10. The dielectric constant ε of the substrate is defined by the product of
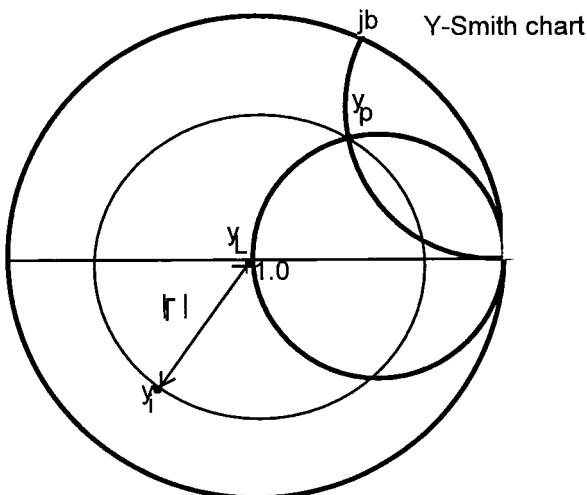
Figure 2.8.   Obtaining y_b using the constant $|\Gamma|$-circle.
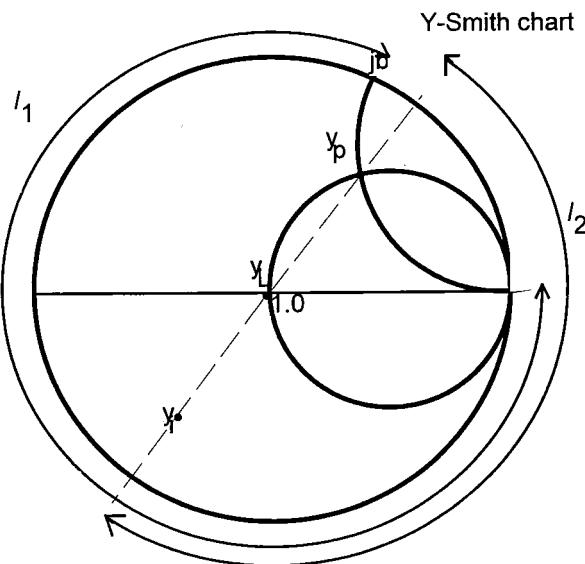
Figure 2.9.   Length of the transmission line can be obtained from the arc length in the Y Smith chart.

the relative dielectric constant ε_r of the material used and the absolute constant $\varepsilon_0 = 8.854 \cdot 10^{-12}$ F/m. As the electromagnetic fields are not entirely contained in the substrate, the propagation mode in the strip line is not a pure
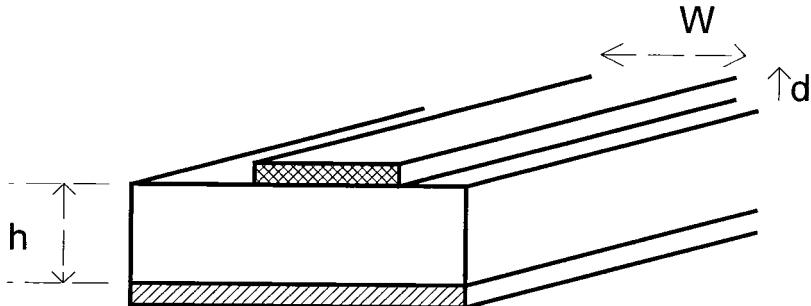
Figure 2.10. Microstrip line with thickness d, width W. The dielectric substrate has height h. Beneath the substrate is the ground plane.

transverse electromagnetic mode. Hence, the concept of an effective relative dielectric constant ε_{eff} is normally used. Formulas for the obtained characteristic impedance Z_o and ε_{eff} are obtained, depending on the ratio between the height of the dielectric substrate and the width of the strip line, W/h [2]. For instance, for $W/h \leq 1$, the following set of relationships define the characteristic impedance, assuming zero or negligible thickness d of the strip line,

$$Z_0 = \frac{60}{\sqrt{\varepsilon_{eff}}} \ln\left(8\frac{h}{W} + 0.25\frac{W}{h}\right) \tag{2.27}$$

$$\varepsilon_{eff} = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2}\left[\left(1 + 12\frac{h}{W}\right)^{-1/2} + 0.04\left(1 - \frac{W}{h}\right)^2\right] \tag{2.28}$$

where $d/h < 0.005$. As an example of the applications of microstrip lines, we can consider the breadboard of Figure 2.11. To obtain optimal performance, the input and output impedances of a low noise amplifier have been matched to 50 Ohms. On the right side, the $\frac{1}{4}\lambda$ lines are visible at the top (V_{dd}) and bottom (V_{gs}) of the breadboard. These lines are used for biasing the amplifier. For RF, the $\frac{1}{4}\lambda$ line will ensure that the circuit sees an open circuit to the biasing circuit, so no loading effects from the biasing circuit to the amplifier can take place. The transmission lines with the stubs connecting "OUT" and "IN" to the amplifier are based on the matching circuit as shown in Figure 2.6. This amplifier will be discussed in more detail in the chapter on low noise amplifiers.

When the microstrip lines are laid out on silicon, designers use lumped element replacement circuits to model microstrip lines. A very accurate model for a metal-over-substrate microstrip line is given in Figure 2.12. The model is valid for line lengths of up to 1 mm [3]. The component values are expressed
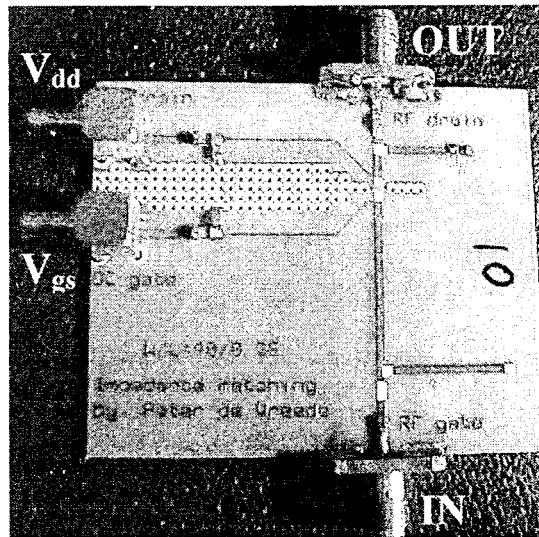
Figure 2.11. Microstrip lines on a printed circuit board to obtain an optimal impedance match between the circuit on chip and the outside world with its 50 Ohms impedance.
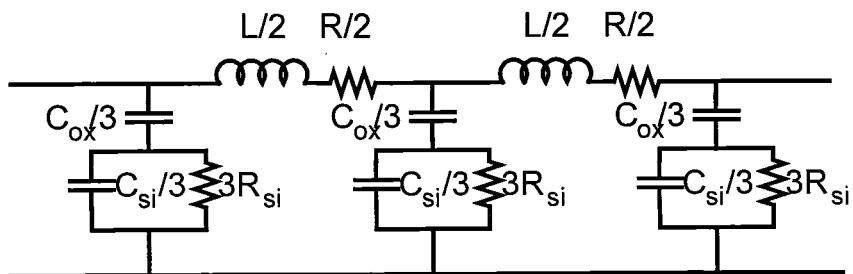
Figure 2.12. Lumped circuit for a metal-over-substrate microstrip line.

as functions of the width w, length l, the dielectric medium thickness h (i.e. the oxide thickness between metal 3 and the Si substrate), the metal 3 thickness d, the metal 3 sheet resistance ρ_{m3}, the resistivity of the Si substrate R_{si}, and the thickness of the substrate h_{sub},

$$D_s = 0.2235(w + d) \tag{2.29}$$

$$p_1 = \log\left(\frac{l}{D_s} + \sqrt{1 + l^2/D_s^2}\right) \tag{2.30}$$

$$D_m = 2\,(h_{sub} + h) \tag{2.31}$$

$$p_2 = \log\left(\frac{l}{D_m} + \sqrt{1 + l^2/D_m^2}\right) \tag{2.32}$$

$$L_s = L_{cor}l\left(p_1 - \sqrt{1 + D_s^2/l^2} + \frac{D_s}{l}\right) \tag{2.33}$$

$$L_m = 200l\left(p_2 - \sqrt{1 + D_m^2/l^2} + \frac{D_m}{l}\right) \tag{2.34}$$

$$L = (L_s - L_m) \cdot 1e^{-9} \tag{2.35}$$

$$R = \rho_{m3}\frac{l}{w}\left(1 + k\left(\frac{f}{1e^9}\right)^{1/2}\right) \tag{2.36}$$

$$C_{ox} = \varepsilon_{r,ox}\varepsilon_o l\left(1.15\frac{w}{h} + C_{fr}\left(\frac{d}{h}\right)^{0.222}\right) \tag{2.37}$$

$$C_{si} = C_{sil}l \cdot 1e^3 \tag{2.38}$$

$$R_{si} = \varepsilon_{r,si}\varepsilon_o\frac{R_{si}}{C_{si}} \tag{2.39}$$

The parameters k, C_{fr}, C_{sil} and L_{cor} are fitting parameters. For long lines, the ground plate of the microstrip line should be replaced by metal 1 or polysilicon layers to reduce the resistive losses.

Coplanar strip lines are used when the frequencies are in the lower RF band, i.e., up to between 5 and 10 GHz. As an example of a coplanar strip line, consider Figure 2.13 where the electromagnetic waves vary only in the horizontal plane. The dielectric is normally the oxide, and metal layer 1 is used for the ground and signal lines.

2.4   BOND PADS AND ESD DEVICES

The bond pad and ESD devices are normally not a point of discussion in digital designs. The designers do not ask themselves if these components might influence the performance of their circuits, because it is not the case. For RF applications, however, this assumption no longer holds. For radio frequencies, both the bond pad and ESD device can be considered to be bad capacitors; they have a bad quality factor and normally have a large capacitance value. They consequently have a major impact on the performance of the RF circuitry. The value needed for the quality factor is determined by the quality factor of the input device following the bond pad and the ESD device. The input capacitance of a bipolar device has a much lower quality factor than for a CMOS device. The quality factor needed for the bond pad and ESD device are therefore much lower for bipolar designs than for CMOS designs. A quality factor of 10 at 1
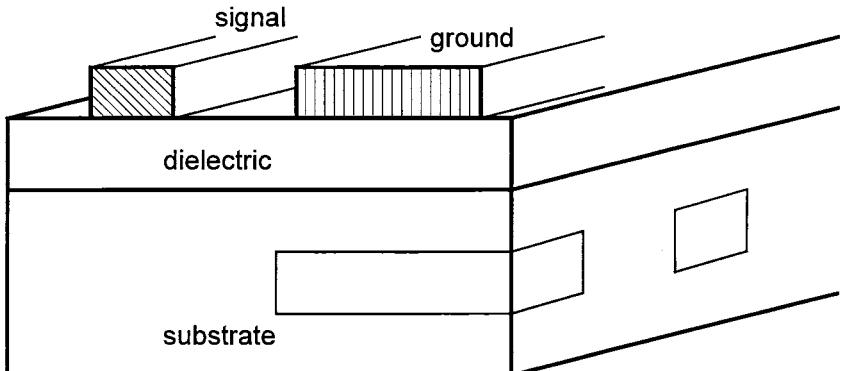
Figure 2.13. Coplanar strip line. The laminates in the substrate are to lower the currents flowing in the substrate. These currents can be considered as losses.

GHz is good enough for bipolar designs, whereas a designer for CMOS designs needs a quality factor better than 50. The bond pads and ESD device structures found in digital-oriented libraries do fulfil the requirements needed in bipolar RF designs, but not for CMOS designs.

In this section we discuss bond pads and ESD devices for RF CMOS applications.

2.4.1 Bond Pads

A bond pad is normally a large metal plate to which the bond wire can be connected in the vertical direction. The connection to the circuits is made in the horizontal direction using one or more metal layers. A typical bond pad is shown in Figure 2.14. Normally, all metal layers are connected together to form the metal plate. By using several metal layers, the plate gains a certain mechanical strength, necessary during bonding. The bonding machine will attach the bond wire to the metal plate and, during this process, the plate can crack if it does not have a certain stiffness. The peeling-off of the metal layers is prevented by using several metal layers; the vias "glue" these layers into one thick package. It would be very difficult for a metal layer to peel off from such a package. The capacitance value is now determined by the oxide thickness t_{ox} between the bottom metal plate and the substrate, which serves as the ground plane. Typical values for a geometry such as the one in Figure 2.14 are a capacitance value of 300 fF, a series resistance of 20 Ω and a quality factor of 30 at 900 MHz [4], [5].

The quality factor is too low for CMOS RF applications. Besides, the bond pad has no shielding against substrate bounces. To improve the RF performance,
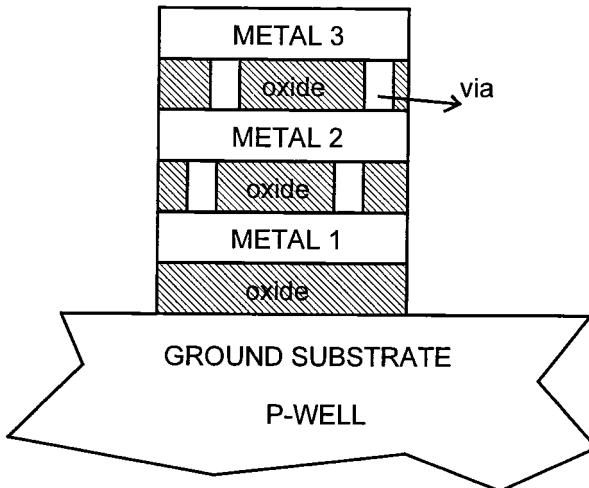
Figure 2.14.   Bond pad made of several metal layers connected to each other. The white regions between the metal layers represent the vias.

the capacitance value first needs to be decreased. This can be realized by a larger t_{ox}, which can be obtained by removing the lower metal layers. By using only the top metal layers, the capacitance value can be decreased to a value of around 50 fF, as can be seen in Figure 2.15. The series resistance also needs to be decreased. In the structure of Figure 2.14 the resistance is determined by the p-well. A buried polysilicon layer has a much lower resistance, and a side effect is that this layer will also shield the bond pad from substrate bounces. The resulting bond pad has typical specifications such as a series resistance value of 4 Ω and a quality factor of 500 at 900 MHz. These specifications mean that this bond pad can be considered as a near-perfect capacitor for RF designs: it has no power losses and a minimum loading for the input device.

2.4.2   ESD Devices

Electrostatic Discharge (ESD) structures are normally included at the periphery of circuits to prevent circuit damage by external electrostatic pulses [6]. These electrostatic pulses can occur everywhere, and are not always caused by human actions. One commonly accepted test model for these pulses is the Human Body Model (HBM) [7]. According to this test model, a voltage pulse of 2 kV is generated and provided to the device under test (DUT) by means of a switch action. A current of at least 1 A must be handled by the DUT and it should be still operational after this large current. If this is so, the DUT is said to be ESD proof under HBM conditions.
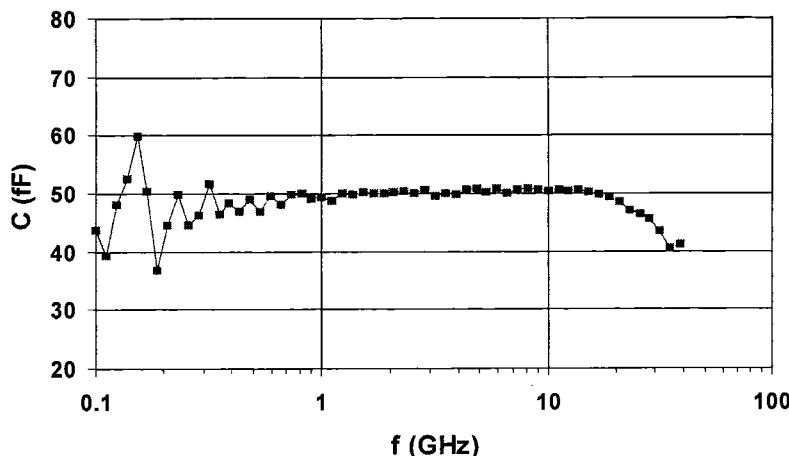
Figure 2.15.   The measured capacitance versus frequency for a dedicated RF bondpad.

If no special circuitry is placed in front of the actual circuit, it is clear that the high voltage and large current will damage the circuit. We therefore need an ESD protection circuit. The ESD circuitry must not influence the performance of the actual circuit, however. Loading effects and power losses should be kept to a minimum. Consequently, for RF applications in particular, the Q-factor of these ESD structures has to be high enough to prevent that this ESD device absorbs the RF power. Both the ESD-capacitance and series resistance should therefore be low.

In the literature, only a few publications are known on this subject: [8], [9], [10], [11]. They only cover the digital application area, however, and not the RF field.

ggNMOST ESD Device

In CMOS technology, a commonly used ESD protection device is the grounded gate NMOST (ggNMOST), as shown in Figure 2.16. The ESD protection device is placed between the bond pad, where the electrostatic pulse enters the chip, and the actual circuit. The ggNMOST must prevent a large voltage pulse causing the gate voltage of the input transistor of the circuit to rise too much so that this transistor is permanently damaged. This protection must be functional for positive and negative pulses.

To understand the working of a ggNMOST, consider the illustration in Figure 2.17. The NMOST acts like an np-diode, and will start conducting for negative pulses when the diode is set in forward mode. Positive pulses will be handled by the parasitic NPN bipolar transistor, located as shown in Figure 2.17. The
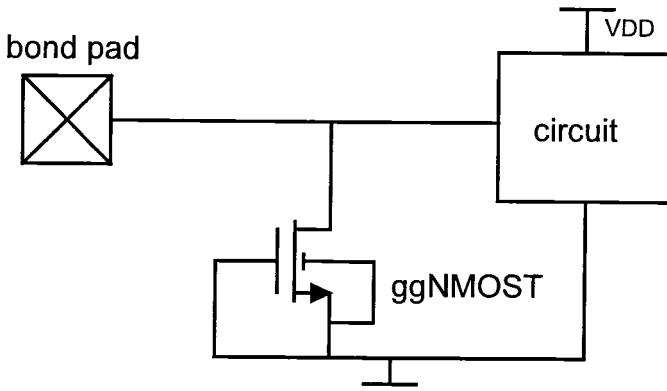
Figure 2.16.   The ggNMOST ESD structure.

positive voltage at the bond pad, V_{in}, can be treated as a reversed bias voltage for the collector-base diode of the bipolar device. There is an upper limit for the reverse bias voltage, otherwise avalanche breakdown will take place [12], [13]. When the electric field built up across the np-junction is high enough, electron-hole pairs are generated. Electrons in the conduction band have gained that much energy so that they can give an electron in the valence band at least that the amount of energy it needs to jump towards the conduction band itself[4]. This process of generating an electron-hole pair is known as impact ionization. If the electric field is high enough, these secondary electrons and holes can again cause impact ionization, and so a process of rapidly increasing numbers of electron-hole pairs is started. Electron and hole currents therefore increase. The pn-diode breaks down when this multiplication process runs away, an avalanche process starts (hence the name avalanche breakdown), and the terminal current rapidly increases.

Clearly, it takes some time before this avalanche breakdown process is activated; the positive pulse at V_{in} needs to pass a certain threshold voltage before the ESD device becomes active. A ggNMOST is better suited to handle negative pulses than positive pulses.

For RF applications, an ESD device can be treated as a capacitor C_{ESD} in series with a resistor R_{ESD}, determining the quality factor the capacitor. For sake of clarity, both components are schematically drawn in Figure 2.17. Obviously, the quality factor should be as high as possible to minimize the

---

[4]Obviously, this process of generating an electron-hole pair can also be initiated by a hole in the valence band.
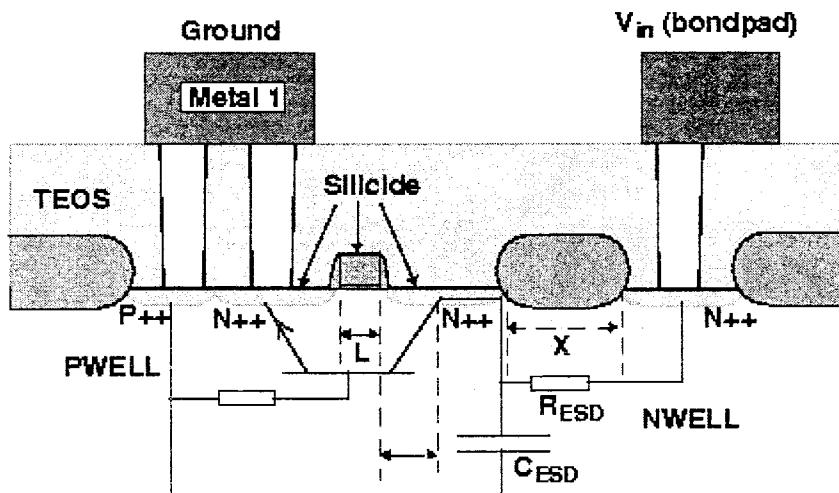
Figure 2.17.   Illustration of a ggNMOST in a p^- epi layer CMOS technology.
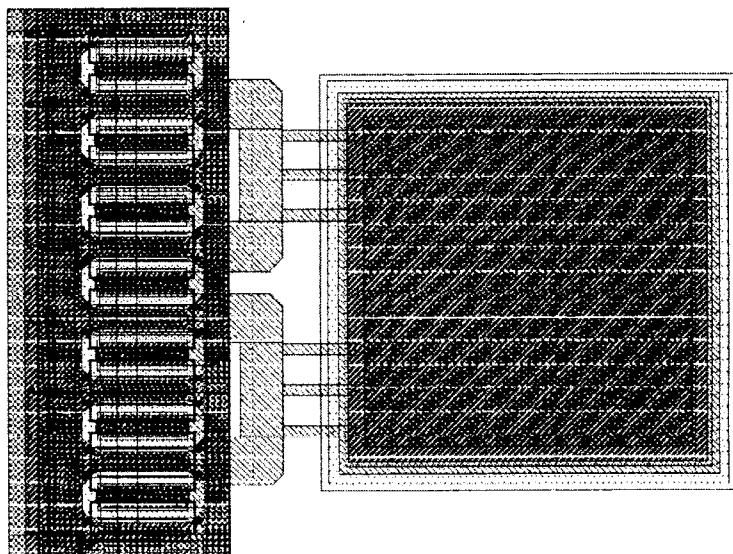
Figure 2.18.   Layout of a ggNMOST ESD device (left) including the bond pad (right). The MOST is laid out in a finger structure.
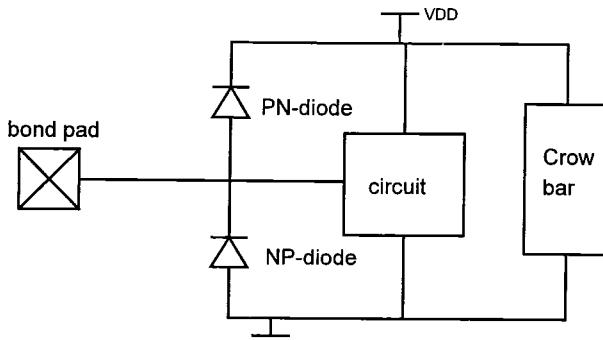
Figure 2.19. ESD device based on two diodes, each operating in forward mode when being activated by a electrostatic pulse at the bond pad.

power loss into the ESD device, and to maximize the power transfer into the circuit behind the ESD device. In addition, the capacitance of the ESD device should be as small as possible; it must not influence the frequency behavior of the circuit. A typical ESD structure including the bond pad is shown in Figure 2.18, where the ggNMOST has a finger like structure due to its geometrical dimensions: $W = 100 \ \mu\text{m}$, $L = 0.25 \ \mu\text{m}$. Performance values for a 0.25 μm CMOS technology are: $C_{ESD} = 270$ fF and $R_{ESD} = 5 \ \Omega$. The quality factor is typically in the range of 40 at 900 MHz.

pn and np-Diode ESD Device

One way to overcome the avalanche breakdown process for negative pulses is to also have a diode which is forward biased for this pulse. This solution is shown in Figure 2.19. The np-diode is forward biased for a negative pulse, and the pn-diode is forward biased for a positive pulse. The difference between the two diodes is that for the pn-diode the p-doped side is connected to the input voltage and must therefore be laid out in a n-well.

The crowbar in Figure (2.19) represents a circuit to handle the large current flowing when the pn-diode is in forward mode. Without a crowbar, the large ESD current (up to 1 A) will flow into the circuit and will damage it. The crowbar must prevent this by short circuiting the circuit. A crowbar normally consists of a trigger circuit, triggered by the ESD pulse, and a huge ggNMOST device which is then turned on.

During HBM testing, several questions have to be solved:

- An important question is, of course: can the ESD device sustain the 1 A current? Figure 2.20 shows the results of an HBM test on an np-diode ESD
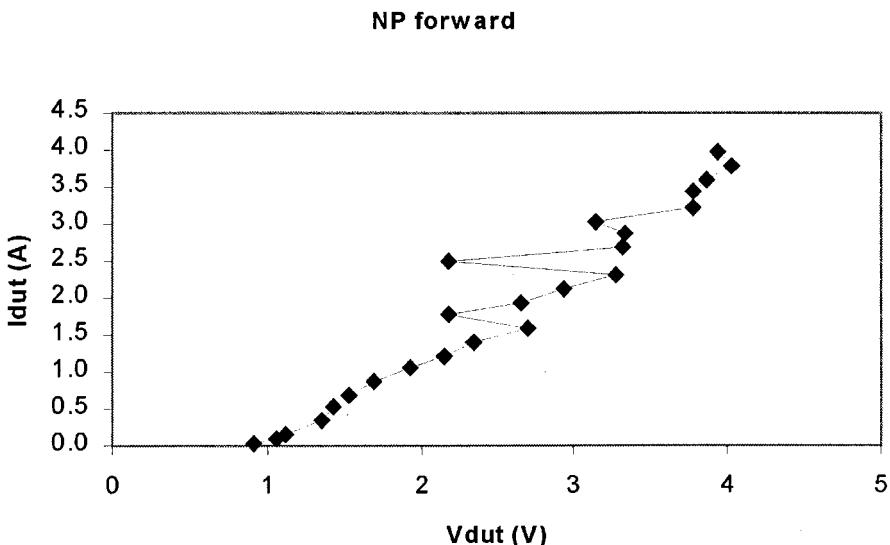
NP forward

Figure 2.20.   Forward current as a function of the forward voltage across the np-diode ESD.

in 0.25 μm CMOS technology. We can see that the ESD device is capable of handling currents of up to 4 A. Even more important is the fact that for the 1 A condition, the voltage drop over the ESD device is only 2 V. This is a very important aspect, since this voltage is also across the gate of the input transistor of the circuit. If this voltage is too large for the technology used, the circuit could therefore be permanently damaged. For example, for a 1.2 V process, the ESD diode of Figure 2.20 will not prevent the circuit from becoming damaged. This damage due to large voltage spikes can be prevented by adding a resistor in the signal path between the ESD and the circuit, as is normally done in digital circuitry. This would complicate RF designs, however, since additional noise is generated and the input resistance is influenced. In a 0.25 μm CMOS technology with a tolerable voltage of 2.5 V, the 2 V voltage drop is not a problem and no additional resistor is necessary in the signal path.

■ Another question to solve is whether the ESD device can still function after an HBM pulse, or can it only handle one such pulse? This question can be answered by looking at the leakage current of the ESD device after handling an HBM pulse. The measured leakage currents of the device used above are
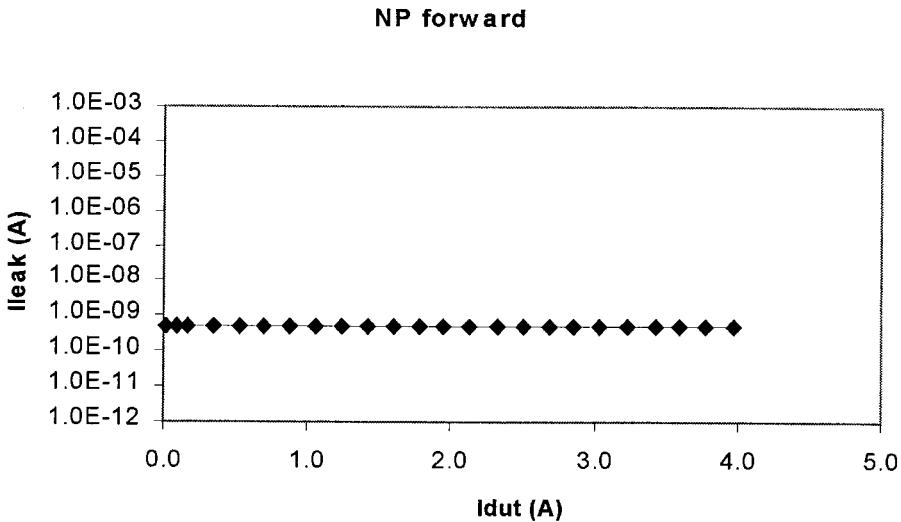
NP forward

Figure 2.21.   The leakage current of the np-diode ESD as used in Fig. (2.20)

shown in Figure 2.21. The leakage current is low for all forward currents
(Idut) and it can be concluded that this device is not damaged during HBM
tests. If it were damaged, then a large current would flow, and the diode
would go into forward operation mode.

- Finally, it is important to know when the ESD device breaks down in reverse
  mode. The test results for this condition are shown in Figure 2.22. The np
  diode can handle up to 35 V in reverse mode before it breaks down and starts
  conducting current at 15 V.

The conclusion from all these measurements is that the np-diode used is a
very good ESD device in the sense that it complies with the HBM rules. For
RF applications, in a similar way as for the ggNMOST ESD device, it is also
important that the quality factor is high, but with a small capacitance value. The
following figures show a comparison between the presented ggNMOST device
and the np and pn-diode devices, all designed in 0.25 μm CMOS technology.
All devices are HBM proof. Measurements of the capacitance values show that
the diode-based ESD devices have a substantially lower capacitance value than
the ggNMOST device (Figure 2.23). The ggNMOST will therefore have more
impact on the input impedance of the circuit, which is not nice from an RF point
of view. Figure 2.24 shows that the pn/np-diodes have a much higher quality
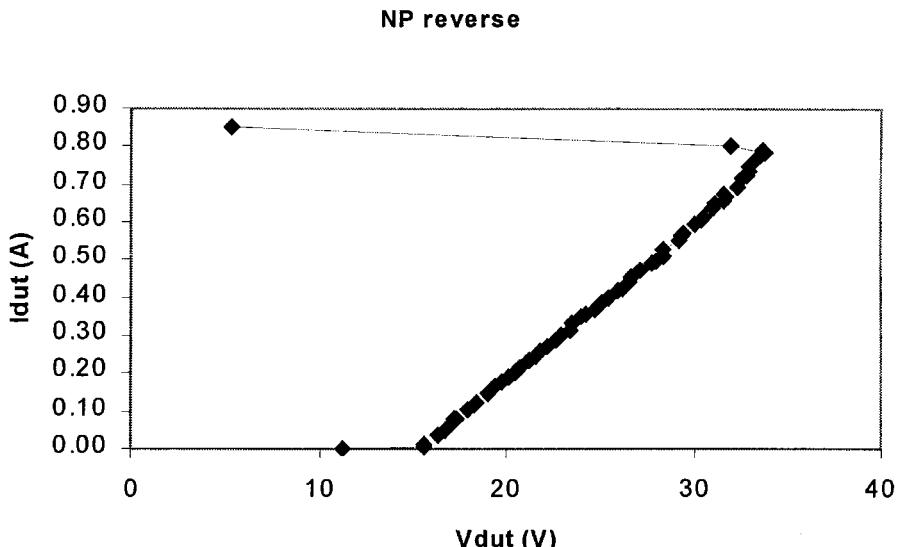
NP reverse

Figure 2.22. Reverse voltage across the np-diode during HBM testing. Breakdown takes place at 35 V.

factor than the ggNMOST device. The quality factor for RF applications in the 1 to 2 GHz range is high enough to allow the ESD device to be treated as a pure capacitance.

From these measurements, we can conclude that ESD devices based on pn and np-diode structures are most suitable for RF applications.

2.5 SUBSTRATE

Given a particular technology, the designer sometimes has the freedom to choose the substrate (or bulk). Although this option is sometimes provided in bipolar technologies (see Silicon-on Anything technology), the type of substrate is mainly a point of discussion in CMOS technologies. Low-ohmic substrates together with a high ohmic epitaxial layer were used in the past; mainly to increase the immunity to the latch-up problem. Low-ohmic substrate has a resistivity of several milli Ohms per cm ($\Omega \cdot$cm). Due to the process scaling and the related lower voltage supplies, high-ohmic substrate can be used without latch-up problems. High-ohmic substrate is available in several gradations, from a few Ohm per cm up to several 100 $\Omega \cdot$cm. As the backgate of a MOS
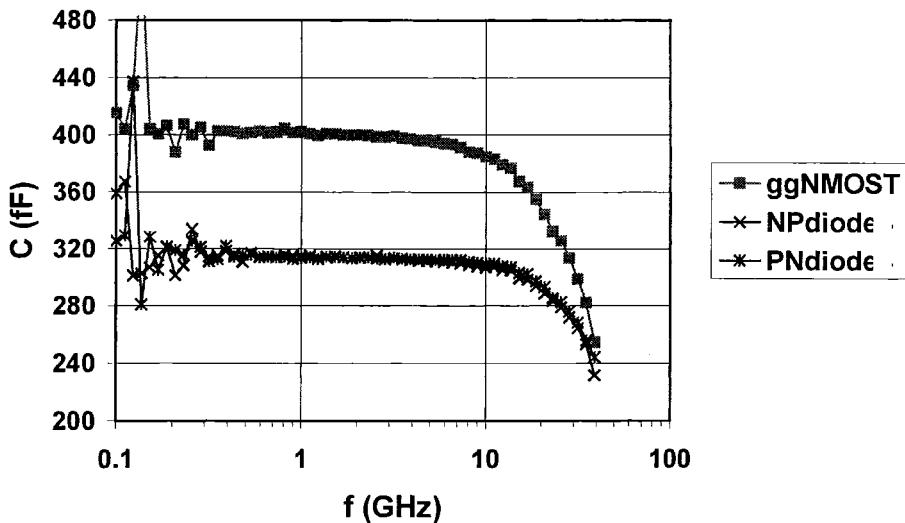
Figure 2.23. Comparison between ggNMOST and np/pn diode ESD devices. The capacitance value also includes the test bond pad.
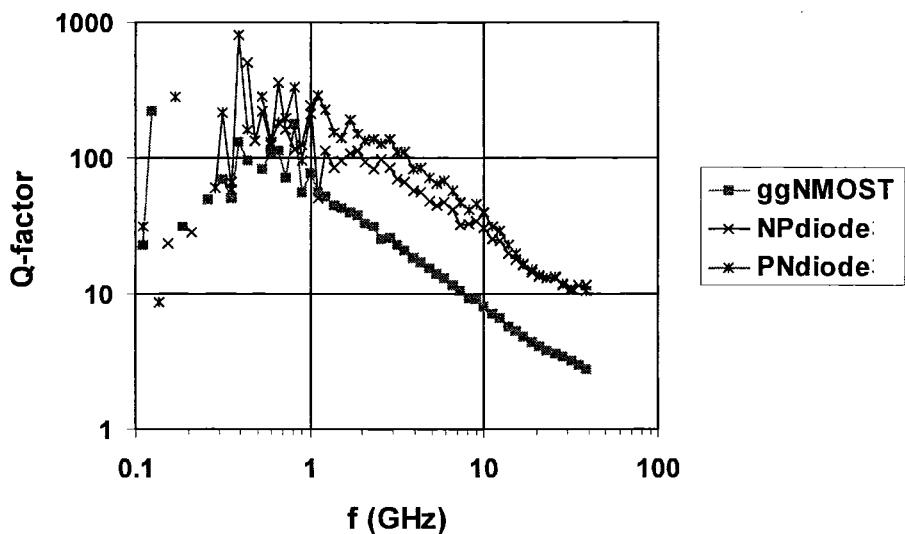
Figure 2.24. Measured Q-factors as a function of the frequency for the different ESD devices as mentioned in the text.

device is connected to the substrate, it is clear that the type of substrate plays a major role in the RF performance of a CMOS circuit.

2.5.1 Substrate Bounces

The substrate is the connecting layer between all the circuits on a single die. If one circuit generates small (often spiky) signals, and hence the name noise, on the substrate, other circuits will be influenced by this noise. The amount of generated noise on the substrate is therefore a main concern in the design, especially when large parts of digital circuitry are laid out on the die. There are several paths to inject noise into the substrate [14]:

- If a switching ground bus is used in the digital domain, switching transients will couple directly into the bulk through the bulk contacts. Parallel summation of all these contacts with their epitaxial resistances provides a very low ohmic path to the substrate.

- The source and drain of a MOS device form a p-n junction to the bulk. This p-n junction realizes a capacitive coupling for which the capacitance of many MOS junctions can be significant; several hundred pico Farad is possible for large digital circuitry. This capacitance can be reduced or even eliminated when the source and bulk are tied together.

- The n-well also introduces a very large p-n junction with the bulk. This causes a capacitance between the positive voltage supply, thus biasing the n-well and the ground line, and biasing the substrate.

- The major source of noise injection is normally due to the bond wires. The MOS logic is switching and produces current spikes through both the positive and ground supply lines, and hence through the supply pins. In digital standard cells, substrate contacts are normally present for latch-up reasons. The substrate is very well connected to the digital ground as a result of the multiple substrate contacts. The self inductance of a bond wire causes a voltage drop proportional to the current spikes, $v_{drop} = L_{bondwires} \cdot dI_{spikes}/dt$ (see Figure 2.25). Off-chip, the digital ground is normally connected to the substrate and the substrate noise is therefore equal to v_{drop} [15], [16].

The noise introduced on the substrate in this way can also be picked up by the other circuits in several ways:

- The bulk is biased with the ground, then the substrate noise will couple resistively through the bulk contacts.
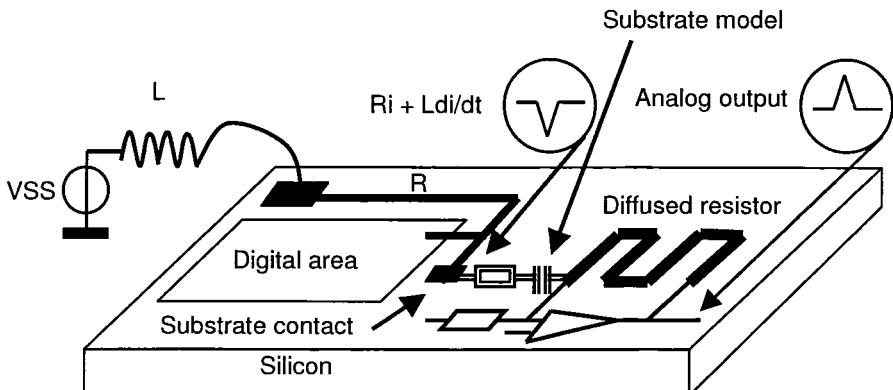
Figure 2.25. Illustration of the cause of substrate bounces. Current spikes from digital switching activities are injected into the substrate via bond wires.

- The p-n junctions of the MOS device and of the n-well will capacitively couple substrate noise into the voltage supply lines (even if analog and digital circuits have separate supply lines).

- The backgate of the MOS device will pick up the noise on the substrate and inject this into the signal paths.

To demonstrate the reality of the problem of substrate bounces, consider the design of a second-order low pass sigma-delta modulator. The design is realized in a 0.25 μm CMOS technology with a 10 mΩ·cm substrate and a 1.8 V power supply with separate analog and digital supply lines. The substrate noise produced by this modulator, including a decimation filter, has been measured. The sigma-delta modulator is depicted in Figure 2.26. All design rules (see next section) to minimize effects of substrate noise have been applied. An external clock at 12 MHz is supplied, and an internal PLL converts this clock to 216 MHz clock signals. The I/O of the circuit is an 8 bit data stream at 13.5 MHz, and a clock line at the same frequency. The signal present at the substrate is measured. Figure 2.27 shows the spectrum of this signal when the power supply of the chip and the clock are turned off. The spectrum is flat above 1 GHz (except that we measure some interference from the GSM/DCS band).

Figure 2.28 shows the spectrum of the substrate signal when the system is active. Odd multiples of 13.5 MHz are visible within the whole band up to the RF frequencies. The substrate noise has been increased by 20 dB in the RF band, and even more at lower frequencies. According to these measurements, the substrate noise is concentrated at multiples of the digital clock frequency,
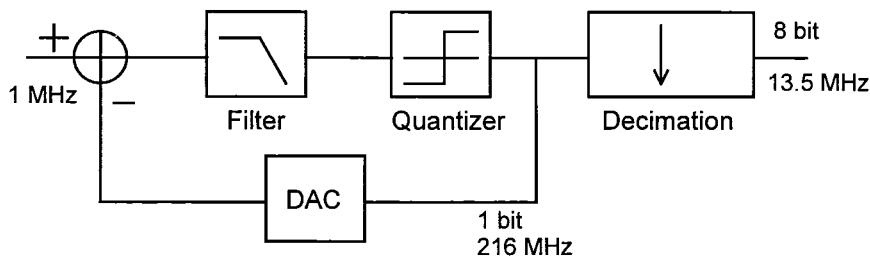
Figure 2.26.   Second order low pass sigma delta modulator including the decimation filter.
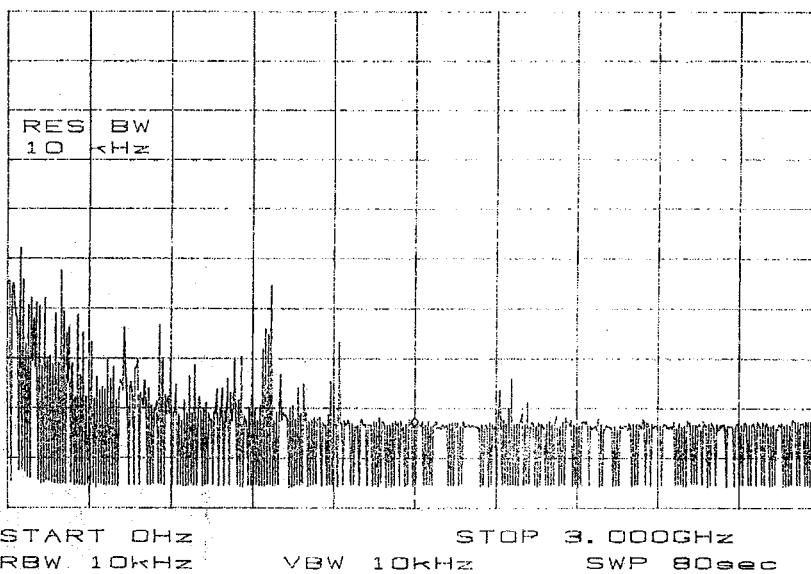
Figure 2.27.   Spectrum of the signal present at the substrate when the system is turned off. (RL: -10 dBm, 10 dB/div)

also observed in other measurements. From the measurements, it is clear that substrate noise is a major design issue when analog and digital circuitry are laid out on the same die.

2.5.2   Design Techniques to Reduce the Substrate Bounces

Substrate noise cannot be prevented. However, by using some design tricks we can minimize the noise or, to a certain extent, make the circuits immune to the noise. Consider the simple current mirror in Figure (2.29). If the analog ground is separated from the substrate, the substrate noise on the backgate of the MOS
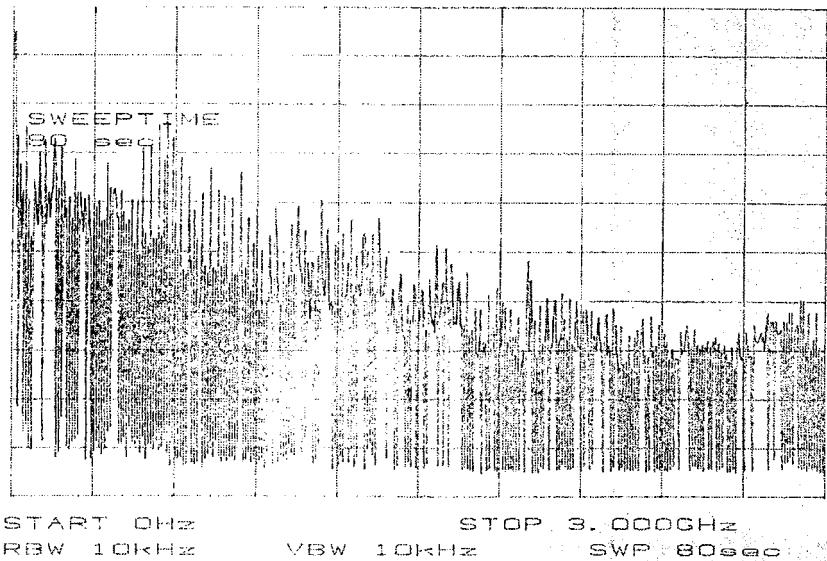
Figure 2.28.   Spectrum when system is active. (RL: -10 dBm, 10 dB/div)

transistors will directly modulate the drain current. This is due to the parasitic capacitance of the MOS device. The supply of the analogue circuits should therefore be referred to the substrate. The substrate noise is then present at the gate, source and backgate of the transistor and will therefore not (or to a lesser degree) modulate the drain current [15]. For noise and matching reasons, the dimensions of both transistors are normally far from the minimum dimensions, still resulting in a large drain-bulk capacitance. Cascode transistors may also be used to overcome this problem. A few design rules can be derived from the example:

- Make analog circuits fully differential with a possible clean common mode signal. The accuracy of balancing is mainly determined by matching properties.

- Use NMOS transistors only as DC current sources, and refer them only to the substrate and not to a clean ground supply. Use PMOS transistors as differential pairs and other signal handling blocks. These transistors have an n-well which can be used to shield the transistor from the substrate. In RF applications, this rule is difficult to fulfil.

- Use different supply lines for analog, digital and RF circuits. In an RF frontend, there are two main circuits which generate noise on the substrate:
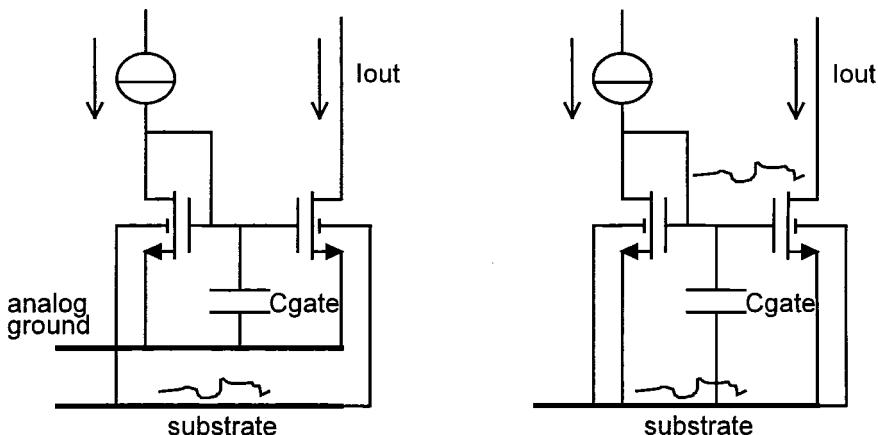
Figure 2.29. Illustration of the effect of substrate noise on analog circuits. On the left side is a current mirror with a clean analog ground. On the right is noise immunity because the analog ground is connected to the substrate.

the voltage controlled oscillator (VCO) and the power amplifier (PA). If transmission and reception do not occur at the same time, the PA and the receive path can share the same supply lines. However, the VCO must have a separate supply line to prevent pulling from the PA and to prevent noise on the supply lines (especially the ground line) contributing to the phase noise of the VCO.

- The use of guard rings is related to the type of substrate. In the case of a low ohmic substrate with a high ohmic epitaxial layer on top, guard rings do not keep the noise outside the ring. As an example, consider the following test chip, where a digital clock oscillator is injecting noise onto the substrate (Figure 2.30). The test chip is realized in a 0.25 μm CMOS technology with five metal layers, and a 10 mΩ·cm (i.e. low-ohmic) substrate beneath a 3 μm epi-layer of 11 Ω·cm. The clock circuit consists of an 11-stage ring oscillator based on simple inverters. Each inverter has the following dimensions: $(W/L)_n = 1.6\,\mu$m/0.5 μm and $(W/L)_p = 2.4\,\mu$m/0.5 μm. The frequency can be tuned using the supply voltage. The die microphotograph of the test chip is shown in Figure 2.31. There is also a guard ring on the chip, laid out around the LNA, to see if it is possible to prevent interference of the substrate noise generated near the clock circuit by the LNA. This guard ring is connected to the ground via a DC probe. There is also a test point available in the guard ring to measure substrate noise. As the measurements have been
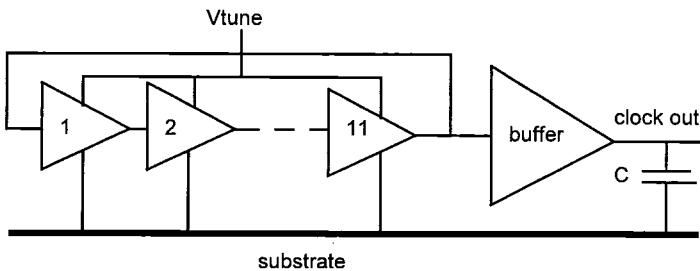
Figure 2.30.   The clock circuit consists of an eleven stage ring oscillator followed by a five stage buffer. The output is by means of a large capacitor connected to the substrate.
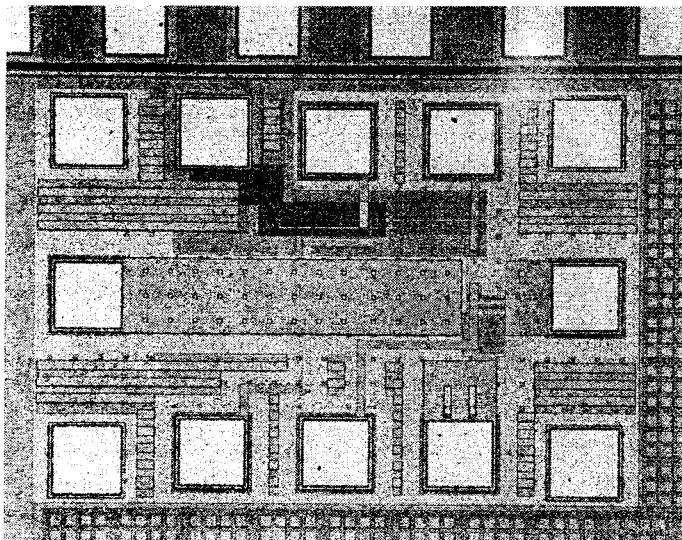
Figure 2.31.   Test chip to measure substrate noise. The clock circuit is on the top, the LNA is on the right. Connections to the substrate in and outside the guard ring have also been added.

performed on wafers, no bond wires have been used and the measured noise at the substrate is due to the p-n junctions [17], [18]. The frequency is set to 772 MHz during the measurements. A five stage buffer that is loaded by a MOS capacitor of 160 fF follows the clock circuit. The capacitor is tightly connected to the substrate to make sure that a maximum signal is injected into the substrate. Although this will never be the case in real designs, it will make measurements easier for this test chip. There are several locations on the test chip to measure the substrate noise. One such measurement point is near the clock output. The measured spectrum of the substrate noise is
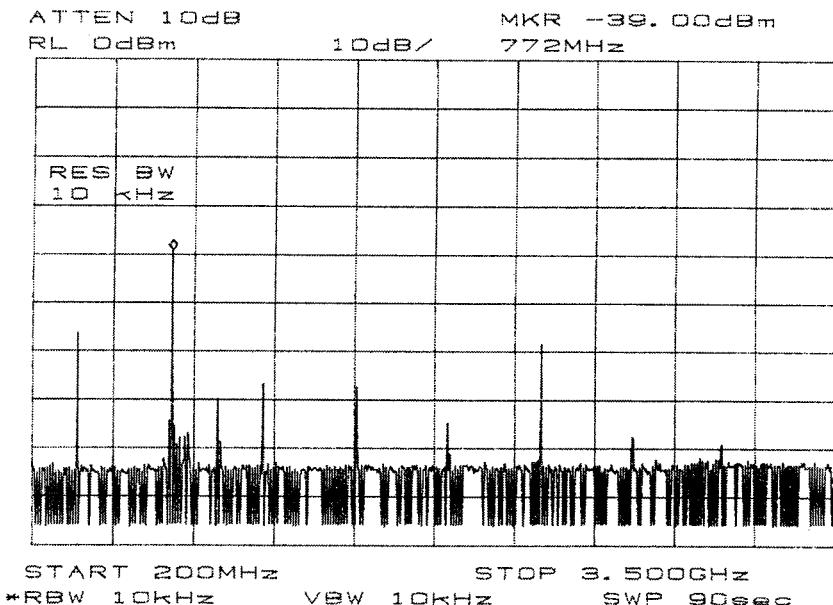
Figure 2.32.   Measured spectrum of the substrate inside the guard ring.

shown in Figure 2.32. The fundamental tone and several of its harmonics are clearly visible. Even the 4th harmonic is still at -80 dBm, which is above the sensitivity level of many mobile and wireless telecommunication standards. The measured spectrum of the substrate noise within the guard ring is identical or almost identical to the spectrum measured near the clock, outside the guard ring. Even measuring the substrate noise at a distance of 1 cm from the clock circuit gave the same spectral results. We can therefore conclude that guard rings will not guard the circuit on low-ohmic substrates against substrate bounces. The substrate can be considered to be a very good conducting plate, and it can therefore be modeled in the circuit as a single node. Guard rings may help to collect "walking" electrons, i.e. electrons which are present just beneath the oxide and are remedies from the switching activities in the digital circuits. Using an n-well around the analog circuit, connected to the positive supply line, these electrons can be captured and the n-well will prevent these electrons from entering the analog domain. Recent studies have shown that the problem of substrate noise does exist even in high ohmic substrate [19]. The substrate noise levels in a 8.5 Ω·cm bulk process are approximately four times lower than in a 0.7 mΩ·cm substrate with a 12 μm thick epitaxial layer and a resistivity of 8.5 Ω·cm. Although
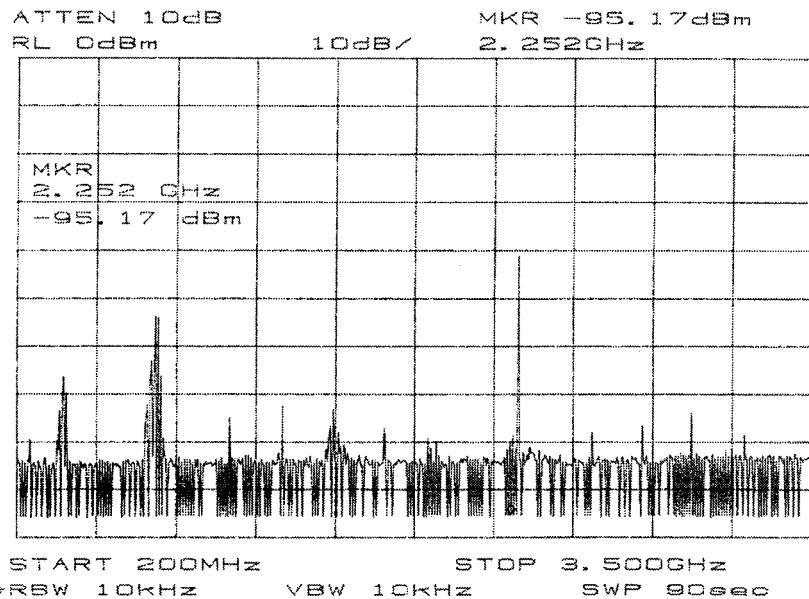
Figure 2.33. Measured spectrum at the output of the LNA when both the LNA and clock are operational.

high ohmic substrates provide reduction of noise levels on the substrate, the use of high ohmic material does certainly not eliminate the problem, a factor four reduction is too low.

■ As can be seen from the measurement results in Figure 2.32, the clock frequency and multiples of this frequency are the dominant presence on the substrate. These signals will be injected into the input and output nodes of an analog or RF circuit. As an example, consider the test chip of Figure 2.29 once again. The measured spectrum available at the output of the LNA is shown in Figure 2.33. The LNA has an input signal at 2.3 GHz, that is three times the clock frequency. The fundamental frequency of the clock and its harmonics are visible in the spectrum. It is clear from the measurement results that it is not only the amplified information signal that is available at the output of the LNA, but also the unwanted third harmonic of the clock. In fact, any substrate noise, with spectral components in the same frequency range as that in which the LNA is operating, will be picked up by the LNA. The information signal will therefore be distorted by this substrate noise. The gain of the LNA is not influenced by the substrate noise, simply because the injected substrate signals are too weak to shift the bias operating point

of the LNA. The s-parameters were measured both when the clock was both active and when it was turned off. These measurements gave the same results. Hence, the performance of the LNA is not affected by the substrate noise. This is to be expected since the RF performance is mainly determined by the bias operating point and the matching network. The 3rd harmonic of the clock is too weak to disturb this operating point. These measurements show us that clock planning is essential in a system with RF and digital circuitry on the same die.

The above list of design issues to cope with substrate bounces is certainly not complete. It is merely mentioned to illustrate the problem and to make it clear that designers should tackle this problem, otherwise performance degradation will occur.

REFERENCES

[1] D.B. Rutledge, *The Electronics of Radio*, Cambridge University Press, New York, 1999.

[2] G. Gonzales, *Microwave Transistor Amplifiers: analysis and design*, Prentice-Hall, Inc, New Jersey, 2nd edition, 1997.

[3] S.P. Voinigescu *et al.*, "Process- and Geometry-scalable Bipolar Transistor and Transmission Line Models for Si and SiGe MMICs in the 5-22GHz Range," in *International Electron Devices and Modeling*, 1998, vol. 17, pp. 307–310.

[4] J. Craninckx and M. Steyaert, "A 1.8-GHz Low Phase Noise CMOS VCO using Optimized Hollow Inductors," *IEEE Journal of Solid-State Circuits*, vol. 32, no. 5, pp. 736–744, May 1997.

[5] J. Craninckx, *Low-Phase-Noise Fully Integrated CMOS Frequency Synthesizers*, Ph.D Thesis, Katholieke Universiteit Leuven, Leuven, 1997.

[6] A. Amerasekera and C. Duvvury, *ESD in Silicon Integrated Circuits*, John Wiley, Chichester, 1995.

[7] EIA/JEDEC STANDARD, *Electrostatic Discharge (ESD) Sensitivity Human Body Model (HBM), EIA/JESD22-A114-A*, 1997.

[8] C-J Wu *et al.*, "A new On-Chip ESD Protection Circuit with Dual Parasitic SCR-Structures for CMOS VLSI," *IEEE Journal of Solid-State Circuits*, vol. 27, no. 3, pp. 274–280, Mar. 1992.

[9] I.E. Opris, "Bootstrapped Pad Protection Structures," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 2, pp. 300–301, Feb. 1998.

[10] M-D. Ker, "ESD Protection for CMOS Output Buffer by using Modified LVTSCR Devices with High Trigger Current," *IEEE Journal of Solid-State Circuits*, vol. 32, no. 8, pp. 1292–1296, Aug. 1997.

[11] E. Fujishin *et al.*, "Optimized ESD Protection Circuits for High-speed MOS/VLSI," *IEEE Journal of Solid-State Circuits*, vol. 20, no. 4, pp. 594–596, Apr. 1985.

[12] S.M. Sze, *Physics of Semiconductor Devices*, John Wiley & Sons, New York, 2nd edition, 1981.

[13] Y. Taur and T.H. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge University Press, New York, 1998.

[14] M. Felder and J. ganger, "Analysis of Ground-Bounce Induced Substrate Noise Coupling in a Low Resistive Bulk Epitaxial Process: Design Strategies to Minimize Noise Effects on a Mixed-Signal Chip," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 11, pp. 1427–1436, Nov. 1999.

[15] B. Nauta and G. Hoogzaad, "Substrate Bounce in Mixed-mode CMOS IC's," in *Workshop on Advances in Analog Circuit Design (AACD)*, 1998.

[16] D.W.J. Groeneveld, "Groundbounce in CMOS," in *Workshop on Advances in Analog Circuit Design (AACD)*, 1993.

[17] D. Leenaerts and P. de Vreede, "Influence of Substrate Noise on RF Performance," in *European Solid-State Circuits Conf. (ESSCIRC)*, 2000, pp. 300–303.

[18] D.M.W. Leenaerts and P.W.H. de Vreede, *Analog Circuit Design*, chapter Mixed-mode telecom design, Kluwer, Norwell,MA, 2000.

[19] X. Aragones and A. Rubio, "Experimental Comparison of Substrate Noise Coupling using Different Wafer Types," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 10, pp. 1405–1410, Oct. 1999.

Chapter 3

Low Noise Amplifiers

One of the key building blocks in a receiver path of a cellular or wireless system is a low noise amplifier. This circuit must increase the signal level while hardly increasing the signal-to-noise ratio of the incoming signal. These two tasks are not always easily achieved, mainly because noise impedance matching and input impedance matching are not always obtained for the same source impedance. The aim of this chapter is to explain the design steps required to design a low noise amplifier in various technologies.

3.1   SPECIFICATION

The low noise amplifier (LNA) is the first gain stage in the receiver path. Therefore, according to Friis' formula (see (1.26)), the noise figure (NF) of this circuit directly adds to that of the overall system. Remember that the noise figure is a measure of the degradation of the signal-to-noise ratio (SNR) at the output of the LNA compared to that at the input. Another main performance parameter of the LNA is its gain. The signal should be amplified as much as possible with hardly lowering the SNR, while also maintaining linearity. The last performance parameter is therefore represented by IP_3. Obviously, the obtained NF, gain and IP_3 should be within specification with a minimum power consumption. This is the dilemma for an RF designer. The design specifications usually have to be met for a certain source and load impedance, while achieving a minimum power consumption.

The complete LNA can be thought of as a cascade of three stages, as depicted in Figure 3.1. The first stage is needed to transform the source impedance to an impedance which will match with the input impedance or optimal noise impedance of the actual LNA, for instance. The second stage is the actual LNA.

79
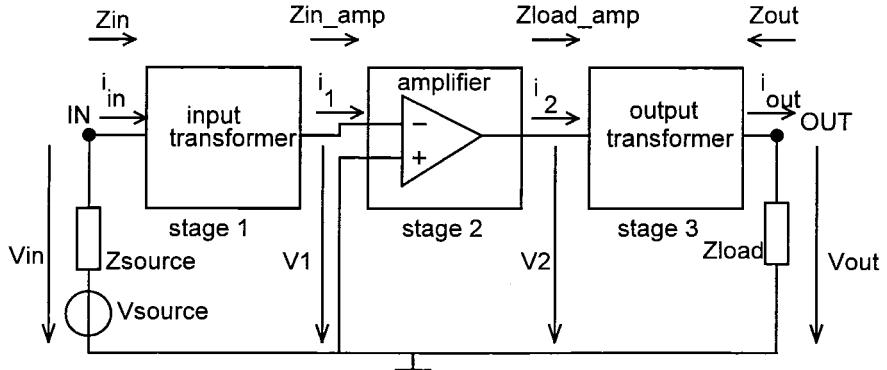
Figure 3.1. The LNA and the input/output transformer stages to understand the overall analog behavior.

It may consist of several stages itself, but it could also be a single transistor as in microwave applications. The last stage is again an impedance transformer to match the output impedance of the actual LNA to a specified load impedance.

To each of the stages, the five gain definitions, as explained in Chapter 1, can be applied. The difference in voltage gain A_v and the power gain G_p, for instance, can now easily be seen. Voltage gain, according to Figure 3.1 is defined as

$$A_v = \frac{v_2}{v_1}$$

and does not take into account losses due to the input transformer[1]. This loss is normally not important in analog design because the source impedance of the voltage source is much lower than the input impedance of the circuit. The power gain is defined as

$$G_p = \frac{P_{l,del}}{P_{i,del}} = \frac{P_{load}(\text{real})}{P_{in}(\text{real})} = \frac{\text{real}(v_{out} \cdot i_{out}^*)/2}{\text{real}(v_{in} \cdot i_{in}^*)/2}$$

and shows dependency on voltage and current. Obviously, the voltage gain can be much lower than the power gain due to the losses at the input of the first stage.

---

[1] The loaded voltage gain is defined as

$$A_{v,loaded} = \frac{v_{out}}{v_1}$$

In many analog designs, there is normally no impedance matching and stage 1 in Figure 3.1 is left out. The design usually has to have a broadband behavior. This means that low noises figures can only be achieved by changing the width of the input transistor. For example, if the actual noise figure for an LNA design has to be the same as the minimum noise figure, without using impedance matching, the width of the transistor will be extremely large (in CMOS technology). This can result in a broadband low noise amplifier with a noise figure of around 1 dB from 0 to 2 GHz. However, the power used is extreme. This is why RF designers use the transformer stage as the first stage. Matching impedances using additional components allow the use of small transistor widths for the actual LNA, thus enabling low power consumption. The price paid is a narrow band low noise amplifier; the impedance match is only valid for a small frequency band.

The input transformer normally consists of passive components such as inductors and capacitors. Resistors are not used because they add noise. As explained in Chapter 1, noise impedance matching and input impedance matching (to achieve maximum power transfer) are normally not achieved for the same source impedance. In most designs, noise is the key design parameter and noise impedance matching, i.e. optimal noise input impedance equal to source impedance, is preferred. A (small) power transfer loss is accepted, as the input impedance is not matched to the source impedance.

The input transformer not only adapts the impedance, but inherently also adapts the amplitude of the signal. The increase of signal amplitude at the input of the actual amplifier influences the linearity of the LNA. The larger the input signal, the more this signal will be influenced by the non-linear voltage-current characteristic of the amplifier. The antenna signal can easily have a dynamic range from −90 dBm to −20 dBm. Large dynamic ranges make it difficult to fulfil the linearity specification without consuming too much power.

Several LNA designs are treated in detail in the following sections. Designs steps are explained and trade-offs are discussed to come to an overall design. To get a good "feeling" for what a noise figure of 3 dB means, consider the following simulations using a behavioral model of an LNA. The used Matlab routines can be found in Appendix A. The antenna signal first has an amplitude of 1 V at a frequency of 2.25 GHz. The LNA has a power gain of 20 dB (10 times voltage gain), a noise figure of 0 dB (noiseless) and is highly linear. The input and output signal of the LNA in time and the frequency domain are shown in Figure 3.2. As power matching for the input impedance at 50 Ω is assumed, the actual input amplitude is half that of the antenna signal. The available power at the input is therefore $(1/\sqrt{2})^2/50$ W or −20 dB, and the delivered power at the input is −26 dB or 4 dBm. This value and the amplification are clearly
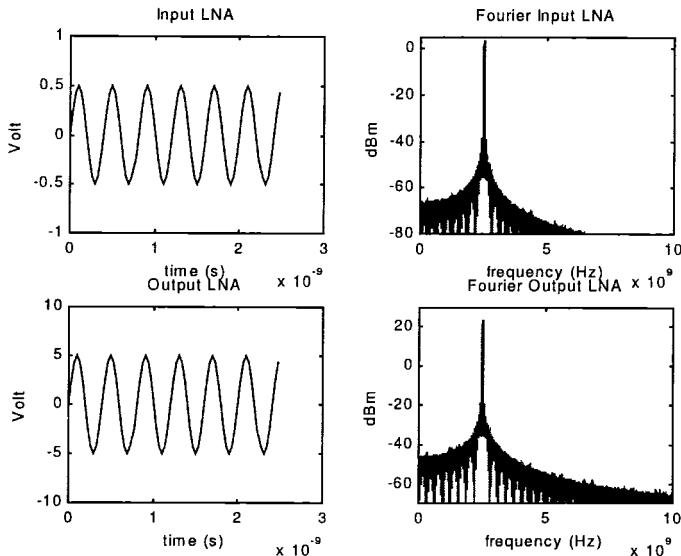
Figure 3.2. Input and output signals in time and the frequency domain of the LNA. The LNA is almost ideal except for the gain, which is set to 20 dB.

visible in Figure 3.2 (please note that powers are based on the rms value of a sinusoidal signal).   Now, let us add noise in such a way that the NF equals 10 dB and lowers the input amplitude to 1 mV. The result is visible in Figure 3.3. Although the input amplitude is still very large for a received signal at the antenna, the output signal of the LNA is due to the noise figure already affected. The noise level at the output of the LNA is equal to -70 dBm, an increase of 30 dB due to the power gain and the noise figure. If the amplifier is made non-linear with an OIP_3 equal to 20 dBm for an input amplitude of 0.1 V, the results shown in Figure 3.4 are obtained. The third harmonic is at 7.5 GHz with a power of -30 dBm. The IM_3 is therefore also -30 dBm, which leads to an IIP_3 of $-15 + (-30/2) = 0$ dBm. The output IP_3 is therefore $0 + 20 = 20$ dBm.

The input transformer often consists of an LC tank with a certain quality factor. The amplitude of the voltage is increased due to this tank circuit. If we increase the quality factor of the input matching stage of the LNA from $Q = 1$ to $Q = 3$, the results in Figure 3.4 are changed into this of Figure 3.5. The input signal has been increased due to the quality factor, and hence the LNA becomes more non-linear. The IIP_3 is reduced by 10 dB. This simulation shows that the input matching network is very important and needs to be designed with care.
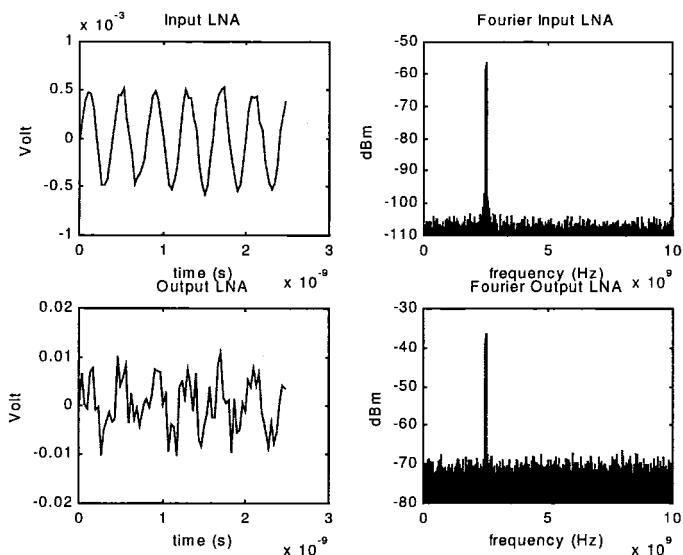
Figure 3.3. Input and output signals in time and frequency domain of the LNA. The LNA has a gain of 20 dB and an NF of 10 dB.
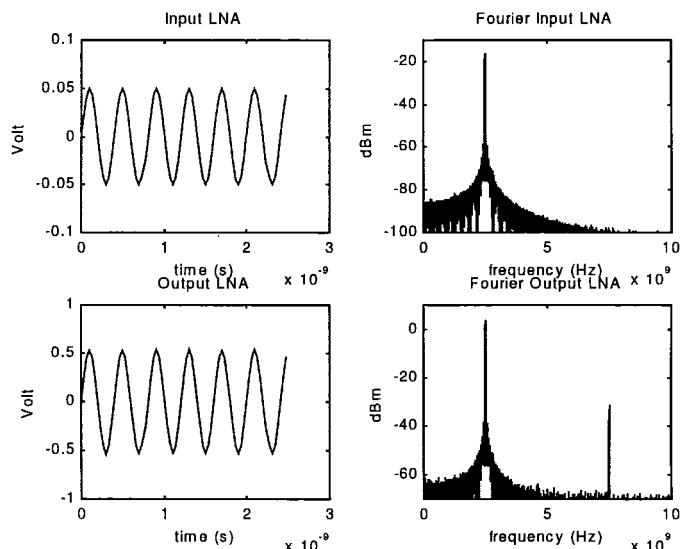
Figure 3.4. Input and output signals in time and frequency domain of the LNA. The LNA has a gain of 20 dB and an NF of 10 dB. The OIP_3 is set to 20 dBm.
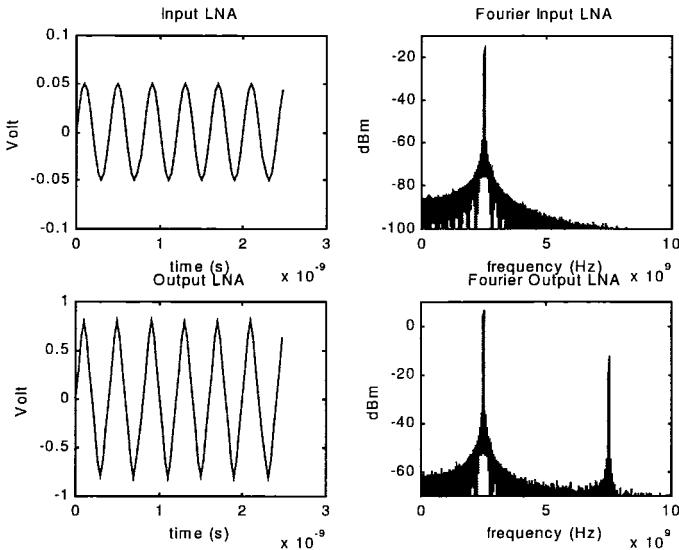
Figure 3.5. Simulation results for the LNA with a quality factor for the input matching network of $Q = 3$.

The simulation results on the behavioral model give designers a good insight into the impact of certain design specifications. Several LNA designs will be discussed in the following sections.

3.2   BIPOLAR LNA DESIGN FOR DCS APPLICATION IN SOA

In this section, we discuss the design of a bipolar LNA for DCS 1800 applications. The frequency range for this application is 1.805 to 1.88 GHz. The required noise figure is set to 4.5 dB, the output linearity specification to 0 dBm, and the required power gain is 15 dB; opting for a fully differential design, the source impedance is set to 100 Ω. To ensure the minimum power consumption, Silicon-on-Anything technology has been chosen with a supply voltage of 1.8 V. In this technology, the bipolar device has an emitter width of 1.0 μm and a collector length of 1.2 μm. The f_T of this transistor is approximately 6 GHz for a collector current of 20 μA.

3.2.1   Design of the LNA

The "cascode" stage is chosen as the topology for the LNA design. The basic cascode stage consists of a single transistor. The collector of the input transistor (in a common emitter connection) is directly connected to the emitter of the
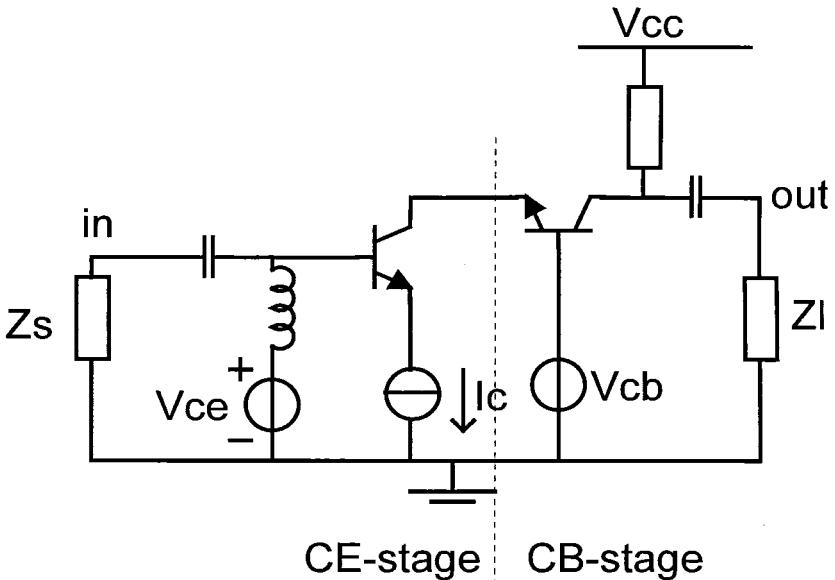
Figure 3.6.   Basic configuration for a cascoded CE stage.

cascode transistor (in a common base connection). The most important feature of such a cascode transistor is that a common-base stage results in a good isolation between the input and the output of the amplifier, defined by the collector of the CB-stage. This isolation is due to the low collector-emitter capacitance of the transistor. This circuit is therefore attractive for higher frequencies. Another reason is that a cascode results in a higher power gain compared to a single stage. Figure 3.6 gives a schematic of a cascoded amplifier. The first step in the design is to find the optimal collector current I_c for the CE-stage. This current setting is dependent on the noise figure and the required input impedance for optimal noise and power matching. Assume that the CE-stage is loaded with the CB-stage, and hence with an impedance equal to $1/g_m$ with $g_m \approx 40I_c$. The source impedance of the CE-stage is 50 Ω in single ended mode. From Figure 3.9, we can see that the current for optimal noise matching is 3.5 μA. In this design, we are aiming at an optimal noise match, i.e. $Z_s = Z_{opt}$, and an optimal power transfer, i.e. $Z_s = Z_{in}^*$. Consequently, we are looking for a collector current yielding $Z_{opt} = Z_{in}^*$, but Figures 3.7 to 3.9 show that this optimum is not reached.

Degeneration can be applied to achieve better matching. Placing an inductor in the emitter will increase the real part of the input impedance. The best match
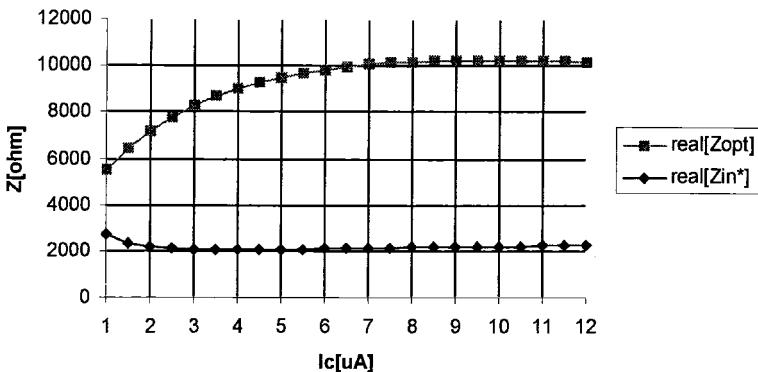
Figure 3.7.   Input impedance Z_{in} compared to the optimal noise impedance Z_{opt} for the real part as a function of the collector current.
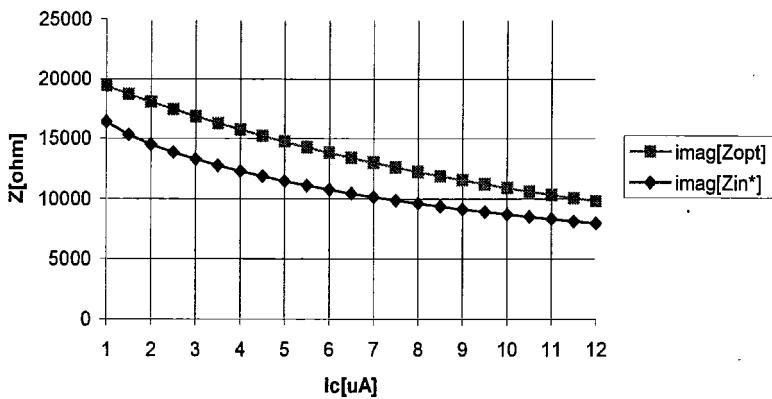
Figure 3.8.   Input impedance Z_{in} compared to the optimal noise impedance Z_{opt} for the imaginary part as a function of the collector current.

can be achieved using an inductor of 500 nH with an quality factor of 20 at 1.85 GHz, as shown in Figures 3.10 to 3.11.

A collector current of approximately 4 μA is a good trade-off between the noise matching and input power matching. The resulting source impedance is $11 + j11$ kΩ. These values are only valid if the emitter degeneration inductor is 500 nH for a single transistor. Note that there is a value for this inductor
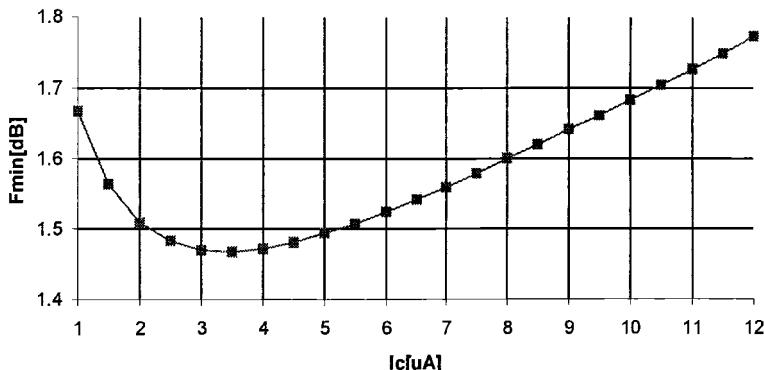
Figure 3.9.   The optimal noise figure F_{min} as function of the collector current.
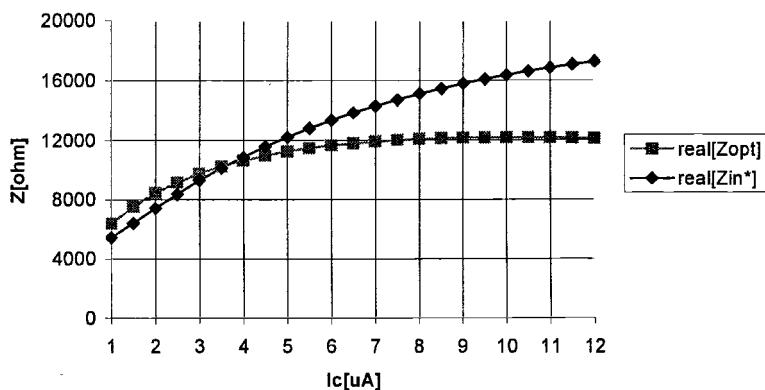
Figure 3.10.   The real part of the optimal noise impedance compared to the real part of the input impedance as a function of the collector current for a degenerated CE stage by using an inductor.

for each current, resulting in the best match between noise and power, but changing the optimum source impedance. This inductor value is obviously not the value which we need in the final design. The 500 nH inductor is needed for a single transistor, but in the actual design many transistors are placed in parallel to achieve the required total current. The ratio of transistors in parallel compared to a single one is known as the "mult" factor. The inductor value
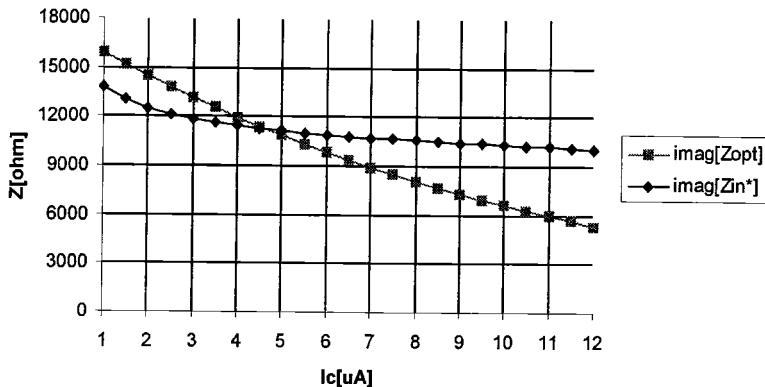
Figure 3.11. The imaginary part of the optimal noise impedance compared to the imaginary part of the input impedance as a function of the collector current for a degenerated CE stage by using an inductor.
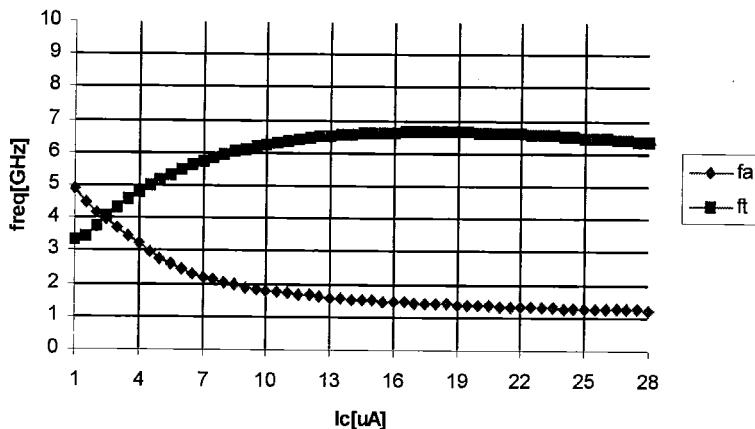
Figure 3.12. Cut-off frequency f_T and available frequency f_A as a function of the collector current.

needed is therefore reduced by this "mult" factor. The same holds for the source impedance.

The next design step is to investigate for which collector currents f_T and f_A are high enough. For the CE-stage, the simulation results of these quantities are given in Figure 3.12.
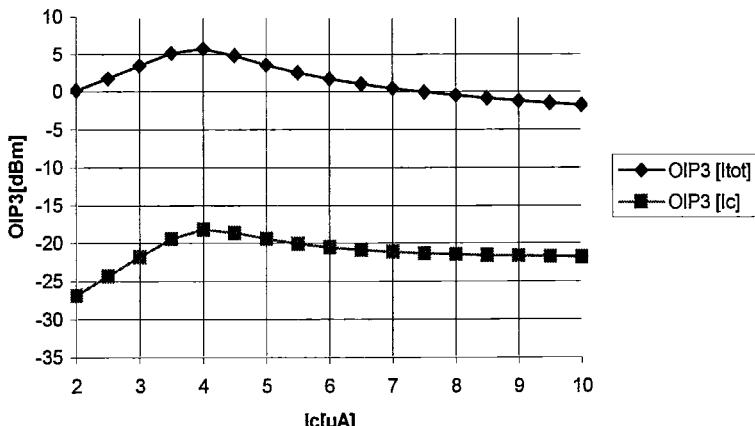
Figure 3.13. Output IP_3 as a function of the collector current for a single transistor (I_c) and with a large number of transistors in parallel (I_{tot}).

For a good design, the bandwidth has to be higher than the RF frequency, and the cut-off frequency has to be higher than two times the RF frequency. Consequently, the current for a single transistor has to be somewhere between 2 μA and 10 μA. For the CE-stage, a fair assumption is that the current gain is approximately 2.5 (f_T/f) and the voltage gain is about 1. The voltage gain in the CB-stage then has to be approximately 12 to realize a power gain of approximately 30 (i.e. 15 dB). We can see that the optimal current with respect to frequency behavior is in the same range as the optimal current for noise and power matching. In both cases, the optimal current is determined for a single transistor.

The last step in the design of the CE stage is to determine the actual current through the stage. Because the OIP_3 of a single transistor is much too low (see Figure 3.13), a number of transistors have to be designed in parallel. The OIP_3 increases with the mult factor. This mult factor is determined by the current through mult transistors (I_{tot}) divided by the current through a single transistor (I_c). Figure 3.13 also shows the OIP_3 for a total current of 1 mA. In this case, there are 250 transistors in parallel. The total current of 1 mA is chosen to scale the OIP_3 to a higher level that meets the specification. Due to the number of transistors in parallel, the overall impedance Z_o is also scaled according to $Z_o = Z/\text{mult}$. The resulting source impedance of $11 + j11$ kΩ for a single transistor now becomes $44 + j44$ Ω for a single ended input and, hence, roughly 100 Ω for a differential input, exactly as required.
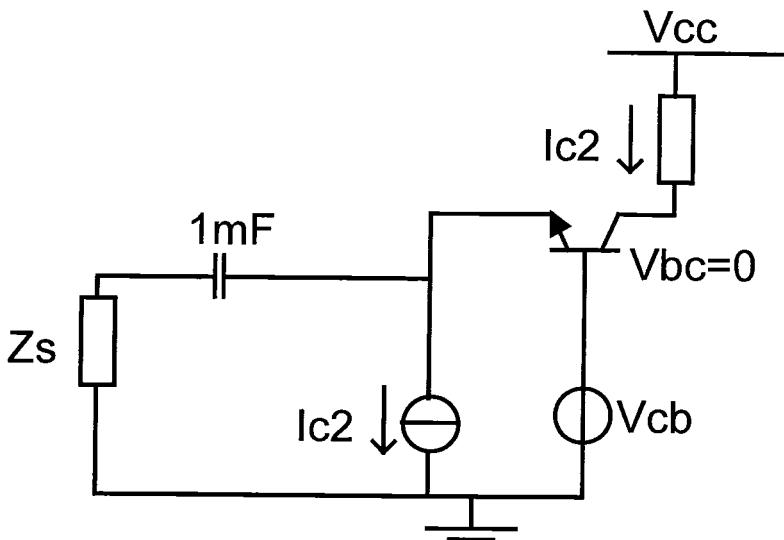
Figure 3.14.   The source and load impedance for the CB-stage.

The CB-stage can now be designed (Figure 3.14). The source impedance Z_s can be found as the output impedance of the CE-stage, and is fitted as a function of the collector current I_{c1}. This results in

$$Z_s = R_s + jX_s \tag{3.1}$$

$$R_s \approx \frac{-7500}{\log(I_{c1})} \tag{3.2}$$

$$X_s = 7500\log(I_{c1}) + (I_{c1}) \cdot 10^9 \tag{3.3}$$

The load impedance Z_L is determined by the required voltage gain. The current gain of the CE-stage was 2.5. To obtain a power gain of 30, i.e. 15 dB, the voltage gain must be 12. The voltage gain of a CB-stage is given as $g_m Z_L$, and therefore $Z_L = 12/g_m$ where g_m is the transconductance of the CB-stage ($g_m = 40I_{c2}$). The CB-stage must also achieve the correct frequency behavior and, therefore, f_T and f_{out} are simulated as a function of the collector current I_{c2} under the conditions as shown in Figure 3.14. A current of at least $7\mu A$ is mandatory to achieve an output bandwidth in the same range of the required RF setting (see Figure 3.15). Assuming the currents I_{c1} and I_{c2} to be equal, the output IP_3 of the CB-stage can be simulated. Figure 3.16 shows the OIP_3 for a single transistor and several transistors in parallel.

From Figure 3.16 it is evident that the best OIP_3 is obtained at current levels around 25 μA. However, the current for the best OIP_3 in the CE-stage was
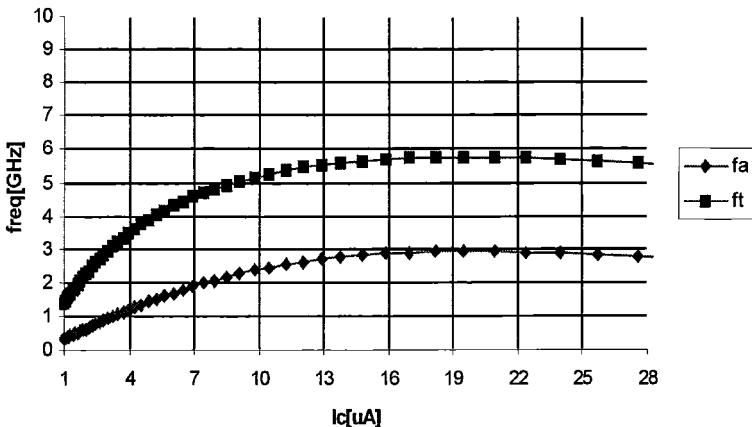
Figure 3.15. Cut-off frequency and output frequency for the CB-stage as a function of the collector current through the stage.
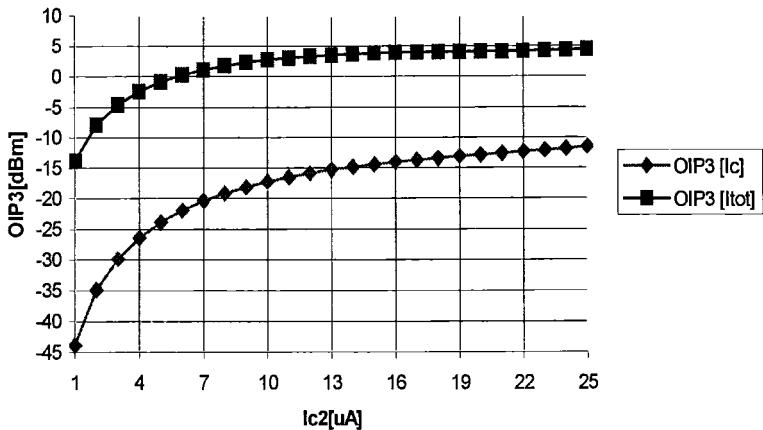
Figure 3.16. Output IP_3 as a function of the collector current I_{c2} for a single transistor (I_c) and for several transistors in parallel to obtain a total current of 1mA.

around 4 μA. Therefore, the top of the linearity curves for the CE-stage and the CB-stage are at different current levels. As the assumption is made that I_{c1} is equal to I_{c2}, the total OIP_3 might not be as high as possible. Knowing that the
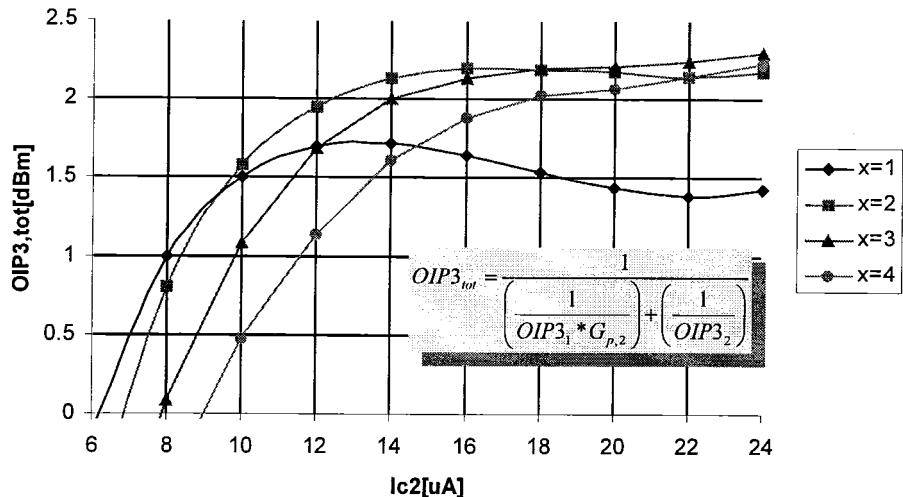
Figure 3.17. Overall output linearity as a function of the current through a single transistor in the CB-stage for several ratios of the mult factor.

overall output linearity can be derived as

$$OIP_{3,tot} = \frac{1}{\frac{1}{OIP_{3,CE} \cdot G_{p,CB}} + \frac{1}{OIP_{3,CB}}}$$

we see that it is not only a high OIP_3 for the CB-stage that is important, but also a high OIP_3 for the CE-stage. The power gain for the CB-stage is also important to get a high overall OIP_3. The difference in optimal current for the CE-stage and CB-stage can be corrected by using different mult factors. Defining $x = mult_{CE}/mult_{CB}$, then we can plot OIP_3 as a function of I_{c2}, as can be seen in Figure 3.17. From this result, an near-optimal choice is $x = 3$ and, hence, $I_{c1} = 8.5 \ \mu A$ and $I_{c2} = 24 \ \mu A$. Choosing 0.85 mA as the total current we have 100 as the mult factor for the CE-stage and 36 for the CB-stage. For the real part of the input impedance we obtain $11000/100 \approx 100 \ \Omega$. Because the final design is fully differential, the final input impedance is 200 Ω, while the source impedance is 100 Ω, resulting in a power loss at the input; instead of a 15 dB power gain we obtain roughly 12 dB. The OIP_3 of the final design is approximately 2 dBm.
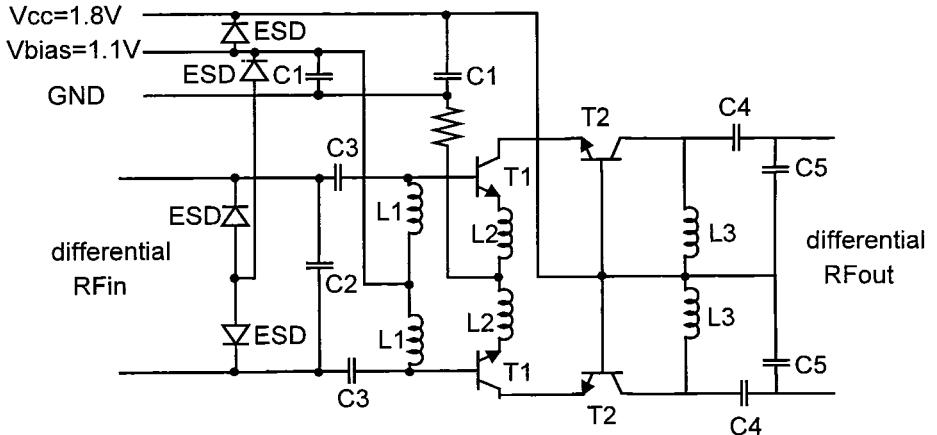
Figure 3.18. Fully integrated, differential input and differential output LNA, based on a CE-stage (T1) and a CB-stage (T2).

3.2.2   Measurements

The complete design of the fully differential LNA is shown in Figure 3.18. The CE-stage is formed by T1, and the CB-stage is realized by T2. The emitter degeneration inductor is realized using L2. The inductor has an inductance of L=3.4 nH, Q=18 at 1.8 GHz. The inductance is close to the value obtained from simulations, i.e. 500 divided by the mult factor of 100, yielding 5 nH. Inductors L1 together with capacitors C2 and C3 form a resonance tank at the center frequency of the LNA and provide DC blocking. Inductor L1 has a measured inductance of L=6.5 nH with a measured Q of 23 at 1.8 GHz. The load of the CB-stage is realized by inductors (L3) and partly by capacitors (C4,C5) to reduce noise while achieving the proper output impedance at 1.8 GHz. All inductors are integrated on the chip and have a diameter of 600 μm. ESD protection diodes (D1, D2) are also included in the design as decoupling capacitors (C1), which are realized using high density capacitor structures. Resistor R1 is used for biasing the CE-stage.

   The die microphotograph is shown in Figure 3.19. Measurements have been performed on wafers using Cascade RF probes. The LNA draws 1.7 mA. Capacitances C2 to C5 could be trimmed using lasers and when there is no trimming, the center frequency is 1.5 GHz which was also simulated. The measured transducer gain is $G_T = 11.2$ dB, which is close to expectations, as well as the noise figure $NF = 3.8$ dB. The OIP_3 is measured to be -3.5 dBm
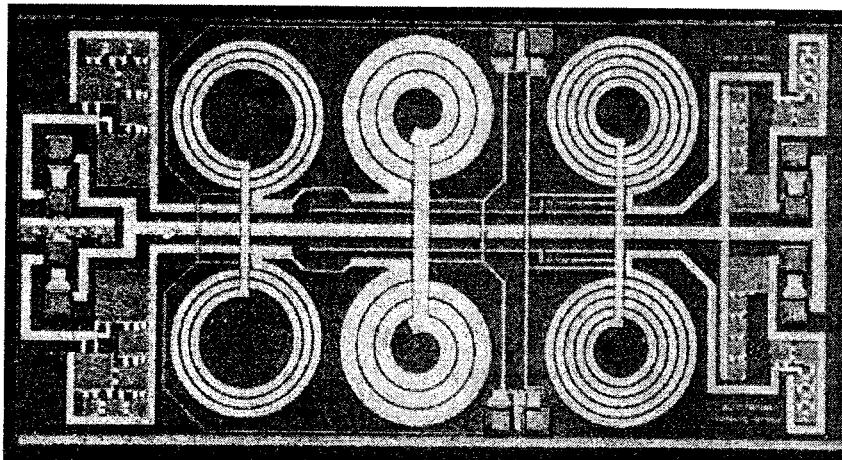
Figure 3.19.   Die microphotograph of the LNA. The inductors L1, L2 and L3 from left to right are clearly visible. The input is located on the left side of the chip, and the output is on the right side, both implemented in ground-signal-signal-ground pad settings for differential measurements.

using a two tone measurement with −25 dBm input power. Because 0 dBm was required, the transistors are more non-linear than predicted in the simulations. More current consumption is needed to increase the linearity.

3.3   CMOS LNA DESIGN

In the past few years, CMOS technologies have entered the field as an alternative technology for RF-design. The driving force behind this trend is the demand to increase functionality on a single die, hence the integration of RF functionality with digital circuitry. Because logic is always implemented in CMOS technology, the question is whether RF circuitry can be implemented in this technology and still have a performance similar to that when using other technologies. However, before that point is reached, we have to show that the performance of RF circuitry in MOS technology is similar to that in bipolar designs. From a historical point of view, the low noise amplifier was the first building block to test the capability of MOS in RF applications.

3.3.1   Single Transistor LNA

Although LNAs can be built up from many transistors, from a noise performance point of view, the fewer transistors the better. In this section we show that a single transistor LNA can perform very well.

Design Steps

There are a few design parameters to determine in the design of an LNA: the gate-source voltage V_{gs} to bias the transistor, the voltage supply V_{dd}, the width W of the transistor, the length L of the transistor and the matching networks at the input and output. The goal is to obtain the required performance on noise figure NF, linearity IP_3 and power dissipation for a given source and load impedance. The design cycle consists of several steps. They are explained for a 2 GHz design in a 0.25 μm CMOS design in a 10 mΩ·cm substrate beneath an epi-layer.

1. **Choice of L.** The length of the transistor is chosen as the minimum feature size, $L = 0.25$ μm. This will give the best frequency performance of the transistor due to minimum parasitic capacitances.

2. **Choice of V_{ds}.** From the three remaining design parameters, i.e. the gate voltage V_{gs}, the drain voltage V_{ds}, and the width of the transistor W, V_{ds} is chosen to be 0.9 V. This value is chosen because it will allow for really low power design, and also with a battery supply of 1 V in mind. Note that the drain voltage for a single transistor can be equal to the supply voltage, depending on the type of load used.

3. **Choice of V_{gs}.** As can be seen in Figure 3.20, in principle, the width of the transistor is not set by noise figure considerations. However, this is only true under conditions of perfect noise matching. In the case of 50 Ω source and load impedance, on or off-chip noise matching is mandatory to achieve the minimum noise figure. As an example, the noise figure is also shown for the 50 Ω source and load impedance, without any impedance match. A transistor width of 600 to 700 μm will give the transistor an input impedance close to 50 Ω. Noise matching is then achieved at the cost of a large current. The minimum noise figure depends on the voltage bias conditions as shown in Figure 3.21. From this figure, it follows that the gate voltage V_{gs} has to be 0.7 V, because the lowest possible noise figure is then achieved

4. **Choice of W.** The only design parameter left is the width W of the transistor. Taking only noise into account, a transistor with the smallest possible width ($W = 0.5$ μm) is the best choice, because this leads to the lowest current and thus the lowest power consumption. If this smallest transistor width is taken, the input impedance of the LNA is totally set by the bond pad together with its ESD device, because the impedance of the transistor is small compared to the impedance of the bond pad and ESD. If it still is possible to design a proper impedance matching network in this situation, the smallest
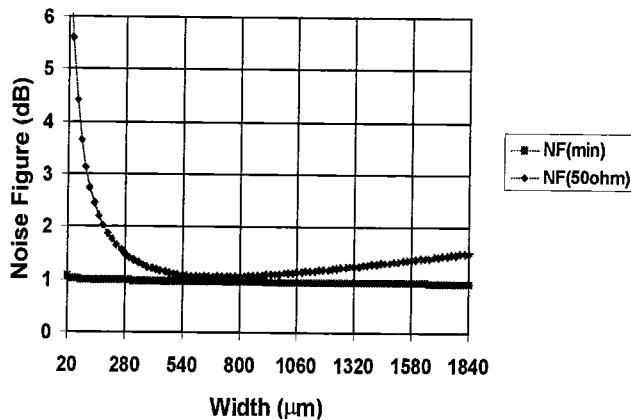
Figure 3.20.   Noise figure as a function of the width of the device under the following conditions: $V_{gs} = 0.7$ V and $V_{dd} = 0.9$ V, $L_{finger} = 2.5 \ \mu$m.
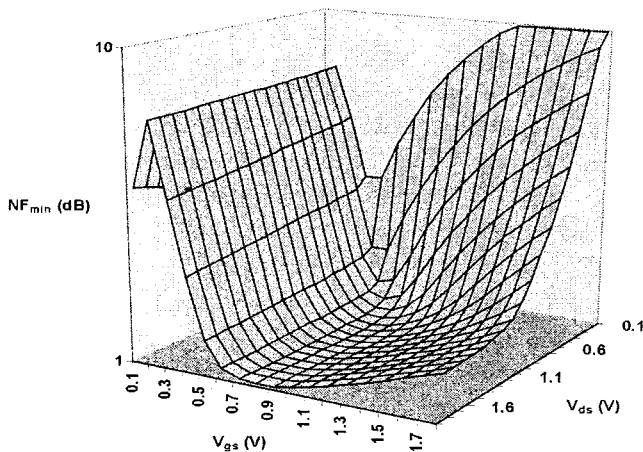
Figure 3.21.   Simulated minimum noise figure of a transistor with a given width, versus the gate and drain voltage for a finger length of $L_{finger} = 2.5 \ \mu$m.

possible transistor is indeed the best choice. However, the output $I P_3$ is not independent of the width of the transistor (Figure 3.22). This means that using the design parameters already given, the width W of the transistor is set by the desired (output) $I P_3$. An output $I P_3$ of about 10 dBm has been chosen, resulting in a transistor width of 40 μm. If input matching has to be achieved on-chip, the width of the transistor is determined by both the required input impedance and the linearity. Figure 3.22 does not take
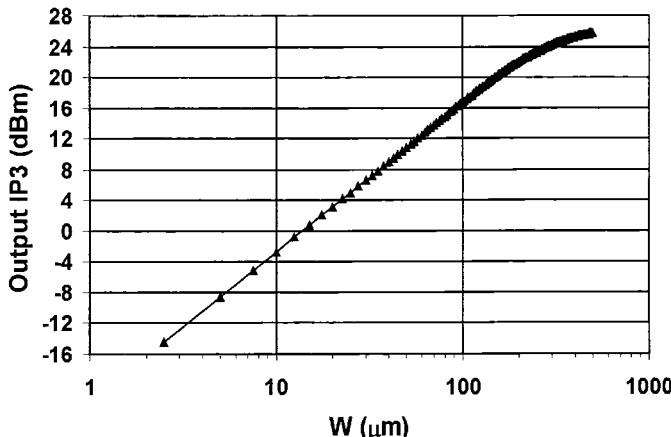
Figure 3.22. Simulated output IP_3 of a transistor, versus the width of the transistor, under the obtained bias settings.

the compression point (CP_{1dB}) of the transistor into account. Normally, the maximum input signal has been defined for each telecommunication standard, and the compression point of the LNA must be better by at least 4 dB. Assuming a power gain of 6 dB, this will give an IIP_3 of 4 dBm for the chosen OIP_3. As the compression point can be found approximately 10 dB lower than the IIP_3, we have a CP_{1dB} of -6 dBm, which is acceptable.

5. **Choice of L_{finger}.** The chosen transistor has the dimensions $W/L = 40\mu m/0.25\mu m$. To obtain good RF performance, the transistor has a folding geometry. The finger length (L_{finger}) has also an influence on the lowest minimum noise figure that can be achieved, see Figure 3.23. This is due to the reduced gate resistance caused by the folding action. The lower the chosen finger length, the lower the minimum noise figure. For reasons of low gate resistance, a transistor with L_{finger} of 2.5 μm has been chosen.

6. **Fitting the model.** The next step in the design is to fit the measured y-parameters of this transistor (which are extracted from measured s-parameters) with simulated y-parameters. This is necessary, as an accurate prediction of the noise behavior of the transistor has to be known for the design of the impedance matching network. This behavior will be extracted from simulations, and these simulations are more reliable if the simulated y-parameters fit with the measured y-parameters. A MOS MODEL 9 mini-set has been extracted from the measured DC behavior of the transistor. This is a standard MOS MODEL 9 mini-set without extensions in the model to model
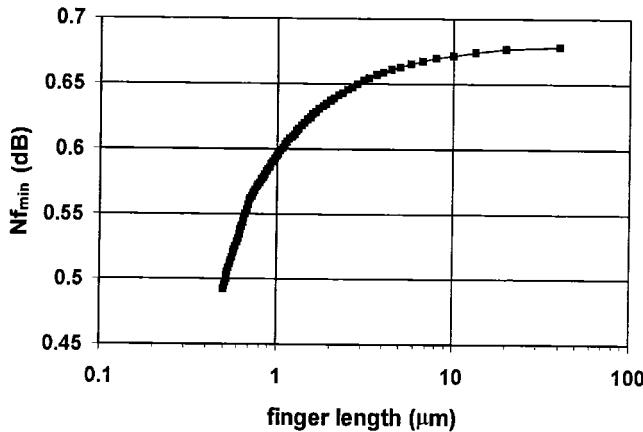
Figure 3.23. Simulated minimum noise figure of a transistor versus the finger length of the transistor for the chosen bias conditions.
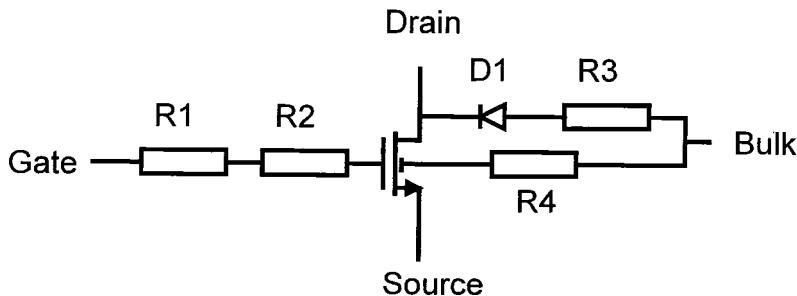
Figure 3.24. Schematic view of the transistor with some extra elements to model the transistor properly. R_1 is the gate resistance of 53 Ω, R_2 is the NQS resistance of 9 Ω, R_3 represents the substrate resistance under the junction capacitor and is 1 Ω, R_4 is the substrate resistance under the bulk and is 1 Ω and, finally, D_1 represents the junction capacitance.

the RF-behavior more accurately. As a supplement to this standard MOS MODEL 9 model, a gate resistance, an NQS (Non-Quasi Static) resistance, and bulk resistances are added. The gate resistance and bulk resistances are calculated by fitting the y-parameters, and the NQS resistance is calculated by using a formula similar to the one described in Chapter 1. The transistor with these added elements is shown in Figure 3.24. After a few iterative steps, the measured and simulated y-parameters are fitted onto each other as can be seen for the y_{11} parameter in Figures 3.25 to 3.26, for example.
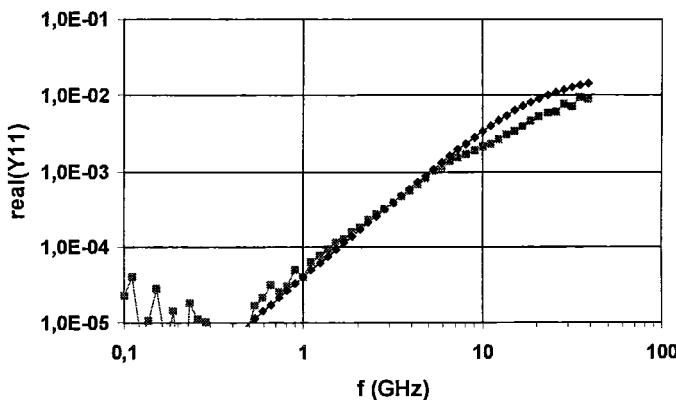
Figure 3.25. The real part of y_{11} of the $W/L = 40\mu m /0.25\mu m$ NMOS transistor for the chosen settings. Top curve: simulations, bottom: measurements.
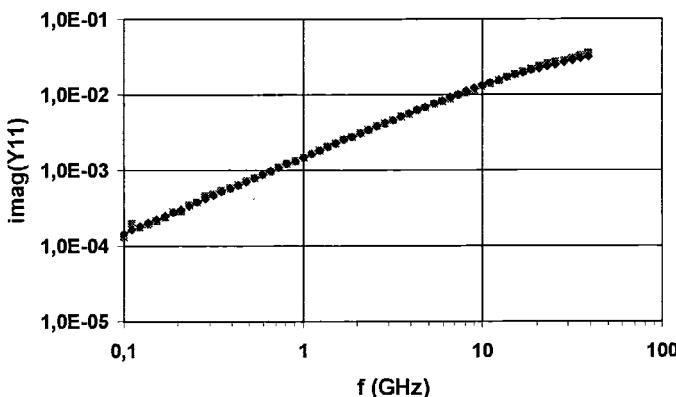
Figure 3.26. The imaginary part of y_{11} of the $W/L = 40\mu m /0.25\mu m$ NMOS transistor for the chosen settings. Measurements and simulations are almost identical.

Figure 3.27 shows the transistor with its parasitic elements in the layout (such as bond pads) with the elements defined as:

- R1=50 Ω : Two port input
- R2=4 Ω : Input parasitic resistance (bond pad)
- R3=2.56 Ω : Input parasitic interconnect resistance
- R4=4 Ω : Output parasitic resistance (bond pad)
- R5=3.34 Ω : Output parasitic interconnect resistance
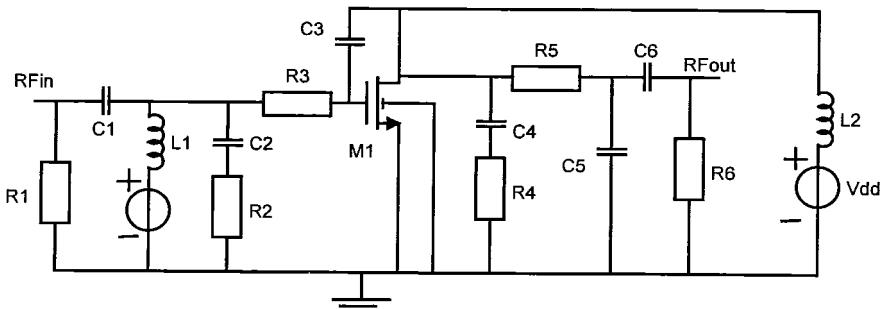
Figure 3.27.   Schematic view of the transistor with its parasitic elements in the layout (C2, C3, C4, C5, R2, R3, R4, R5). See the table for the description of these elements.

- R6=50 Ω : Two port output
- C1: DC blocking
- C2=104 fF: Input parasitic capacitance (bond pad)
- C3=13.4 fF: Parasitic interconnect capacitance between drain and gate
- C4=100 fF: Output parasitic capacitance (bond pad)
- C5=190 fF: Output parasitic interconnect capacitance
- C6: DC blocking
- L1: RF blocking
- L2: RF blocking

Using the obtained RF model, an accurate fit between simulated and measured gain and stability factor can be achieved.

7. **Design of the impedance matching network**. The next step in the design is the impedance matching network, and the biasing of the LNA. Microwave design techniques used to realize this design are explained in Chapter 2. The impedance matching network at the input and output will be performed using transmission lines. The biasing circuit is realized using a $\frac{1}{4}\lambda$ microwave line as explained in Chapter 2. A schematic overview of the design is shown in Figure 3.28. Figure 3.29 shows the Smith chart with noise and gain circles of the transistor. The stability circles shown in this Smith chart are extracted from the measured s-parameters. The simulated s-parameters are fitted on these measured s-parameters (or y-parameters) by adding parasitic elements around the transistor as explained above. The noise parameters used to draw the noise circles in the Smith chart are extracted from these simulations. The
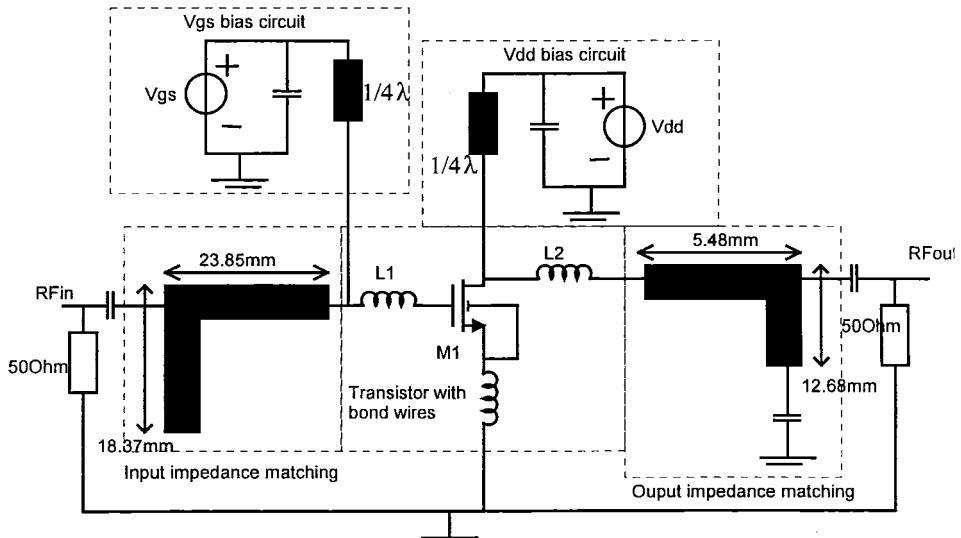
Figure 3.28.   Schematic of the LNA in the case of impedance matching with transmission lines. $L1 = L2 \approx 1$ nH (a single bond wire), $L3 \approx 0.5$ nH (four bond wires) or 3 nH (a single bond wire).

stability circles are shown on the Smith chart. We can be see that care has to be taken to prevent instability of the circuit, i.e. the chosen impedance falls outside these circles. To achieve a noise figure equal to the minimum noise figure, the input impedance should be the optimal noise impedance ("□" in the figure). The output impedance matching network is chosen in such a way that this impedance falls far enough outside the output stability circle ("O"). As a result of these simulations, the lengths of the transmission lines, calculated by using the microwave design techniques as explained in Chapter 2, are given in Figure 3.28.

Simulation and Measurement

The measurement results of the LNA will be discussed in this section. In the case of the transmission line impedance matching, the measured operating frequency is less than 4% away from the expected value. We may conclude that impedance matching with transmission lines is very accurate. The operating frequency of all six measured samples performed within 4% of specification. A die microphotograph of the bonded transistor is shown in Figure 3.30. Grounding of the LNA is performed by means of two bond wires in parallel on each
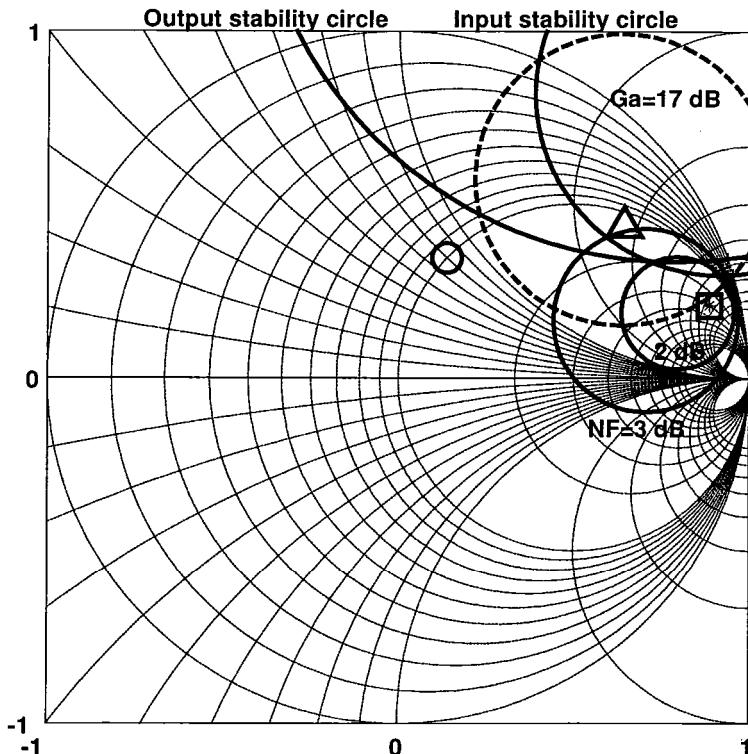
Figure 3.29. Smith chart of the LNA with noise, gain and stability circles. (□): input impedance, (△): output impedance at perfect impedance match, (O): chosen output impedance.

side. Each bond wire has an inductance of 1 nH, two in parallel do have a inductance of $0.5 \cdot \sqrt{2} = 0.7$ nH, where the $\sqrt{2}$ stems from mutual coupling.

The breadboard with the transmission lines is shown in Figure 3.31.

Since the LNA is designed for optimum noise matching, the power matching is not that good (simulated $s_{11} = -3.6$ dB). A degeneration coil can be used to achieve a better power matching. In this case, a single bond wire to the ground is used instead of four bond wires. This single bonding wire has an extra long length. Simulations show that a value of 3 nH is needed, which results in a bond wire of 3 mm. The simulated input power matching is then much better ($s_{11} = -17.4$ dB). The input IP_3 is also better, however, the gain is lower. Table 3.1 shows the measured performances of the two LNAs.

The linearity is measured by a 2-tone measurement with two Marconi signal generators and a spectrum analyzer. The 2-tone frequencies are 2028.8 MHz and 2028.9 MHz. The input power of both tones is −30.5 dBm. The differ-
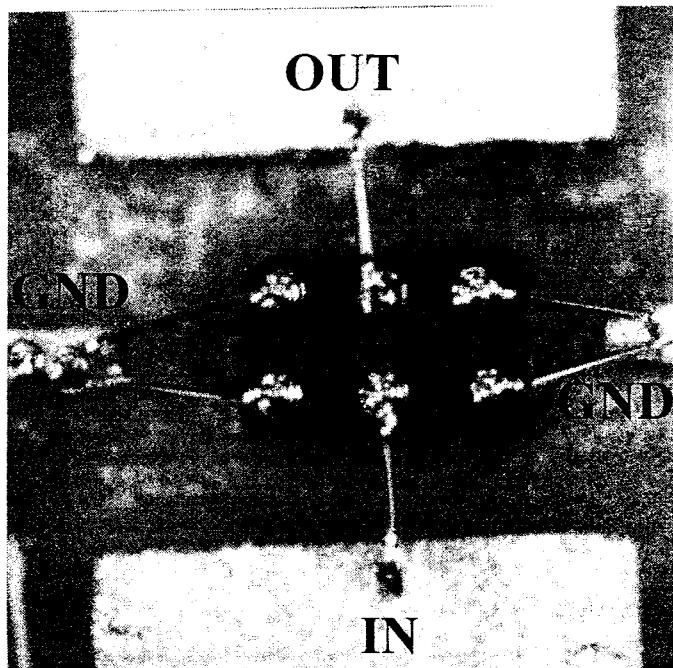
Figure 3.30. A photograph of the bonded LNA. Grounding of the LNA is performed by means of 2 bond wires in parallel on each side of the die.

Table 3.1. Measured performance of the two LNAs

Parameter	measured	measured
Number of bond wires to ground	4	1
Source degeneration coil	No	Yes (3 nH)
Frequency (f)	2.03 GHz	2.05 GHz
Supply current (I_{dd})	4 mA	3.9 mA
Power dissipation (P_{diss})	3.6 mW	3.5 mW
Magnitude s-parameter s_{11}	−6.3 dB	−32.3 dB
Magnitude s-parameter s_{22}	−2.7 dB	−4.8 dB
Voltage Standing Wave Ratio at the input	2.9	1.0
Magnitude s-parameter s_{12}	−23.1 dB	−33 dB
Magnitude s-parameter s_{21}	12.8 dB	7.5 dB
Power gain (G_p)	14 dB	7.5 dB
Available gain (G_a)	16.1 dB	9.2 dB
Output Intercept Point OIP_3	10.5 dB	10.5 dB
Input Intercept Point IIP_3	−3.5 dB	3.0 dB
noise figure (NF)	2.3 dB	3.0 dB
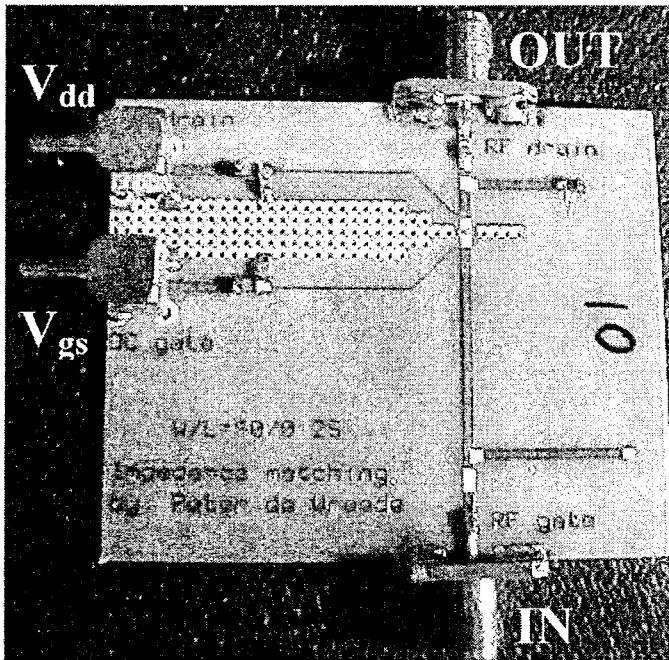
Figure 3.31.   Breadboard with transmission lines.

ence between IIP_3 and OIP_3 is the power gain (G_p). Figure 3.32 shows the measured output power as a function of the input power at the 1st and 3rd harmonic at the standard voltage biasing conditions. The 1 dB compression point is extracted from this measurement, which is -12 dBm.

Although the LNA is made of a single transistor it already performs very well.

3.3.2   Classical LNA Design

One of the most-used structures for an LNA is the input transistor loaded with a cascode transistor, like a CE-stage followed by a CB-stage in the bipolar technique. The cascode transistor is added for isolation purposes. This additional transistor reduces the effect of the collector-base capacitance of the input transistor, which acts as a Miller capacitor. We will treat the design of this LNA structure for a GSM/DCS application. The chosen technology is a 0.25 μm CMOS technology in a 10 mΩ·cm substrate beneath an epi-layer. The power supply is set to 2.5 V.
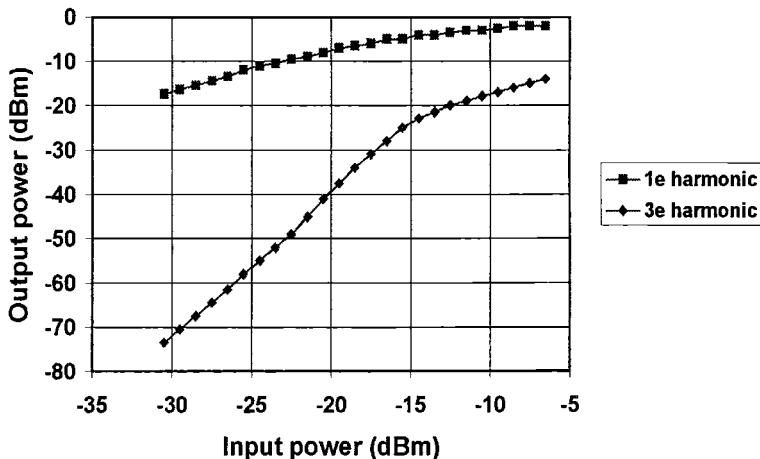
Figure 3.32.   Linearity measurement results using two-tone set up.

The Design

The basic structure of a cascoded input transistor is depicted in Figure 3.33. Transistor M_1 is the input transistor and M_2 acts as the cascode transistor. Inductor L_1 represents the bond wire as a connection between the on-chip ground line and the off-chip ground plate. Inductors L_2 and L_3 are used to tune the LNA. Capacitor C_{gs} represents the parasitic gate-source capacitance of M_1. The bias circuit for M_1 is not yet shown, but will be discussed later. Assuming power matching, then the source impedance of 50 Ω should be equal to the real part of the input impedance which is roughly determined by

$$\text{Re}(Z_{in}) \approx \frac{g_{m1}L_1}{C_{gs}} \tag{3.4}$$

Hence, although the bond wire is a necessity for connection, it can help to increase the real part of the input impedance. In (3.4), g_{m1} represents the transconductance of M_1. Inductors L_1 and L_2 together with capacitance C_{gs} form a resonance tank which should resonate at the proper frequency. Therefore

$$\omega_c^2 = \frac{1}{(L_1 + L_2)\, C_{gs}} \tag{3.5}$$

The source resistance R_s at node RF_{in} will degenerate the quality factor Q of the capacitance according to

$$Q = \frac{1}{R_s \omega C_{gs}} \tag{3.6}$$
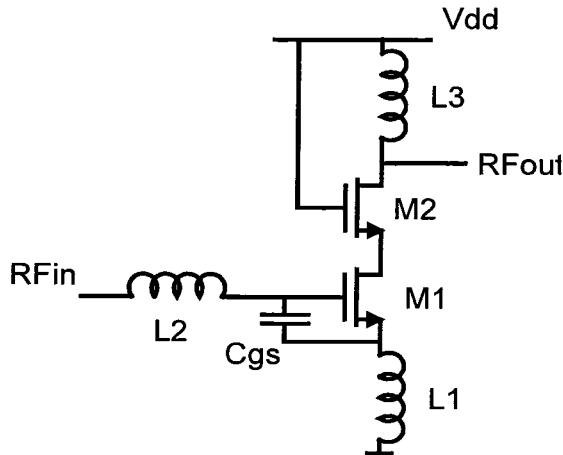
Figure 3.33.   Basic configuration of a cascoded input transistor.

On the other hand, this Q should not be too high, as it will increase the gate voltage of the M_1 and thus lower the IIP_3. Similarly, the loaded Q of the bond wire L_1 is equal to

$$Q = \frac{1}{g_{m1}\omega L_1} \tag{3.7}$$

and to have a reasonable design, (3.6) should be equal to (3.7). While g_{m1} represents the transconductance of M_1, we can also define the transconductance of M_1 together with the input matching network, i.e.

$$G_m = \frac{\partial I}{\partial v_{RFin}} = \frac{1}{\omega L_1} \tag{3.8}$$

We can use this overall transconductance to calculate the input referred noise voltage and, together with (1.31) the noise factor can be obtained as

$$F = 1 + \frac{2}{3}\frac{\omega L_1}{QR_s} + \frac{1}{15}\frac{1}{g_{m1}R_s}\left(1 + \frac{1}{Q^2}\right) \tag{3.9}$$

where the third term represents the contribution due to thermal noise in the channel.

A first order design can be realized by setting the frequency to 2 GHz and using a bond wire of 1 nH inductance for L_1. Allowing a Q of 4, (3.6) gives the capacitance value of roughly 0.4 pF. Consequently, from (3.5) we obtain the inductance value of 14 nH for L_2. The transconductance of M_1 can then be derived as 20 mA/V. Knowing that the technology provides 0.2 fF/μm gate
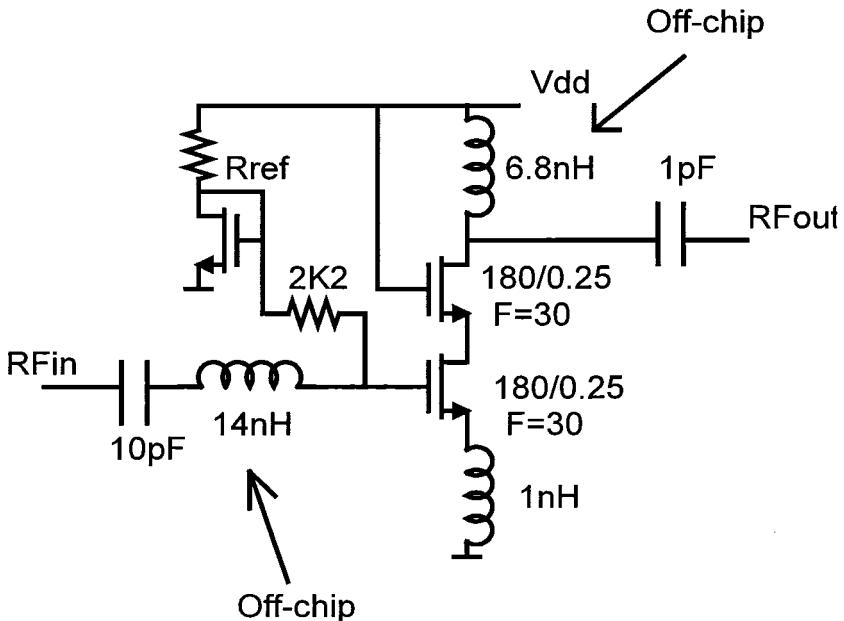
Figure 3.34.   Design of the DCS LNA with a standard cascode configuration.

oxide capacitance, 0.4 pF will be realized with a width of 200 μm for M_1. From the transconductance and the dimension of M_1 (the minimum length has been chosen) follows the current consumption of 5 mA. The final design is shown in Figure 3.34. A folding structure (F equals 30, yielding a finger length $L_{finger} = 6$ μm) for the transistors has been chosen to reduce gate resistance. The cascode transistor M_2 has the same dimensions as M_1. This has the advantage that the drain of M_1 can be merged with the source of M_2, thus reducing parasitic capacitances significantly. The 10 pF capacitance is for DC blocking and hardly changes the resonance frequency of the input tank. Transistor M_1 is biased using a current-mirror concept. An additional resistance of 2 kΩ has been added to increase the input impedance seen from M_1 towards the bias circuit. The higher the impedance, the less the bias circuit will influence the RF performance. The bias current is set with resistor R_{REF}, which is 10 times lower than the current through M_1 to reduce power dissipation. The value of the load inductance is chosen such that the output impedance is roughly 50 Ω at the operating frequency.
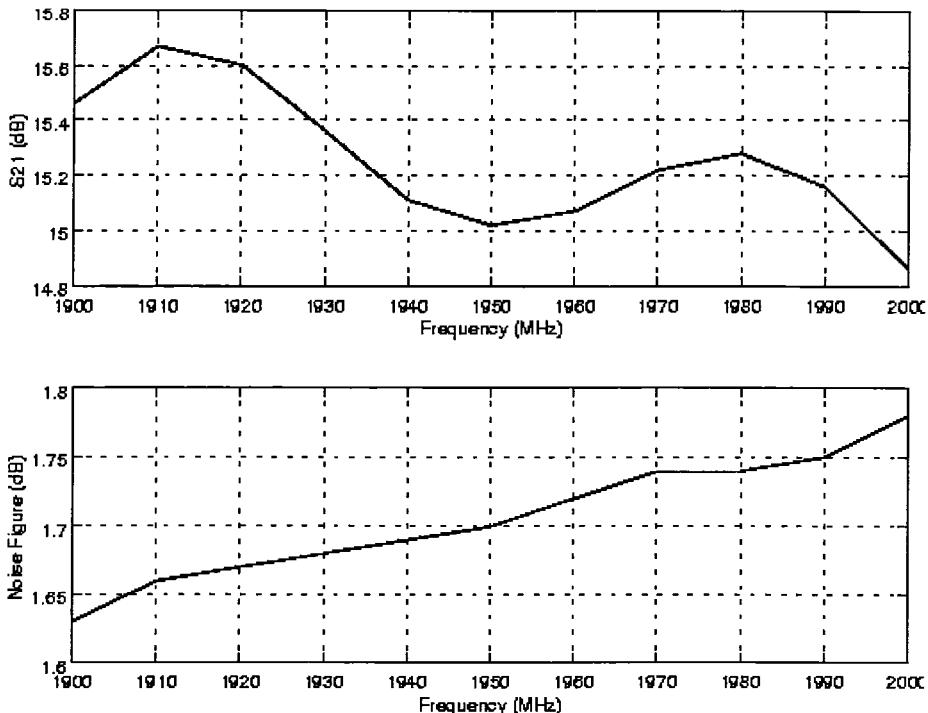
Figure 3.35. Measured performance of the LNA. The top picture shows the measured transducer gain (s_{21}) and the bottom shows the measured noise figure, both as a function of the frequency.

Measurement Results

The measured performance is shown in Figure 3.35. A noise figure of 1.7 dB has been achieved at a transducer gain of 15 dB. The measured IIP_3 was 1.5 dBm at 12 mW power dissipation. The fact that the noise figure of this LNA is lower than the single transistor LNA can be explained by the fact that the technologies used were not exactly the same. The difference was in a higher gate resistance for the single transistor, directly reflecting the higher noise levels.

3.4 EVALUATION

Many LNAs have been presented in the literature: [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], and [11]. Design [2] is a bipolar design, [3] is a 0.25 μm CMOS design, [8] is a 0.35 μm CMOS design, [4] is a 0.6 μm CMOS design, [7] and [9] are designed in SIMOX technology and [11] in SiGe. Finally, [5] is designed in Silicon-on-Anything. Let us also refer to the bipolar LNA of
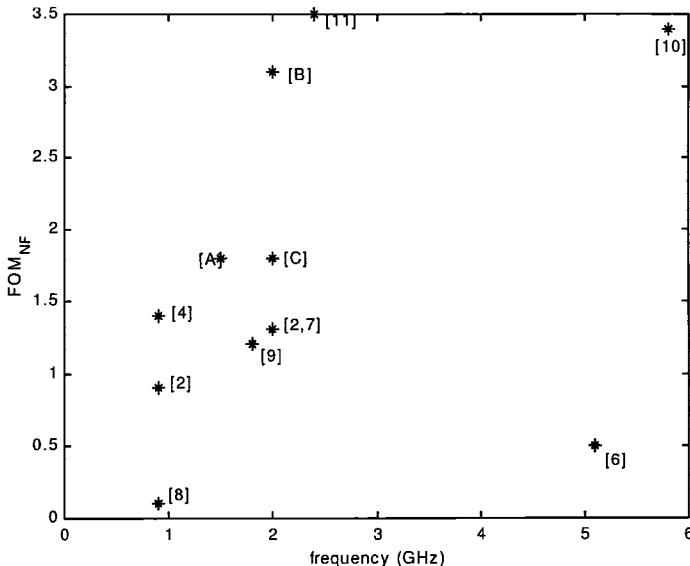
Figure 3.36. Comparison of several LNA designs for the FOM$_{NF}$ as a function of frequency.

section 3.2 as [A], the single MOS LNA as [B] and the cascode CMOS LNA as [C]. A Figure of Merit (FOM) proposed in [10] is used to compare the proposed LNAs with earlier results as a function of frequency. This FOM compares the performance of LNAs based on the noise figure. A modification to this FOM to obtain a more fair comparison leads to

$$FOM_{NF} = \frac{Gain}{F \cdot P_{diss}(mW)}$$

where the gain is defined in its linear dimension. Based on this FOM, the LNAs discussed in this chapter perform competitively with LNAs designed in recent SiGe and SIMOX processes (Figure 3.36). LNA [5], designed in the SOA technology, has an FOM$_{NF}$ of 39 because it is developed in a dedicated low power technology. The SiGe design of [11] shows a good result due to its extreme low noise figure (0.95 dB).

A FOM can also be defined for the linearity

$$FOM_{IP3} = \frac{IIP_3(mW) \cdot Gain}{P_{diss}(mW)}$$

for which the ranking of the LNAs is given as shown in Figure 3.37. Looking at the figure, it becomes clear that the SOA process does not produce high linearity devices. Finally, an FOM can be defined for an overall dynamic
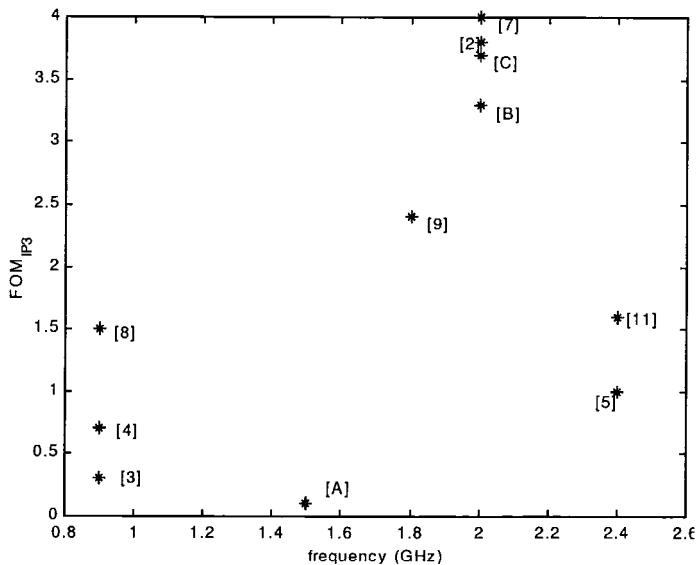
Figure 3.37.   Comparison of several LNA designs for the FOM_{IP3} as a function of frequency.
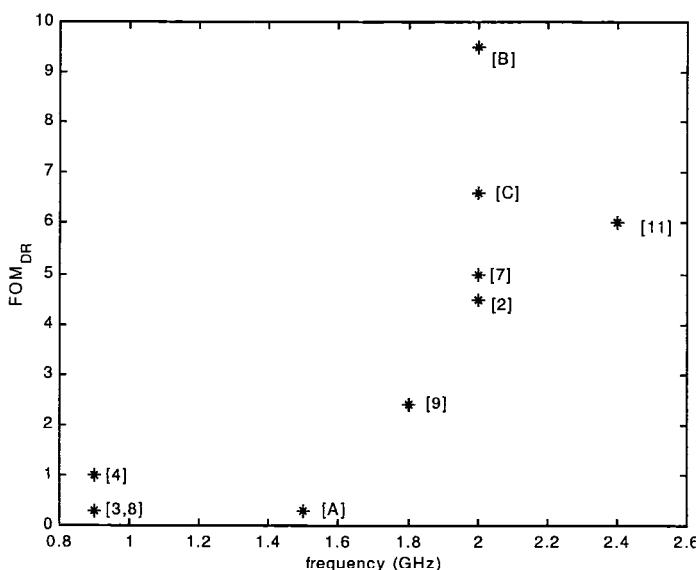
Figure 3.38.   Comparison of several LNA designs for the FOM_{DR} as a function of frequency.

range, $FOM_{DR} = FOM_{IP3} \cdot FOM_{DR}$. The ranking based on this FOM is depicted in Figure 3.38. A single transistor LNA performs very well, however, it is not simple to use in real practical applications, as most applications also demand good isolation. It can also be seen that MOS technology can be very advantageous for RF applications. Passive components are still a major problem in standard MOS technology, but this may change with dedicated RF CMOS technologies. The silicon-on-anything technology is very interesting for low power and low-noise applications, where linearity is not the main issue.

REFERENCES

[1] M. Bopp *et al.*, " A DECT Transceiver Chip Set using SiGE Technology," in *IEEE International Solid-State Circuits Conf. (ISSCC)*, 1999, pp. 68–69.

[2] K. Fong, "Dualband High Linearity Variable-Gain Low-Noise Amplifiers for Wireless Applications," in *IEEE International Solid-State Circuits Conf. (ISSCC)*, 1999, pp. 224–225.

[3] P. Orsatti *et al.*, "A 20mA-receiver, 55mA-trasnmit GSM transceiver in 0.25mm CMOS," in *IEEE International Solid-State Circuits Conf. (ISSCC)*, 1999, pp. 232–233.

[4] J. Zou and D. Allstot, "A Fully Integrated CMOS 900 MHz LNA Utilizing Monolithic Transformers," in *IEEE International Solid-State Circuits Conf. (ISSCC)*, 1998, pp. 132–133.

[5] A. Wagemans *et al.*, "A 3.5mW 2.5GHz diversity receiver and a 1.2mW 3.6 GHz VCO in Silicon-On-Anything," in *IEEE International Solid-State Circuits Conf. (ISSCC)*, 1998, pp. 250–251.

[6] T. Liu, "5 Ghz CMOS Radio Transceiver Front-end chip," in *IEEE International Solid-State Circuits Conf. (ISSCC)*, 2000, pp. 320–321.

[7] M. Harada, "0.5-1V 2GHz RF Front end Circuits in CMOS/SIMOX," in *IEEE International Solid-State Circuits Conf. (ISSCC)*, 2000, pp. 378–379.

[8] G. Gramegna *et al.*, "Ultra-wide Dynamic Range 1.75 dB Noise-Figure, 900 MHz CMOS LNA," in *IEEE International Solid-State Circuits Conf. (ISSCC)*, 2000, pp. 380–381.

[9] K. Takikawa *et al.*, "RF circuits Technique for Dual-band Transceiver IC for GSM and DCS1800 Applications," in *European Solid-State Circuits Conf. (ESSCIRC)*, 1999, pp. 278–281.

[10] H. Ainspan *et al.*, "A 6.25 GHz Low DC Power Low-Noise-Amplifier in SiGE," in *IEEE Custom Integrated Circuits Conf. (CICC)*, 1997, pp. 177–180.

[11] J.R. Long *et al.*, "RF Analog and digital Circuits in SiGe technology," in *IEEE International Solid-State Circuits Conf. (ISSCC)*, 1996, pp. 82–83.

Chapter 4

Mixers

In the receive path, a low noise amplifier is normally followed by one or more mixers to translate the RF frequency into a base band frequency. These down conversion mixers multiply two signals, the RF signal and the local oscillator (LO) signal. The multiplication of these two distinctly different signals produces harmonics, due to non-idealities, influencing the information signals. Up conversion mixers are used in a similar way in the transmit path. Several mixer configurations together with their design problems will be treated in this chapter.

4.1   SPECIFICATION

Mixers are used to perform frequency translation and therefore have two distinctly different input signals. The input signals in the receiver mixer are both RF signals, especially in zero or near-zero IF receiver concepts. The frequency spectrum of the output signal is then at a low IF frequency. At the transmit side, on the other hand, the two input frequencies of the mixer are far apart, one RF signal coming from the oscillator and the other being an IF signal. As frequency translation can be viewed as a multiplication of time signals, linearity is an important design specification. Non-linear multiplication causes harmonics that can fall into the information band.

Suppose that the input signal on the RF port of a down converter mixer is given as

$$v_{RF}(t) = \sin(\omega_{RF}t + \omega_{bb}t) \tag{4.1}$$

where the RF carrier frequency is 1 GHz and the signal information frequency $f_{bb} = 1$ MHz. If the local oscillator (LO) frequency to 1.5 GHz and the mixer is ideal, except for linearity, i.e. $OIP_3 = -5$ dBm, the frequency spectrum
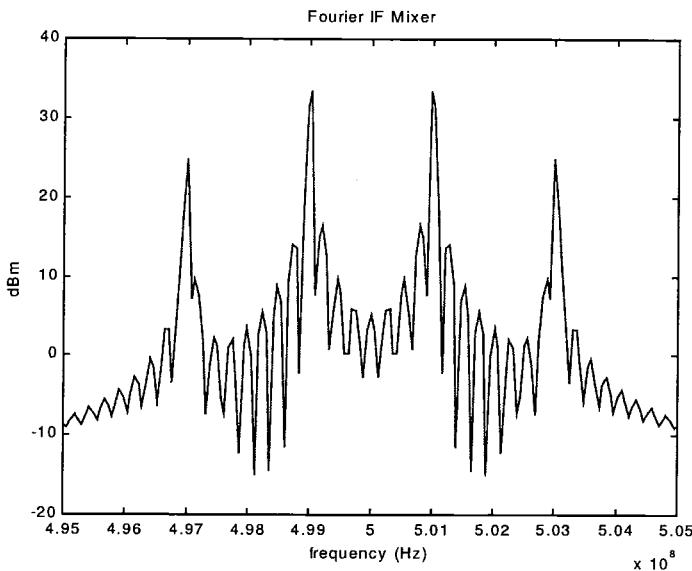
Figure 4.1.   Spectrum at the IF port of a mixer with an OIP_3 of −5 dBm.

around the IF frequency shows the wanted tone at 499 MHz, the image tone at 501 MHz and a strong third harmonic tone at 497 MHz, for instance (see Figure 4.1, obtained using the Matlab routines of Appendix A). If the IF filters following the mixer are not that good, harmonics will be present at the output of the receiver and will thereby distorting the information signal. When a strong interferer is present in a co-channel or adjacent channel, it is clear that these signals will mix with the wanted tones, and an unwanted frequency spectrum arises. As an example, consider the situation when an additional tone is present at 1.0018 GHz. The spectrum at the IF port of the mixer is depicted in Figure 4.2.

Mixers can be divided into several classes, depending on the criteria used. The first commonly used criteria is the provided gain. The term "gain" often leads to confusion when considering a mixer. The *voltage conversion gain* is defined as the ratio of the rms voltage of the IF signal to the rms voltage of the RF signal. The *power conversion gain* of a mixer is defined as the IF power delivered to the load divided by the available RF power from the source. For example, if the treated mixer has a power conversion gain of 10 dBm, the obtained output spectrum changes from the one in Figure 4.1 into that of Figure 4.3. It is not only the wanted tones that are amplified. The harmonics are also amplified, and even have a higher gain, thus showing the increased non-linearity of the mixer.
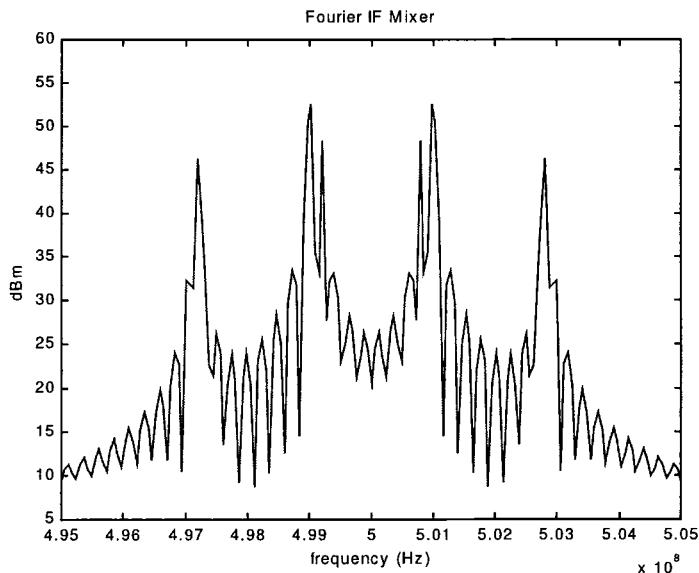
Figure 4.2. Output spectrum of the mixer when a strong interferer is present. The wanted tone at 499 MHz is difficult to distinguish from the unwanted tone at 499.2 MHz.
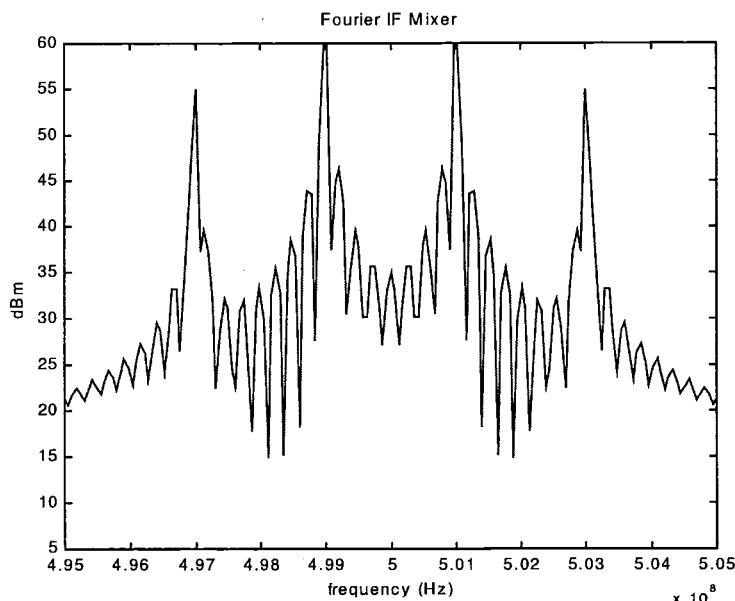
Figure 4.3. Output spectrum at the IF port of the mixer when the mixer has a power conversion gain of 10 dBm.
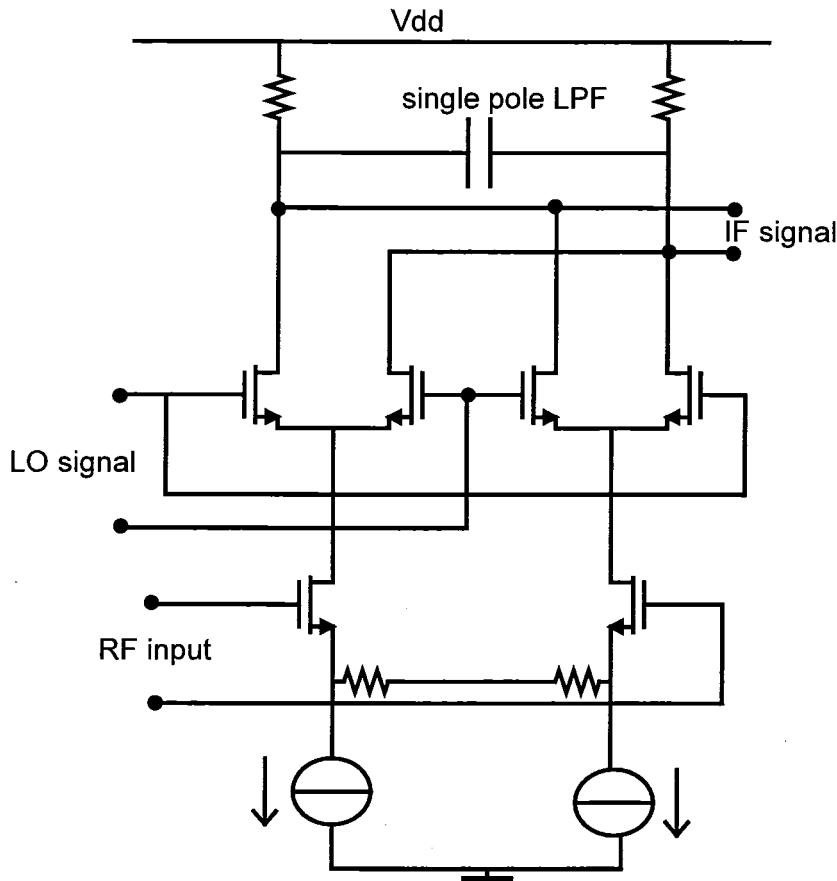
Figure 4.4.   "Gilbert" mixer implemented in MOS technology.

Using the gain criteria, we can distinguish "Gilbert" or active mixers and "passive" mixers.

The Gilbert type mixer is based on a transconductor converting the incoming RF into a current, followed by current steering transistors reversing the sense at the load as controlled by the LO drive (Figure 4.4). The load may be resonant, broadband, or in the case of a down-mixer, formed with a low pass pole. There is usually some modest signal gain through the circuit, allowing a single stage LNA to be used. Note that since the LO switches the signal, there is no attempt at linear multiplication. Consequently, the LO drive should be large enough to ensure that the transconductor output current is fully switched, and this implies a relatively large LO drive (in bipolar implementations, 300 to 400 mV_{pp} is typical on each side). A further reason for using a relatively large LO drive is

related to noise. At first sight, this type of mixer appears to have no obvious source of $1/f$-noise. The transconductor transistors in the signal path have their output currents switched alternately between the positive and negative signal outputs, and hence the noise from these devices appears to be chopped. Thus, any low frequency component is therefore removed. The switching transistors appear to function as cascodes when active, and should therefore also contribute negligible noise. In reality, there must be a finite time when both steering transistors are on, since there is DC as well as signal current to be passed. During this period, the steering devices look like a differential pair without any degeneration, and their noise is visible at the outputs. Note that a larger LO drive amplitude leads to shorter transition times and hence minimizes the switching transistor noise, albeit at the expense of more power in the LO buffering.

The second main type of mixer used is the so-called "passive" mixer. In this configuration, the MOS devices are used as reversing switches in a series signal path connection, with no gain implied in the circuit. As a consequence, the noise of the following stages is more significant and may demand higher gain in the LNA to make it manageable. Large switching transistors are sometimes favored to make their non-linear resistance negligible in the signal path, and hence to improve intermodulation performance. The main purported advantages here are that there is no DC consumption, and that if the circuit is AC coupled, there is no DC through the mixer transistors. According to the literature, there should therefore be no $1/f$-noise present (Figure 4.5). The low frequency output of the mixer may be connected directly to the "virtual earth" input of a transimpedance amplifier. Capacitance can be added to the input of the op-amp to allow some filtering of the RF components. In this way, the mixer transistors can be viewed as either the V-I conversion impedances, or as switching a current (but only AC) to the virtual earth points from another pair of series impedances. In this second context, there is some argument for having very small capacitances in series with the inputs of the mixer to allow their impedance to dominate at RF, and to provide a higher impedance at baseband.

Another way of distinguishing mixers is to see whether or not the signals are applied differentially. When the LO signals are differential, but the RF signal is not, we speak of a *single-balanced mixer*. If the RF signal is applied differentially, then we speak of a *double-balanced mixer*. This type of mixer generates less even-order harmonic distortion while the single-balanced mixers exhibit less input-referred noise for a given power dissipation. Differential output ports are used instead of single-ended output ports to avoid the problem that the output port of the mixer contains a component with no frequency translation (*"direct feedthrough"* effect).
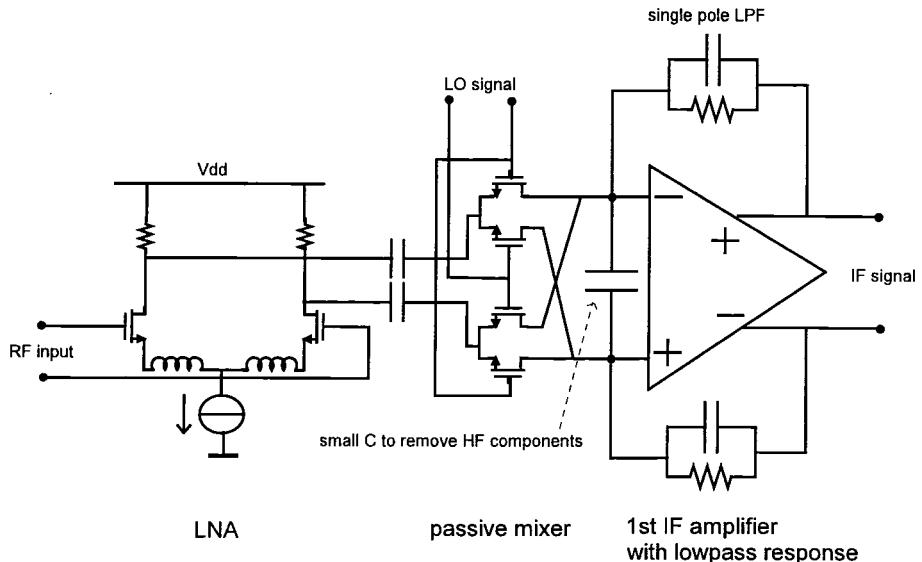
Figure 4.5.   Example of "passive" mixer implementation in MOS receiver path.

A final design parameter of mixers is the noise figure. Consider a noiseless mixer with a gain of unity. The spectrum at the RF port consists of a signal band and an image band, if we assume that the desired signal spectrum resides on only one side of the LO frequency. This is common for heterodyne architectures. Both bands contain the signal component as well as the thermal noise of the source resistance (i.e. the load of the preceding LNA). Due to the frequency translation, the frequency band around IF therefore contains not only the information in the signal band, but also that from the image band. This doubles the noise in the IF band, making the SNR at the output half that at the input, resulting in a noise figure of 3 dB for the noiseless mixer. This measurement provides the *"single-sideband"* (SSB) noise figure. Now suppose that the LO frequency is also situated in the signal band, reflecting a homodyne receiver concept. As no image band exists, the SNR at the output is equal to that at the input and hence the *"double-sideband"* (DSB) noise figure is 0 dB for the noiseless mixer. The 3 dB difference in the noise figure makes it important to mention which measurement set up was used.

4.2   BIPOLAR MIXER DESIGN

Let us first focus on a single-balanced active bipolar mixer as drawn in Figure 4.6. The relatively high input impedance of this circuit might cause problems
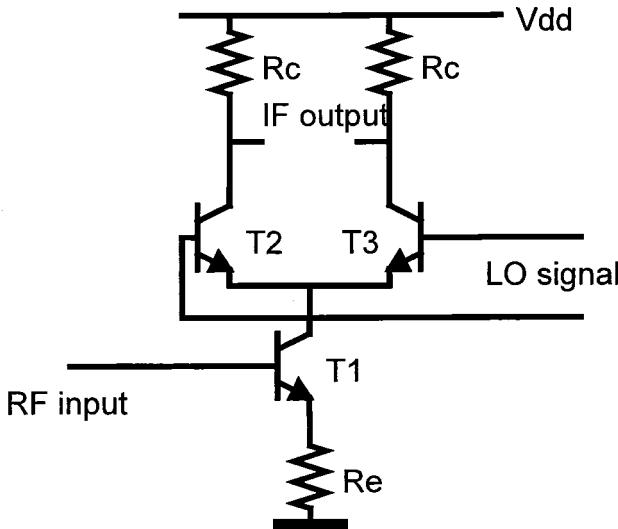
Figure 4.6. Single-balanced active mixer with emitter degeneration in the RF part.

when the mixer is preceded by an image reject filter. Such a filter is usually off-chip and is designed for 50 Ω matching. Impedance converters may be used, but they add additional components. We can also solve the impedance matching problem by using the emitter T1 as an input port. Such a circuit is depicted in Figure 4.7. At the cost of increased noise, a better impedance matching with improved linearity is now obtained.

The input impedance of the circuit in Figure 4.7 is $R_e + 1/g_{m1}$, and can be set to 50 Ω. Assuming a source resistance R_s, the RF current i_{T1} of T1 is given by

$$i_{T1}(t) = v_{RF}(t)/(R_s + R_e + 1/g_{m1}) \qquad (4.2)$$

where v_{RF} is the RF voltage at the RF input of the mixer. Assuming perfect switching, a 50% duty cycle, and neglecting the higher harmonics, the IF output voltage v_{IF} is given as

$$v_{IF}(t) = \frac{v_{RF}(t)R_c}{R_s + R_e + 1/g_{m1}} \cdot \frac{4}{\pi} \cos(\omega_{LO}t) \qquad (4.3)$$

where the LO frequency is set to $\omega_{LO}/2\pi$. Applying a Fourier transformation, the spectrum of v_{IF} is given as

$$V_{IF}(\omega) = \frac{V_{RF}(\omega \pm \omega_{LO})}{R_s + R_e + 1/g_{m1}} \cdot \frac{2R_c}{\pi} \qquad (4.4)$$
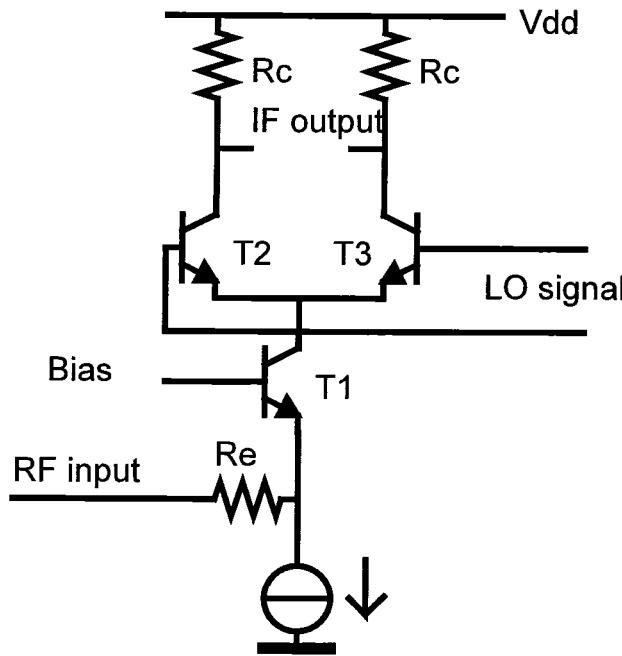
Figure 4.7.   Single-balanced active mixer with the RF input at the emitter of the input device.

The voltage conversion gain is now derived from

$$A_v = \frac{2}{\pi} \cdot \frac{R_c}{R_e + 1/g_{m1}} = \frac{2}{\pi} \cdot \frac{R_c}{R_s} \qquad (4.5)$$

under the assumption of input impedance matching. The power delivered to a load is equal to

$$P_{del,output} = \frac{V_{IF,rms}^2}{2R_c} \qquad (4.6)$$

The available power at the input is given by

$$P_{av,input} = \frac{V_{RF,rms}^2}{4R_s} \qquad (4.7)$$

Combining (4.5), (4.6) and (4.7) finally leads to the delivered power conversion gain

$$A_p = \frac{1}{\pi^2} \cdot \frac{2R_c}{R_s} \qquad (4.8)$$

Note that $A_p \neq A_v^2$, which often leads to confusion when referring only to conversion gain.

The main source of distortion in mixers like the one depicted in Figure 4.7 is the input voltage to output current converter in the RF part, i.e. T1. Large emitter resistances are needed to overcome this distortion which, in turn, cause a significant contribution to the noise. Design trade-offs are needed here. A second source of distortion stems from the non-linear base-emitter capacitance of T2 and T3. Certainly when T2 and T3 are on simultaneously for a fraction of the period, the collector current of T1 flows through these non-linear capacitances and will therefore be distorted [1]. To overcome this problem, non-overlapping sinusoidal LO signals can be used, preventing T2 and T3 from being on simultaneously.

Noise has always been a problem in this type of mixers. A nice explanation about noise contributions in single-balanced bipolar mixers can be found in [2]. The noise contribution of T1 is mainly from to the thermal noise due to the emitter resistor R_e and the collector shot noise. This latter source can be decreased by increasing the current through the device, since the input referred noise voltage $\overline{v_{noise}^2}$ is equal to $2kT/g_m = 2q/I_{c1}$. The noise contributions of T2 and T3 come from the shot noise current $\overline{I_{noise}^2} = 2qI_c$ and the thermal noise of the base resistance. These noise contributions are injected into the output nodes of the mixers via a capacitive path to ground, formed by the base-emitter junctions of T2 and T3 and the base-collector, collector-substrate junctions of T1. The currents through T2 and T3 should be decreased to reduce the shot current noise, which is exactly the opposite what has to be done to lower the noise contribution of T1. A design trade-off is therefore needed here. Large LO signals are needed to further reduce the noise contributions of T2 and T3. To reduce the thermal noise of the base resistance, the sizes of T2 and T3 should be large, leading to large base-emitter capacitances. This latter effect results in large injected voltage noise into the output node. Small capacitances are needed to overcome this, hence small sizes should be used, which in turn leads to large base resistances. A design trade-off is once again necessary. This concludes the discussion on bipolar mixers.

4.3 CMOS MIXERS

In the receiver path, active CMOS mixers are more and more replaced by passive CMOS mixers due to power consumption restrictions. In contrast, the active mixer is still used in the transmit path to get more power gain. In this section, both type of CMOS mixers will be discussed.
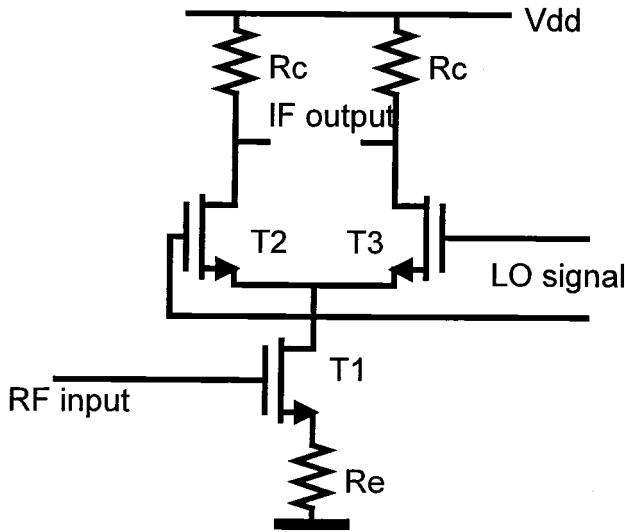
Figure 4.8.   Single-balanced CMOS mixer

4.3.1   Active CMOS Mixers

The need for good switching devices in active mixers leads to the use of MOS device-based mixers. The most simple active CMOS mixer is presented in Figure 4.8. One of the most important design criteria is the noise produced by the mixer. Many literature have been published recently on this topic. Assuming $I_0 = I_2 - I_3$, and under large LO swing conditions this current will be a function of the oscillator voltage output $V_{LO}(t)$ and $I_1 = I_{bias} + i_{RF}$, where I_{bias} is the bias current and i_{RF} is the small signal current due to the signal at the RF input. Consequently, $I_0 = F(V_{LO}(t), I_{Bias} + i_{RF})$ and, since i_{RF} will be small, a Taylor expansion may be applied, yielding

$$I_0 = a_o(t) + a_1(t) \cdot i_{RF} \qquad (4.9)$$

where both $a_o(t)$ and $a_1(t)$ are periodic waveforms.
  We can easily see that

$$a_1(t) = \frac{g_{m2}(t) - g_{m3}(t)}{g_{m2}(t) + g_{m3}(t)} \qquad (4.10)$$

and hence this factor will play an important role in the noise behavior. Also note that a cyclo-stationary behavior may be observed due to the time dependency. The term $a_o(t)$ will disappear in double balanced mixer structures. In [3], this time-varying bias condition approach is used to calculate the overall noise

behavior of a single balanced mixer. They found the single-sideband noise figure to be

$$NF_{ssb} = \frac{\alpha}{c^2} + \frac{(\gamma_1 + r_{g1} g_{m1}) g_{m1} \alpha + 2\gamma_2 \overline{G} + (R_{LO} + 2r_{g2})\overline{G^2} + 1/R_L}{c^2 g_{m1}^2 R_s}$$

(4.11)

where R_s, R_{LO} and R_L are the source and LO and load resistances, respectively. The thermal noise generated by a MOS transistor in saturation has been assumed to be

$$\frac{\overline{i_n^2}}{\Delta f} = 4kT\gamma g_m$$

(4.12)

where $\gamma \approx 2/3$ is for long transistor devices and g_m is the transconductance of the MOS device. The parameter r_g resembles the polysilicon gate resistance. Furthermore, parameter α is the total power of the waveform $a_1(t)$, and $\overline{G}$ is the time average of $G(t)$, the small signal function of $a_1(t)$. The parameter c represents the first order Fourier coefficient and is given as

$$c \approx \frac{2}{\pi} \left(\frac{\sin(\pi \Delta f_{LO})}{\pi \Delta f_{LO}} \right)$$

(4.13)

Although expression (4.11) covers any possible way of switching, it is difficult to use (4.11) in practice. Some simplifications can therefore be considered, like the assumption that the MOS transistors are only hard switched, for instance; they are turned on and off instantaneously [4].

Assuming a current I flowing through transistor T_1, then a current equal to I will flow through T_2 for half a period, and similarly through T_3 for the next period. The output is now a square wave function at a frequency of ω_{LO} with a zero DC value. Assume that the $1/f$-noise sources of T_2 and T_3 are referred to the gate of T_2 as a voltage source V_n. This noise will modulate the time at which the pair T_2 and T_3 switches. Consequently, noise advances or retards the time of zero-crossings, and can be treated as an additional pulse train of random width Δt and amplitude $2I$, superposed with the square function. Note that Δt is dependent on the slope S of the LO voltage at the switching instance. As this happens at the rising and falling edge of the square function, the pulse train has a frequency of $2\omega_{LO}$. In the extreme case, we may approximate the pulse by ideal delta functions. The frequency spectrum of the base noise current of T_1, i_{n1} is then sampled and appears at integer multiples of $2f_{LO}$ at the IF output,

$$i_{o,n1} = \frac{1}{\pi} \frac{I}{A} V_n(f \pm 2nf_{LO})$$

(4.14)
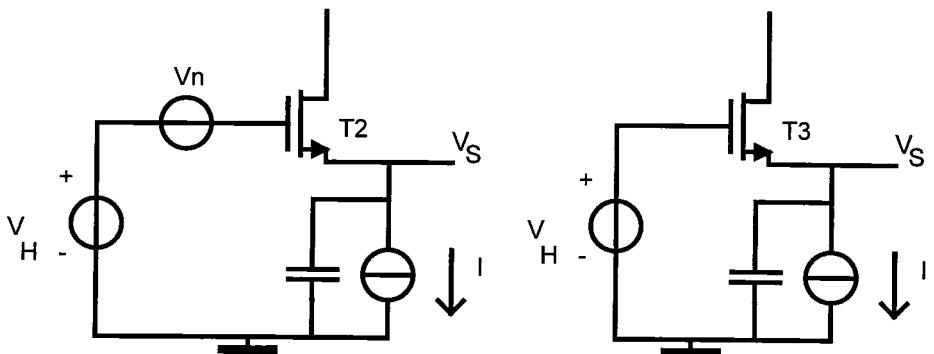
Figure 4.9. Mixer with square-wave function at LO at each half cycle: the first half cycle on the left and the second half cycle on the right. The positive amplitude level of the switching function is denoted by V_H.

where A is the amplitude of the square wave function V_H and $n \in N$. This spectrum therefore also includes the $1/f$-noise, although it is not present around the LO frequency. The above reasoning implies that $1/f$-noise at the output of the mixer may be eliminated if the zero crossing happens perfectly, i.e. $\Delta t = 0$. However, there is another mechanism to introduce the flicker noise at the output. During hard-switching, the gate voltage of T_2 in the first half period consists of the amplitude level of the square-wave function superposed on the noise amplitude. In the second period, T_3 becomes active and only the amplitude of the square-wave function is present at the gate. This mechanism is visualized in Figure 4.9. The voltage V_S charges and discharges exponentially, due to the capacitance C, and hence the output current alternates at twice the LO frequency with a non-zero DC value. This again indicates that baseband $1/f$-noise is present at the mixer output. The amount of output noise current due to T_2 is given by

$$i_{o,n2} = \frac{2}{T_{LO}} C V_n \qquad (4.15)$$

and can be decreased by reducing the tail parasitic capacitance and applying a square-wave function with sharp transitions.

So far we have only discussed the influence of $1/f$-noise sources, but the white noise sources also play an important role. The spectral power density of the white input noise is usually given by:

$$\overline{V_n^2} = \frac{4kT\gamma}{g_m} \qquad (4.16)$$

where γ is the channel noise factor (traditionally 2/3 for long-channel devices), and g_m is the transconductance of the MOS device. In the mixer, the operating point is time dependent and, therefore, g_m should be replaced by $G_m(t)$, a function of time. A LO sinewave $V = 2A\sin(\omega_{LO}t)$ produces a time-periodic $G_m(t)$ which is then nonzero over a time window $\Delta V/2A\omega_{LO} = \Delta V/S$. The white noise current contribution of T_3 can therefore be expressed as

$$\overline{i^2_{o,n3}} = 4kT\gamma\frac{4I}{ST_{LO}} \approx 4kT\gamma\frac{I}{\pi A} \tag{4.17}$$

with the approximation valid for sinusoidal LO signals[1]. If we include noise due to switches, a load R_L and the mixer conversion gain of $2/\pi$, the total white noise at the mixer output is given by

$$\overline{V^2_{o,n}} = 8kTR_L + 8kT\gamma\frac{R_L^2 I}{\pi A} + m \cdot \frac{4kT\gamma}{g_m}\left(\frac{2}{\pi}g_m R_L\right)^2 \tag{4.18}$$

where m represents the accumulation of noise after aliasing.

For a scaled double-balanced mixer with the same total current as a single-balanced mixer, the total output noise is the same as (4.18). As the conversion gain from the differential input is half of a single input, the equivalent input referred noise is twice as large. The main advantage of a double-balanced mixer is therefore not in the noise suppression, but in LO feedthrough suppression.

Consider the double-balanced mixer of Figure 4.10, which was designed for a GSM/DCS/PCS application. Degeneration resistors are used in the sources of T_{1a} and T_{1b} to increase the linearity of the RF port (not shown in the figure). An NMOS device is normally used as the current source at the common source point of T_{1a} and T_{1b},.to set the bias current. A current mirror configuration is then needed to set this bias current. We can choose the current mirror ratio to be more than one to reduce the overall power consumption. However, the $1/f$-noise of the mirror transistor is then increased by the mirror ratio and injected into the RF port of the mixer. To reduce this effect, the current mirror ratio can be reduced to a factor equal to one, but at the cost of more power consumption. A better solution is to use a resistor R_{bias} to set the bias current, as shown in Figure 4.10. T_3 is used together with a resistor R_1 to set the proper bias point for T_{1a} and T_{1b}. R_1 has a high resistance value to ensure that this bias circuit will not affect the RF operation. The resistors are realized using n^+-poly and have

---

[1] The output noise density is therefore not dependent on the transistor sizes. The finite bandwidth of the sampling pulses limits the number of aliases. The integrated rms output noise therefore remains constant. This behavior is similar to the voltage noise on a switched capacitor, kT/C, and hence also independent of the size of the switch.
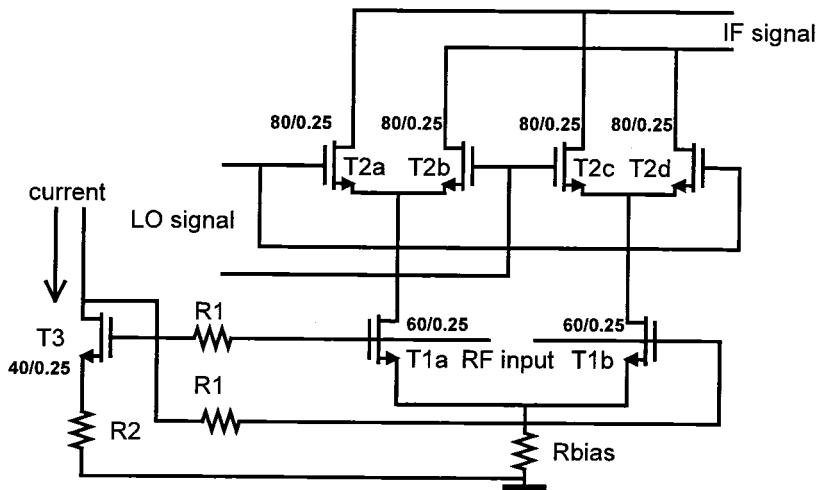
Figure 4.10.   A resistor degenerated active double balanced mixer. Transistor dimensions are in μm.
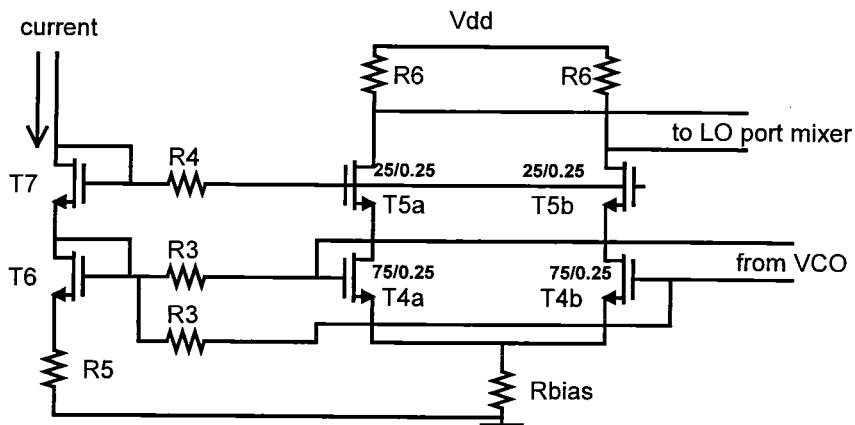
Figure 4.11.   Buffer stage to drive the LO port of the mixer. Transistor dimensions are in μm.

the following values: $R_{bias} = 25 \ \Omega$, $R_1 = 2 \ k\Omega$ and $R_2 = 1700 \ \Omega$. Additional emitter degeneration resistors can be used to increase the linearity of the mixer. A buffer stage is used to achieve isolation between the VCO and the LO port of the mixer. This buffer can then also help in setting the proper DC levels for steering the LO port of the mixer. The buffer stage is shown in Figure (4.11), where a resistor R_{bias} is again used to set the bias current. Cascode transistors T_{5a} and T_{5b} are added to improve VCO-LO isolation.

The design of the mixer, including the buffer, was realized in a 0.25 μm CMOS process. The RF frequency is set to 1.95 GHz, the LO frequency to 2.043 GHz, yielding an IF of 93 MHz. The power in the LO drive is set to -5 dBm. The measured conversion gain is 7.2 dB, the single-sideband noise figure is 8.6 dB, the IIP_3 equals 2.5 dBm and the complete circuit consumes 7.5 mA at 2.5 V supply voltage.

4.3.2 Passive CMOS Mixers

The design of passive mixers seems trivial, because the transistors only act as switches; the larger the transistor sizes, the smaller the on-resistance of the transistor and the better the linearity. However, if the mixer is incorrectly designed, this can lead to an excessive amount of noise; the design steps are more complex than you might imagine.
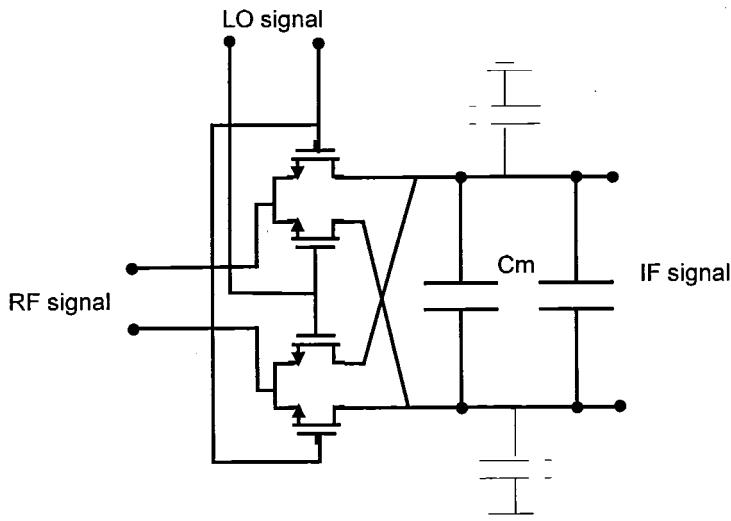
Figure 4.12. Mixer test chip configuration. The W/L ratio of all MOS devices is 50/0.35. Note that there are no ESD structures on the chip.

In order to determine the existence or otherwise of $1/f-$noise in an AC coupled passive mixer, a single mixer structure was laid out and processed in 0.35 μm CMOS technology. The passive mixer includes a metal sandwich capacitor C_m of 12 pF (laid out for symmetry) between the IF-output nodes on-chip. The sizes of the mixer transistors are W/L=50μm/0.35μm and the threshold voltage is 0.5 V. The schematic is shown in Figure 4.12 and the die microphotograph is shown in Figure 4.13.
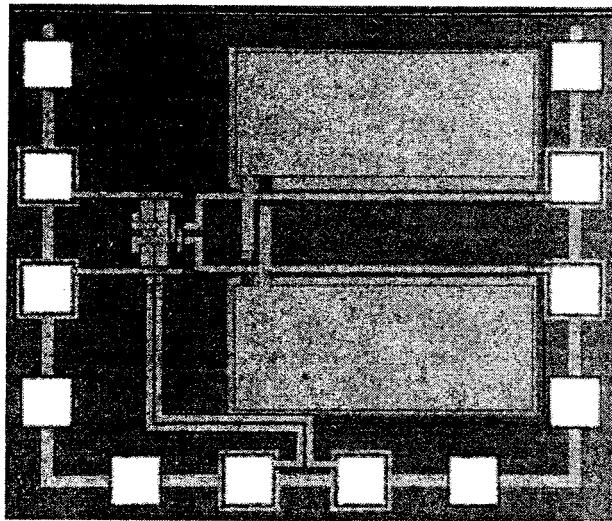
Figure 4.13.   Mixer test chip microphotograph. The RF input is on the left, the LO input is at the bottom, and the output is on the right.

Figure 4.14 shows the spectrum of the mixer in operation. The RF frequency is set to 2 GHz and the LO frequency is 200 kHz from the RF signal (i.e. 2.000, 2 MHz.). The input power of the LO is 0 dBm resulting in 1.2 V_{pp}. The IF output is at multiples of 200 kHz. Fundamental IF and the first three harmonics are visible in the spectrum.

We will use this mixer to show that there are three noise sources in the circuit that contribute to the overall noise, measured at the output of the passive mixer. These noise sources are due to the $1/f$-noise of the transistors of the mixer, due to the implicit existing switched-capacitor behavior, and finally due to the overall gain of the mixer and IF amplifier together.

$1/f$-Noise in Mixer Transistors

Many studies have been carried out on the $1/f$-noise behavior of a single MOS transistor. The most well-known $1/f$-noise model assumes a time-invariant current I flowing through the device and is written as

$$S_{v,f}(f) = c\frac{I^n}{f^m}$$ (4.19)

where c, n and m are constants. This model has a spectral density function which is not dependent on time. For the $1/f$-noise source, this aspect is certainly
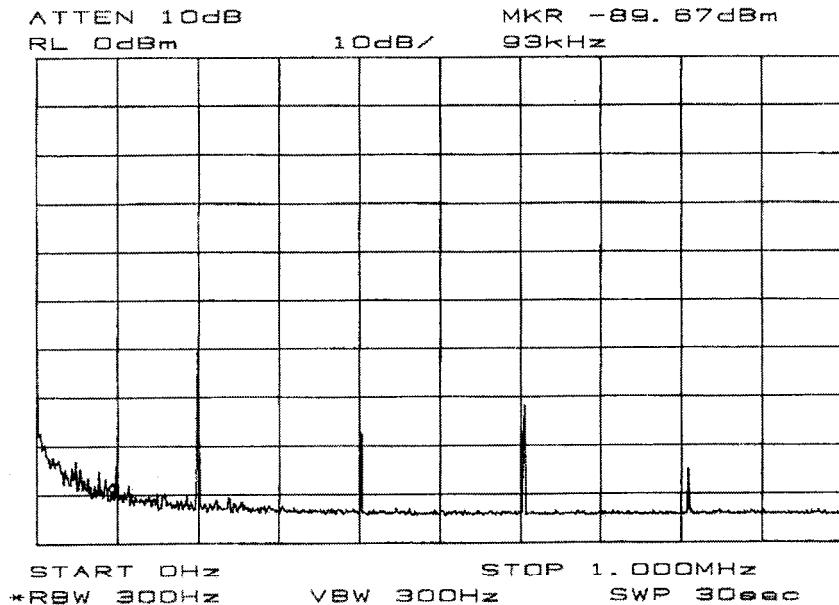
Figure 4.14. Measured output of the mixer.

questionable. Equation 4.19 implies that a purely AC-driven MOS transistor cannot have $1/f$-noise. Intuitively this seems strange, because there is a current flowing from drain to source all the time, and it can thus be argued that this current must contribute to $1/f$-noise. In a passive mixer, the MOS devices are only AC-driven, and the common wisdom is that no $1/f$-noise will be observed.

Suppose that the device in question has a time-varying bias condition with a certain fundamental frequency. It would be nice to obtain a noise model which can inhibit frequency-translated versions of its original noise transfer function at other, or all harmonics of this fundamental frequency. We therefore define the following noise model, which is based on a similar model proposed in [5]. The model is depicted in Figure 4.15. Assume a linear time-invariant (LTI) system with a transfer function $H(f)$. The spectral density of this LTI system is given by $S(f)$. The input to this LTI system is white noise $\varsigma(t)$ with a spectral density $S(f) = 1$. For $1/f$-noise modelling, the transfer characteristic can be

$$H(f) = \frac{k}{\sqrt{f}} \tag{4.20}$$

The output of the LTI is fed into a modulator, which modulates the noise in time with the bias parameter, i.e. the current through the device. The output
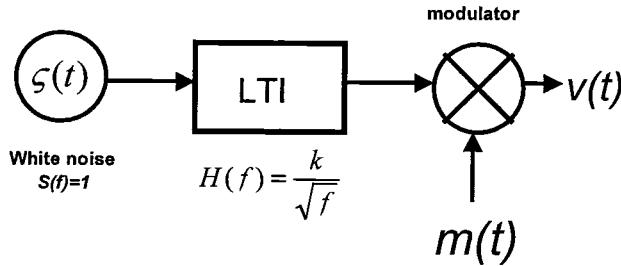
Figure 4.15. Noise model for time-varying bias conditions.

v(t) represents the noise at the output of the device. This means that in the time domain we have

$$v(t) = F^{-1}\left[|H(f)|^2\right] \cdot m(t) \tag{4.21}$$

and in the frequency domain we have

$$S_v(t, f) = |H(f)|^2 \otimes F\left[m(t + \frac{\tau}{2})m(t - \frac{\tau}{2})\right] \tag{4.22}$$

where $\otimes$ is the convolution operator.

Let us first try to clarify the model in terms of a physical process. Suppose the device for which the noise model must hold is highly non-linear in its current-voltage relationship between input and output ports. As the equivalent voltage/current noise levels are small, one can argue that the squared terms directly resulting from the small noise amplitudes are negligible. However, since the bias voltage/current is non-linear, this voltage/current has higher order harmonics and the noise component should be multiplied by each of the harmonics.Consider a MOS device as an example. The current can be taken simplistically as a square function of the gate voltage (in the case of saturation). However, the $1/f$-noise is multiplied by the fundamental and each harmonic of the drain current. These harmonics are caused by the square function. Therefore, in this case, $m(t)$ represents the drain current.

Assuming that the time averaged value of the time varying bias condition is taken, then $m(t) = \overline{m}$, and for $1/f$-noise (4.22) becomes

$$S_v(t, f) = |H(f)|^2 \cdot \overline{m}^2 \simeq \frac{I_{DC}}{f} \tag{4.23}$$

which is the analysis strategy used by many simulators. This assumption means that there is no $1/f$-noise when only AC currents are applied to the electrical

component. When this assumption does not hold, (4.22) must be solved including the convolution operator. Assuming $m(t)$ represents the time varying bias condition with fundamental frequency f_c, the Fourier expansion of $m(t)$ can be used to write

$$m(t) = \sum_{k=-\infty}^{\infty} b_k \exp(j2\pi k f_c t) \qquad (4.24)$$

which, after some mathematical manipulations leads to an expression for the nth order Fourier component as

$$S_v^n (t, f) = \sum_{k=-\infty}^{\infty} b_k b_{n-k} \left| H\left(f - \frac{2k-n}{2} f_c\right) \right|^2 \qquad (4.25)$$

The zeroth order component of the time varying spectral density for this cyclostationary process is defined as

$$S_v^0 (t, f) = \sum_{k=-\infty}^{\infty} b_k b_{-k} |H(f - k f_c)|^2 \qquad (4.26)$$

which basically means that the $1/f$-noise is spread around all harmonics of the fundamental frequency f_c. This model therefore seems suitable for noise modeling in non-linear electronic circuits.

The following measurement set-up was used to identify any $1/f$ device noise with only AC channel current. The mixer test circuit of Figure(4.12) was biased so that two of the transistors were OFF and two were ON, with their gate voltages set to 1.5V DC. Hence, only one of the two possible paths through the mixer was active. DC blocking was performed using two series capacitors of 2.7 μF each. An external low noise instrumentation amplifier unit was employed (EG&G 5113 on a battery power supply to suppress 50 Hz mains interference) to raise the signal level to measurable values. The spectrum analyzer was an HP 4195A, and the signal generator was a Marconi 2042. Passive LC low pass filters with a corner frequency f_c=20 kHz were also used to avoid the risk of undetected blocking and intermodulation in the measurement chain. The complete measurement set-up is shown in Figure 4.16. Figure 4.17 shows the measured spectrum when the gain of the LNA was set to 67.9 dB and the RF input signal was set to -10 dBm at 40 MHz. The spectrum of the measured output shows a 1/f dependency with a noise level of approximately -138 dBm at 2 kHz. If this noise is indeed related to the channel current of the MOS transistors, we should expect the measured level to be linearly related to the input signal magnitude. Increasing this input level from -10 dBm to 0 dBm and then to $+5$ dBm, we would expect the same level of increase in the noise at
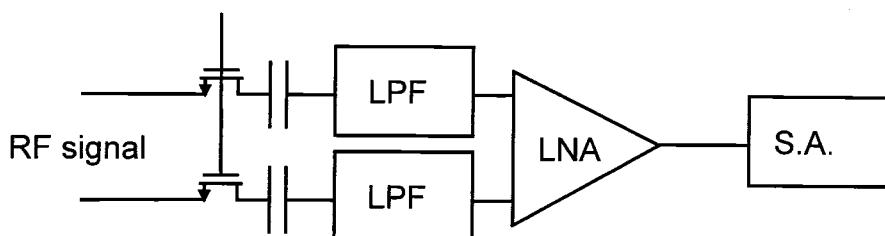
Figure 4.16.   Measurement set-up to investigate $1/f$-noise in a MOS device under time-varying bias conditions.
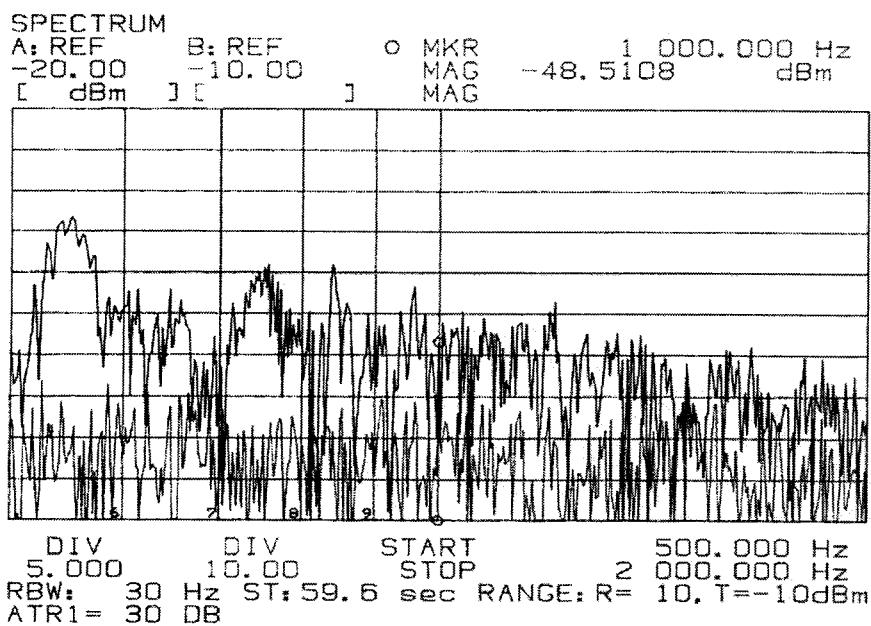
Figure 4.17.   Measured noise spectrum at the output of the LNA. The lower trace is the white noise spectrum of the equipment. The upper trace is the $1/f$-noise spectrum of the transistors. At 2kHz, the noise level is approximately -138dBm.

2 kHz. The measured spectra are shown in Figure 4.18; taking into account the instrumentation LNA gain of 47.9 dB, we can observe the expected increase in the noise levels. The measurements clearly show $1/f$-noise behavior for the transistors under a time-varying bias condition, without DC bias components. Hence, passive mixers exhibit $1/f$-noise, although this noise contribution is small compared to the two other noise sources in a passive mixer.
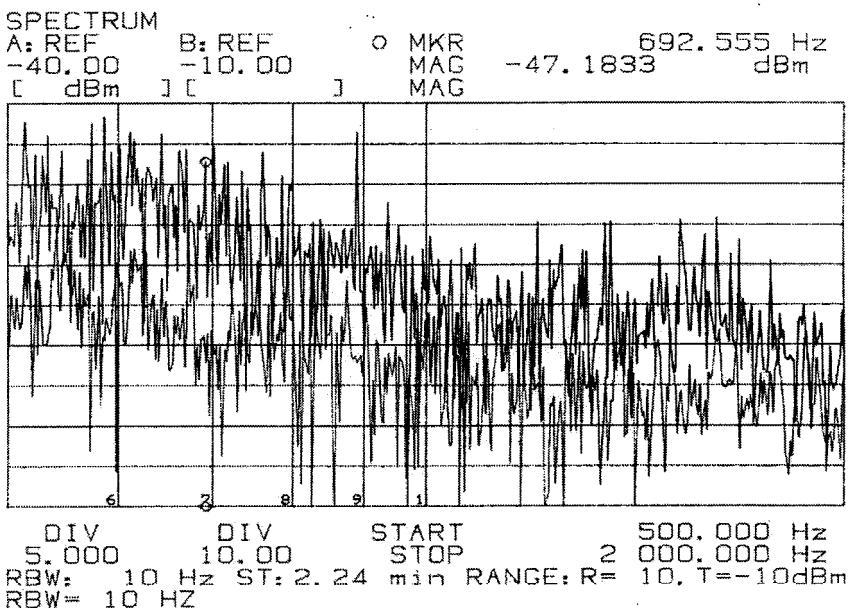
Figure 4.18. Measured noise spectrum at the output of the LNA. The lower trace is the $1/f$-noise spectrum of the transistors for 0dBm input signal, and at 2 kHz the noise level is approx. -128 dBm. The upper trace is the spectrum for 5 dBm input signal with a noise level of approx. -123 dBm at 2 kHz.

$1/f$-Noise due to IF Amplifier

The passive MOS mixer cannot simply be treated in isolation, but must be considered in the context of the associated receive path circuitry. The noise associated with the first low frequency amplifier in the receive path is of particular concern, since there is only a modest degree of gain in front of it, and its signal bandwidth may well also extend down to within the $1/f$-noise range of MOS devices. From a simple viewpoint, we might think that the effect of the amplifier's input noise should be calculated to be less than that of the thermal noise (and $1/f$-noise) of the LNA and mixer, when referred back to the antenna. The op-amp appears to be in a unity gain follower configuration, from the point of view of its own LF input noise, since there is no LF path through the mixer due to the AC blocking capacitors at the input. Hence, a simple scaling of the input g_m and total gate areas should be sufficient to handle the problem.

However, things are not that simple in reality, and we must consider the behavior of the subsystem more closely. To start with, the behavior of the
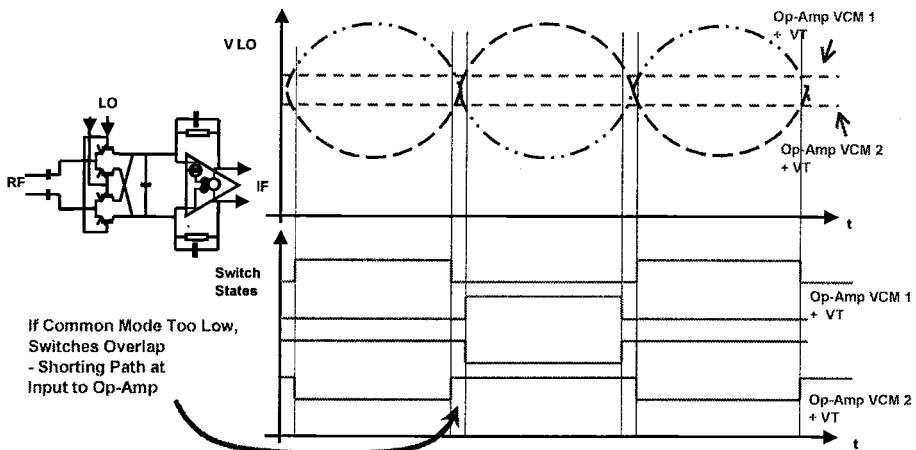
Figure 4.19. "Passive" mixer; Switch ON period overlap due to effect of op-amp common mode input level with respect to local oscillator common mode level.

mixer itself is affected by the common mode levels of both the local oscillator drive signal and the two "virtual earth" inputs of the op-amp.

Ideally, the mixer performs an instantaneous reversal of the input signal paths for the amplifier. At the frequencies of modern integrated wireless systems, the local oscillator waveform clearly cannot approach a square wave, and so there is a more gradual switch over, and the exact point of the crossover is not as certain. The switch on/off points of the MOS devices also depend on several other factors. Clearly, the LO waveform must exceed the turn-on voltages of the MOS transistors, and this, in itself, is dependent on the threshold voltage V_t value and the potential at their source nodes, i.e., the common mode input level of the op-amp. Taking these factors into account, it is now clear that the on/off periods of the switches also depend on the common mode level of the LO wave-form. The situation is set out graphically in Figure 4.19. If the LO common mode value is below the ON voltage of the MOS switches, then at no time will the reversing paths be on at the same time; however, if the common mode is higher than the ON voltage, a crossover path will momentarily occur. When this happens, there is a change in the feedback factor as seen by the input referred noise source of the op-amp, shown in Figure 4.20.

Here we see the equivalent circuit at the instant of change over. The transfer function from the input referred noise source to the output of the amplifier is now determined by the ratio between R_{fb} and R_{sw}, normally a small value. Clearly, the feedback resistors must be large compared with the series impedance of the mixer for there to be any gain for the signal path. During the crossover
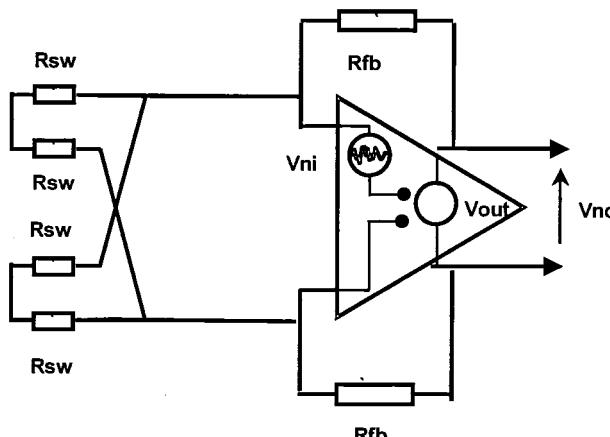
Figure 4.20. "Passive" mixer; op-amp input noise gain during switch crossover.

periods, we may therefore expect the gain seen by the noise to be significantly greater than unity. Values of the order of $20X$ to $50X$ are quite reasonable during this time, making the $1/f$-noise (and possibly also the thermal noise) very significant.

This behavior was verified using the test chip and a proper measurement set-up. In order to have a decent amount of equivalent input noise of the amplifier, the obsolete CA3140[2] devices were used. The amount of noise generated eased the measurement. The measured equivalent input noise was $30\,nV/\sqrt{Hz}$. A pseudo-differential band-limited transimpedance amplifier was constructed using the discrete amplifiers, where the common-mode voltage was set to 0.75 V. The mixer was connected to the amplifiers and a DC bias was applied to both of the LO input pins, to set all of the mixer transistors either to ON or OFF. The DC level of the LO signal is set to 1.0 V, reflecting the OFF state of the mixer ($1.0 < 0.75 + 0.5$). The gain of the LNA (see Figure 4.16) is 2500 times. Figure 4.21 shows the measured noise spectrum at the output. The calculated on-resistance of the mixer is 120 kΩ, and the gain of the noise is 1.33, which is indeed small, as expected.

Figure 4.22 shows the output with the DC bias at the LO ports increased to 1.6 V, i.e. the switches are now ON, and the significant noise increase can be clearly seen. Increasing the DC level indeed increases the noise levels at the output, reflecting an increase in gain and therefore a decrease in resistance of

---

[2]The CA3140 is an operational amplifier, made in an old technology, having high noise levels.
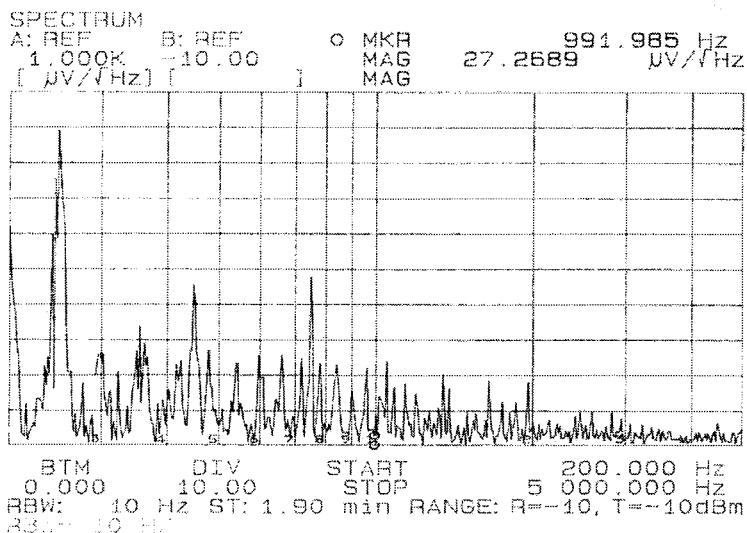
Figure 4.21. Test circuit noise with LO DC bias set to 1.0 V (MOS switches OFF) (LNA gain is x2500). Output noise level @ 1000Hz = 50 $\mu V/\sqrt{Hz}$.
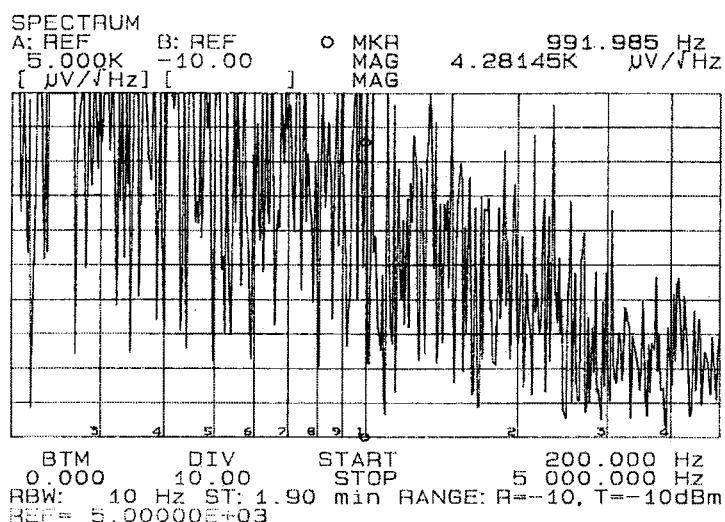
Figure 4.22. Test circuit noise with LO DC bias set to 1.6V (MOS switches ON) (LNA gain is x1000). Output noise level @ 1000 Hz = 3500 $\mu V/\sqrt{Hz}$.

SPECTRUM
A: REF   B: REF   o MKR   991.985 Hz
   5.000K   -10.00   MAG   124.487   µV/√Hz
[µV/√Hz] [   ]   MAG

         BTM   DIV   START   200.000 Hz
       0.000   10.00   STOP   5 000.000 Hz
     RBW: 10 Hz ST: 1.90   min RANGE: R=-10, T=-10dBm
     RBW=   10 Hz
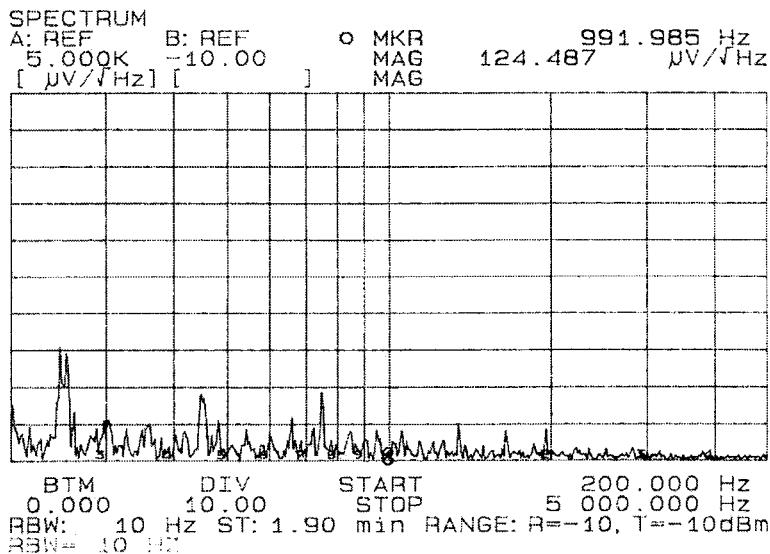
Figure 4.23.   The LO DC is set to 1.2 V, the LO amplitude equals 1 V$_{pp}$ which is equivalent to -2 dBm.

the transistors in the mixer. The resistance is calculated to be 172 Ω, yielding a gain of 233 times.

We can see that proper settings of the common mode voltage and the LO DC drive are mandatory to keep the noise levels low. This behavior can also be observed under dynamic conditions.

Verification of this noise amplification mechanism under dynamic conditions required a differential LO input variable amplitude, whose DC common mode could be adjusted with respect to the op-amp virtual earth levels, and hence with the turn ON points of the MOS switches. This was achieved using a signal generator (Marconi 2032), a balun (H-183 $-$ 4, 30 $-$ 3000 MHz) and a bias tee. The effective overlap of the LO could be set by adjusting the common mode and the input amplitude. For the measurements, the same common mode DC level was used at the transimpedance amplifier (0.75 V). The LO frequency was set to 500 MHz, and filtering to select the baseband region was achieved in the instrumentation LNA as well as with the external filter. The results are shown in Figure 4.23 and Figure 4.24.

Further verification can be obtained by noting that the peak signal in Figure 4.24 is 2 V; this same maximum gate drive can be achieved in a non-overlapping fashion by using a larger LO amplitude on a lower common mode DC level.
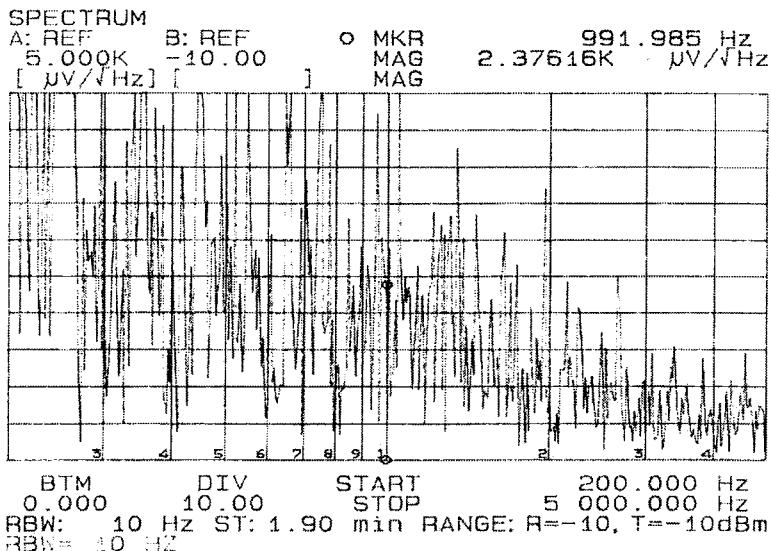
Figure 4.24.   The LO DC is set to 1.5 V, the LO amplitude equals 1 V_{pp} which is equivalent to -2 dBm.

Figure 4.25 shows the noise observed with the DC level set to 1.2V but the amplitude raised to 1.6 V_{pp}. The measured noise drops from 2300 $\mu V/\sqrt{Hz}$ to 870 $\mu V/\sqrt{Hz}$ at 992 Hz, roughly 2.6 times lower, confirming the effect of the overlap compared to simple maximum drive.

From a design point of view, the obtained results are significant. A mixer is normally designed for high linearity, and hence the mildly non-linear switch ON resistance is made low compared with the other impedances. However, when the LO drive levels are such that there is any cross-over within the mixer, low resistance switches exacerbate the noise gain problem. Clearly, care must be taken to ensure that LO signals are non-overlapping. Hence, the DC level of the LO signal should be chosen carefully with respect to the common mode level of the following amplifier input. However, the required drive voltage is also important; lower common mode DC levels at the LO port demand larger AC swings to achieve the same turn ON resistance, making additional demands on the VCO design.

$1/f$-Noise due to Switched-Capacitor Behavior

Consider the network when all the parasitic capacitances are included. There are the MOS switch source/drain-substrate junctions, the parasitic term to ground for the coupling capacitors, and some wiring terms. The drive from the LO can
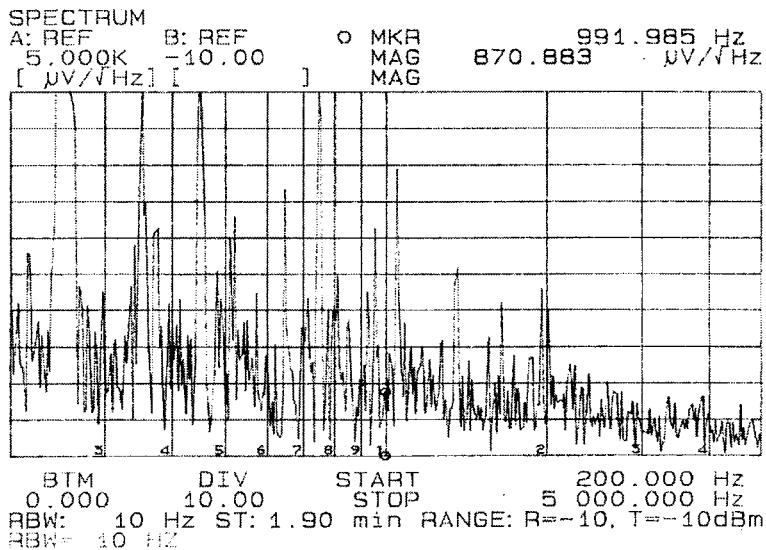
Figure 4.25. The LO DC is set to 1.2 V, the LO amplitude equals 1.6 V$_{pp}$ which is equivalent to +4 dBm.

be considered as the equivalent of a switched capacitor network, with the signal storage due to these elements, and the signal in question being the op-amp input referred noise - mostly 1/f type in this case. Figure 4.26 shows the principle.

From inspection, it is clear that there will be larger parasitic capacitances C_{par} present with larger MOS switches, leading to larger samples of noise being stored. As a consequence, there will be a lower equivalent resistance through the mixer as seen by the op-amp, and thus the amplification factor will be increased. In addition, as the LO frequency f_{LO} is increased there will also be an increase in the rate of transfer of the sampled noise, and hence a lower equivalent resistance, and similarly a higher noise gain, since the equivalent crossover resistance is:

$$R_{eq} = \frac{1}{C_{par} f_{LO}}$$

The experimental set-up was similar to that in the previous experiments. The op-amp common mode input level was set to 0.75 V, the LO common mode was set to 1.2 V (to ensure no overlap) and the LO amplitude was set to 1.79 V$_{pp}$ (+3 dBm at the generator). Capacitors of 6.8 pF and 15 pF were used. For each capacitor value, the LO frequency was varied from 30 MHz to 240 MHz. A few measurement results are shown in Figure 4.27 and Figure 4.28.
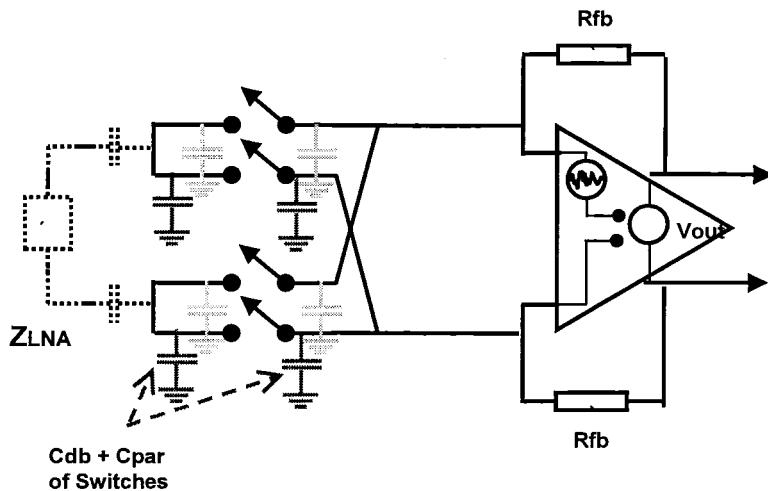
Figure 4.26. Equivalent switched capacitor network in passive mixer with non-overlapping LO signals.
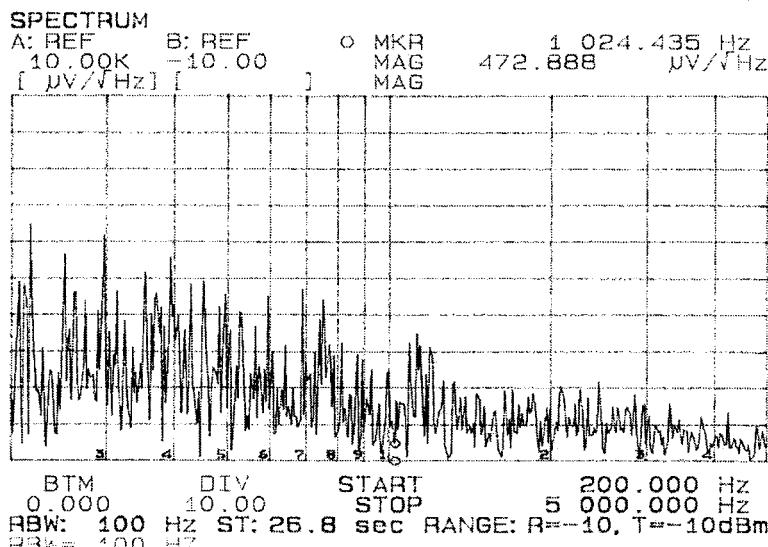
Figure 4.27. Op-Amp noise amplification due to SC effects. $C_{ex}t$ = 6.8 pF, LO Frequency = 30MHz, LO DC=1.2V, LO amplitude = 1.79 V_{pp} (+4 dBm), LNA gain = 2500.

```
SPECTRUM
A: REF      B: REF        O  MKR              976.148 Hz
   10.00K    -10.00          MAG    3.55863K   · μV/√Hz
 [ μV/√Hz] [              ]  MAG
```

```
  BTM           DIV        START              200.000  Hz
  0.000        10.00        STOP            5 000.000  Hz
RBW:   100  Hz  ST: 26.8  sec  RANGE: R=-10. T=-10dBm
RBW=   100  HZ
```
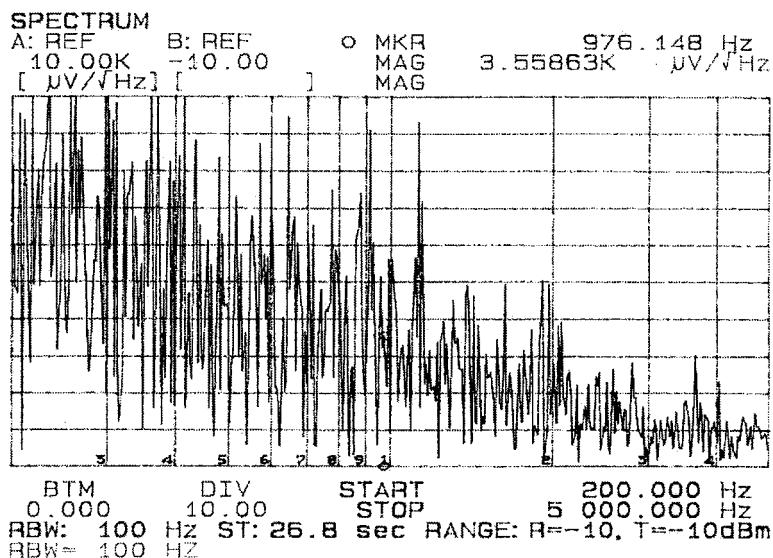
Figure 4.28. Op-Amp noise amplification due to SC effects. C_{ext} = 6.8 pF, LO Frequency = 120 MHz, LO DC=1.2 V, LO amplitude = 1.79 V$_{pp}$ (+4 dBm), LNA gain = 1000.

All the measured results are colated and in Table 4.1.The results in Table 4.1 show that, within the tolerances of the measurement system, the expected trends in the increase of noise with both capacitance and LO frequency are confirmed. Once again, we can see that if large devices are used in the mixer to achieve good linearity, then there is a noise penalty to be paid.

Table 4.1. Measured noise level in $\mu V/\sqrt{Hz}$ for several settings of the frequency and capacitances.

f (MHz)	C = 6.8 pF	C = 15 pF
30	0.4	0.88
60	0.88	1.6
120	2	4
240	3.5	

4.3.3 Concluding Remarks

These studies have reinforced the view that in an integrated low or zero IF receiver chain, the mixer, LO drive and first IF amplifier must be considered together as a complete subsystem.

The use of passive mixers as a means of achieving good $1/f$-noise performance has been critically examined, as have several design issues. In addition

to the obvious input referred noise of the IF amplifier, we have shown that the ON-resistance of the mixer, the common mode voltage of the amplifier, the mean DC level of the LO signal and the LO frequency are all important in determining the noise performance.

The following observations can be made:

1. A passive MOS mixer has $1/f$-noise, even when there is no DC passing through the transistors

2. Increasing the size of the MOS switches increases the overlap capacitances, and hence RF currents due to the LO circulating in the mixer. Therefore the noise will increase.

3. The input referred noise of the first IF op-amp can be strongly amplified if the LO signal causes the opposing mixer paths to be ON at the same time. This is a direct function of the common-mode DC level in the mixer (as set by the IF op-amp) and the mean DC of the LO drive. It is not a strong function of LO frequency.

4. Using larger MOS switches in the mixer increases the noise amplification if this crossover takes place.

5. Even if the LO drive does not allow the switches to be ON at the same time, the IF op-amp noise can also be amplified by a crossover path arising from switched capacitor action in the mixer. The use of larger MOS switches increases the parasitic capacitances present, lowering the effective resistance of the crossover path, and hence increasing the noise gain. This effect increases directly with increasing LO frequency.

6. These noise gain effects are also modified by the output impedance of the LNA driving the mixer.

REFERENCES

[1] R.G. Meyer, "Intermodulation in High-Frequency Bipolar Transistor Integrated-Circuit Mixers," *IEEE Journal of Solid-State Circuits*, vol. 21, no. 8, pp. 534–537, Aug. 1986.

[2] C.D. Hull and R.G. Meyer, "A Systematic Approach to the Analysis of Noise in Mixers," *IEEE Transactions on Circuits and Systems I*, vol. 40, no. 12, pp. 909–919, Dec. 1993.

[3] M. Terrovitis and R.G. Meyer, "Noise in Current-Commutating CMOS Mixers," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 6, pp. 772–783, June 1999.

[4] H. Darabi and A.A. Abidi, "Noise in RF-CMOS Mixers: A simple Physical Model," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 1, pp. 15–25, Jan. 2000.

[5] A. Demir and A. Sangiovanni-Vincentelli, *Analysis and Simulation of Noise in Non-linear Electronic Circuits and Systems*, Kluwer Academic Publishers, New York, 1998.

Chapter 5

RF power amplifiers

The last stage in a transmitter path is the power amplifier (PA). No more signal processing takes place after the PA, thus putting stringent demands on the PA. Furthermore, efficient RF PAs are highly desirable in battery-operated systems, since PAs typically dominate the power consumption of the system. This chapter focuses on the design of bipolar and MOS PAs and discusses principles to improve the linearity of a PA.

5.1 SPECIFICATION

In an RF front end, the last step between up-conversion of the baseband signal to the radio frequency and the antenna is amplification of the RF signal. One or more levels of *delivered* power to the antenna have been specified for each telecommunication standard. For instance, for the Bluetooth wireless standard, transmit class 2 refers to 0 dBm output power, while for UMTS, the specification is 24 dBm. The latter specification means that the amplifier must deliver a serious amount of power to the load (i.e. the antenna), hence the commonly used name power amplifier.

5.1.1 Efficiency

When a specification reads that an output power P_{ant} must be delivered to the load, the amplifier itself must generate much more power. The front end will normally be followed by a diplexer or duplexer with some losses ($\approx$ 3 dB) that must be overcome by the power amplifier. Furthermore, the amplifier has a certain *efficiency* η; it costs power from the supply line to produce power at the output node. Define the DC input power as P_{DC} and the RF output power as
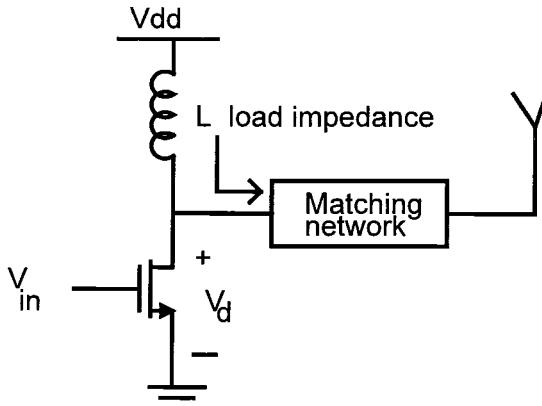
Figure 5.1.   Simplified output stage of an RF power amplifier

P_o, then the output efficiency is defined as

$$\eta = \frac{P_o}{P_{DC}} \tag{5.1}$$

A more realistic measure of efficiency is the *power added efficiency* (PAE) that takes into account the amount of power that is delivered to the input, P_{in},

$$PAE = \frac{P_o - P_{in}}{P_{DC}} = \eta \left(1 - \frac{1}{G_p}\right) \tag{5.2}$$

where G_p is the power gain. Nowadays, amplifier efficiencies range from 35% to 55%. Consequently, for a specification of 0 dBm delivered power, the amplifier should be able to dissipate 5 dBm. When considering battery-operated handheld sets, the power efficiency of the amplifier is definitely a main concern.

5.1.2   Generic Amplifier Classes

Power amplifiers can be divided into classes based on their achieved efficiency [1], [2], [3]. Consider the simplified NMOS power amplifier of Figure 5.1. The behavior of the amplifier is determined by the input signal, the load impedance and conduction angle. The six most well known classes will be discussed briefly with reference to Figure 5.2.

- **Class A**. The transistor of the amplifier is biased and driven so that it is always in active mode. In this way, the distortion will be small at the cost of large currents and voltages. This latter aspect results in a high power consumption, reducing the efficiency of the amplifier. This is even more

true because the amplifier consumes power when there is no output signal. The maximum theoretically obtainable efficiency is 35% for resistive load and 50% for inductive load.

- **Class B.** To increase the efficiency, the transistor can be made active only half of the time, at the cost of an increased distortion. To still have low distortion levels, the bias currents are chosen to be small and the transistor is normally in its saturation mode instead of in the linear mode [4], [5].

- **Class C.** A further increase in efficiency can be obtained by leaving the transistor unbiased and driving it far into linear mode. Class C is usually considered when the conduction angle of the transistor is smaller than 180 degrees. The transistor is still operating as a current source. The efficiency is depends on the conduction angle and as the angle reduces the efficiency increases. However, a reduction in conduction angle means that the current drive capability is reduced, resulting in a lower output power. Theoretically, a conduction angle of 180 degrees, i.e. Class B, gives 78.5% efficiency. A conduction angle of 90 degrees gives an efficiency of 90%. The output signal does not follow the input signal, the amplifier behaves *non-linearly* and the distortion levels are high. This type of amplifier can be used in the case of frequency modulated signals. To obtain a reasonable output sine wave, the single transistor amplifier is loaded with an RLC tank, where the R represents the load, i.e. the antenna input impedance.

- **Class D.** The difference between Class C and D lies in the fact that Class D uses at least two transistors, but neither is forced to simultaneously support both voltage and current. In theory, one can reach 100% efficiency, but in practical applications the efficiency is comparable to Class C. The disadvantage of class D compared to Class C is in the synchronization of the two (or more) switches.

- **Class E.** Basically, a resonance network is used to allow switching when the voltage is low [6]. More precisely, the switching device, i.e. the power transistor, becomes active when the slope of the voltage and current are both almost zero or almost zero. Consequently, even with mis-timing, the loss is low and therefore high efficiency rates can be achieved. The resonance network is placed between the output of the transistor and the load, and the resonance frequency is at the fundamental RF frequency. The drawback of this approach is the sometimes extreme output peak voltage. This class of operation is most often used in RF mobile transmitters [7], [8].

- **Class F.** If we combine the single switch of Class C with the square wave voltage approach of Class D, we obtain Class F. A resonance circuit at
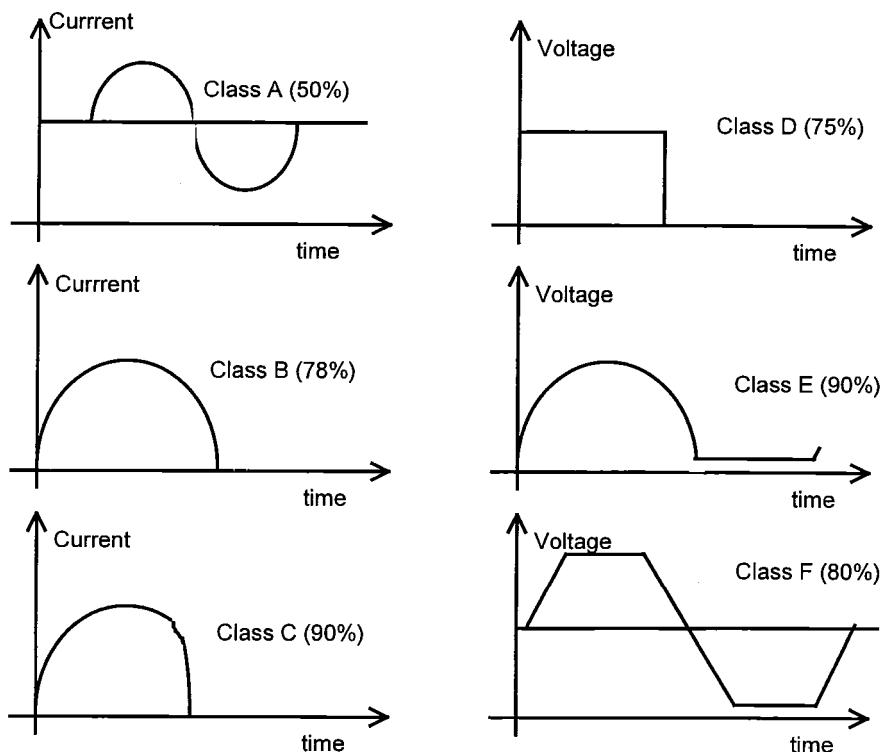
Figure 5.2.   Classes of power amplifiers with characteristic waveforms.

the third harmonic of the RF frequency is placed at the output of the single transistor to flatten out the voltage and "shape" it like a square wave voltage. Class F can also be considered as Class B with resonances on both the fundamental and the third harmonic, shaping the voltage waveform to a square waveform.

The classes discussed can also be related to the conduction angle and input signal overdrive. This is indicated in Figure 5.3. For small input signals, the RF PA of Figure 5.1 can operate in class A, AB, B or C, depending on the conduction angle. The conduction angle is determined primarily by the DC gate bias. The efficiency can be improved by reducing the conduction angle and moving in the direction of class C at the expense of lower output power. An alternative is to increase the gate overdrive until the PA operates as a switch, while keeping the same conduction angle.
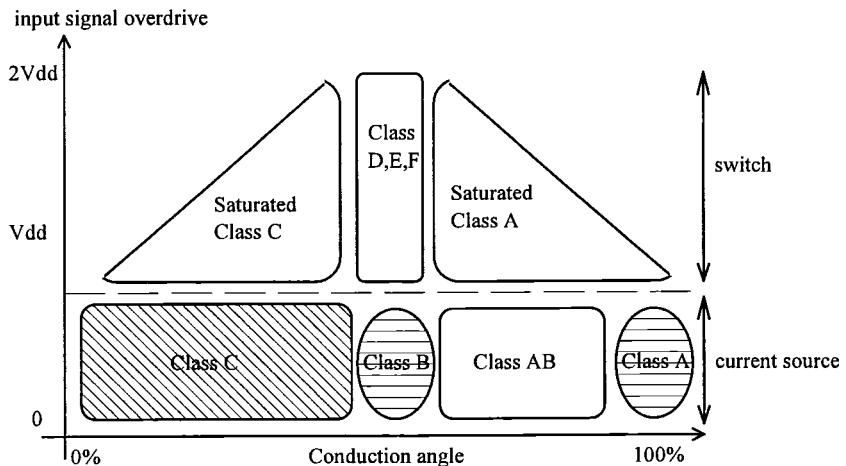
Figure 5.3.   Definition of PAs based on conduction angle and signal overdrive.

5.1.3 Heating

Frequent battery operation is one obvious reason behind the need to increase power efficiency. The higher the efficiency, the longer the standby time and talk time. One other important reason to search for optimal efficiency is the heating problem . The maximum output power is preferably not limited by the heat it can dissipate. The ratio between the maximum output power and the dissipated power is $\eta/1 - \eta$. Suppose a power transistor is able to dissipate 1 W safely, without any heating problems. When the device is used in a Class A operation, the maximum output power is 0.54 W. However, if the device operates in class C, the maximum output power becomes 3 W, but at the same heating. A much larger output power with the same heating constraint is therefore possible in Class C compared to Class A.

The thermal resistance is defined as

$$\theta_{jc} = \frac{T_j - T_c}{P_{diss}} \tag{5.3}$$

where T_j is the maximum peak junction temperature and T_c is the case temperature, normally $25^{\circ}C$. Parameter T_j is determined by the reliability of the transistor in the process. From T_j, T_c and θ_{jc} the maximum P_{diss} is calculated. The parameter θ_{jc} is less in a ceramic package than in a plastic package. The die in a ceramic package can therefore have a higher P_{diss} than in a plastic package, in both cases T_j being the same. For plastic packages the maximum temperature is 150° to $175^{\circ}C$ and $200^{\circ}C$ for ceramic, assuming that T_j is higher. It is

important to notice that the thermal resistance of a power amplifier depends on the layout of the chip (number of active areas), the thickness of the silicon chip ($100 - 200 \ \mu$m), the die attachment process, and the thickness of the package.

5.1.4   Linearity

Some telecommunication standards modulate the data in the frequency domain, like the GFSK modulation scheme used in Bluetooth, for instance. For the amplifier, the output amplitude does not have to follow the input amplitude; non-linear amplification is allowed. However, other standards only modulate the data or also modulate it in the amplitude domain, (see UMTS with the $3/8\pi$-QAM modulation scheme, for instance). It then becomes important for the amplifier to be linear; the amplitude information may not be distorted. We have already touched on the fact that linearity becomes an issue in the case of amplitude modulated signals. The trade-off between efficiency and linearity for the simplified amplifier of Figure 5.1 is determined by the gate voltage V_{in}. The output drain voltage V_d is a linear relation of V_{in} in the case of class A operation. The power dissipation is found to be $P = V_d \cdot I_d$, where I_d is the current through the device. In this case we have a highly linear, low efficiency power amplifier. Increasing the gate overdrive will reduce the power dissipation. The transistor then acts as an on-off switch and, assuming there is no overlap between the drain current and the voltage, the power dissipation becomes zero, and is therefore 100% efficient. However, the output voltage and current are distorted square waves and the amplifier is therefore no longer linear. A switched mode amplifier will be highly efficient, but also highly non-linear.

A non-linear amplifier causes amplitude modulation to amplitude modulation (AM-AM) behavior for an amplitude modulated input signal. Assume we can model the amplifier's gain as

$$A = a_1 + a_2 V_{in}(t) + a_3 V_{in}^2(t) \tag{5.4}$$

The amplitude of the input signal will also be amplified with the higher order coefficients, resulting in

$$V_d = A \cdot V_{in}(t) = a_1 V_{in}(t) + a_2 V_{in}^2(t) + a_3 V_{in}^3(t) \tag{5.5}$$

and thus AM-AM behavior occurs. If the coefficients are complex, amplitude modulation to phase modulation (AM-PM) behavior will also occur. These two behaviors obviously should be minimized to allow proper detection of the real amplitude information.

A specification which is closely related to linearity is the *adjacent channel power ratio* (ACPR). This specification defines the amount of power transmitted

at a certain offset frequency $P(\Delta f)$, normally the adjacent channel, compared to the power transmitted in the channel of interest, P

$$ACPR(\Delta f) = \frac{P}{P(\Delta f)} \tag{5.6}$$

For the EDGE cellular standard, at an offset of 30 kHz, the maximum allowed transmitted power is -30 dB, while the transmitted power is set to 0 dB, yielding an ACPR of 30 dB. Obviously, the more non-linear the amplifier, the higher the power levels in adjacent channels will be due to the cross modulation.

5.1.5   Ruggedness

Ruggedness is the ability to withstand electrical overstress without failure or degradation. The ruggedness is normally tested under some prescribed conditions: output overvoltage, input overdrive and mismatch load conditions. For bipolar RF power amplifiers, the two main techniques known to improve the ruggedness of the circuit, are collector ballasting and emitter ballasting. The former technique uses a thicker collector epitaxial region to support collector base depletion. This results in series resistance under each active area before these active areas are connected in parallel by the low ohmic substrate. There is consequently more voltage drop over, and better current distribution through, the active area. The price to pay is a decrease in efficiency. In emitter ballasting, several lumped resistors are added on each emitter site before they are placed in parallel. The effect is a lower gain but higher collector-emitter saturation voltage.

To obtain reasonable efficiency figures and high linearity, we can apply linearization techniques. Conceptually we can use a highly non-linear amplifier to amplify the phase information. The amplitude information is added by modulating the power supply of the amplifier or by adding in phase shifted signals. One other approach is to pre-distort the input signal to correct for the non-linear behavior of the amplifier. A few possible implementations of these approaches will be discussed in the following sections. We will start with the discussion on silicon integrated RF power amplifiers.

5.2   BIPOLAR PA DESIGN

Monolithic RF power amplifiers have traditionally been the territory of GaAs technologies. Recently, efforts have been made to design power amplifiers for the lower mobile-phone frequency band using Si technologies. It is the GSM bands at 900 and 1800 MHz that are the first targets here [9], [10]. Although some of the designs are still hybrid solutions, the demands are for monolithic

implementations of the power amplifiers in standard bipolar processes. The advantages are clear: low cost and ease of integration with other mainstream Si-based circuits. In this section we discuss a monolithic silicon-based power amplifier [11]. The application field is the USA PCS-CDMA standard. The modulation scheme of the CDMA system requires the PA used in the handset to be highly linear. This makes it very challenging to design a highly efficient monolithic PCS-CDMA PA using Si technologies, due to their inherently high substrate loss and parasitics.

A simplified schematic diagram of a standard biased PA is shown in Figure 5.4. The PA consists of two common-emitter stages ($Q0$ and $Q1$). The output-matching network consists of capacitors $C3$ to $C5$, inductors $L2$, and two transmission lines (CPWs). The input matching network consists of capacitors $C6$ and $C7$ and a transmission line (CPW). In Figure 5.4, $Lgnd1$, $Lgnd2$ and $Lgnd3$ are ground inductors from bonding wires. $L1$ and $L2$ are off-chip inductors. The interstage-matching network consists of an on-chip capacitor $C1$ and the off-chip inductor $L1$ realized using a bonding wire and a PCB trace. $C2$ is an on-chip bypass capacitor.

This biasing scheme for $Q1$ provides a near-constant low-frequency (1 MHz) small-signal impedance presented at the base of $Q1$, and a nearly linear control of the collector's quiescent current of $Q1$ as shown in Figure 5.5 where $Rb2 = 15\ \Omega$. When $Q0$ and $Q1$ are biased into class AB operation, however, its large-signal RF impedance presented at the base of $Q1$ at 1.9 GHz is capacitive, and is much larger than its counterpart as shown in Figure 5.6. Also shown in Figure 5.6 is the PAs output power versus RF input signal levels. We can see that the two-stage PA starts to saturate at the output power close to 600 mW. This is due to the fact that as the output power increases, the average voltage drop across the bias impedance increases. This, in turn, causes a reduction in the base-emitter voltage of $Q1$ and therefore pushes it into saturation [12].

Figure 5.7 shows a simplified schematic of a PA using an impedance-controllable biasing scheme to bias $Q1$. It is identical to the PA shown in Figure 5.4 except for the biasing scheme for $Q1$. The biasing scheme comprises two current-mirror subcircuits: one consisting of transistors $Q2$, $Q4$ and $Q7$, and the other consisting of transistors $Q5$ and $Q6$. If $Iclass$ and $Ibias$ are chosen correctly, this biasing scheme is capable of providing independent control of bias impedance and class of operation of $Q1$. Whereas $Ibias$ controls the output impedance of the bias circuit, $Iclass$ controls the quiescent current of the output stage. This proposed biasing scheme allows the output stage to be adjusted for optimum efficiency and linearity.
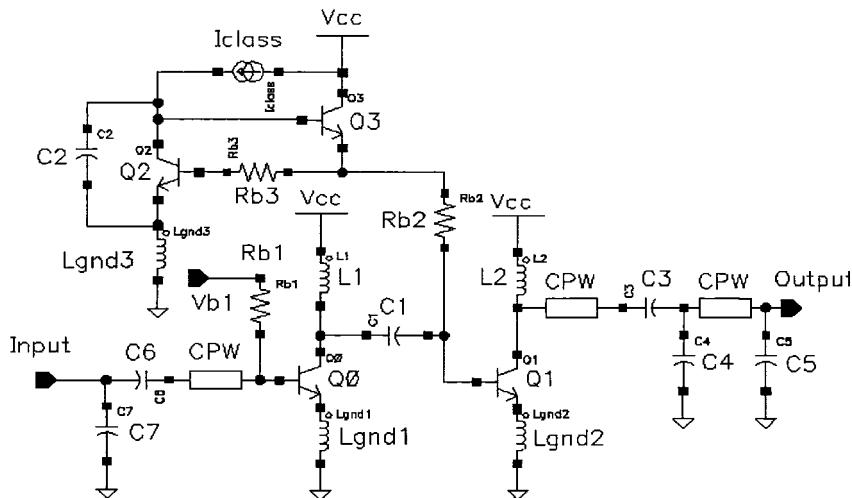
Figure 5.4.   Simplified schematic of the power amplifier.
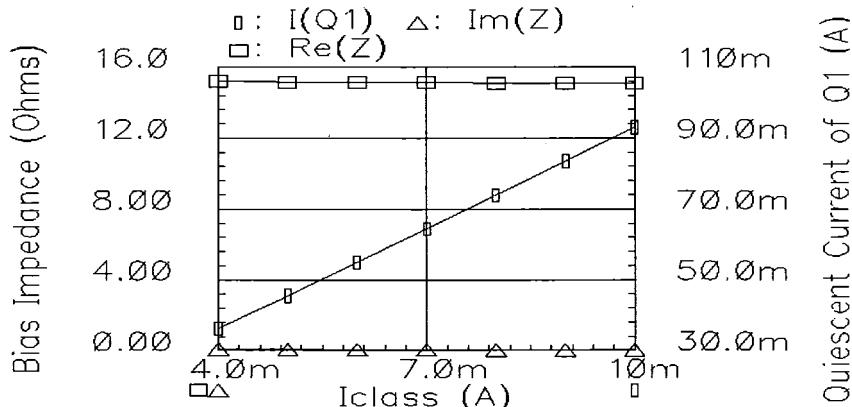
Figure 5.5.   Quiescent current of $Q1$ and small-signal impedance (1 MHz) of the conventional biasing scheme as functions of $Iclass$ ($Rb2 = 15\ \Omega$).

Its mechanism of controlling the quiescent current is explained below. Neglecting base currents, we have

$$Vbe(Q1) + Vbe(Q3) + V(Rb2) = Vbe(Q2) + Vbe(Q4) + V(Rb3) \quad (5.7)$$

and since $Q2$ and $Q3$ have the same current, we then have

$$Vbe(Q3) = Vbe(Q2) \quad (5.8)$$
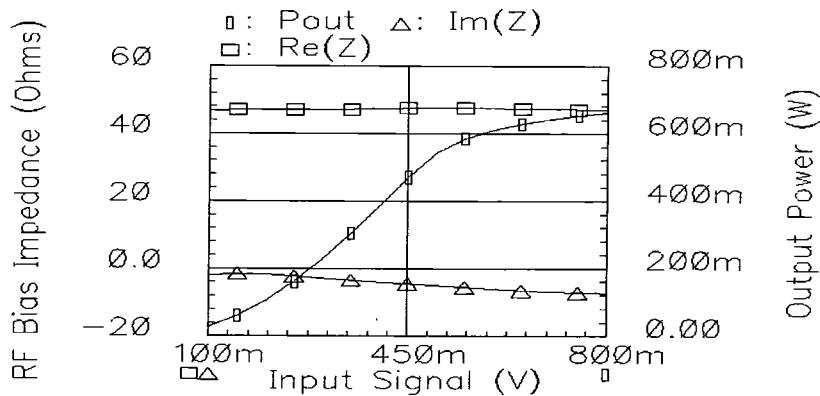
Figure 5.6.   RF impedance of the conventional biasing scheme and output power of the PA as functions of the input signal level ($Rb2 = 15 \, \Omega$, $Iclass = 8$ mA).
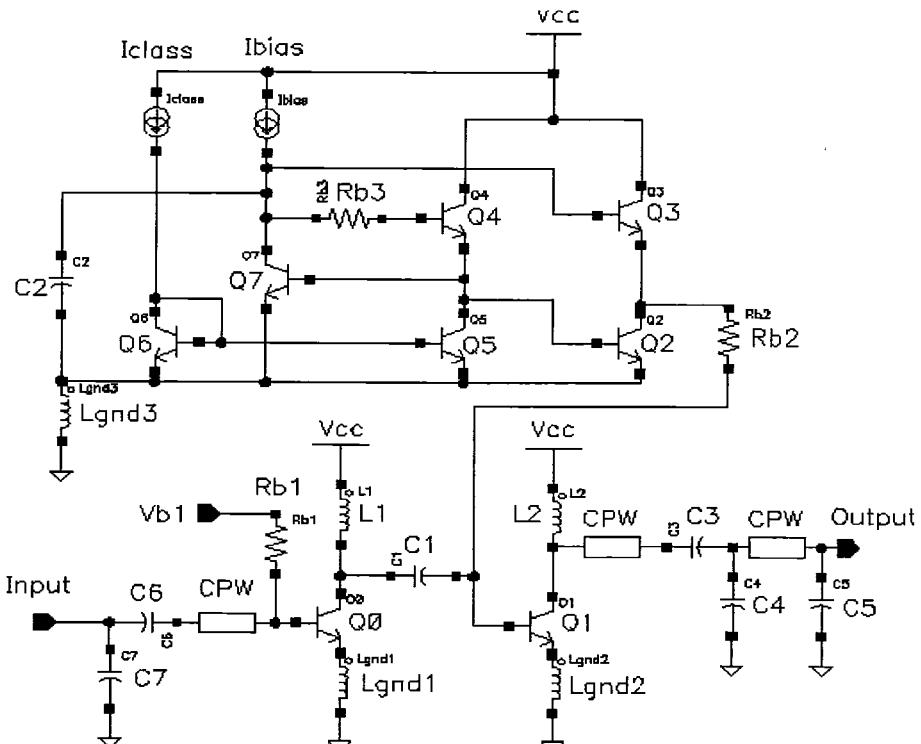
Figure 5.7.   Schematic of the PA with an impedance-controllable biasing scheme.

By choosing $Rb2$ and $Rb3$ properly, this leads to

$$Vbe(Q1) = Vbe(Q4) \tag{5.9}$$

when

$$V(Rb2) = V(Rb3) \tag{5.10}$$

The current flowing in $Q6$ must flow in $Q5$ and $Q4$ because $Q5$ and $Q6$ form a current mirror. Since $Iclass$ controls the current flowing in $Q6$, it therefore dictates the quiescent current in $Q4$ that in turn controls the quiescent current in the output transistor $Q1$. The mechanism of $Ibias$'s controlling the output impedance of the biasing circuit can be explained in a similar fashion. By properly scaling the emitter area ratios between transistor pairs, we can readily control the quiescent current of the output stage and the output impedance of the bias circuit. This helps us to optimize the efficiency of the output stage while maintaining the required linearity.

Figures 5.8 and 5.9 show the collector's quiescent current of $Q1$ and low-frequency (1 MHz) small-signal impedance of the impedance-controllable biasing scheme presented at the base of the output stage as functions of $Iclass$ and $Ibias$. For simplicity, $Rb3$ is set to zero ($Rb3 = 0$) in this analysis. We can see that the control current Iclass in both biasing schemes has a similar effect on the quiescent current of $Q1$ and the bias impedance: there is near-constant bias impedance and nearly linear control of the quiescent current. Furthermore, the control current $Ibias$ provides an additional means to adjust the bias impedance, and only slightly changes the quiescent current.

It is of more interest to see its large-signal impedance presented at the base of $Q1$ at 1.9 GHz. Figure 5.10 shows the RF impedance of the impedance-controllable biasing scheme and the output power, as a function of the input signal level. For a fair comparison, the sizes of $Q0$ and $Q1$ and the bias conditions have been set identical to those for the PA discussed above. There are several advantages of using the impedance -controllable biasing scheme. Firstly, it presents low impedance at the base of $Q1$. Secondly, its large-signal impedance is inductive and not far off from its low-frequency small-signal values. This helps interstage matching by cancelling part of the capacitive base impedance of $Q1$. Thirdly, it helps increase output power due to its low RF impedance. As seen in Figure 5.10, the PA only starts to saturate at the output power close to 1 W. This is a significant improvement over that of the PA with the conventional biasing scheme.

The PA with the impedance-controllable biasing scheme was implemented in a Philips BiCMOS process (QUBiC3) featuring 30 GHz f_t NPNs, 0.5 μm CMOS and high value poly resistors. The emitter area of $Q1$ is 128 times
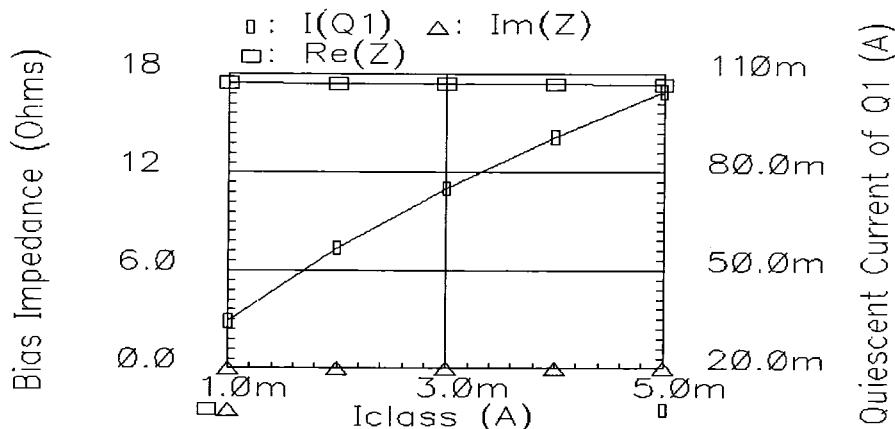
Figure 5.8. Quiescent current of $Q1$ and small-signal impedance (1 MHz) of the impedance-controllable biasing scheme as functions of $Iclass$ ($Rb2 = 15\ \Omega$, $Ibias = 3$ mA).
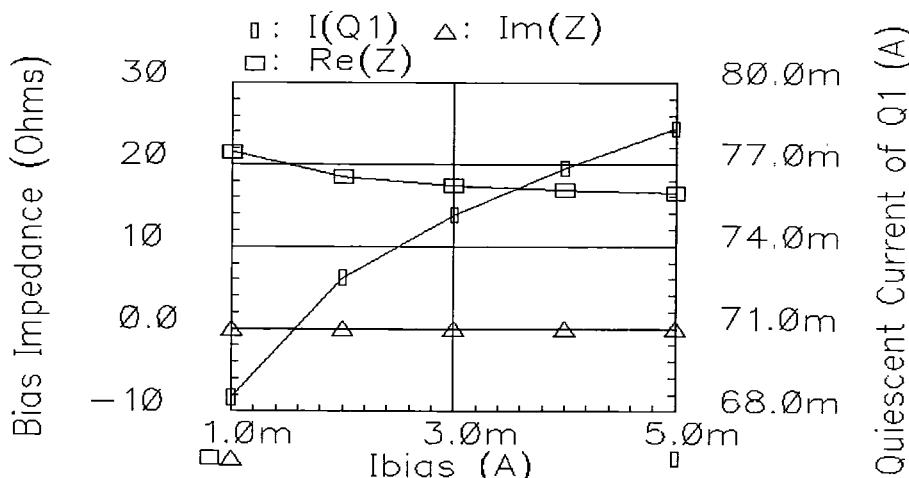
Figure 5.9. Quiescent current of $Q1$ and small-signal impedance (1 MHz) of the impedance-controllable biasing scheme as functions of Ibias ($Rb2 = 15\ \Omega$ and $Iclass = 3$ mA).

$0.7\mu m \times 20.2\mu m$. A photomicrograph of the die with an area of about 0.9 mm^2 is shown in Figure 5.11. The PA was first tested at 1.9 GHz using a three-stub tuner in the output to benchmark its performance. Both the input and output stages were biased in a class AB operation. The PAs ACPR and PAE versus the output power are shown in Figure 5.12. A spectrum analyzer from Rohde & Schwarz was used in the ACPR measurements. With the output matched using
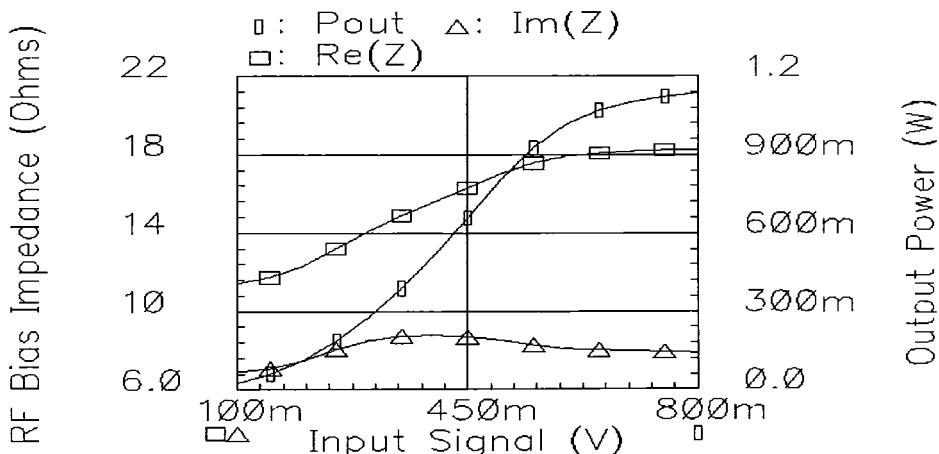
Figure 5.10. RF impedance of the impedance-controllable biasing scheme and output power of the PA as functions of the input signal level ($Rb2 = 15\ \Omega$, $Iclass = Ibias = 3$ mA).
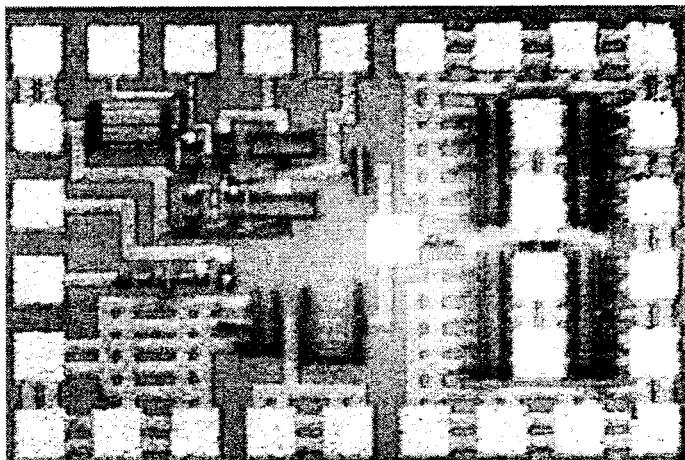
Figure 5.11. Photomicrograph of the die.

the tuner, the PA delivers 28.1 dBm output power with 24.5 dB gain, 31.5% PAE and −45 dBc ACPR. It is worth noting that in this PA, there is a valley with quite a steep slope in the ACPR curve. This indicates that we can improve ACPR margins with a slight back-off from the highest output power level.

The tuner was then replaced by surface-mount components on the FR-4 test board. The PA was again tested at 1.9 GHz with a slightly lower base

PCS CDMA PA with a Tuner
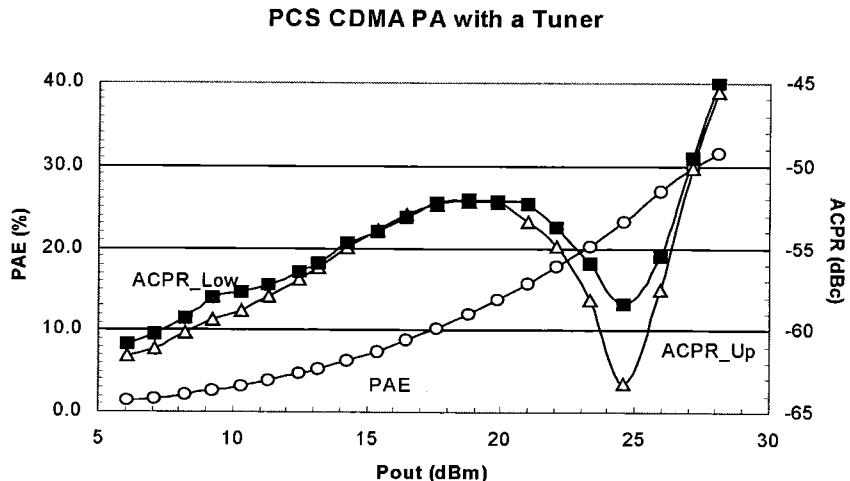
Figure 5.12.   Measured ACPR and PAE of the PA with the tuner.

bias voltage for the input stage to enhance its PAE. *Iclass* and *Ibias* were readjusted to compensate drops in PAE and gain due to the insertion loss from the surface-mount components. The PAs ACPR and PAE versus the output power are shown in Figure 5.13. With the input and output matching networks using the low-cost surface-mount components on the FR-4 test board, the PA delivers 28.2 dBm output power with 21.5 dB gain, 30% PAE and −45 dBc ACPR. The same steep slope can be seen in the ACPR curve near the peak output power level. At lower output power levels, there is a peak almost reaching the ACPR limit (−45 dBc) in the ACPR curve due to the lower bias in the input stage.

Figure 5.14 shows an actual ACPR measurement of the PA matched with the surface-mount components at its highest linear output power level. A 10 dB attenuator and a coaxial cable were used in the output. Their combined measured attenuation is 10.8 dB. This attenuation plus the reading of 17.45 dBm channel power adds up to 28.25 dBm output power. We can see that whereas the ACPR reaches its specified limit of −45 dBc, the alternate-channel-power ratio has large margins over its specification of −56 dBc. The effectiveness of the impedance-controllable biasing scheme has been verified in these measurements.

In this section, we have shown that it is possible to design a linear power amplifier in standard Si-technology. A monolithic Si PCS-CDMA power amplifier was designed which was capable of delivering 28.2 dBm output power

PCS CDMA PA with SMD Matching
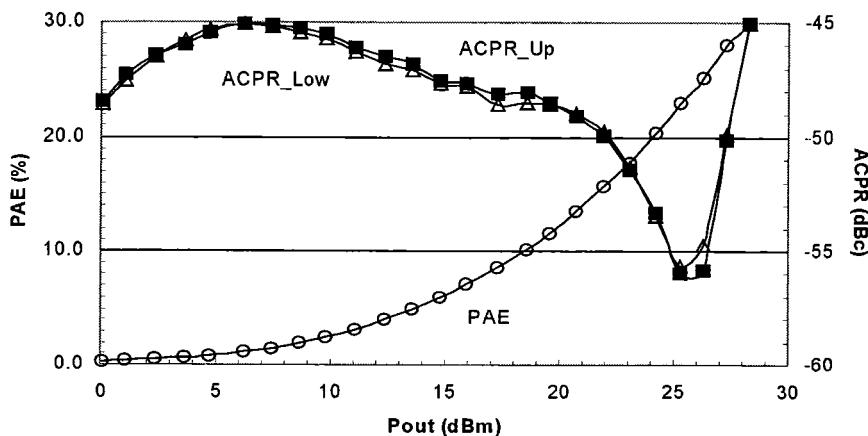
Figure 5.13. Measured PAE and ACPR of the PA with matching networks using the surface-mount components.
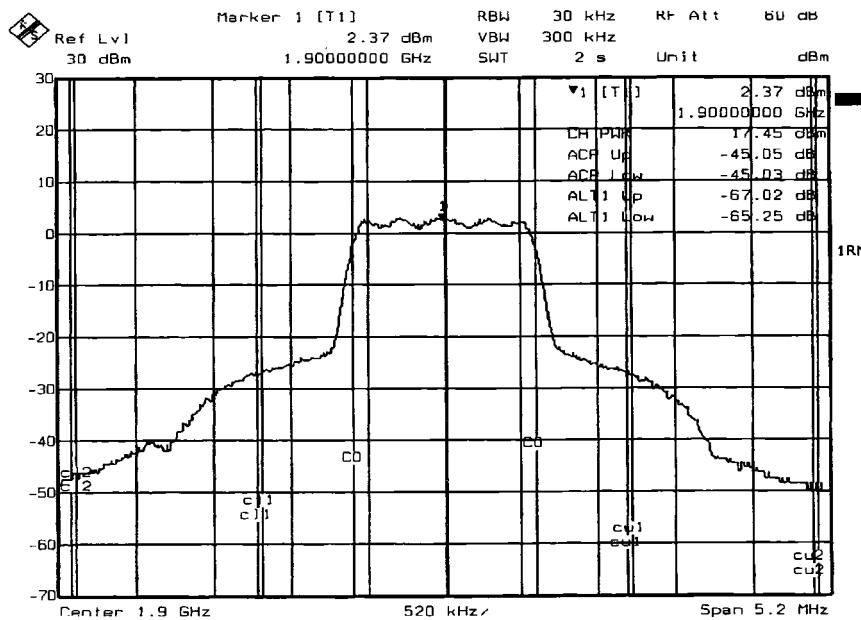
*Figure 5.14.* A snapshot of the ACPR measurement of the PA with the matching networks using surface-mount components.
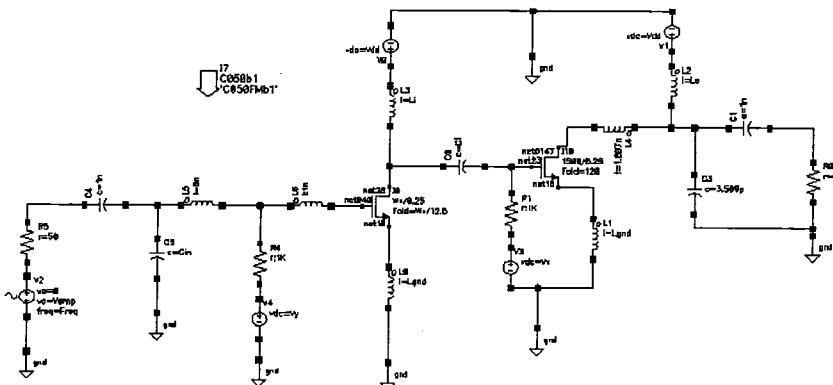
Figure 5.15.   Class AB power amplifier schematic.

with 30% power-added efficiency (PAE) and −45 dBc adjacent-channel-power ratio (ACPR) at 1.9 GHz and 3.6 V supply voltage. The PA was implemented in a 30 GHz BiCMOS process.

5.3   CMOS PA DESIGN

We have seen in the previous section that monolithic Si bipolar power amplifiers are capable of delivering output powers in the range of 25 dBm to 30 dBm. Many wireless LAN standards require, however, lower output powers. The Bluetooth communications protocol, for instance, has three classes of transmitter power requirements: Class 1 (20 dBm), Class 2 (4 dBm) and Class 3 (0 dBm). The low radio power requirements for this protocol have prompted heavy research interest in the area of using CMOS to implement the RF power amplifier.

We will demonstrate a CMOS RF power amplifier capable of meeting the requirement specifications of the Bluetooth Class 1 standard [13]. Since Bluetooth employs a constant envelope modulation scheme, we have focused on achieving the required high saturated power while maximizing PAE. We have set our goal to attain 20 dBm saturated output power with a gain of 20 dB and an associated PAE of more than 35%.

We have employed a single-ended two-stage common-source topology with on-chip interstage matching for the design of our power amplifier (Figure 5.15). For accurate simulation purposes, we have used the RF MOS transistor models as discussed in Chapter 1.

The supply voltage limit is 2.5 V for the 0.25 μm CMOS technology. In the design enough headroom has been alloted to allow for process/chip assembly variations. This headroom is realized by designing the PA to achieve a required
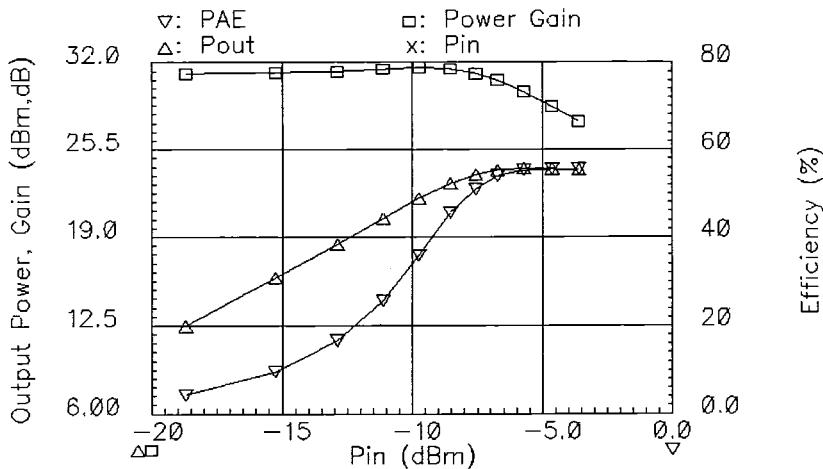
Figure 5.16. Simulated PA performance without taken into account the parasitics due to the layout process.

output power of 24 dBm at 2 V supply. A 1.5 mm output stage biased close to class A region has been used along with a 0.5 mm driver biased in deep class AB region. These transistor dimensions are needed to deliver the output power at a 50 Ω load impedance. Meeting the above specification would allow direct interface on the same chip to the 0 dBm transceiver to meet the Class 1 transmission requirements. For a cost effective solution, we have minimized the number of off-chip components and also avoided a flip-chip configuration. The ground inductance on the MOSFET sources due to the bond wires was set at an achievable value of 0.2 nH for both stages while performing these simulations. Resistive gate biasing is adopted to prevent unwanted oscillations. We have used the bondwire inductance for the driver stage drain connection along with an on-chip ring capacitor structure to provide the interstage matching. Input and output matching was done using lumped element components for simulation purposes.

The simulated performance of this two stage design at 2 V supply voltage is shown in Figure 5.16. The design can provide 31.5 dB of small-signal gain and a saturated power of 24 dBm can be obtained with 55% PAE.

Special care has been taken to generate the layout of the chip. Both the driver and output stages have been built up from a 20 finger interdigitated unit cell of width 250 μm resulting in a gate finger length of 12.5 μm (see also Chapter 1) . The substrate is tied to the source of the unit cell to prevent any substrate bias effects. The RF signal path is kept as short as possible between the stages
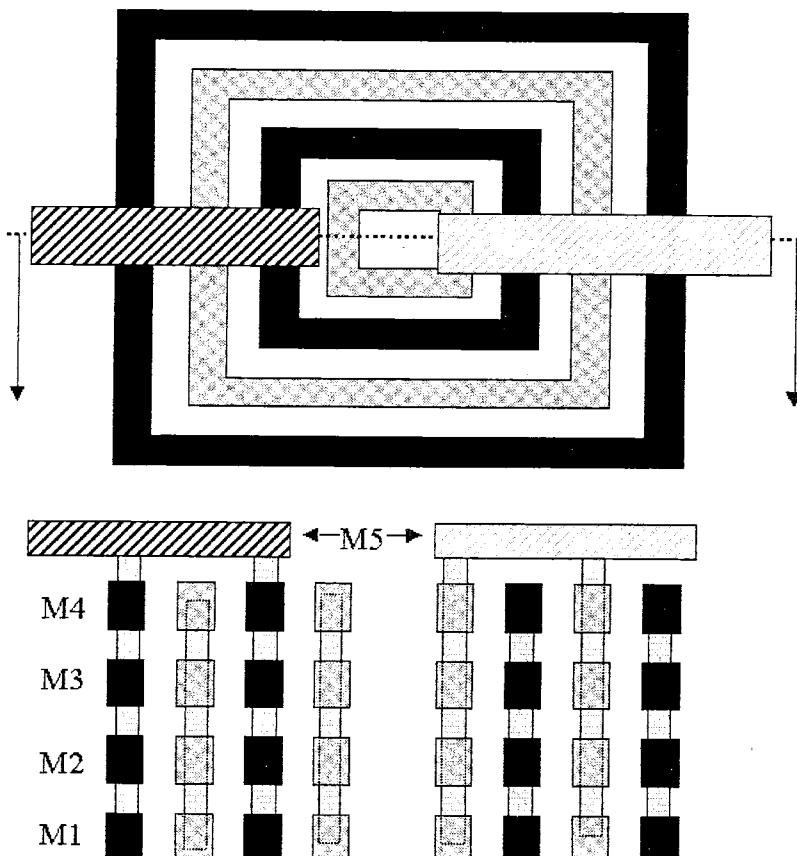
Figure 5.17.   Ring capacitor top and side view. ■ denotes node A and □ represents node B.

and multiple ground pads are used to minimize the ground inductance from bondwires for the stages.

The 10 pF interstage matching capacitance has been formed with five 2 pF unit capacitors. We have designed a ring capacitor instead of using the multilayer fringe capacitor structure as discussed in Chapter 1. The ring shaped structure minimizes the occupied chip area (Figure 5.17). The ring capacitor utilizes both the cross-over capacitance and the fringe capacitance between multi-level interconnect metals in minimum geometry configurations in our CMOS technology to achieve a high capacitance density (2 to 3 times) compared to a conventional metal-sandwich structure.

Post-layout simulation results, which include the effect of parasitic capacitances extracted from the layout are shown in Figure 5.18. As seen from this figure, the most prominent effect of the parasitics is a degradation in the small
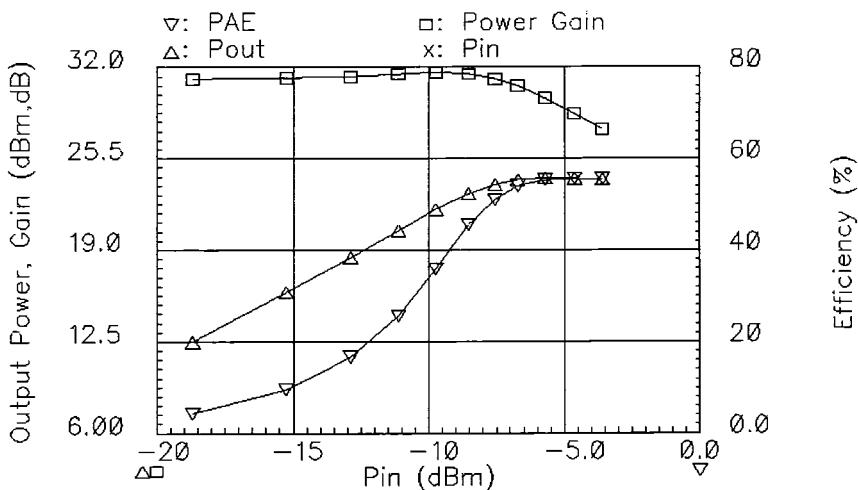
Figure 5.18.   Post-layout simulation results.

signal gain by almost 5 dB due to extra gate-drain capacitance for the two stages from the layout. However, the original saturated power and PAE can still be obtained using a higher input level.

The power amplifier was fabricated in a five metal level 0.25 μm low resistivity substrate (0.01 Ω·cm) CMOS process. The silicon was thinned down to 280 μm to help reduce the ground bondwire lengths. The die was then bonded to the PCB using a conducting glue which also provides a well defined ground to the backside of the chip. Bond wires were used to connect the bondpads to the respective traces on the PCB as shown in Figure 5.19.

Large signal loadpull measurements were then performed on the power amplifier using a Focus Microwaves Loadpull system. Figure 5.20 shows the measured output power and gain at 2.4 GHz for the power amplifier as a function of the input power level for three supply voltage levels. It can be seen that the saturated output power of 22 dBm for a supply voltage of 2 V is about 2 dB lower than simulated results with an associated PAE of 47%. This is due to a higher than designed value for the drain bondwire inductance of the driver along with a slightly undersized driver stage which results in a premature power gain drop-off. Both of these shortcomings can be easily remedied by a simple redesign of the PCB and an increase in the size of the driver stage. For the current design, at a supply voltage to 2.5 V, we can achieve a saturated output power of +24 dBm. The output stage and the driver draw supply currents of 184 mA and 28 mA respectively at this output power level resulting in an asso-
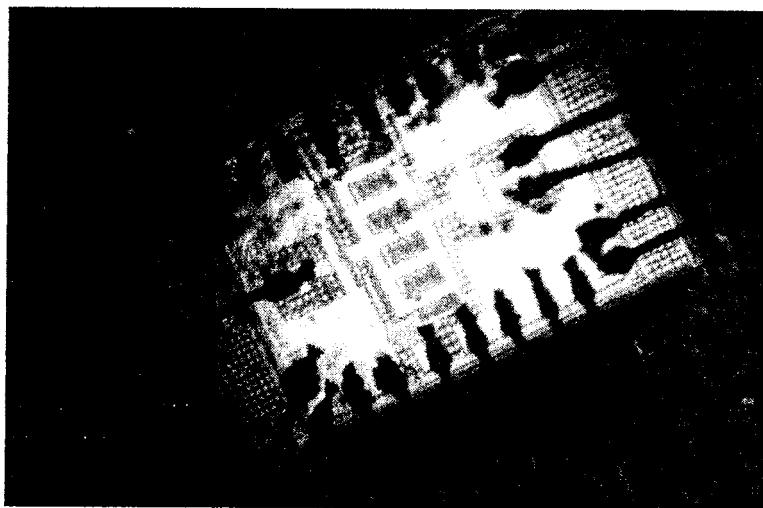
*Figure 5.19.*   Photomicrograph of the die, while mounted on the PCB.
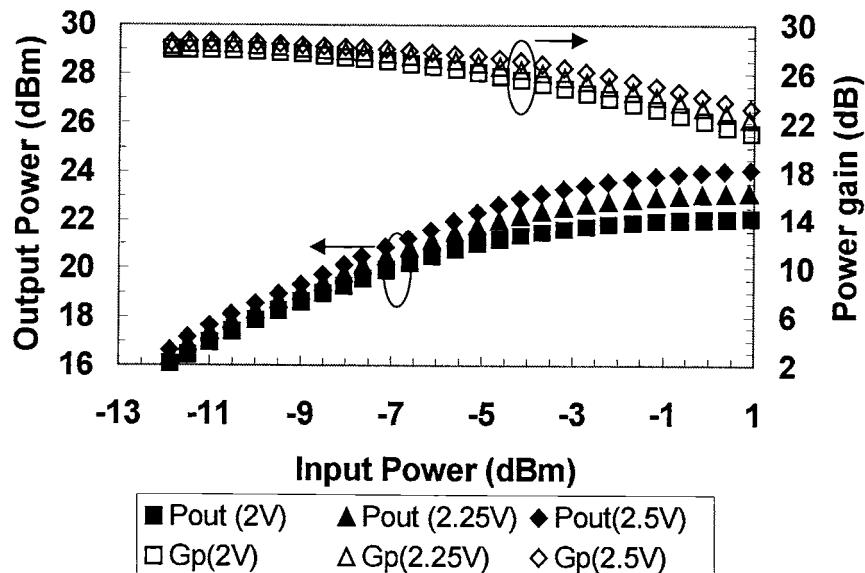
*Figure 5.20.*   Measured output power and power gain
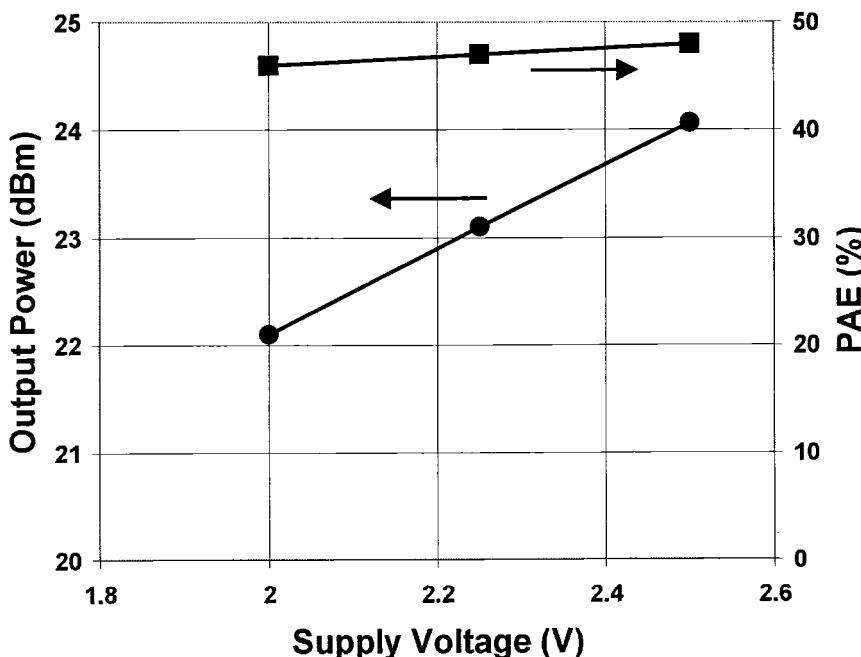
Figure 5.21.   Pout and PAE versus supply voltage.

ciated PAE of 48%. This is the same as the total drain efficiency since the gain of the circuit is larger than 20 dB. A plot of the saturated output power and the associated PAE is shown in Figure 5.21 as a function of the supply voltage.

Issues of significance in the area of RF CMOS power amplifier design include catastrophic gate oxide breakdown and the effect of hot carrier induced device degradation on the RF performance of the amplifier. The hot carrier effect is a phenomenon where under high electric fields near the drain, the channel electrons can cause damage to the Si-SiO2 interface thereby increasing threshold voltage, decreasing transconductance and degrading MOSFET performance. So far, CMOS RF power amplifier papers have not reported on the hot carrier effect even though they are operated at DC+RF voltage levels exceeding the recommended DC limit. Although DC/transient reliability testing has been performed to study hot carrier induced degradation [14], [15], [16], there have been no studies on hot carrier effects under large signal steady state RF operation in CMOS.

We have studied the effect of hot carriers on the saturated output power of the designed power amplifier. The supply voltage limit for this technology from a hot carrier standpoint is 2.5 V DC. We operated our power amplifier at
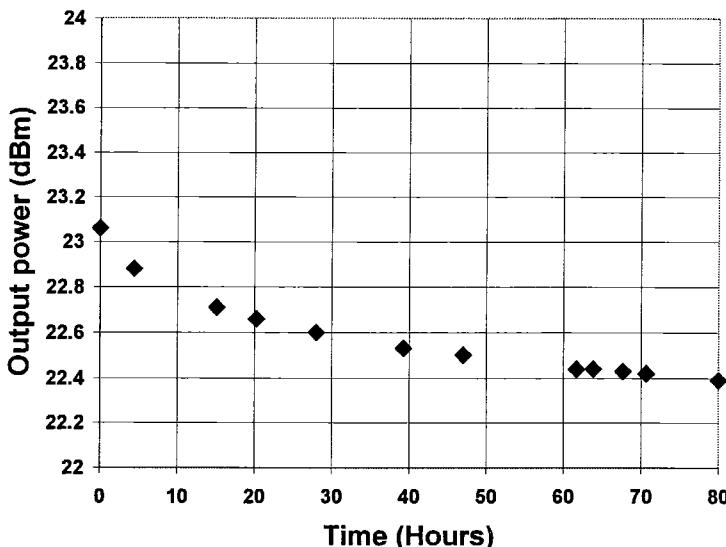
Figure 5.22.   RF Hot Carrier reliability measurement.

Vdd=2.25 V to deliver 23 dBm initial output power. However, under saturated large signal operation, the instantaneous RF voltage swing at the drain of our transistor can reach upto 4.5 V (2 times V_{dd}), well exceeding the DC limit for this technology.

As seen from Figure 5.22, there is no catastrophic destruction of the power amplifier. However, output power of the amplifier decreases with an exponential time constant indicative of classical hot carrier degradation and levels off after about 70 hours of testing indicating that all the created trap sites at the interface have been filled by electrons. The original performance can be recovered by increasing the gate bias. This indicates that the degradation is mainly due to an increase in threshold voltage of the MOSFETs. An increase of the gate bias by 0.2 V results in a Pout of 22.9 dBm and a PAE of 46.5%.

5.4   LINEARIZATION PRINCIPLES

Several linearization techniques have been developed over the past few years to enhance the linearity of existing power amplifiers. Linear power amplifiers have become important, mainly due to the use of spectrally efficient modulation schemes, where the envelope is varying. The need for multi-carrier base-station transmitters instead of single-carrier base-station transmitters also raised the need for linearization techniques. Multi-carrier systems are more flexible and have lower costs than their single-carrier counterparts.
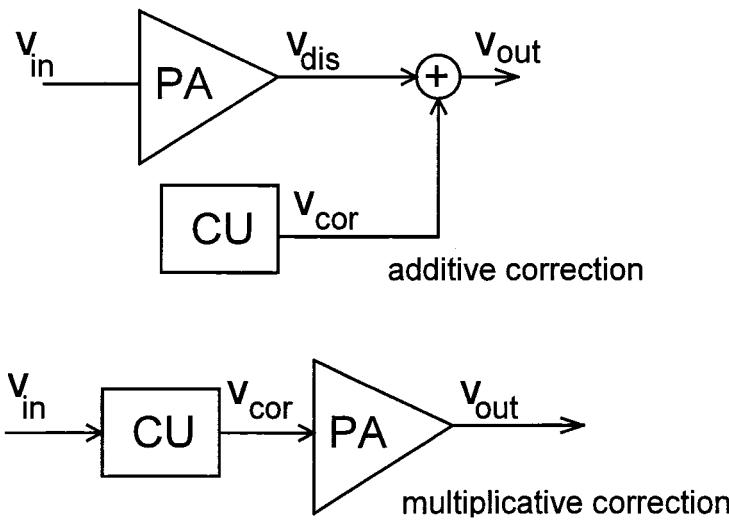
Figure 5.23. Examples of the basic linearization techniques additive and multiplicative corrections.

There are basically two principles for linearization: feedforward or additive correction and feedback or multiplicative correction. Both techniques are shown in Figure 5.23. In additive correction, the correction unit (CU) produces a signal equal to $V_{cor} = -V_{un}$, assuming that $V_{dis} = A \cdot V_{in} + V_{un}$ and A is the gain of the power amplifier. Consequently, the output signal is the linear amplified input signal. The main drawback of this technique is that the summing of signals takes place when the signals have been amplified. Summing of high powers or voltages/currents is therefore needed, but is often difficult to realize. Because additive correction is a feedforward loop, PA and CU must track each other, meaning that one must monitor the PA and adapt the CU when necessary, otherwise the summing will be on the wrong signals in time. An advantage of this technique is the fact that the order of the total non-linearity does not increase. This is in contrast to multiplicative correction, where the order increases, because the CU is now in front of the PA and thus the non-linearity of the CU is also amplified. The output signal is determined by $V_{out} = F(V_{in}) \cdot A$ and, if F is implemented as a feedback loop, we obtain $F = (V_{in} - kV_{out})$ where k is the feedback gain.

It is worth noticing that, assuming a fixed maximum output power, linearization gives lower efficiency values than when no linearization techniques are used. Another important parameter is the adjacent channel power ratio, reflecting the unwanted amount of power which is delivered in the adjacent channel.

RF predistortion

Modulation predistortion
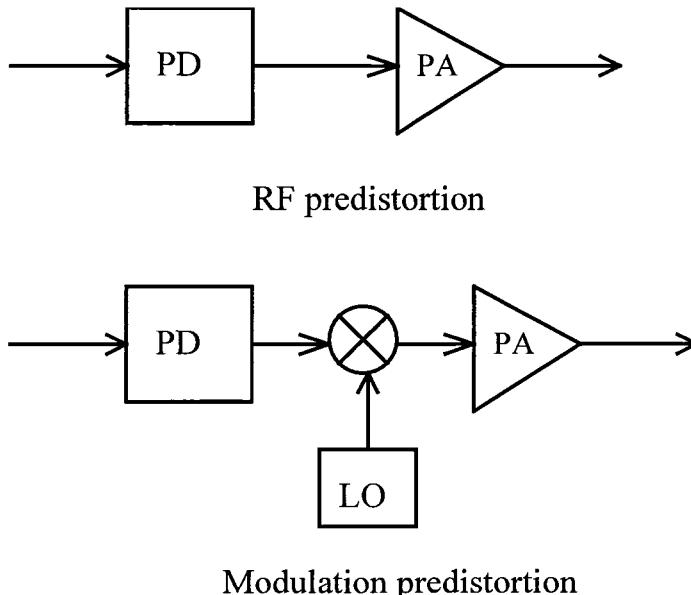
Figure 5.24.   Schematic block diagram for predistortion in the RF domain and for modulation predistortion at IF or baseband.

If we fix this ratio, linearization techniques give better efficiency and allow higher output powers.

In the following sections, we will discuss a few promising linearization techniques, such as predistortion, phase-correction and envelope elimination and restoration (EER). The list of examples is not exhaustive, but it gives a good overview of the state-of-the-art in RF power amplifiers and transmitter concepts.

5.4.1   Predistortion Technique

The basic idea behind predistortion is to distort the signal with the image of the spectral distortion behavior of the power amplifier. Amplifying this predistorted signal using the PA results in a clean spectral output signal. This approach was proposed by Stapleton [17]. The simplest method would be to use a non-linear device in front of the amplifier, to produce the inter-modulation products in anti-phase to those of the amplifier in open loop configuration. This non-linear device can be placed in the RF part of the transmitter, and we refer to this principle as RF predistortion (see Figure 5.24). The advantage of RF predistortion is in the simplicity of the circuits needed at the cost of limited performance. Due to the wide band operation, both the intermodulation and harmonic distortion can be tackled. As can be seen in Figure 5.24, the predistortion can also be
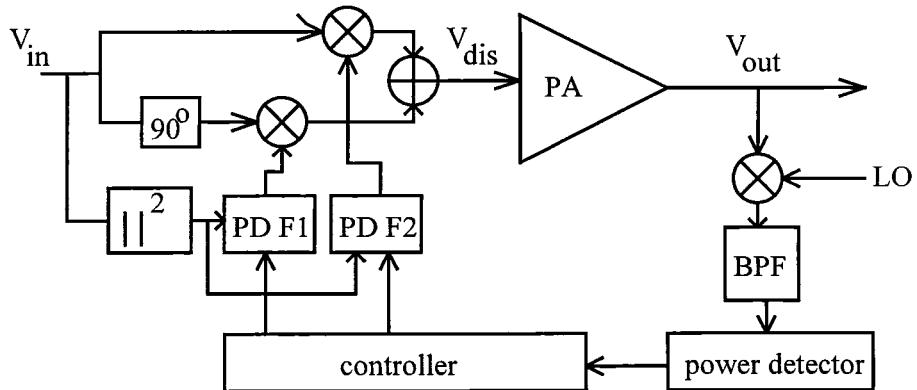
Figure 5.25.   Block diagram of baseband predistortion.

applied at IF or baseband, and is as such referred to as *modulation predistortion*. In general, the predistortion works on complex signals and reduces the inter-modulation harmonics. The main advantage is that the correction can be done in the digital domain, resulting in low power consumptions. If the PD is applied to the base band signal, and thus to the I and Q path, this technique is sensitive to gain and phase unbalance. DC offset in the quadrature modulator also influences the performance. If the PD is applied at IF, and thus after the quadrature modulator, the imbalance of gain and phase do not play a significant role. However, a costly IF and/or RF filter is required to shape the spectrum.

Let us assume that complex baseband predistortion is applied to linearize the power amplifier. A possible block diagram is shown in Figure 5.25, where different non-linear functions have been used for the in-phase and quadrature-phase. The predistorter creates a signal $V_{dis}(t)$, the distorted version from the modulated input signal $V_{in}(t)$. The output of this signal due to amplification is $V_{out}(t)$, and this signal is fed back to the baseband by means of a bandpass filter (BPF) to separate the out-of-band signal power from the wanted signal. The out-of-band power is then averaged by the power detector and used by the controller to generate the predistortion correction signals F_1 and F_2, both also depending on the magnitude of the input signal. Assume

$$F_1(t) = f_{11} + f_{13}x(t) + f_{15}x^2(t)$$
$$F_2(t) = f_{21} + f_{23}x(t) + f_{25}x^2(t)$$

(5.11)

where

$$x(t) = |v_{in}(t)|^2$$

(5.12)

The controller controls the values for f_{11} to f_{25}. The predistorter's complex gain can now found to be

$$F = F_1 + j F_2 \tag{5.13}$$
$$= F\left(|v_{in}(t)|^2\right) \tag{5.14}$$
$$= f_{11} + j f_{21} + (f_{13} + j f_{23})\, x(t) + (f_{15} + j f_{25})\, x^2(t) \tag{5.15}$$
$$= c_1 + c_3 x(t) + c_5 x^2(t) \tag{5.16}$$
$$= F\left(x(t)\right) \tag{5.17}$$

and signal $v_{dis}(t)$ is derived as

$$v_{dis}(t) = v_{in}(t) \cdot F\left(v_{in}(t)\right) \tag{5.18}$$
$$= c_1 v_{in}(t) + c_3 v_{in}(t) x(t) + c_5 v_{in}(t) x^2(t)$$
$$= c_1 v_{in}(t) + c_3 v_{in}(t)\, |v_{in}(t)|^2 + c_5 v_{in}(t)\, |v_{in}(t)|^4$$

From this expression, we can observe the dependency of $v_{dis}(t)$ to $v_{in}(t)$ up to the fifth degree. If the coefficients c_3 and c_5 are chosen properly, the third and fifth order intermodulation distortion can be reduced. In a similar way as in (5.12), let us define

$$y(t) = |v_{dis}(t)|^2 \tag{5.19}$$
$$= x(t) \cdot |F(x(t))|^2 \tag{5.20}$$

for the envelope of a predistorted signal. We can model the amplifier's complex gain as

$$A = A_r(y(t)) + j A_i(y(t)) \tag{5.21}$$
$$= A\left(|v_{dis}(t)|^2\right) \tag{5.22}$$
$$= a_1 + a_3 y(t) + a_5 y^2(t) \tag{5.23}$$
$$= A\left(y(t)\right) \tag{5.24}$$

where coefficients a_1, a_3 and a_5 stem from the complex power series. Now the output of the amplifier can be written as

$$v_{out} = v_{dis}(t) \cdot A\left(|v_{dis}(t)|^2\right) \tag{5.25}$$

Note that both $F\left(|v_{in}(t)|^2\right)$ and $A\left(|v_{dis}(t)|^2\right)$ depend only on the power of their input signals, but not on their phase. As $y(t)$ in (5.21) contains only modulated amplitude information, the equation describes amplitude modulation to amplitude modulation (AM-AM) behavior. Because the coefficients are complex, the equation also describes an amplitude modulation to phase modulation (AM-PM) characteristic.

Combining (5.25) with (5.18) yields

$$v_{out} = v_{in}(t) \cdot F(x(t)) \cdot A(y(t)) \tag{5.26}$$
$$= v_{in}(t) \cdot F(x(t)) \cdot A\left(x(t) \cdot |F(x(t))|^2\right)$$

and the complex gain of the overall system is found to be

$$G_{overall}(x(t)) = \frac{v_{out}(t)}{v_{in}(t)} = F(x(t)) \cdot A\left(x(t) \cdot |F(x(t))|^2\right) \tag{5.27}$$

On the other hand, we may model the whole amplifier by a complex power series, and if we truncate after the fifth order term, we obtain

$$G_{overall}(x(t)) = G_1 + jG_2 \tag{5.28}$$
$$= g_1 + g_3 x(t) + g_5 x^2(t)$$

The output voltage is therefore also given by

$$v_{out} = v_{in}(t) \cdot G_{overall}(v_{in}(t)) \tag{5.29}$$
$$= g_1 v_{in}(t) + g_3 v_{in}(t) x(t) + g_5 v_{in}(t) x^2(t)$$
$$= g_1 v_{in}(t) + g_3 v_{in}(t) |v_{in}(t)|^2 + g_5 v_{in}(t) |v_{in}(t)|^4$$

and therefore, by equating (5.26) and (5.29)

$$G_{overall}(v_{in}(t)) = F(x(t)) \cdot A\left(x(t) \cdot |F(x(t))|^2\right)$$
$$g_1 + g_3 x(t) + g_5 x^2(t) =$$
$$\left[c_1 + c_3 x(t) + c_5 x^2(t)\right] \cdot \left[a_1 + a_3 x(t) \cdot |F(x(t))|^2 + a_5 x^2(t) \cdot |F(x(t))|^4\right] \tag{5.30}$$

Solving the right hand side of (5.30) and re-ordering the zeroth, first- and second-order terms yields the relations

$$g_1 = c_1 \cdot a_1$$
$$g_2 = c_3 \cdot a_1 + c_1 \cdot a_3 \cdot |c_1|^2 \tag{5.31}$$
$$g_3 = c_5 \cdot a_1 + c_3 \cdot a_3 \cdot |c_1|^2 + c_1 \cdot a_5 \cdot |c_1|^4 + 2c_1 \cdot a_3 \cdot \mathrm{Re}\left(c_1 c_3^*\right)$$

where x^* denotes the complex conjugate of x. The third order term depends on the amplifier gain coefficients a_1 and a_3, and the same holds for the fifth order term plus the additional coefficient a_5. Therefore, by an appropriate choice of the predistortion coefficients c_1, c_2 and c_3 it is possible to reduce or eliminate the third- and fifth-order inter modulation products. The need of the feedback loop in Figure 5.25 is now clear: the loop finds the appropriate values for the coefficients c_1, c_2 and c_3 and automatically takes care of process and temperature variations.
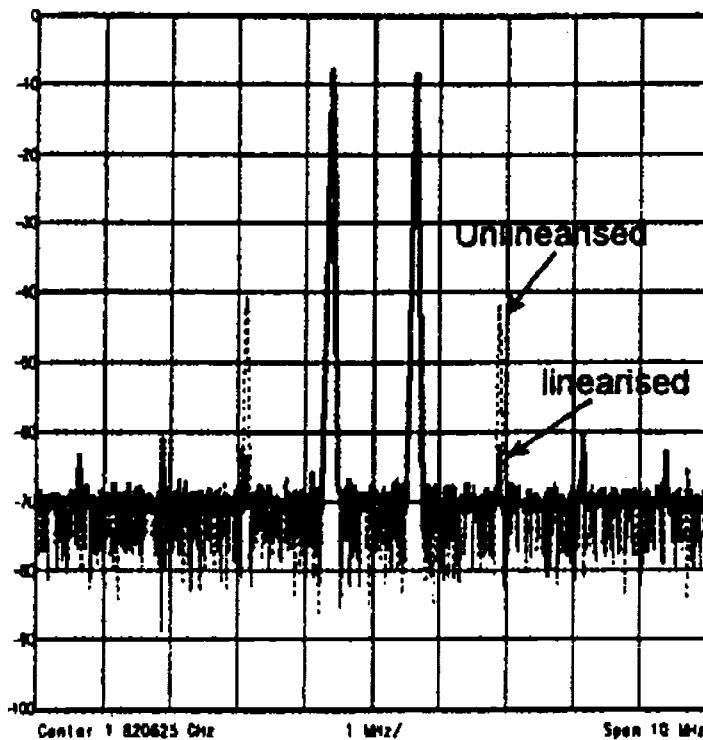
Figure 5.26.   A broadband predistortion technique is applied to a 30 W power amplifier and measured under two tone conditions at 1.8 GHz.

Figure 5.26 shows an example of measurement results when a broadband predistortion technique is applied to a 30 W power amplifier and measured under two tone conditions at 1.8 GHz. A 20 dB suppression of the third harmonic is achieved. In Figure 5.27, a 120 W power amplifier for IS-95 CDMA was tested. A 10 dB suppression is gained at an offset of 750 kHz. Both measurements show that predistortion is a possible linearization technique for modest degrees of linearity improvement.

5.4.2   Phase-Correcting Feedback

Suppose that the input to the power amplifier has a constant envelope and a varying phase. The amplitude modulation is achieved by varying the supply voltage. In such a configuration, the power amplifier can operate in class E. As the supply voltage is varied to modulate the output voltage, the drain voltage of
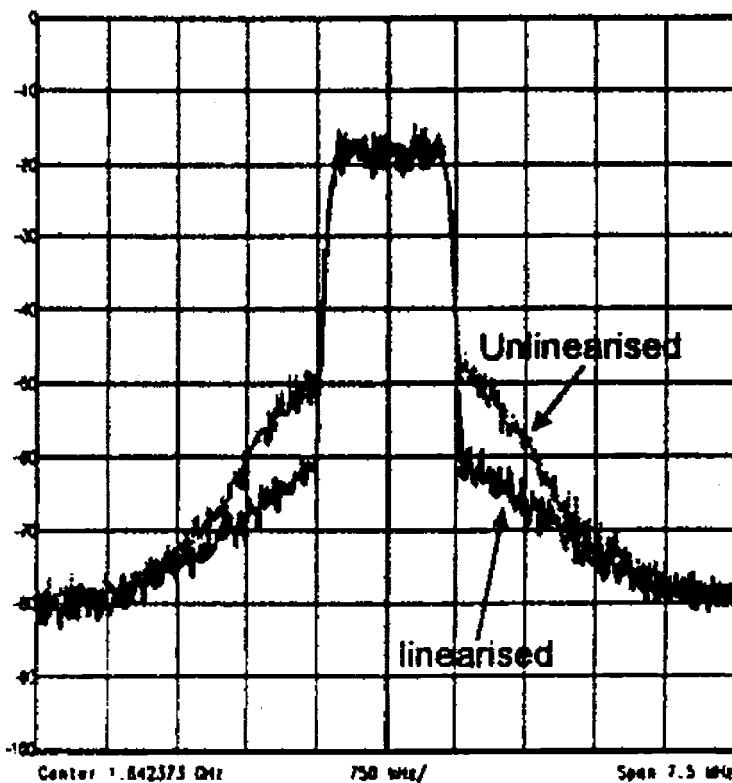
Figure 5.27. A 120 W RF power amplifier for IS-95 CDMA standard has been tested with and without predistortion techniques. A 10 dB improvement has been achieved.

the output transistor will also vary, see Figure 5.1. Consequently, the transistors' capacitances will change, and thus the phase of the output signal changes. This mechanism is an example of amplitude-to-phase modulation, and the resulting phase distortion can be in the range of 20 to 40 degrees.

Utilizing a phase correcting feedback mechanism can help to reduce the phase distortion to the level of a few degrees. A block diagram of this feedback principle is shown in Figure 5.28. In this block diagram, it is not only the PA which introduces AM-to-PM distortion, but also the limiting amplifier, needed to limit the output of the PA to a fixed value. Let us denote the AM-to-PM distortion for the PA by $\theta_{PA}(s)$, and for the limiting amplifier by $\theta_{LA}(s)$, where s is the Laplace operator. Define the phase of the input signal and output signal as $\theta_{in}(s)$ and $\theta_{out}(s)$, respectively. Finally, assume a first order low pass loop
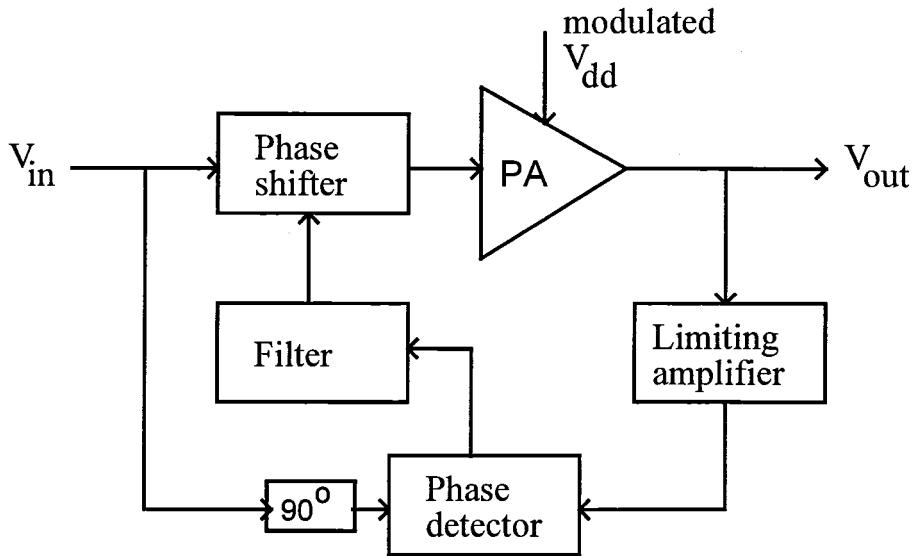
Figure 5.28. Block diagram of a phase correcting feedback loop. The input of the amplifier is constant in amplitude and varies in phase. Amplitude modulation is performed by modulating the supply voltage of the amplifier.

filter to have a first order feedback loop with a transfer function $H(s)$, resulting in a feedback loop transfer equal to

$$K(s) = K \cdot H(s) \tag{5.32}$$

The system is therefore stable. Constant K represents the sensitivity of the phase detector and the phase shifter together. The relationship between the input and output phase is now given as

$$\theta_{out}(s) = \theta_{in}(s) + \frac{\theta_{PA}(s)}{1 + K(s)} - \frac{\theta_{LA}(s) \cdot K(s)}{1 + K(s)} \tag{5.33}$$

with the error signal

$$\theta_{error}(s) = \frac{\theta_{PA}(s)}{1 + K(s)} - \frac{\theta_{LA}(s) \cdot K(s)}{1 + K(s)} \tag{5.34}$$

Because $H(s)$ has a lowpass transfer characteristic, the loop appears as a high-pass filter for the distortion of the PA. The pole of the filter must be placed such that the bandwidth of the filter is larger than the bandwidth of the amplitude

Limiting cell   Saturated gain stage
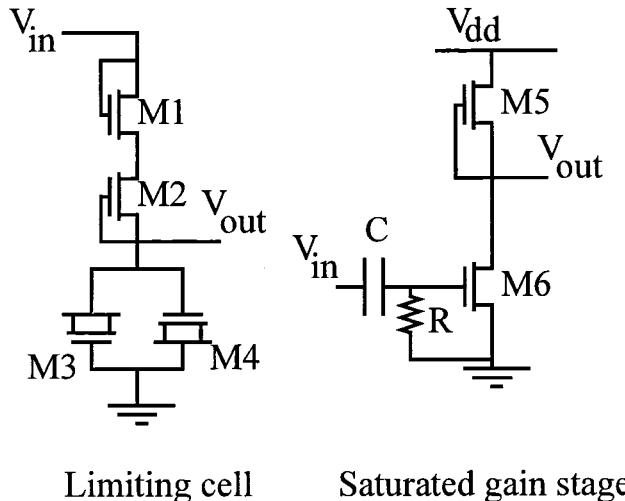
Figure 5.29.   MESFET implementations of a limiting cell and a saturated gain stage.

modulated signal. Suppose that this amplitude modulated signal has a bandwidth of 50 kHz. Then setting the loop filter bandwidth to 200 kHz means that the loop transfer characteristic can be approximated by $K \cdot H_o = K_o$ where H_o is the filter gain. The PA distortion is therefore reduced by a factor $1 + K_o$. The phase error due to the limiting amplifier only has a minor attenuation of $(K_o/1 + K_o)$. Assuming K_o to be equal to 20, then the phase distortion of 20 and 40 degrees reduces to 0.95 and 1.9^o. Assuming a maximum phase error at the output, then the maximum allowed phase error of the limiting amplifier can be found from (5.34).

An example of phase-correcting feedback can be found in [18]. The design was realized in a 0.8 μm GaAs MESFET process. The limiting amplifier consists of a series-parallel limiting cell followed by several saturating gain stages as shown in Figure (5.29). The limiting cell contains a passive stage consisting of two back-to-back parallel diodes (M3 and M4) and a resistor (M1 and M2). A passive stage is needed, since the output of a PA can swing higher than the supply voltage. For positive voltages, M1 acts as a diode-connected MESFET while M2 behaves like a current source. Similarly, for negative voltages, M2 acts as a diode-connected MESFET while M1 behaves like a current source. For small signal levels M1 and M2 behave like a resistor. Phase errors in the order of 1 to 2° can be achieved at a reasonable power dissipation. The saturated gain stage must ensure that the output voltage has a fixed level for all input values. Several stages are usually needed. Each stage is
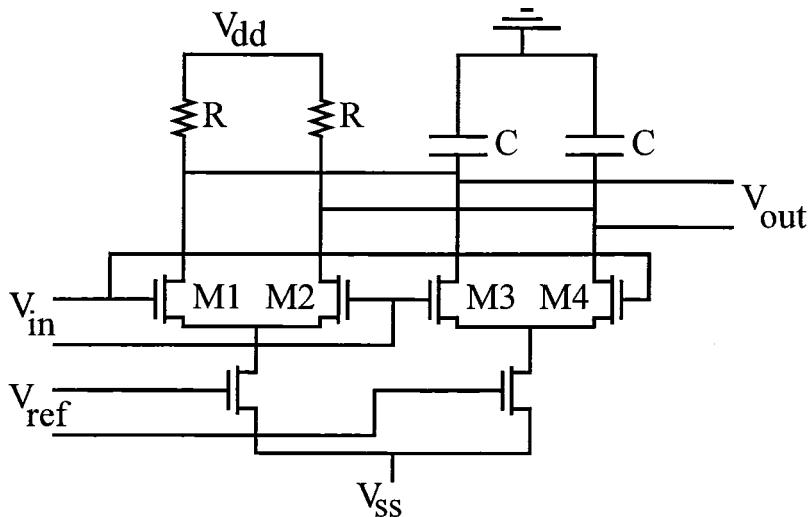
Figure 5.30.   Simplified circuit diagram for the phase detector.

based on an inverter and is AC coupled to the next stage by means of capacitor C. Together with resistor R, the capacitor determines the load of the previous stage and provides a relatively constant load under amplitude variations.

The phase detector is based on a Gilbert active mixer configuration as depicted in Figure 5.30. Capacitors C are implemented on-chip to provide small high frequency loads for M1 to M2 and M3 to M4, which act as switches. The resistors R provide large loads for the DC component of the switched current. If the phase difference between V_{in} and V_{ref} is 90°, then the differential output V_{out} equals zero. When both input signals are in-phase or 180 degrees out of phase, each mixer is completely turned on and the output reaches its maximum absolute value. The sensitivity of the phase detector is given by

$$K = \frac{\Delta V}{\Delta \varphi} = \frac{R \cdot I}{90^o} \qquad (5.35)$$

where I is the current through the resistor.

The phase shifter is based on a varactor-tuned LC network to retain the amplitude of the signal waveform. Two such networks are used, one connected to the source and one to the drain of a MESFET. The resonant frequency in the drain-tuned circuit will increase while that of the source-tuned network decreases. Consequently, a negative phase shift is obtained in the source network and an equal positive phase shift in the drain network.
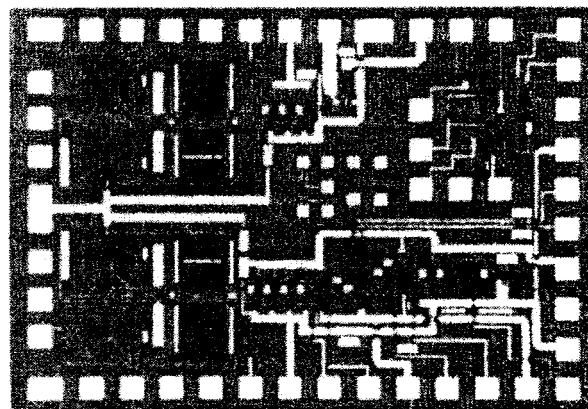
Figure 5.31.   Microphotograph of the phase correcting feedback loop. The chip has dimensions 2.4mm x 1.7mm.

A microphotograph of the feedback loop is shown in Figure 5.31. To reduce parasitics, interconnections were laid out in the highest metal layer. The power amplifier, in this case a class E end stage preceded by a class F stage, is on a separate module. Measurement results on the complete system are shown in Figure (5.32). For a 50 Ω load, the amplifier delivers more than 400 mW at 2.4 V supply voltage. The efficiency is above 60%, but drops rapidly for the phase-corrected system if the supply voltages decreases The power amplifier has a phase error of more than 20 degrees. The feedback loop reduces this error to 4 degrees. Phase corrected feedback loops clearly help to reduce the resulting phase error due to the power amplifier. The additional power consumption due to the loop is in the order of 20 mW.

5.4.3   Envelope Elimination and Restoration (EER)

The principle of first eliminating the envelope of the RF input signal to generate a constant-amplitude signal which is passed through a non-linear PA, while in the meantime extracting the amplitude information, amplifying it separately and then recombining it with the phase information, dates back to the fifties [19], [20]. Because a constant amplitude signal passes through the non-linear amplifier, the phase information will hardly be distorted. A switched mode power amplifier can therefore be used, allowing high efficiency figures. The extracted amplitude information passes through a highly linear amplifier, but because this signal does not contain any RF component, the realization of such an amplifier is simplified. A simplified block diagram of envelope elimination
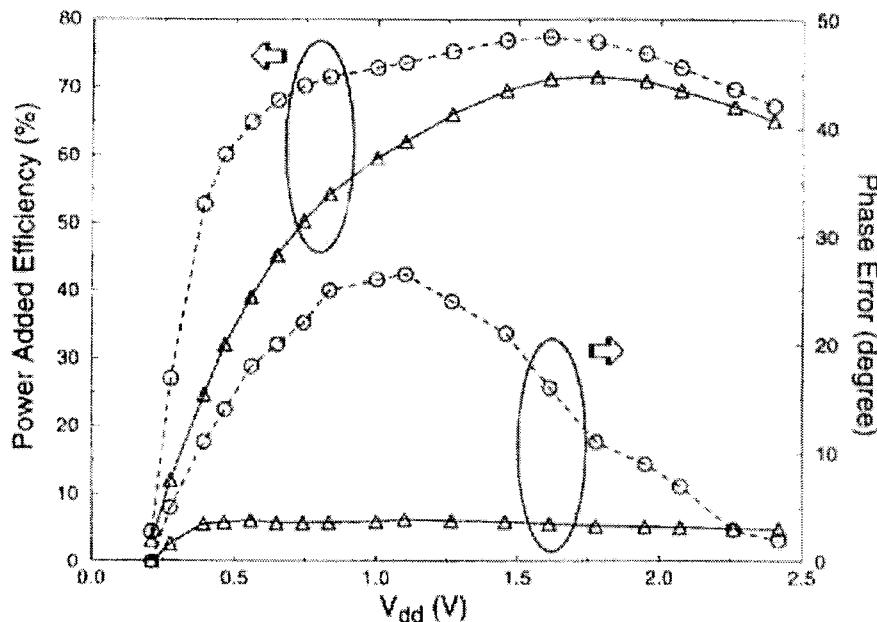
Figure 5.32.   Measurement results. The graphs marked with the − o − are the system without phase correction, results with phase correction are marked as −△−. The power added efficiency and phase error are shown, both as functions of the supply voltage.

and restoration (EER) is shown in Figure 5.33. A switching power supply has been used to amplify the low frequency amplitude information. Recombination of phase and magnitude has been realized by direct modulation of the power supply of the power amplifier. A feedback path closes the loop and guarantees tracking between the RF output amplitude and RF input signal. In this way, errors introduced by the switching power supply block and mismatches between the phase and magnitude paths, can be eliminated.

There are some possible drawbacks in this concept. Firstly, the limiting amplifier can introduce AM-to-PM errors, from which this system cannot recover. Secondly, there are two signal paths and the delay in the RF signal path (i.e. the phase path) is substantially shorter than that of the low frequency magnitude path. This delay causes intermodulation distortion and, for two equal input tones, this distortion is expressed by

$$IM = 2\pi BW^2 \Delta \tau^2$$
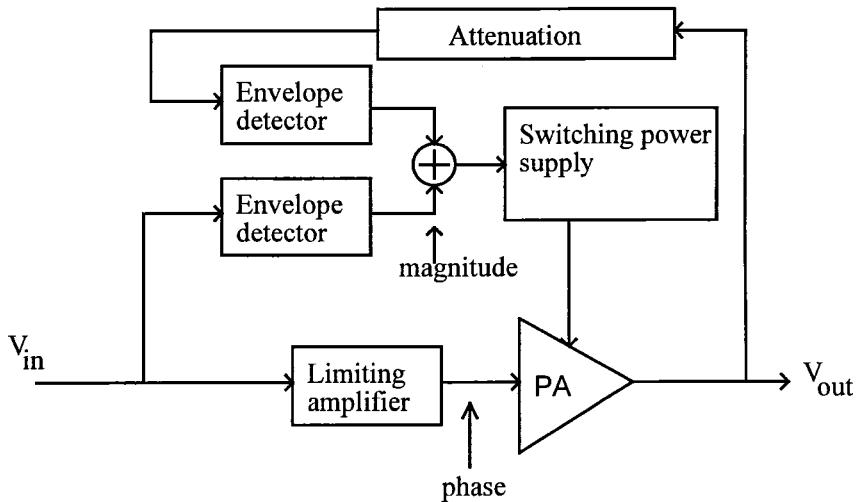
Figure 5.33. Conceptual block diagram of an EER system with feedback.

where BW is the bandwidth of the RF signal, and $\Delta\tau$ is the delay mismatch. For two tones that are 200 kHz apart and a required IM of -40 dBc, this yields a maximum delay of 200 ns. Thirdly, the bandwidth of the switching power amplifier is related to its efficiency. For high efficiency figures, this bandwidth is rather small, in the order of a few tenths of kHz.

Su demonstrated that the concept of EER can work pretty well [21]. They implemented the envelope detector, limiting amplifier, attenuator and switching power supply in a 0.8 μm CMOS technology and tested the concept on three different amplifiers. They considered the performance of a 3.3 V GaAs PA under AMP conditions (Figure 5.34). Under normal conditions, this PA can reach 40% efficiency (peak PAE). However, to fulfil the linearity requirements of the North American Digital Cellular (NADC) standard (employing $\pi/4$-QPSK modulation), the PA must be backed off by 4 dB, yielding a PAE of 16% at 25 dBm output power,.(see cross mark in Figure 5.34). Using the EER concept, this PA can reach a PAE of 35% at 29 dBm output power. A similar improvement can be seen with a 4.8V GaAs PA (see Figure 5.34). In [21] a CMOS PA was also tested under EER conditions. A measured $\pi/4$-QPSK constellation diagram under normal conditions was compared to the diagram in the case of EER. The linearized system achieved a phase error of 1.3^o rms, and a magnitude error of 2.5% rms.
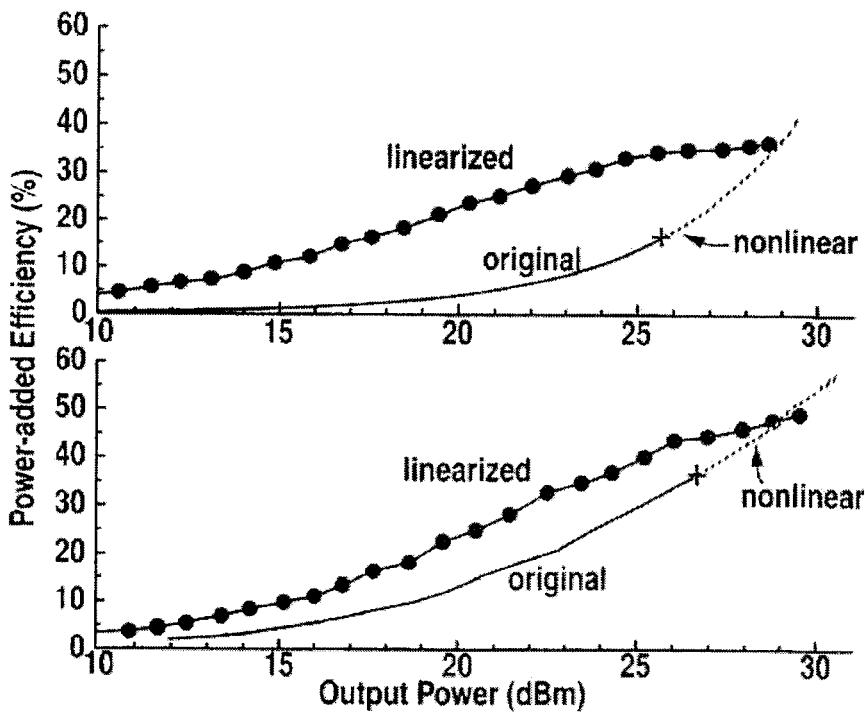
Figure 5.34. Measured efficiency of a 3.3V GaAs PA with and without EER (top). A similar measurement result is shown for a 4.8V GaAs Pa at the bottom.

5.4.4   Cartesian Feedback

The last technique to be discussed is an example of modulation feedback called Cartesian feedback. At the output of the amplifier, the transmitted signal is decomposed into the distorted I and Q signals This information is then used to adapt the I and Q signals at the input of the amplifier. A schematic of this approach is shown in Figure 5.35.

    The feedback loop has a delay compared to the direct signal path. A phase shifter is therefore needed in the feedback path. The delay in the loop limits the bandwidth over which the loop can operate. A few 100 kHz bandwidth can normally be achieved, allowing this technique to be used in telecommunication standards with a relatively small channel bandwidth and spacing, such as DAMPS $\pi/4$-DQPSK, for instance. As an example, measurement results for this standard are depicted in Figure 5.36 and 5.37. A 30 dB improvement has been achieved.
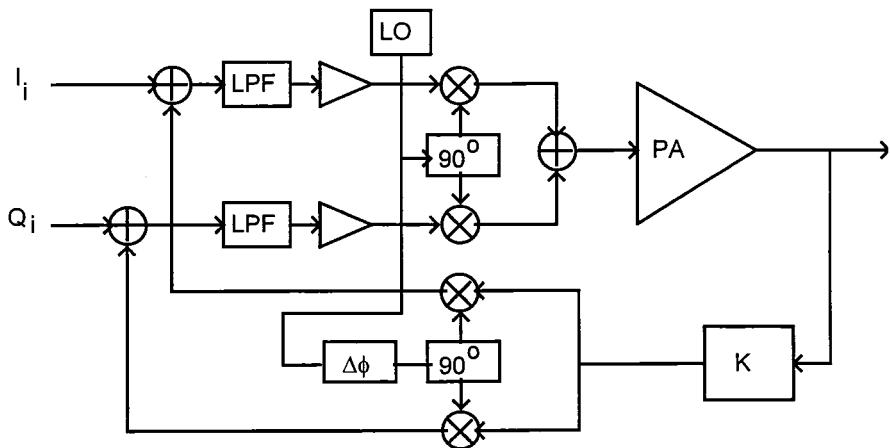
Figure 5.35. Schematic drawing of Cartesian feedback linearization technique.
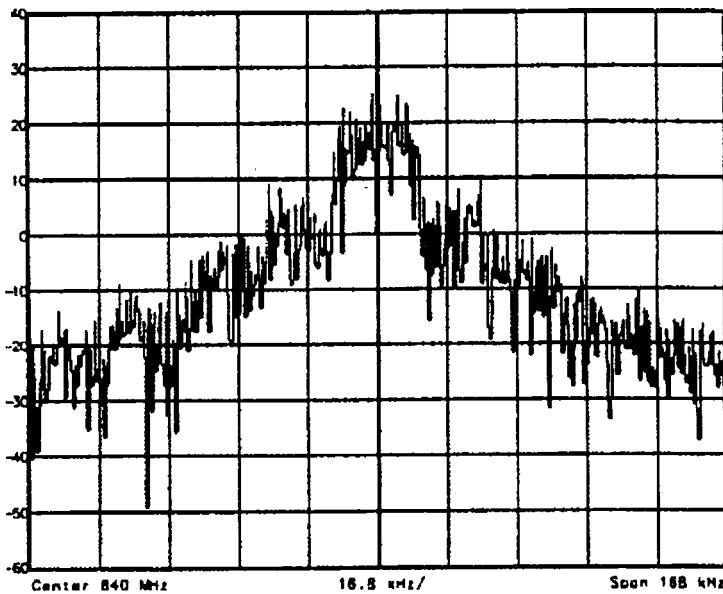
Figure 5.36. Uncorrected response of the RF amplifier for DAMPS $\pi/4$-DQPSK.

Although measurements show good results, there are some major problems when using Cartesian feedback loops. Classical Cartesian feedback is a homodyne solution technique. As in Figure 5.35, the VCO operates at the same frequency as the PA, which gives the potential problem of pulling. The output
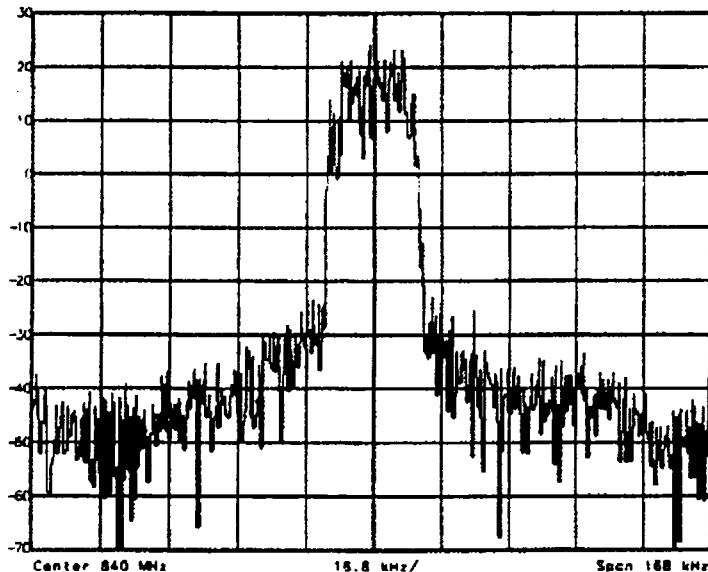
Figure 5.37. Cartesian loop response for same amplifier. Suppression is 30dB in adjacent channel.

of the PA couples back to the VCO. A super heterodyne solution can be used to prevent this, requiring a band-pass filter. Due to the required suppression, SAW-filters are needed which introduce large delays and we end up with a small modulation bandwidth of a few hundred Hertz. Besides the problem of the loop bandwidth, SAW filters are discrete components and are expensive.

REFERENCES

[1] J.B. Hagen, *Radio-Frequency Electronics, Circuits and Applications*, Cambridge University Press, New York, 1996.

[2] H. Krauss *et al.*, *Solid State Radio Engineering*, Wiley, New York, 1980.

[3] F. Raab, "High Efficiency Amplification Techniques," *IEEE Circuit Syst. (Newsletter)*, , no. 12, pp. 3–11, Dec. 1985.

[4] F. Sechi, "Linearized Class-B Transistor Amplifier," *IEEE Journal of Solid-State Circuits*, vol. 11, no. 4, pp. 264–270, Apr. 1975.

[5] S.L. Wong and S. Luo, "A 2.7-5.5V, 0.2-1 W BiCMOS RF Driver Amplifier IC with Closed Loop Power Control and Biasing Functions," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 12, pp. 2259–2264, Dec. 1998.

[6] N.O. Sokal and A.D. Sokal, "Class E - A New class of high Efficiency Tuned Single-ended Switching Power Amplifiers," *IEEE Journal of Solid-State Circuits*, vol. 10, no. 6, pp. 168–176, June 1975.

[7] M. Kazimierczuk, "Class E Tuned Power Amplifier with Shunt Inductor," *IEEE Journal of Solid-State Circuits*, vol. 16, no. 2, pp. 2–7, Feb. 1981.

[8] T. Sowlati *et al.*, "Low Voltage High Efficiency GaAs Class E Power Amplifier for Wireless Trasnmitters," *IEEE Journal of Solid-State Circuits*, vol. 30, no. 10, pp. 1074–1080, Oct. 1995.

[9] W. Simburger *et al.*, "A monolithic Transformer Coupled 5-W Silicon Power Amplifier with 59% at 0.9GHz," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 12, pp. 1881–1892, Dec. 1999.

[10] X. Zhang *et al.*, "A SiGe HBT Power Amplifier with 40% PAE for PCS CDMA Applications," in *proc. of 2000 IEEE MTT-S Digest*, 2000, pp. 857–860.

[11] S. Luo and T. Sowlati, "A monolithic Si PCS-CDMA Power Amplifier with an Impedance-Controllable Biasing Scheme," in *proc. of 2000 IEEE MTT-S Digest*, 2001.

[12] T. Sowlati and S. Luo, "Bias Boosting Technique for a 1.9GHz Class AB RF Amplifier," *Proc. Int. Symp. on Low Power Electronics and Design (Italy)*, July 2000.

[13] V. Vathulya *et al.*, "Class 1 Bluetooth Power Amplifier with 24dBm Output Power and 48% PAE at 2.4 GHz in 0.25 um CMOS," in *European Solid-State Circuits Conf. (ESSCIRC)*, 2001.

[14] E. Takeda and N. Suzuki, "An Empirical Model for device Degradation due to Hot Carrier Injection," *IEEE Electron Device Letters*, vol. 34, no. EDL-4, pp. 111–, June 1983.

[15] J. Choi *et al.*, " Hot Carrier-Induced MOSFET Degradation: AC versus DC Stressing," in *VLSI Technology Symposium Digest*, 1987, pp. 45–47.

[16] W. Weber *et al.*, "Lifetimes and Substrate Currents in Static and Dynamic Hot-Carrier Degradation," in *IEDM Technical Digest*, 1986, pp. 390–.

[17] S. Stapleton and F. Costescu, "An Adaptive Predistorter for a Power Amplifier based on Adjacent Channel Emissions," *IEEE Trans. on Vehicular Technology*, vol. 41, no. 2, pp. 49–57, Feb. 1992.

[18] T. Sowlati *et al.*, "Phase-Correcting Feedback System for Class E Power Amplifier," *IEEE Journal of Solid-State Circuits*, vol. 32, no. 4, pp. 544–549, Apr. 1997.

[19] L. Kahn, "Single-sideband transmission by envelope elimination and restoration," *proc. IRE*, pp. 803–806, July 1952.

[20] L. Kahn, "Comparison of linear single-sideband transmission with envelope elimination and restoration single-sideband transmitters," *proc. IRE*, pp. 1706–1712, Dec. 1956.

[21] D.K. Su and W.J. McFarland, "An IC for Linearizing RF Power Amplifiers using Envelope Elimination and Restoration," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 12, pp. 2252–2258, Dec. 1998.

Chapter 6

Oscillators

Oscillators are intriguing building blocks. An oscillator is one of the very few building blocks in a transceiver that has a built-in timing reference. When the power supply is switched on, DC power is somehow translated into a periodic signal, which forms the heartbeat of many systems. Ongoing world-wide research contributed to the design of low phase noise oscillators, which makes it more and more a science rather than an art. Advances in IC-technology, especially in passives, have simplified complete oscillator integration. This chapter starts by discussing oscillator basics and specifications. Behavioral models, properties and circuit topologies of LC and RC oscillators are then highlighted. The chapter ends by discussing two monolithic oscillator realizations.

6.1 INTRODUCTION

This section starts by discussing the ideal oscillator, followed by an oscillator with non-idealities which are encountered in practice. The third part highlights a practical oscillator classification. This section ends by discussing the oscillation conditions and amplitude stabilization mechanisms.

6.1.1 The Ideal Oscillator

The basic function of an oscillator is to generate a periodic signal with certain properties. An ideal oscillator generates a signal which only has wanted properties. The output of an ideal harmonic oscillator (Figure 6.1(a)) with angular frequency ω_{osc} [rad/s] and peak amplitude $V_{carrier}$ [V] can be written as

$$V_{out}(t) = V_{carrier} \, \cos(\omega_{osc} \, t) \tag{6.1}$$

In the frequency domain, this is equivalent to a discrete spectral line with height $V_{carrier}$ at angular frequency ω_{osc}. This means that all carrier power is located
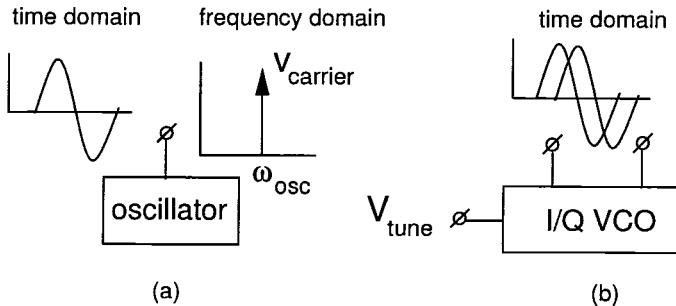
Figure 6.1.   A single phase oscillator (a) and a tunable I/Q VCO (b).

in an infinitely small bandwidth around ω_{osc}. A tunable ideal oscillator (Figure 6.1(b)) can be represented by (6.2).

$$V_{out}(t) = V_{carrier} \cos(2\pi \ (K_{vco} \ V_{tune} + f_{center}) \ t) \qquad (6.2)$$

In (6.2) the initial phase of $V_{out}(t)$ is assumed to be zero. Tuning voltage V_{tune} [V] controls the frequency, and tuning constant K_{vco} [Hz/V] determines the tuning slope. f_{center} is the oscillation frequency with a zero tuning voltage. Equation (6.2) does not include the "integrating" property of a controlled oscillator. The phase is the integral of the oscillator frequency with respect to time. Use of (6.2) for non-continuous V_{tune} results in phase jumps in the oscillator output signal. The following expression models a practical oscillator more accurately. The output phase is always continuous, even if a non-continuous V_{tune} is modulating the oscillator:

$$V_{out}(t) = V_{carrier} \cos(2\pi \ (K_{vco} \int_{-\infty}^{t} V_{tune} \, dt + f_{center} \ t)) \qquad (6.3)$$

As the tuning input is a voltage, the oscillator in Figure 6.1(b) is a Voltage Controlled Oscillator (VCO). In the case of a Current Controlled Oscillator (CCO), the tuning constant K_{cco} will have the unit [Hz/A]. Instead of generating one output signal, the VCO in this figure generates a sine and a cosine. Many modern receiver and transmitter architectures require these so-called "In-phase and Quadrature" (I/Q) signals in their signal-processing part.

6.1.2   The Non-ideal Oscillator

In practice, anything that can change will change due to non-idealities. The oscillator and its properties are no exception. An oscillator will never have the exact center frequency required, due to the processing spread in the IC process. Some additional tuning range will therefore always be required on
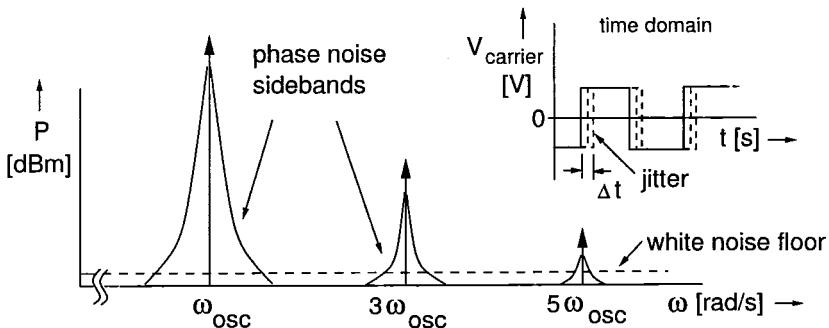
Figure 6.2. The spectrum (fundamental and two harmonics) of a square wave with phase noise sidebands. The inset shows the effect of the sidebands in the time domain: jitter.

top of the required range. Noise from the oscillator circuitry and externally generated noise (on the power supply, for example) corrupt the spectral purity of an oscillator signal. This means that the carrier power is now distributed in a finite bandwidth around ω_{osc} and its harmonics. Figure 6.2 shows the fundamental and two harmonics of a square wave. The application determines whether the harmonics of an oscillator output signal are unwanted or wanted. For example, if an oscillator is used as a clock generator, the harmonics are generally wanted. In that case, zero-crossings in the time domain are less sensitive to noise from the circuits where the clock-signal is used. An oscillator output signal usually also has even harmonics. In balanced architectures, these are usually lower than the odd harmonics.

The most important distinction between an ideal and a practical oscillator are phase noise sidebands, as illustrated in Figure 6.2. Although most power is present at ω_{osc}, some power is also present at small offsets from ω_{osc}. These phase noise sidebands decrease with increasing offset frequency from ω_{osc} or its harmonics. White noise becomes dominant at a certain offset frequency. The white noise floor can originate from the oscillator but often originates from cascaded circuits. In the time domain, phase noise is referred to as jitter, and is shown in the inset of Figure 6.2. Due to the presence of jitter, the exact moment of a zero-crossing of the square wave (in other words, the phase) is stochastic. The ideal oscillator described by (6.1) can be extended to model this phase uncertainty by stochastic variable $\theta(t)$.

$$V_{out}(t) = V_{carrier} \cos(\omega_{osc} t + \theta(t)) \tag{6.4}$$

In general, multiplicative amplitude noise will also be present. However, unlike phase noise, this amplitude noise can be removed by a limiter at the expense
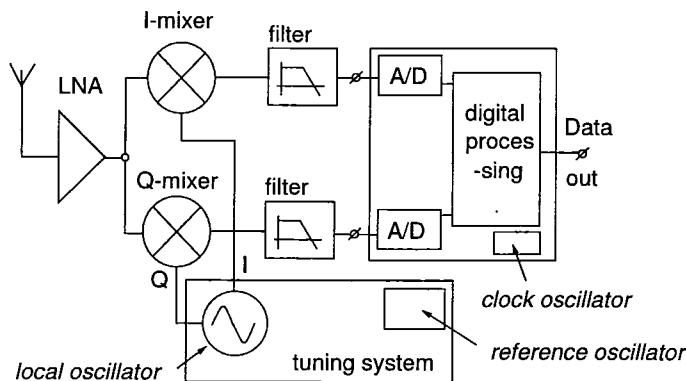
Figure 6.3.   Simplified block diagram of a Zero-IF receiver.

of generation of harmonics. Section 6.2 discusses the influence of oscillator phase noise on transceiver performance and its specification.

6.1.3   Application and Classification

Oscillators are used at various places in an RF transceiver system. Figure 6.3 illustrates this with a block-diagram of a zero-IF receiver. The Local Oscillator (LO) in the tuning system provides the in-phase and quadrature mixers with a signal which is equal, in frequency, to the average received RF carrier frequency. The mixers convert the wanted signal directly to baseband. The LO is part of a phase locked loop (PLL), and uses a crystal oscillator as reference. PLLs will be discussed in the next chapter. Another oscillator signal is needed in the demodulator to provide a clock signal for the A/D-converters and digital circuitry. In practice, the crystal oscillator needed in the PLL is also used for clock generation in a transceiver.

Oscillator classification may be based on one of the properties of an oscillator, such as, is it a distributed type or a lumped type oscillator, is it a single phase, quadrature or multi-phase type oscillator, etc. The practical classification adopted in this chapter is shown in Figure 6.4.

On an abstract level, it is possible to distinguish between LC and RC oscillators. Losses are always present, in both in LC and RC oscillators. In order to start-up and sustain stable oscillation, active devices which compensate the losses are needed. Apart from these "regenerative" elements, implemented by active devices, LC oscillators (or more accurately "LC-like" oscillators) are constructed using inductors (L) and capacitors (C), or passive devices which can be modeled with equivalent circuits using inductors and capacitors. The
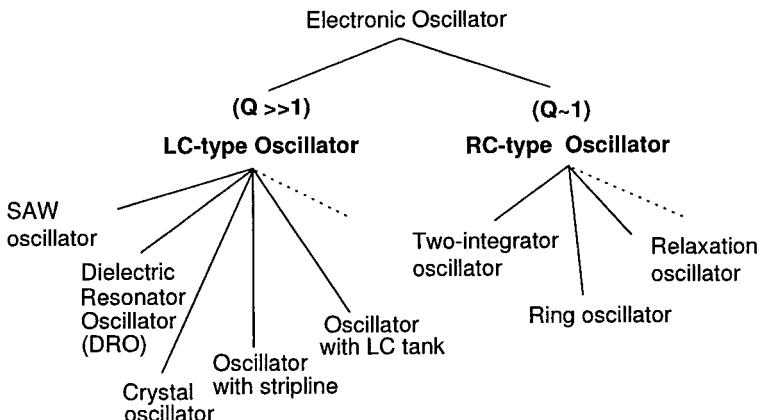
Figure 6.4.   Oscillator classification.

presence of both inductors and capacitors[1] makes it possible to preserve energy in an oscillation period. In Section 6.3.1 we will see that this improves the spectral purity of oscillators. This is indicated by the figure of merit "quality factor" Q in Figure 6.4, which will be defined on page 210. Q is normally much larger than unity for LC-type oscillators.

No inductors are present in RC oscillators. Only resistors (R) and capacitors are used together with the active devices. This means that no energy is preserved per oscillation period, which leads to poorer spectral purity compared to LC oscillators (see Section 6.4.1). Q is close to unity for RC oscillators.

Figure 6.5 shows three LC-type oscillator configurations from the classification. The LO in RF transceivers is often implemented using an LC oscillator with lumped LC resonator (the resonator is also referred to as a "tank" circuit). In that case, the resonator is implemented using lumped inductors and capacitors which can be on-chip [1] or off-chip [2]. Another way of constructing an LC-type oscillator is to use a strip-line or transmission line as a delay element or distributed resonator [3]. Piezoelectric material is used in crystal oscillators, to implement an inductor or complete resonator [4, 5]. Crystal oscillators have an excellent spectral purity and also very good long-term frequency stability. Another example of a very stable resonator is a dielectric resonator (DR). In a Dielectric Resonator Oscillator (DRO), the DR is placed adjacent to a micro-strip line, which implements the coupling between the resonator and the oscillator [6]. In contrast to a crystal oscillator, where the piezoelectric material operates in bulk vibrational mode, a Surface Acoustic Wave (SAW) resonator

---

[1] With a quality factor Q> 1.
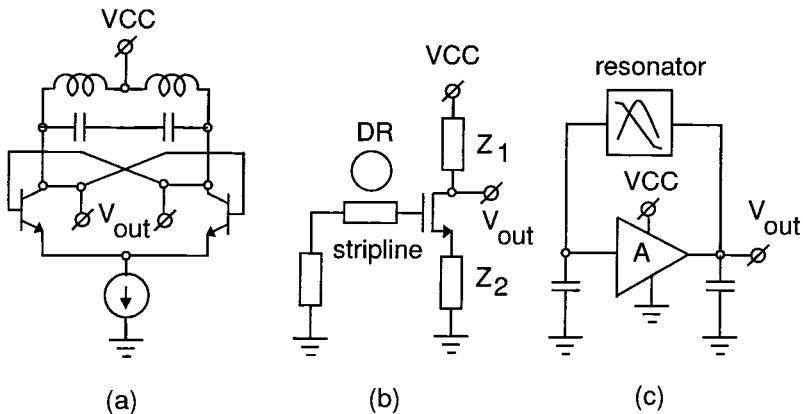
Figure 6.5.   Three examples of LC-type oscillators. A differential bipolar oscillator with LC tank (a), A MOSFET DRO (b) and a general resonator based oscillator (c). The resonator can be a crystal or SAW device, for example.
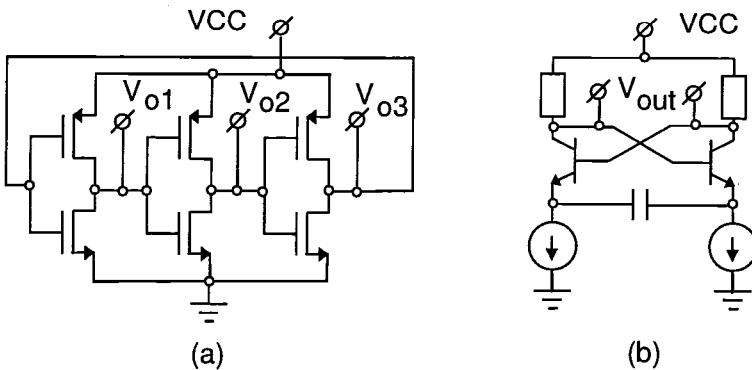
Figure 6.6.   Two examples of RC-type oscillators. A single-ended CMOS inverter ring oscillator (a) and a bipolar relaxation oscillator (b).

operates in surface vibrational mode. This allows higher oscillator operation frequencies when compared to crystals.

Figure 6.6 shows two RC-type oscillator configurations from the classification. Ring oscillators are known for their ease of integration. No coils are needed and simply connecting the output of a few cascaded inverters to the input already implements such an oscillator (Figure 6.6(a)). For the single-ended ring oscillator, an odd number of stages is required to prevent latch up.

A second example of an RC-type oscillator is the two-integrator oscillator [7] (see Figure 6.15 for its behavioral model). Like any even-stage ring oscillator, it provides I/Q signals. This oscillator can generate sine waves with low harmonic distortion.

An example of a bipolar relaxation oscillator is shown in Figure 6.6(b). The principle of a relaxation oscillator is based on the charging and discharging of a capacitor. A current charges a capacitor until a positive voltage threshold is reached and the direction of the current is reversed. Once the negative threshold voltage is reached, the current direction is again reversed, and so on.

The reader should note that this list of oscillator types is far from exhaustive as indicated by the dotted lines in Figure 6.4. Many other types and combinations are possible, but are not as widely used in conjunction with integrated transceivers.

6.1.4 Oscillation Conditions

It is important for an oscillator designer to predict when the oscillator under design is properly dimensioned, so that the oscillator will start and produce a periodic signal. Most oscillators can be analyzed by modeling them as a feedback system. Figure 6.7 shows a general block diagram of a linear feedback system with transfer functions $H_f(j\omega)$ and $\beta(j\omega)$. As will become clear later on, any oscillator is a nonlinear system. The conditions needed for oscillation can be analyzed using linear models, however.

The transfer function Y_{out}/X_{in} of Figure 6.7 is the general equation for a feedback system (6.5).

$$\frac{Y_{out}(j\omega)}{X_{in}(j\omega)} = \frac{H_f(j\omega)}{1 - H_f(j\omega)\,\beta(j\omega)} \tag{6.5}$$

The necessary conditions for steady-state oscillation are known as the Barkhausen conditions [8]. The first condition is called the "gain condition" and specifies that the open loop gain must be unity (see (6.6)). The second condition is referred to as the "phase condition" (see (6.7)). This condition states that the total open loop phase shift must be k times 360 degrees, where k is an integer value including zero. When both conditions are met, the denominator in (6.5) becomes zero, which results in a non-zero Y_{out} for a zero input signal X_{in}.

$$X_{in}(j\omega) \quad\quad Y_{out}(j\omega)$$
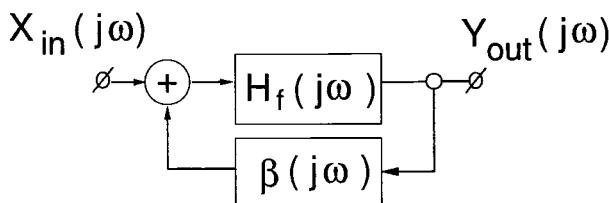
Figure 6.7. Block diagram of a feedback system.

$$|H_f(j\omega)\,\beta(j\omega)| = 1 \tag{6.6}$$

$$\angle\,H_f(j\omega)\,\beta(j\omega) = k\,360° \tag{6.7}$$

The Barkhausen conditions state the necessary conditions for stable oscillation, but not for start-up. In any transceiver system it is vital that the oscillators autonomously start oscillating, usually triggered by noise, when the system is switched on. In order to guarantee oscillator start-up, the open loop gain must initially be larger than unity. This requirement is called the "start-up condition": $|H_f(j\omega)\beta(j\omega)| > 1$. Note that the open loop gain $|H_f(j\omega)\beta(j\omega)|$ ranges typically from 2 to 5 for self-limiting oscillators (see Section 6.1.5) to ensure start-up.

Figure 6.8 is used to illustrate the oscillation conditions. Transfer function $H_f(j\omega)$ is defined as transconductance g_m in the LC oscillator model. The transfer function ($\beta(j\omega)$) of the LC resonator formed by C_p, L_p and loss resistance R_p can be described as

$$\beta(j\omega) = \frac{R_p}{1 + jvQ_p} \tag{6.8}$$

where

$$v = \frac{\omega}{\omega_{osc}} - \frac{\omega_{osc}}{\omega} \tag{6.9}$$

$$Q_p = R_p\sqrt{\frac{C_p}{L_p}} \tag{6.10}$$

$$\omega_{osc} = \frac{1}{\sqrt{L_p C_p}} \tag{6.11}$$

At ω_{osc}, the open loop gain $|H_f(j\omega)\beta(j\omega)|$ is equal to $g_m R_p$. As shown in Figure 6.8, the oscillator does not start if g_m is smaller or equal to $1/R_p$ at ω_{osc}. The lower graph in Figure 6.8 illustrates oscillator start-up when the start-up condition is met.

The open loop gain can be determined with simulation, by cutting open the feedback loop in the oscillator at a certain point, for example at point P in Figure 6.8. This is easy in a behavioral model. When performing open loop simulations at transistor level, the same DC and loading conditions must be enforced as in the closed loop situation in order to get the correct results.

The oscillation conditions can conveniently be checked using bode plots [9]. Figure 6.9 shows the bode plot of the LC oscillator model under discussion. The gain and phase plot shows that the oscillations conditions are exactly met for g_m
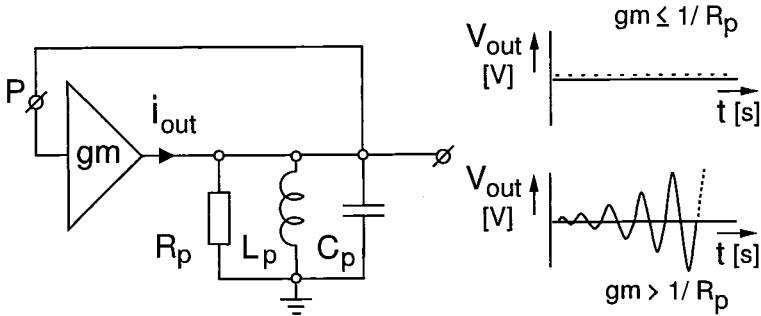
Figure 6.8. Behavioral model of an LC oscillator. To start-up, g_m must be larger than $1/R_p$.
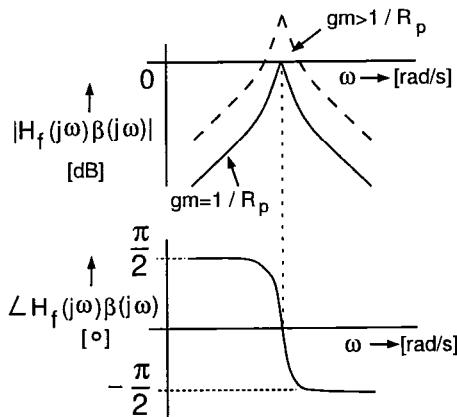
Figure 6.9. Bode diagram of the behavioral model in Figure 6.8.

equals $1/R_p$. The bode plots predicts that the oscillator will start oscillating for any setting of $g_m > 1/R_p$. The discussed oscillation conditions are necessary conditions for oscillation, but can also be met by stable circuits as will be shown in the following paragraphs. If in doubt, the root locus method can be used to assess whether the circuit under investigation will oscillate or not [10] .

Using the oscillation conditions, the oscillation frequency of the classical Colpitts and Hartley oscillators can be derived. Figure 6.10(a) shows a general three terminal oscillator. The active part of the oscillator is a transistor (e.g. a bipolar or a MOS transistor) modeled with its transconductance. One of the three terminals should be grounded. All losses are lumped in resistor R. The open loop model of the three terminal oscillator in Figure 6.10(b) can be used to derive the oscillation frequency. The open loop gain $H_f(j\omega)\beta(j\omega)$ is equal
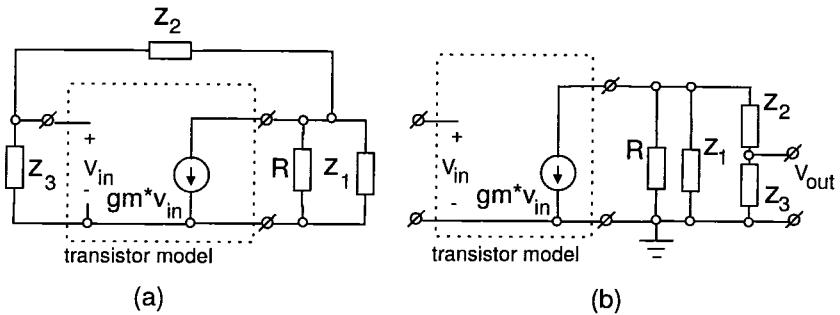
Figure 6.10. General three terminal oscillator. One of three terminals should be grounded (a) and its open loop model with one terminal grounded (b).

to

$$\frac{V_{out}(j\omega)}{V_{in}(j\omega)} = \frac{g_m(Z_1(Z_2 + Z_3) + R(Z_1 + Z_2 + Z_3))}{RZ_1(Z_2 + Z_3)} \frac{Z_3}{(Z_2 + Z_3)} \quad (6.12)$$

If Z_1 and Z_3 are capacitors, and Z_2 is an inductor, a Colpitts oscillator is obtained whose oscillation frequency obeys:

$$f_{osc-Colpitts} = \frac{1}{2\pi} \frac{1}{\sqrt{L_2 C_{tot}}} \quad (6.13)$$

where C_{tot} is the series combination of C_1 and C_3. Note that if Z_3 is replaced by an inductor, and Z_2 by a capacitor, the gain condition can be met but not the phase condition. This configuration will therefore not yield an oscillator. A Hartley oscillator is obtained if Z_1 and Z_3 are replaced by inductors and Z_2 is replaced by a capacitor. The oscillation frequency of the oscillator is now equal to

$$f_{osc-Hartley} = \frac{1}{2\pi} \frac{1}{\sqrt{C_2 (L_1 + L_3)}} \quad (6.14)$$

The circuit diagrams of the Colpitts and Hartley oscillator are shown in Figure 6.11. Since any of the three terminals in Figure 6.11 can be grounded, three types of Colpitts and Hartley oscillators exist. Obviously, the Colpitts oscillator can also be constructed using a MOS transistor and the Hartley type using a bipolar device. In practice, the Colpitts oscillator is encountered more often, since it only requires one inductor. Without detailed analysis, it is difficult to say which type is best for which application. In any case, the choice of the ground terminal determines which parasitics of the active device are shorted and which terminal is best as the output terminal.
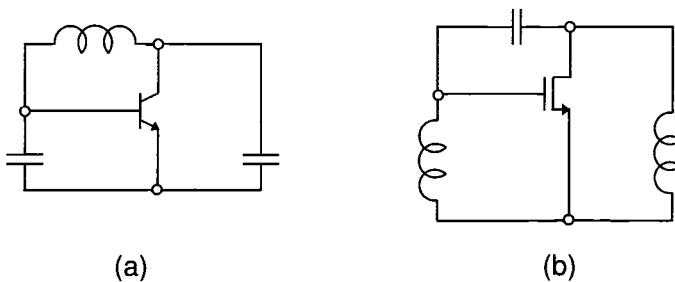
Figure 6.11.   The Colpitts (a) and Hartley (b) oscillator.
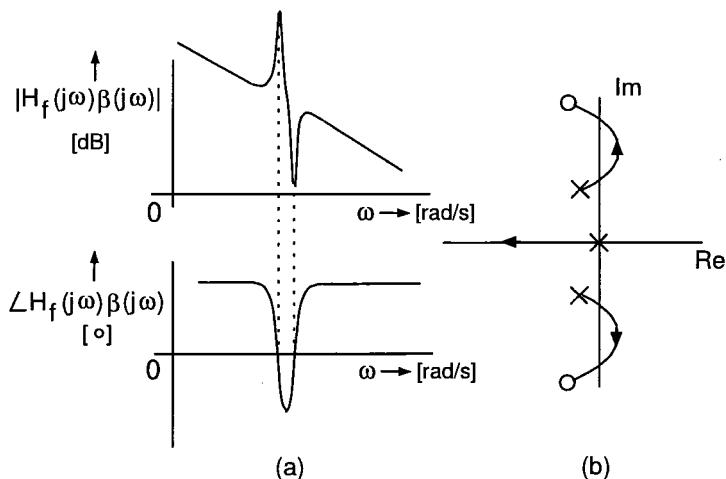
Figure 6.12.   Bode diagram (a) and root locus (b) of a Pierce oscillator with some parasitic capacitance.

Figure 6.12 shows the bode diagram and root locus of a Pierce oscillator (Colpitts-type with grounded emitter or source) with some losses and a parasitic capacitor in parallel with its inductor. Although the oscillator has two complex conjugated poles for any gain setting, these poles enter the left half plane for very high transistor gain, despite the fact that the oscillation conditions are met (as indicated by the bode plot). Hence, the root locus method should be used to see whether the oscillator will start. Alternatively, a transient simulation can be performed if calculations become too complex.

In the bode plots of Figure 6.12(a), the oscillation conditions are met for two frequencies. In a practical oscillator circuit with parasitics this can indicate that there is a risk of multi-oscillation [11]. When multi-oscillation occurs, more than one oscillation coexists in a steady state. The unwanted oscillations generally distort the wanted periodic signal, which then becomes useless for
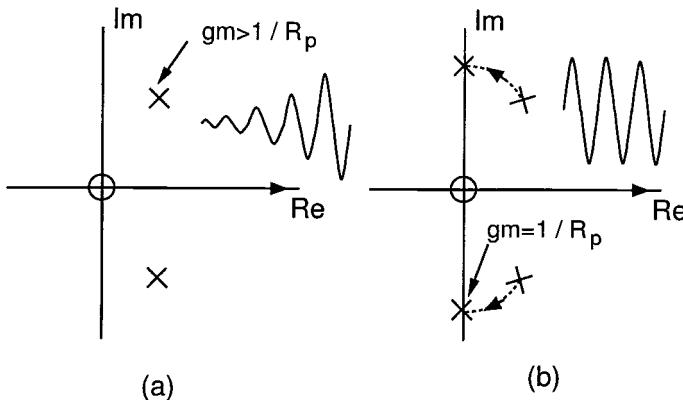
Figure 6.13.   The effective g_m in the LC oscillator model is reduced by a nonlinear mechanism.

application in a transceiver. If this is the case, and if the parasitic cause of the multi-oscillation can not be removed, the gain at the parasitic oscillation frequencies must be sufficiently reduced below unity to eliminate the unwanted oscillation modes.

6.1.5   Amplitude Stabilization

The previous section discussed linear oscillator models and their analysis. None of these models give any information regarding the amplitude of the oscillator output signal. When the start condition for oscillation is met, the poles of a harmonic oscillator are in the half right plane. This illustrated in Figure 6.13(a) for the LC oscillator model presented in Figure 6.8.

The time response of a feedback system with these pole locations corresponds to a growing sine wave. Without adding nonlinearity in an oscillator model, the oscillator signal would keep on growing and growing. Nonlinearity causes the poles of the oscillator to move to the imaginary axes, which corresponds to a steady oscillation level in the time domain. In the case of the LC oscillator from Figure 6.8, a nonlinear mechanism reduces g_m after start-up from a value larger than $1/R_p$ to exactly $1/R_p$ (Figure 6.13(b))[2].

---

[2]This is a simplification. During start-up an oscillator may be modeled using poles and zeros. However, as soon as the carrier is so large in an oscillator that nonlinearities start having influence, the pole/zero description is no longer valid. If linear modeling is applicable, it can be shown that the poles in Figure 6.13(b) are noisy and are located in the left half plane, very close to the imaginary axis. Hence g_m will be a tiny amount smaller than $1/R_p$ [12].
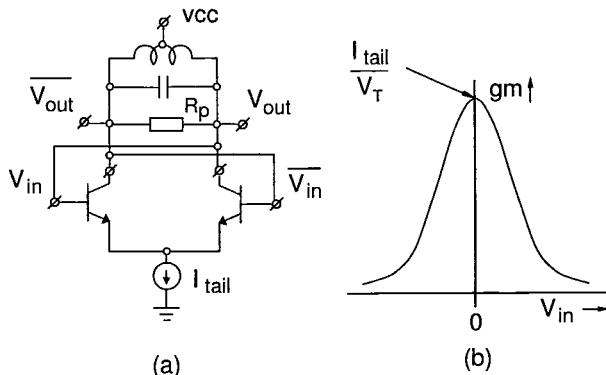
Figure 6.14.   LC oscillator with differential pair (a) and the derivative of its transfer function (b).

Any electronic oscillator contains a nonlinear control mechanism which re-duces the loop gain to unity after start-up, at which point steady state oscillation is reached. This nonlinear mechanism can be implemented in various ways.

One possibility, called self-limiting, is to use the nonlinear characteristic of an active element in the oscillator. Since this method does not need additional circuitry, it is often used in high frequency oscillators. In the literature the simplest self-limiting oscillator is often referred to as the "van der Pol" oscillator [13], in which the active element has the transfer function $I_{out} = \alpha V_{in} - \gamma V_{in}^3$.

More readily available in practice is the limiting characteristic of a MOS or bipolar differential pair, for example. In Figure 6.14(a) a cross-coupled differential pair is shown which is used to realize a simple LC oscillator. The transfer function of this differential pair is $I_{out} = \tanh(V_{in})$. Its derivative to V_{in} ($g_m = \text{sech}(V_{in})$) is shown in Figure 6.14(b). When the oscillator starts up V_{in} is small and g_m is equal to I_{tail}/V_T, where V_T is the thermal voltage kT/q. As can be seen in Figure 6.14(b), g_m decreases for an increasing V_{in}. Since the oscillator output signal traverses the g_m versus V_{in} characteristic, the average g_m is smaller than the small-signal g_m needed for start-up at $V_{in} = 0$. The oscillation amplitude will stabilize when the average g_m exactly compensates all losses (modeled by R_p), and the amplitude condition for oscillation is met. If the oscillator signal becomes large enough, the transistors will act like switches and the output peak amplitude will become $2/\pi \cdot R_p I_{tail}$. The amplitude will therefore increase with increasing I_{tail} (current limited region) until, eventually, the supply voltage (or another saturation mechanism) will limit the amplitude (voltage limited region). The latter region is normally not a good operating region due to increased noise and harmonics.
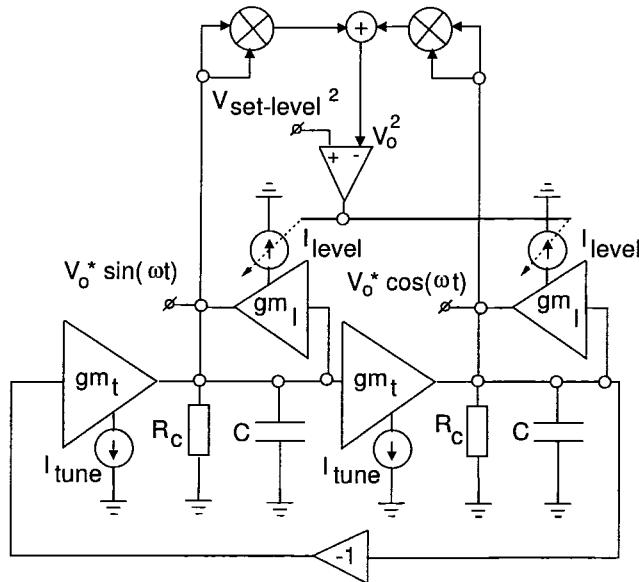
Figure 6.15.   Two-integrator RC oscillator with AGC-control.

A second method for amplitude stabilization is using an Automatic Gain Control (AGC). The oscillation amplitude is measured and used in a negative feedback control loop, which stabilizes the oscillator amplitude to a set value after start-up. AGC can be useful to reduce power dissipation. Consider the example where a large start-up current is needed for the quick start-up of a crystal oscillator. After start-up, this current and thus the power dissipation can be reduced significantly during steady state operation. The AGC control is usually slow, that is, it has a large time constant in order to minimize modulation of the oscillator by noise and spurious signals via the AGC loop. Figure 6.15 shows the behavioral model of a two-integrator oscillator with AGC. Note that after start-up (when steady-state is reached) this AGC-loop can be modeled by a linear model. If the set-level is such that the transconductances g_{mt} and g_{ml} work in the linear region, this oscillator is practically linear. This is especially interesting for applications where harmonics need to be very low.

The frequency g_{mt}/C can be varied by controlling g_{mt} with current I_{tune}. Transconductance g_{ml} is needed to compensate the losses modeled by R_c and can be used for amplitude control. As the oscillator provides I/Q signals, AGC can be implemented using two mixers and a comparator, since $V_o^2 \sin(\omega t)^2 + V_o^2 \cos(\omega t)^2 = V_o^2$. Notice that frequency and amplitude control for this oscillator are orthogonal. In the case of an LC oscillator, AGC can be implemented using a peak detector which measures the amplitude and is used

Table 6.1. Tuning range of several standards.

Standard	Mobile receiver [MHz]	Mobile transmitter [MHz]
DECT	1880 - 1920	same
Bluetooth	2400 - 2483	same
GSM	890 - 915	925 - 960
UMTS TDD	1900 - 1920	same
	2010 - 2025	same
UMTS FDD	2110 - 2170	1920 - 1980

as an error signal in a negative feedback loop, which in turn controls the tail current of the oscillator.

6.2 SPECIFICATIONS

There are many different specifications for oscillators depending on the application. In this section, the most important and the most general oscillator specifications will be discussed.

6.2.1 Frequency and Tuning

The frequency and tuning range are important specifications which can have a significant impact on the level of difficulty when designing an oscillator. Table 6.1 shows the minimum and maximum frequencies of several telecom standards. Exact values may be slightly different, depending on the country in which the system is used. The tuning range is considered to be small compared to the center frequency in these standards, in other words, these are narrow range systems. A satellite receiver with a tuning range of at least 950 MHz to 2150 MHz is considered to be a wide range system, as the tuning range is more than an octave.

As for any specification, the tuning specification must be met under worst case conditions. Therefore frequency deviations due to temperature changes, process spread and power supply variations should be added to the tuning range. For an RC oscillator, the process spread can easily be 20% to 40%. In an LC oscillator the (planar) inductor tolerance is usually very good (e.g. 1%) because of the accurate lithography of the IC process. However, the varactor and fixed capacitances can vary up to 20% in value, and the center frequency can therefore change by more than 10%.
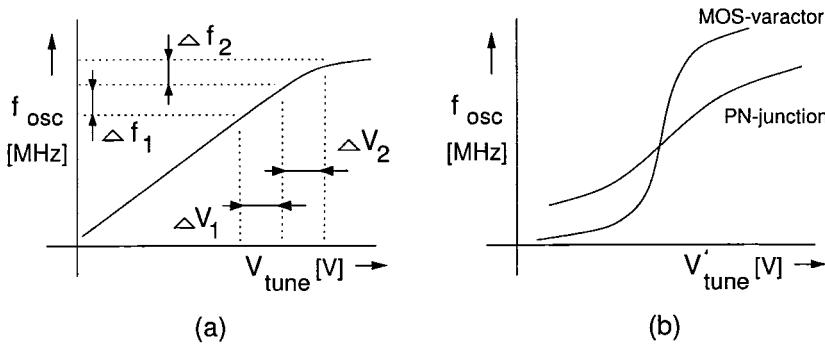
Figure 6.16. Typical tuning curve of an RC oscillator (a), and the tuning curve of an LC oscillator with MOS-varactor and PN-junction varactor (b).

6.2.2   Tuning Constant and Linearity

A VCO (or CCO) has a tuning constant K_{vco} which is specified in Hz/V (or K_{cco} in Hz/A). This constant is sometimes also referred to as VCO gain. For example, if the tuning range is 2.4 GHz to 2.483 GHz (Bluetooth standard) and the available tuning voltage range is 2.7 V, the tuning constant is about 30.7 MHz/V. This constant will become larger if the tuning range is increased, to cover process spread, for example. If the supply voltage decreases, for example in the case of a redesign using a newer CMOS process, the tuning voltage range will also decrease, and thus K_{vco} will increase. In the following chapter it will become clear why minimization of K_{vco} leads to a more robust PLL design.

The phrase "tuning constant" is actually somewhat deceptive, since it is rarely a constant. Figure 6.16(a) shows the typical tuning curve of the two-integrator oscillator from Figure 6.15. For a large part of the tuning range, K_{vco} is equal to $\Delta f_1 / \Delta V_1$. However, at the end of the tuning range, where parasitics start to dominate, K_{vco} decreases ($\Delta f_2 / \Delta V_2 < \Delta f_1 / \Delta V_1$).

Figure 6.16(b) shows the typical tuning characteristics of an oscillator tuned by a MOS-varactor and a PN-junction varactor, respectively. The derivatives of the tuning curves vary significantly, especially for the MOS-varactor. So, instead of specifying the gain constant using one figure, the deviation from this nominal value must also be specified, for example: $K_{vco} \pm 30\%$. In the following chapter on PLLs it will become clear that K_{vco} is an important parameter in PLL design, and that a constant value is preferable.

6.2.3   Power Dissipation

An oscillator is normally part of a larger system with a restricted power budget. Low power design is important, especially for portable applications. Minimum
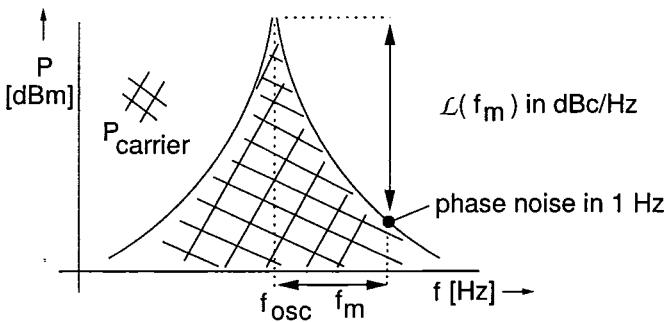
Figure 6.17. $\mathcal{L}(f_m)$ definition.

power dissipation is also important for applications which are connected to the mains, since the whole system has to fit into a cheap package with a certain thermal resistance. In such a case, low power means the possibility of a higher degree of integration. The power budget of an oscillator is specified in milliwatts, or by the available current given a supply voltage. The maximum supply voltage can be dictated by the application or the technology (e.g. breakdown voltages in CMOS). Additional current is needed to realize output buffers, which form the interface between the oscillator and cascaded blocks such as mixers and dividers.

6.2.4   Phase Noise to Carrier Ratio

A practical oscillator suffers from frequency instabilities. These frequency instabilities are called phase noise. Short term instability of an oscillator (frequency changes less than seconds) is characterized by the single-sideband (SSB) phase noise to carrier ratio $\mathcal{L}$ at offset frequency f_m, which is defined in dBs as:

$$\mathcal{L}(f_m) = 10 \log \left(\frac{\text{Phase noise density at } f_m \text{ in 1 Hz}}{P_{carrier}} \right) \tag{6.15}$$

Figure 6.17 illustrates the definition of $\mathcal{L}(f_m)$. Its units are commonly given in dBc/Hz, indicating that the phase noise is measured relative to the carrier and in a 1 Hz bandwidth. Alternatively, the phase noise of an oscillator can be characterized by the Carrier to Noise Ratio (CNR(f_m)), which is simply $-\mathcal{L}(f_m)$, with $\mathcal{L}(f_m)$ in dBc/Hz.

$\mathcal{L}(f_m)$ can be measured directly using a spectrum analyzer. The spectrum analyzer measures the power spectrum of the oscillator which includes not only the phase variations but also amplitude variations. However, the phase noise component is dominant in many practical oscillators, because the ampli-
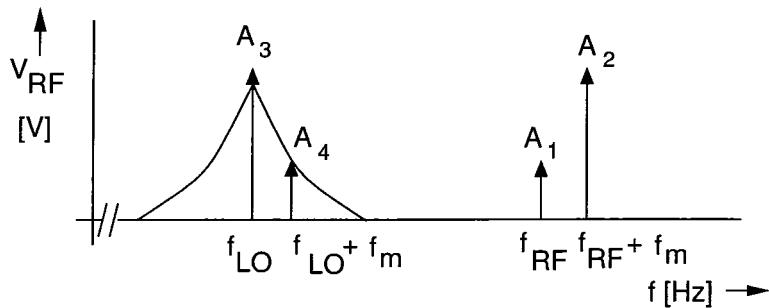
Figure 6.18. Reciprocal mixing: the wanted signal A1 and adjacent channel A2 are both converted to the same frequency.

tude component is reduced by limiting mechanisms. In the literature, $\mathcal{L}(f_m)$ is therefore commonly used with the assumption that it includes only the phase noise component. This assumption is adopted in this chapter. Several methods exists to unambiguously measure the phase noise of an oscillator [14]. These are utilized in dedicated commercial phase noise measurement equipment. Alternatively, $\mathcal{L}(f_m)$ of the oscillator under test can be measured with a limiter in front of a spectrum analyzer.

Phase noise deteriorates transceiver performance or, in other words, transceiver performance requirements will lead to a certain $\mathcal{L}(f_m)$ specification. In the following subsections, the effect of phase noise on transceiver performance will be discussed, and how the required phase noise to carrier ratio can be calculated.

Reciprocal Mixing

One mechanism which deteriorates the signal to noise ratio of a wanted channel is reciprocal mixing. This is illustrated in Figure 6.18. The wanted channel at f_{RF} has an amplitude of A_1. A strong adjacent channel is present at an offset f_m with amplitude A_2. The oscillator with amplitude A_3 has phase noise sidebands with an amplitude of A_4 at f_m. To simplify the discussion on reciprocal mixing, the phase noise sideband is replaced with a single tone interferer at $f_{LO} + f_m$. After mixing, it is not only the wanted signal with amplitude $\frac{1}{2}A_3 A_1$ that is converted to $f_{RF} - f_{LO}$ but also unwanted signal with amplitude $\frac{1}{2}A_4 A_2$ is also converted. This lowers the signal to noise ratio and hence A_4 should be sufficiently small to achieve a certain Bit Error Rate (BER) or signal-to-noise ratio (SNR).

In a GSM system, the in-band interferer levels are specified, thus allowing the maximum tolerable $\mathcal{L}(f_m)$ to be calculated. For example, with the frequency band from 600 kHz to 800 kHz the interferer level can be -43 dBm under

minimum wanted signal conditions. More information is needed to calculate the phase noise specification of the oscillator: the reference sensitivity (-105 dBm), minimum signal to noise ratio for the demodulator (12 dB) and effective noise bandwidth (180 kHz). The unwanted signal $\frac{1}{2}A_4A_2$ at the wanted frequency should be no larger than -105 dBm minus 12 dB. This means that compared to the -43 dBm interferer, the power difference is 74 dB in a 180 kHz bandwidth. Therefore, $\mathcal{L}$ (600 kHz) should be -74 dB minus 10 log(180 · 10³), which is -126.6 dBc/Hz.

Signal to Noise Degradation of FM Signals

In frequency and phase modulated signals, information is carried in the phase of the carrier signal. When this signal is down-converted in a receiver using an oscillator, the oscillator phase noise is superposed on the wanted signal and degrades the SNR. For example, consider an FM modulated signal $V_{RF} = V_{carrier}\cos[2\pi(f_{carrier} + \Delta f \cos(2\pi f_m t))t]$ with $f_{carrier}$ equal to 1 GHz, Δf is $50/\sqrt{2}$ kHz and modulation frequency f_m is 1 kHz. Assuming an oscillator phase noise spectrum which has a 6 dB per octave (or 20 dB per decade) slope, (6.16) relates the rms frequency deviation δf resulting from phase noise and $\mathcal{L}(f_m)$ in a noise bandwidth NBW [15].

$$\mathcal{L}(f_m) = -10\log\left(\frac{2}{\delta f^2(f_m)} f_m^2\, NBW\right) \tag{6.16}$$

If a system requires a signal-to-noise ratio of 80 dB after frequency conversion, a maximum δf of 3.5 Hz rms is allowed. This results in $\mathcal{L}(1kHz)$ equal to -89 dBc/Hz using (6.16) with a NBW equal to 5 kHz.

Spurious Emission

A transmitter of a telecom transceiver will emit power at unwanted frequencies in addition to the wanted output signal, due to the phase noise of the used oscillator. This is why maximum permitted emissions are specified in telecom standards. For example, in the type-approval template of a GSM system, the spurious emission at 20 MHz offset (GSM RX-band) from the 915 MHz carrier (worst case location in TX-band) must be -79 dBm down in the 100 kHz bandwidth relative to the +33 dBm carrier. This means that $\mathcal{L}(20MHz)$ must be better than -162 dBc/Hz[3].

---

[3]Filtering in the power amplifier and in the duplex filter of a transceiver may ease the oscillator specification.

6.2.5   Harmonics

In practice, even a harmonic oscillator will generate more frequencies in addition to the wanted fundamental frequency. Due to implementation aspects, there will always be some power at harmonics of the fundamental frequency. These harmonics can be specified relative to the carrier in dBc. In many cases, nonlinear operation of an oscillator improves the phase noise, but also generates harmonics. When an oscillator is used to drive mixers, these harmonics may improve the switching of the mixer and thus the mixer noise figure. However, unwanted channel power or noise power located at harmonics may fall into the wanted channel when the oscillator-mixer combination is used for frequency conversion. Care must be taken to ensure that the conversion products of the oscillator signal are sufficiently low relative to the wanted channel which was mixed with the fundamental. Second order harmonics can be suppressed to some extent (typically 40 dB) by a balanced design.

6.2.6   I/Q Matching

Many receiver architectures which allow a high degree of integration require quadrature signals. An example is the zero-IF architecture. In this architecture, the RF signal is mixed with an I/Q oscillator signal and the resulting baseband signals are processed further in the in-phase and quadrature branches. Quadrature errors (both in phase and amplitude) in the branches or in the quadrature relation of the I/Q oscillator will reduce the SNR at the output of a receiver front-end.

Consider the simplified zero-IF receiver from Figure 6.3, for example. The in-phase signal provided by the tuning system can be written as $\cos(\omega_{osc}t)$. When allocating all imperfections to the quadrature signal, this signal can be expressed as $(1 + A_e)\sin(\omega_{osc}t + \phi_e)$. The amplitude error is A_e in this expression, and the phase error ϕ_e is in radians. The resulting SNR in dB due to non-zero A_e and ϕ_e is

$$\frac{S}{N} = 10\log\left(\frac{4}{((1+A_e)\cos(\phi_e)-1)^2 + ((1+A_e)\sin(\phi_e))^2}\right) \qquad (6.17)$$

Figure 6.19 shows the allowable amplitude error and phase error for 50, 40 and 30 dB SNR.

Equation (6.17) can be approximated by (6.18) for small A_e and small ϕ_e. The inverse of this approximation also describes the image rejection ratio (IRR) of image-reject architectures, such as the Hartley architecture [16], as a result
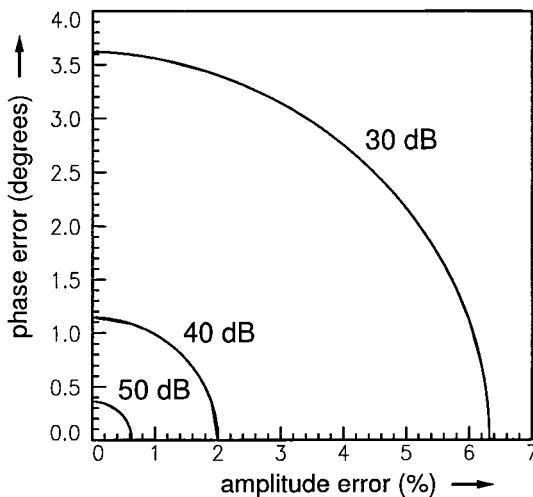
Figure 6.19. SNR resulting from imperfect quadrature signals of the local oscillator in a Zero-IF architecture.

of imperfect quadrature signals.

$$\frac{S}{N} = 10 \log \left(\frac{4}{A_e^2 + \phi_e^2} \right) \qquad (6.18)$$

Quadrature signals can be generated by using multi-stage oscillators (RC or LC) with an even number of stages. An example is the two-integrator oscillator from Figure 6.15. Such an oscillator provides "correct-by-construction" I/Q signals. The symmetry dictates a 90° phase shift per section. Practical I/Q matching is limited by the device matching and the symmetry of the layout.

6.2.7 Technology and Chip Area

The technology plays an important role in the performance and cost of an oscillator. For example, the inductor quality is much better on high-ohmic substrates compared to low-ohmic substrates. It is therefore much easier to realize high-performance LC oscillators on high-ohmic substrates. The availability of good varactors is also of importance for sufficient tuning range and low phase noise. Technology parameters (see Chapter 1) such as f_T, f_{max} and transistor parasitics like the collector (or drain) capacitances give an indication of how big the influence of parasitics will be on the LC or RC oscillator design, at a certain target oscillation frequency. If chip area minimization is important, then the use of a large number of coils should be avoided. RC oscillators are usually much more compact than LC oscillators, but are also much noisier and consume more power.
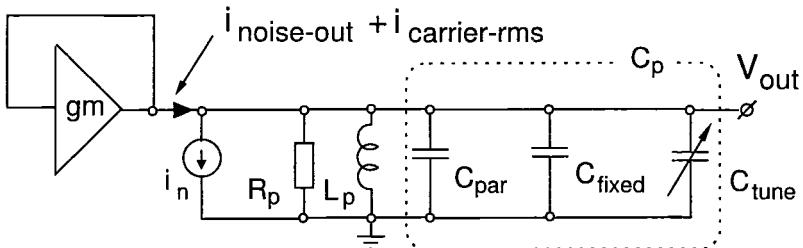
Figure 6.20.   LC oscillator model tunable with varactor C_{tune}.

6.3   LC OSCILLATORS

Many properties of LC oscillators can be investigated using a feedback model. This section starts with a linear feedback model of an LC oscillator which will be used to gain insight into important LC oscillator properties. These properties are frequency, tuning range and phase noise to carrier ratio. Nonlinear modeling is needed to understand all phase noise generation mechanisms. Although an in-depth quantitative analysis of phase noise inducing mechanisms is beyond the scope of this chapter, most important theories will be outlined briefly. The final part of this section highlights some LC oscillator topologies.

6.3.1   Frequency, Tuning and Phase Noise

Figure 6.20 shows the linear LC oscillator model which will be used to investigate LC oscillator properties. All noise sources are modeled by noise current source i_n. Varactor C_{tune}, fixed capacitance C_{fixed} and parasitic capacitance C_{par} together form the total tank capacitance C_p. The output noise current is $i_{noise-out}$ and the rms carrier current is denoted by $i_{carrier-rms}$. The carrier current is shown in the small-signal model for clarity, but it can only be determined assuming one of the amplitude stabilization mechanisms described in Section 6.1.5. For example, if transconductance g_m is implemented with a transistor, self-limiting will determine the amplitude of $i_{carrier-rms}$.

In Figure 6.20, all losses are lumped in resistor R_p. This includes the series resistance of practical coils and capacitors (see also Chapter 1) as shown in Figure 6.21(a).

Using the component quality factors of an inductor $Q_{ls} = \omega L_s / R_{ls}$ and of a capacitor $Q_{cs} = 1/(\omega C_s R_{cs})$, R_p, L_p and C_p can be expressed with the circuit elements from Figure 6.21(a) as denoted in (6.21), (6.19) and (6.20), respectively.
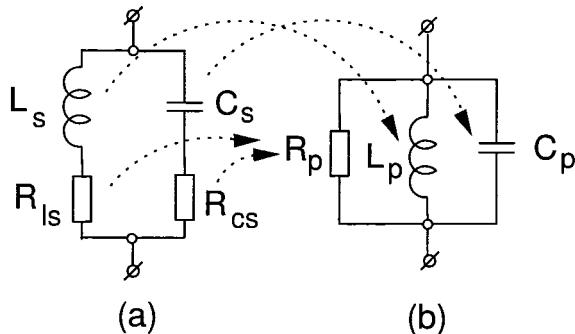
Figure 6.21. A simple model of a practical resonator (a) and the resonator with losses lumped in R_p (b).

$$R_p = \frac{R_{cs}\left(1 + Q_{cs}^2\right) R_{ls}\left(1 + Q_{ls}^2\right)}{R_{cs}\left(1 + Q_{cs}^2\right) + R_{ls}\left(1 + Q_{ls}^2\right)}$$

$$L_p = L_s \frac{1 + Q_{ls}^2}{Q_{ls}^2} \qquad (6.19)$$

$$C_p = C_s \frac{Q_{cs}^2}{1 + Q_{cs}^2} \qquad (6.20)$$

Note that for high component quality factors Q_{ls} and Q_{Cs}, L_s and C_s are accurately represented by L_p and C_p, respectively.

Frequency

The oscillation frequency of the behavioral model in Figure 6.20 is equal to:

$$\omega_{osc} = \frac{1}{\sqrt{L_p \left(C_{tune} + C_{fixed} + C_{par}\right)}} \qquad (6.21)$$

Both L_p and C_p (the sum of all three capacitances) can be used to tune the oscillator. Inductor L_p is usually fixed. C_{fixed} is an intentionally added capacitor, to improve the quality factor of C_{tune}, for example (fixed capacitance normally has a better quality factor than a varactor). If this capacitor and C_{tune} are made zero, the maximum frequency $1/\sqrt{L_p\, C_{par}}$ is reached. C_{par} consists of several contributions: the active oscillator part, parasitic capacitance in the resonator (for example, from the inductor), capacitance of cascaded circuits and
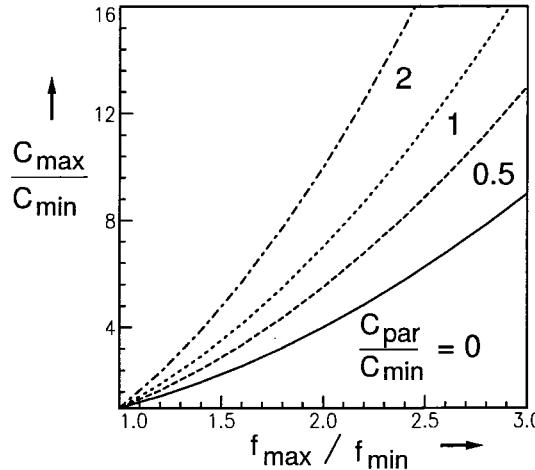
Figure 6.22. $Ratio_{var}$ versus f_{max} over f_{min} for four values of C_{par} over C_{min}.

inter-connect capacitance. Clearly, L_p and C_{par} must be made small for a high maximum frequency.

The frequency for the practical tank circuit in Figure 6.21(a) is equal to:

$$\omega_{osc} = \frac{\sqrt{L_s - C_s R_{ls}^2}}{\sqrt{C_s L_s} \sqrt{L_s - C_s R_{cs}^2}} \tag{6.22}$$

For high quality factors, this reduces with reasonable accuracy to $1/\sqrt{C_s L_s}$.

Tuning

In general, tuning of LC oscillators can be performed by varying L_p or C_{tune} in (6.21). Discrete tuning can be accomplished by switching between different inductors [17]. In practice , L_p is a (balanced) coil on the silicon die, with a fixed value and tuning is realized by changing C_{tune}. The maximum capacitance C_{max} to minimum capacitance C_{min} ratio of varactor C_{tune} can be expressed as a function of maximum and minimum frequency (f_{max} and f_{min}, respectively) and C_{par} and C_{min}. The formula for this ratio is shown in (6.23). If C_{par} is zero, $Ratio_{var}$ is simply the squared ratio of f_{max} over f_{min}. However, if C_{par} is substantial compared to C_{min}, the required $Ratio_{var}$ to realize the required tuning range increases.

$$Ratio_{var} = \frac{C_{max}}{C_{min}} = \left(\frac{f_{max}}{f_{min}}\right)^2 + \frac{C_{par}}{C_{min}}\left(\left(\frac{f_{max}}{f_{min}}\right)^2 - 1\right) \tag{6.23}$$

Equation (6.23) shows that at high frequencies, where parasitics start to dominate, a higher $Ratio_{var}$ is needed to realize a certain tuning range compared to low frequencies. This trend is shown graphically in Figure 6.22.

Typical $Ratio_{var}$ values are smaller than two for a PN-junction type (e.g. one junction of a transistor) and smaller than three for a MOS-type varactor (MOSFET with source and drain connected). In practice, due to the presence of C_{par}, a tuning range of more than 30% is difficult to achieve. If more tuning range is required of an LC oscillator, one possibility is to switch between different oscillators at the expense of more chip area. A second possibility for tuning range extension is the use of several bands defined by switched capacitors, and using a continuous varactor to reach the frequencies within one band.

Phase Noise to Carrier Ratio

Linear Modeling: Single-Phase LC Oscillators

One of the first phase noise models for oscillators was proposed by Leeson in 1966 [18]. Leeson's formula includes many characteristics of real oscillators, such as the 6 dB per octave decay of the phase noise sidebands close to the carrier. Since no formal proof was given by Leeson, and his formula includes a noise figure as "fit"-factor, this formula is generally considered to be heuristic. Since Leeson, many publications and phase noise theories have brought the understanding of phase noise generation mechanisms to a mature level. Nevertheless, many of the insights provided by recent phase noise theories, which include nonlinearity and time-varying aspects, can be calculated and explained using a linear and non time-varying oscillator model. Hence, the phase noise to carrier ratio of LC oscillators will be calculated using a linear model followed by a discussion of neglected nonlinear effects.

In general, the phase noise sidebands of oscillators can be divided into three regions with different slopes as indicated in Figure 6.23. The phase noise with a 6 dB/octave slope will be considered first. This region, with a $1/f^2$ slope is normally the biggest part of the oscillator sideband. The $1/f$-noise part with a $1/f^3$ slope will be explained in the nonlinear modeling part of this section. The white noise floor usually arises from circuits, such as buffers which are connected to the oscillator.

The oscillator in Figure 6.20 can be modeled with a linear feedback model. The transfer function around oscillation frequency ω_{osc} of a general feedback system shown earlier in Figure 6.7, can be approximated by (6.24), assuming oscillation conditions are met and the offset frequency $\Delta\omega$ is much smaller than ω_{osc} [19].
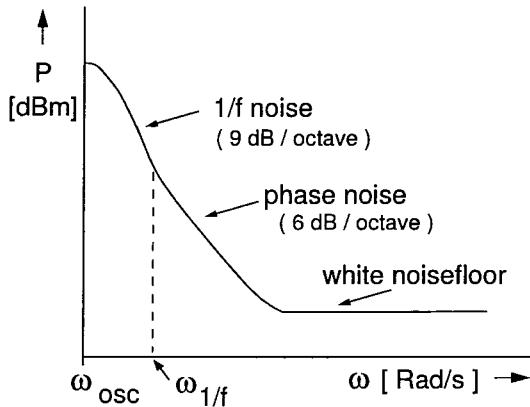
Figure 6.23.   The three main phase noise regions of an oscillator sideband.

$$\left| \frac{Y_{out}}{X_{in}}(j(\omega_{osc}+\Delta\omega)) \right|^2 = \frac{1}{4\,Q^2}\left(\frac{\omega_{osc}}{\Delta\omega}\right)^2 \tag{6.24}$$

in which Q is defined as:

$$Q = \frac{\omega_{osc}}{2}\sqrt{\left(\frac{\delta A}{\delta\omega}\right)^2 + \left(\frac{\delta\phi}{\delta\omega}\right)^2} \tag{6.25}$$

with $A = |H_f(j\omega)|$ and $\phi = \arg(H_f(j\omega))$.

Since $\delta A/\delta\omega$ is zero for the LC resonator in Figure 6.20 at resonance, the quality factor Q_{LC} of a single phase LC oscillator can also be expressed as:

$$Q_{LC} = \frac{\omega_{osc}}{2}\left|\frac{\delta\phi}{\delta\omega}\right| = R_p\sqrt{\frac{C_p}{L_p}} \tag{6.26}$$

The first expression for Q_{LC} with $\delta\phi/\delta\omega$ is particularly interesting, because it provides intuitive insight into phase stability. Consider the example when the phase condition for oscillation in Figure 6.20 is momentarily not met due to some perturbation, for example due to noise source i_n. If $\delta\phi/\delta\omega$ is large, then the frequency change required in the oscillator need only be very small in order to once again comply with the phase condition. Therefore, $\delta\phi/\delta\omega$ is a direct measure of the phase stability. Equation (6.26) can also be derived from the classical quality factor definition in (6.27).

$$Q = \frac{\omega \cdot \text{Maximum stored energy in one period}}{\text{Dissipated power}} \tag{6.27}$$

Knowing the stored energy in Figure 6.20, which is $C_p V_{out}^2$, with V_{out} in rms voltage and the dissipated energy V_{out}^2/R_p, the second expression for Q_{LC} in (6.26) can be derived using (6.27). Since the dissipated energy in an oscillator must somehow be restored by noisy energy from the active part, it is clear that the quality factor of an oscillator is an important parameter and should be maximized.

Now the quality factor Q_{LC} has been defined, we can apply (6.24) directly to get the squared output noise current $\overline{i_{noise-out}^2}$ (Y_{out}^2) of the LC oscillator in Figure 6.20 resulting from $\overline{i_n^2}$ (X_{in}^2):

$$\overline{i_{noise-out}^2} = \frac{1}{4\,Q_{LC}^2} \cdot \left(\frac{f_{osc}}{f_m}\right)^2 \cdot \overline{i_n^2} \qquad (6.28)$$

Equation (6.28) shows that noise current $\overline{i_n^2}$ is shaped by the transfer function of the oscillator.

In order to derive the phase noise to carrier ratio $\mathcal{L}(f_m)$, $\overline{i_{noise-out}^2}$ needs to be divided by the squared signal current $i_{carrier-rms}^2$. An extra factor of $1/2$ also has to be included, since $\mathcal{L}(f_m)$ only takes into account the phase noise at the output, whereas $\overline{i_{noise-out}^2}$ includes both phase and amplitude noise [15]. The resulting expression is given in (6.29).

$$\mathcal{L}(f_m) = \frac{1}{2} \cdot \frac{1}{4\,Q_{LC}^2} \cdot \left(\frac{f_{osc}}{f_m}\right)^2 \cdot \frac{\overline{i_n^2}}{i_{carrier-rms}^2} \qquad (6.29)$$

At this point, it is instructive to see whether we can rewrite (6.29) into Leeson's equation for the phase noise of an oscillator [18]. Indeed if the noise factor F is defined as $\overline{i_n^2}/(4kT/R_p)$ and P_{RF} as $i_{carrier-rms}^2 \cdot R_p$, Leeson's equation appears:

$$\mathcal{L}(f_m) = \frac{1}{2} \cdot \frac{1}{Q_{LC}^2} \cdot \left(\frac{f_{osc}}{f_m}\right)^2 \cdot \frac{FkT}{P_{RF}} \qquad (6.30)$$

where k is Boltzmann's constant and T is the absolute temperature in Kelvin.

The terms $\overline{i_n^2}$ and $i_{carrier-rms}^2$ in (6.29) can be expanded if we assume an implementation of the transconductance g_m (the active oscillator part) in Figure 6.20. Since a cross-coupled differential pair is often used as the active oscillator part, this circuit is used as an example. The cross-coupled differential pair has a tail current I_{tail} and a total transconductance of $-g_{mQ}/2$, as shown in Figure 6.24.
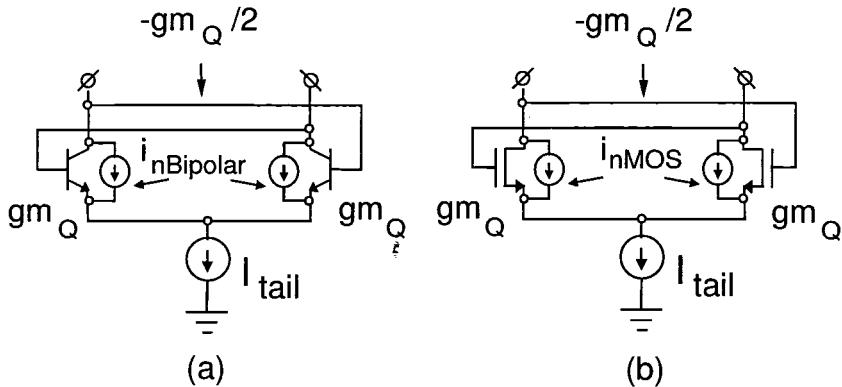
Figure 6.24.   A cross-coupled bipolar (a) and MOS differential pair (b).

$$\overline{i_n^2} \approx \frac{\overline{i_{n\,Bipolar}^2}}{2} + \frac{4\,k\,T}{R_p} = \frac{4\,k\,T}{R_p}(\frac{1}{2}\alpha + 1) \qquad (6.31)$$

$$\overline{i_n^2} \approx \frac{\overline{i_{n\,MOS}^2}}{2} + \frac{4\,k\,T}{R_p} = \frac{4\,k\,T}{R_p}(\frac{2}{3}\alpha + 1) \qquad (6.32)$$

In (6.31) and (6.32), $\overline{i_n^2}$, the total noise current in the LC oscillator, is expanded for a bipolar and a MOS cross-coupled differential pair, respectively. In order to assure start-up of the oscillator, the transconductance $-g_{mQ}/2$ is always larger than the value required to compensate all losses. This is taken into account in (6.31) and (6.32) by defining $g_{m-startup} = 1/R_p$ and $g_m = g_{mQ}/2 = \alpha \cdot g_{m-startup}$. The factor α is one or larger and, in practice, in the range of two to five for self-limiting oscillators. Alternatively, α can be close to unity when an AGC-loop is used to stabilize the oscillator amplitude. Parameter α is in fact the open loop gain of the oscillator.

Equation (6.31) is an approximation. Assuming that the bipolar transistor can be scaled to make the base resistance noise small compared to the collector shot noise and $\beta \gg 1$, the noise generated by the base resistance and base current can be neglected. The factor 2/3 in (6.32) is only accurate for long-channel devices and can be significantly higher in short-channel devices [20]. Furthermore, (6.32) only holds when the differential pair is operating in the strong inversion region [21]. This equation is also an approximation since noise sources such as gate and bulk resistive noise are neglected.

As can be seen from (6.31) and (6.32), the noise sources of a bipolar and MOS cross-coupled LC oscillator differ only by a small factor when using

first order modeling. In the following formulas, only the MOS cross-coupled pair will be considered, but adaptation to a bipolar oscillator implementation is straightforward.

Equation (6.32) can now be substituted in (6.29). The signal current in the oscillator, $i_{carrier-rms}$, can also be replaced. The maximum peak output current of the differential pair is $4/\pi \cdot I_{tail}$. Note that this is $2I_{tail}$ for a Colpitts oscillator [22]. The differential pair will no longer switch at high frequencies, and the current is better approximated by I_{tail}. This value will be used in the following expressions. The following first order formula for the phase noise to carrier ratio in an LC oscillator with MOS cross-coupled differential pair is now obtained:

$$\mathcal{L}(f_m)_{MOS} = \frac{1}{2} \cdot \frac{1}{Q_{LC}^2} \cdot \left(\frac{f_{osc}}{f_m}\right)^2 \cdot \frac{k\,T(\frac{2}{3}\alpha + 1)}{R_p I_{tail}^2} \tag{6.33}$$

Alternatively, η_{LC}, the efficiency of converting DC power to RF power for an LC oscillator, can be introduced:

$$\eta_{LC} = \frac{P_{RF}}{P_{DC}} = \frac{R_p\,I_{carrier-rms}^2}{I_{tail}\,VCC} \tag{6.34}$$

Therefore, $\mathcal{L}(f_m)_{MOS}$ can also be denoted as:

$$\mathcal{L}(f_m)_{MOS} = \frac{1}{2} \cdot \frac{1}{Q_{LC}^2} \cdot \left(\frac{f_{osc}}{f_m}\right)^2 \cdot \frac{k\,T(\frac{2}{3}\alpha + 1)}{\eta_{LC} P_{DC}} \tag{6.35}$$

After several pages of phase noise calculations, it is time for some conclusions and remarks:

- Equation (6.35) models the 6 dB per octave slope (proportional to $1/f^2$) of an oscillator. This slope continues for large offset frequencies until it hits the white noise floor of the physical resistors R_{ls} and R_{cs} in Figure 6.21. In practice, the white noise floor is determined by buffers cascaded after the oscillator. Hence, the point where the oscillator sideband hits the white noise floor is dependent on the buffer design. The measurement equipment used to characterize an oscillator can also introduce a dominant white noise floor.

- Assuming the offset frequency of an oscillator is in the range where the oscillator sideband has a $1/f^2$ slope, extrapolation is possible. For example, if the $\mathcal{L}(10\text{kHz})$ is -80 dBc/Hz, $\mathcal{L}(20\text{kHz})$ will be -86 dBc/Hz and $\mathcal{L}(2\text{MHz})$ will be -126 dBc/Hz.

- From (6.33) it is clear that a doubling of the Q_{LC} in principle gives 9 dB improvement in $\mathcal{L}(f_m)$. The term Q_{LC}^2 contributes 6 dB, but R_p is also doubled, which accounts for 3 dB improvement. In practice, the improvement can also be 6 dB as indicated by (6.29). If $\overline{i_n^2}$ is dominated by device noise, the reduction by a factor two of the tank noise source $4kT/Rp$ caused by the doubling of R_p will be negligible, and $\mathcal{L}(f_m)$ will improve by 6 dB if the quality factor is doubled.

- As one would intuitively expect, the phase noise becomes 3 dB lower if the power is doubled (see (6.35)). Conceptually this can be understood by putting two identical oscillators in parallel. In such a case, properties like tuning range and center frequency remain unchanged, but the power is doubled. Since the signal will add voltage-wise and the noise power-wise, a 3 dB improvement in signal-to-noise is obtained.

- From the latter two conclusions and (6.35), it is clear that the quality factor and the signal power should be maximized for a minimum $\mathcal{L}(f_m)$. This means that both the inductor and varactor should be optimized for minimal losses.

- In practice, the quality factor Q_{LC} is the loaded quality factor. The unloaded quality factor is the quality factor of the tank circuit as defined in (6.26). In an oscillator circuit, the resonator is "loaded" by the impedances of the active oscillator part (such as the input impedance) and by any buffer connected to the oscillator. This loading can be taken into account by incorporating the resistive losses into loss resistance R_p, thus forming the loaded quality factor Q_{LC}. The capacitive part of circuits loading the tank effectively adds to capacitor C_p of the resonator.

Linear Modeling: Multi-Phase LC Oscillators

It was pointed out earlier that many modern transceiver architectures require quadrature signals. Some architectures may even require more than two phases. There are several methods of generating quadrature signals. An LC oscillator can be used together with a divider. If the oscillator has a 50% duty cycle, quadrature signals are obtained. In this case, the oscillator operates at a frequency of twice the wanted output frequency. Poly-phase networks are another means of generating quadrature signals. For example, an RC-network can be used together with limiters to obtain reasonable phase and gain matching. For good matching, the input signal to the poly-phase network should have low harmonic distortion. A third possibility is the use of even-stage RC oscillators. As will become clear in the next section on RC oscillators, LC oscillators have superior phase noise performance compared to RC oscillators. For this reason,
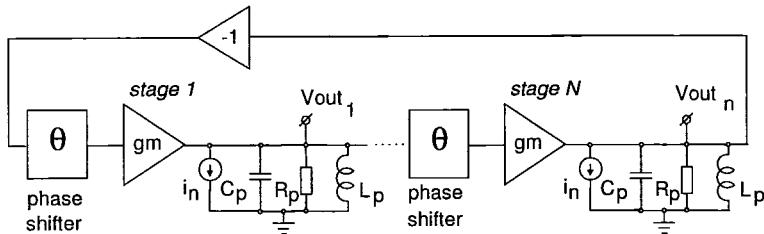
Figure 6.25. Multi-phase LC oscillator model.

if multi-phase signals are needed in combination with a low phase noise to carrier ratio, the use of multi-phase LC oscillators can be the best option for a transceiver. In the paragraphs that follow the phase noise of a multi-phase LC oscillator will be analyzed using linear modeling.

Figure 6.25 shows the multi-phase oscillator model which will be used to calculate $\mathcal{L}(f_m)$. The model is constructed using N identical LC oscillators. A phase shifter with phase shift θ is included in each section in addition to a single phase oscillator. Phase shift θ can be zero, which means that the block can be omitted. However, in the following paragraphs it will become clear that it is best to make θ deliberately non-zero.

The quality factor for a multi-phase oscillator can be derived using the general quality factor definition from (6.25):

$$Q_{LC_{multi-phase}}(\phi) \approx N \cdot Q_{LC} \cdot \cos(\phi) \tag{6.36}$$

where ϕ is the LC resonator phase shift, N is the number of LC oscillator stages and Q_{LC} is the quality factor of the LC resonator in a stage as defined in (6.26). Equation (6.36) is an *approximation* of a more complex expression, but has an error less than 1% for ϕ ranging from 0 to 70° [23]. At 80° and 89° the error is 3.8% and 70%, respectively. Equation (6.36) reaches its maximum when the phase shift ϕ of the resonator in each stage is zero. In that case, the maximum quality factor of an N-stage LC oscillator is:

$$Q_{LC_{multi-phase-max}} = N \cdot Q_{LC} \tag{6.37}$$

Intuitively this result could be expected, since an LC resonator has its maximum phase slope $\delta\phi/\delta\omega$ at zero phase shift. Now the importance of the phase shifter with phase shift θ in Figure 6.25 becomes clear. The quality factor of an N-stage multi-phase LC oscillator can be maximized by incorporating a phase shift θ in each stage, with θ equal to $\pm 180°/N$. The value $\pm 180°/N$ follows from the phase condition for oscillation. The phase shift θ can be deliberately implemented by a phase shifter, but can also be partly or completely

implemented by parasitic phase shift of active devices, especially at high fre-
quencies.

Applying (6.24) and making the same assumptions as were made for the sin-
gle phase LC oscillator, the phase noise of a multi-phase oscillator implemented
by MOS differential pairs can be calculated:

$$\mathcal{L}(f_m)_{MOS} = \frac{1}{2} \cdot \frac{1}{N^2 Q_{LC}^2 \cos(\phi)^2} \cdot \left(\frac{f_{osc_\phi}}{f_m}\right)^2 \cdot \frac{N\,k\,T(\frac{2}{3}\alpha + \cos(\phi))}{\eta_{LC}\,\cos(\phi)\,\frac{P_{DC}}{N}}$$

$$(6.38)$$

with:

$$f_{osc_\phi} = \frac{-\tan(\phi) + \sqrt{4\,Q_{LC}^2 + \tan^2(\phi)}}{2Q_{LC}} \cdot \frac{1}{2\,\pi\,\sqrt{C_p\,L_p}}$$

$$(6.39)$$

In addition to the conclusions regarding phase noise in a single phase LC
oscillator, the following remarks and conclusions can be made for a multi-phase
LC oscillator:

- From the quality factor definition in (6.36) and from (6.38), it follows that
  the resonator phase shift must be minimized for a minimum $\mathcal{L}(f_m)$. This can
  be achieved by making the phase shift θ equal to $\pm 180°/N$ for an N-stage
  LC oscillator. Note that (6.38) is also valid for $N = 1$. Equation (6.38) with
  $N = 1$ is identical to (6.35) if the resonator phase shift ϕ is zero. It follows
  that for a single phase oscillator, the phase shift of the active part should be
  minimized in order to minimize $\mathcal{L}(f_m)$.

- Figure 6.26(a) shows the $Q_{LC_{multi-phase}}$ as a function of the resonator phase
  shift ϕ and the related degradation in $\mathcal{L}(f_m)$ for a quadrature LC oscillator
  $(N = 2)$. As mentioned in the discussion of the phase noise of a single phase
  oscillator, $\mathcal{L}(f_m)$ can be proportional to $1/Q_{LC}^2$ or $1/Q_{LC}^3$. Both cases are
  plotted in Figure 6.26(b). At around 45° and less, the degradation is still
  smaller than 5 dB. For large phase shifts, the degradation amounts to more
  than 25 dB. Hence, the phase shift θ in Figure 6.25, which takes care of
  the required phase shift per section $(\pm 180°/N)$ to obey the phase condition
  for oscillation, is needed for optimal performance. At high frequencies,
  where the parasitic phase shift of the active devices is relatively large, part
  of the required phase shift per section can be already implemented by these
  devices.

- When $\mathcal{L}(f_m)$ is normalized for total power dissipation, $\mathcal{L}(f_m)$ is independent
  of the number of stages. This is made explicit in (6.38) which takes the total
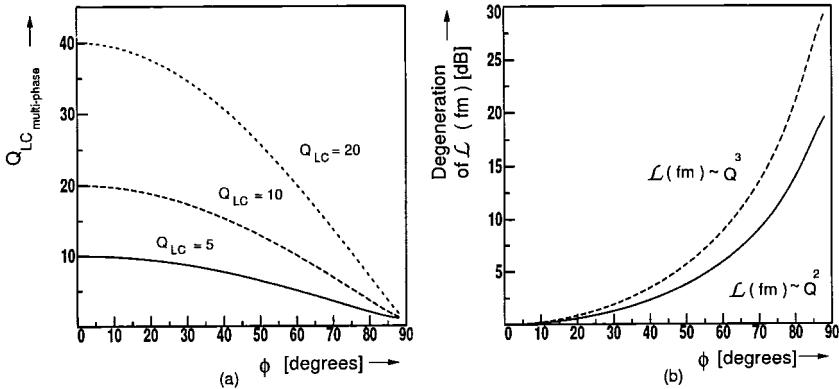
Figure 6.26. Effective quality factor, $Q_{LC_{multi-phase}}$ versus the resonator phase shift, for Q_{LC} is 5,10 and 20 and $N = 2$ (a). On the right, the degeneration of the $\mathcal{L}(f_m)$ versus the resonator phase shift, for $N = 2$ and $Q_{LC} = 5$ (b).

power P_{DC} into account. The number of stages N falls out because N^2 cancels out in the numerator and the denominator.

■ In the behavioral model of Figure 6.25, an additional amplifier with positive feedback can be introduced in each section to partly compensate loss resistance R_p. This provides a control input which, together with transconductance g_m, can be used to control the oscillator output voltage. This new behavioral model is readily obtained (e.g. for two stages) by coupling two single phase oscillators with two additional transconductances. The derived quality factor $Q_{LC_{multi-phase}}$ and conclusions for optimum coupling are also valid for this architecture.

■ The phase shifters with phase shift $\theta = \pm 180°/N$ lower the risk of spurious oscillation in a multi-phase LC oscillator. If the LC resonators in each stage are not operating close to zero phase shift, more than one oscillation mode may exist which can cause multi-oscillation and can hamper the operation of the circuit. If and when this phenomenon occurs depends on the value of the coupling current and the degree of nonlinearity in the multi-phase LC oscillator.

Nonlinear Modeling

At very high frequencies, relative to the cut-off frequency of a process an LC oscillator can operate in the linear region. In general, however, most active devices in the oscillator core operate in the (strongly) nonlinear region. In principle, all conclusions outlined using linear modeling are still valid, but the results cannot be used for accurate quantitative predictions concerning $\mathcal{L}(f_m)$.
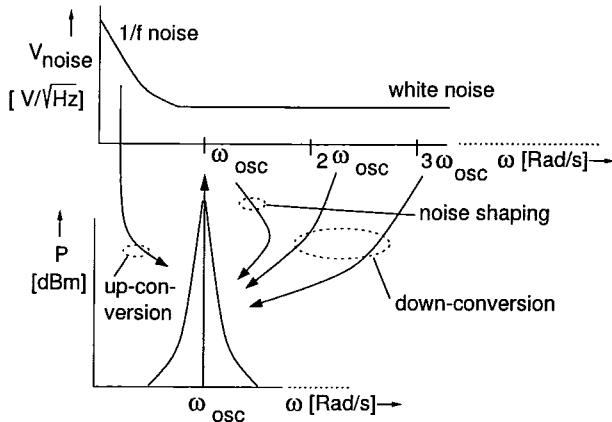
Figure 6.27.   Up and down-converted noise add to the total $\mathcal{L}(f_m)$.

Additional nonlinear phase noise mechanisms can increase the phase noise significantly. On the other hand, some nonlinear effects, like modulation of noise sources, can reduces the phase noise. The most important mechanisms and considerations are the following.

- **Up-conversion of noise:** The $1/f$-noise of the devices in the oscillator core contribute to $\mathcal{L}(f_m)$ via up-conversion. This is illustrated in Figure 6.27. For example, in the case of an LC oscillator with a cross-coupled pair, the transistors may be fully switching on and off and will mix the low frequency noise to ω_{osc} and higher harmonics. As discussed later, the $1/f$-noise corner of the devices in the oscillator does not have to coincide with the $1/f$-noise corner $\omega_{1/f}$ of the oscillator spectrum (see Figure 6.23).

- **AM-PM conversion:**   The easiest way to visualize this mechanism is to consider an LC oscillator with tail current I_{tail}. The transistor which implements the tail current source has $1/f$-noise. In the first instance, this $1/f$-noise only causes AM-modulation. However, nonlinear junction capacitance and especially varactors with high tuning slopes convert the AM noise into PM noise. The $\mathcal{L}(f_m)$ as a result of AM to PM conversion can be expressed using modulation theory:

$$\mathcal{L}(f_m) = 10 \log \left(\frac{K_{tail-rms}^2 \; \overline{i_{n-tail}^2}}{(2 \; f_m)^2} \right) \tag{6.40}$$

in which $K_{tail-rms}$ is the sensitivity of the frequency for tail current variations in Hz per A rms, and $\overline{i_{n-tail}^2}$ is the noise of the tail current source in $A/\sqrt{Hz}$.

Every oscillator control or power supply node can be modeled with a certain (parasitic) K_{vco} or K_{cco}. This constant can be used in (6.40) to calculate the maximum noise that can be allowed (on the power supply or tuning input, for example) given a certain $\mathcal{L}(f_m)$ specification. If AM to PM conversion is dominant, the contribution of this effect can be lowered by adding fixed linear capacitance and lowering L or the varactor capacitance. Lowering the varactor capacitance will also lower the available tuning range. Alternatively, the tuning range can be divided into multiple bands which lowers the varactor slope within one band.

- **Down-conversion of noise:** Similar to up-conversion the noise at higher harmonics of ω_{osc} contributes to the total phase noise at ω_{osc} due to down-conversion or inter-modulation between the carrier and the wide band noise at multiples of the oscillation frequency (Figure 6.27). This mechanism is also referred to as noise folding. The number of harmonics that should be taken into account for an accurate estimation of this mechanism to the total $\mathcal{L}(f_m)$ depends on the bandwidth of the active devices and the frequency of operation. In practice, if noise contributions up to the fifth harmonic are included, results are sufficiently accurate.

  Tail current noise in LC oscillators at $2\omega_{osc}$ also contributes to the phase noise at ω_{osc} [24]. An useful technique to reduce this contribution is to filter the tail current noise at $2\omega_{osc}$ by connecting the output of the tail current source via a series inductor to the VCO core, and at the same time with a capacitor to ground [25, 26]. In this way, the noise has a low-ohmic path to ground and a high-ohmic path to the oscillator core without disturbing the DC conditions.

- **Modulation of noise sources:** In contrast to noise generated by resistors, which is stationary, the channel noise of a MOS transistor and the shot noise (base and collector) of a bipolar transistor are cyclo-stationary. The amplitude of the noise of the latter two sources will vary, since the current through devices varies significantly every oscillation period. This must be taken into account for accurate $\mathcal{L}(f_m)$ predictions.

- **Saturation effects:** The oscillator designer usually tries to maximize the carrier level in order to maximize $\mathcal{L}(f_m)$. This results in a large voltage swing across the varactors and active devices in the oscillator. Phase noise performance can degrade by many dBs if the oscillator enters the voltage limited region during the oscillation period, or, in other words, if the oscillation amplitude is no longer tail current limited. One possible cause can be junctions in a bipolar (e.g. base-collector junction) or MOS transistor

which are no longer reversed biased and which start to contribute additional shot noise. In addition, if tuning diodes are used, these should be used in the reverse region only in order to minimize leakage currents and noise contribution.

- **Carrier level:** If a self-limiting oscillator is used, the amplitude is determined by nonlinear mechanisms, and can only be calculated by taking the nonlinearities of the active devices in the VCO core into account. Alternatively, the carrier can be simulated using transient analysis. Since the rms carrier level is a parameter in $\mathcal{L}(f_m)$, it must be determined accurately for predictions.

There are several theories which take into account the discussed nonlinear phase noise mechanisms, and allow accurate quantitative predictions of $\mathcal{L}(f_m)$ of oscillators working (strongly) in the nonlinear region. A detailed discussion is beyond the scope of this chapter. Instead, a brief overview and some references will be given.

A general theory on phase noise in electrical oscillators is described in [27, 28, 29]. It takes the time-varying nature of oscillators into account and also accommodates cyclo-stationary noise sources such as the MOS channel noise and bipolar shot-noise. The theory assumes that the noise current to phase transfer is linear, that is, if the noise current is doubled then the phase change also doubles. The heart of this model is the impulse sensitivity function (ISF) denoted by $\Gamma(x)$. $\Gamma(x)$ is a dimensionless function with a period 2π which describes how much the phase of the oscillation signal varies when a unit pulse is applied. Since it is a periodic function, it takes the variation in bias conditions during one cycle into account, which is neglected using linear modeling. The effective $\Gamma(x)$ can be used to describe both the $1/f$-noise region as well as the white phase noise region of $\mathcal{L}(f_m)$. For each noise source $\Gamma(x)$ must be calculated (or simulated), and output phase noise powers can then be added power-wise if they are uncorrelated. An interesting design insight derived by this theory is that the $\omega_{1/f}$ (see Figure 6.23) can be made significantly smaller than the $1/f$-noise corner of the active devices in the oscillator. In order to minimize $\omega_{1/f}$, the waveform of the oscillator should be made as symmetrical as possible. For example, the rise and fall times should be as similar as possible. As the symmetry is important for each noise source, every half-circuit, and not only the differential circuit, should be as symmetrical as possible.

A different approach which is illustrated by the calculation of phase noise in a CMOS Colpitts oscillator is outlined in [22]. Without making assumptions about the amplitude level or the existence of AGC control, both the carrier level and the phase noise as a result of the various noise currents are calculated,
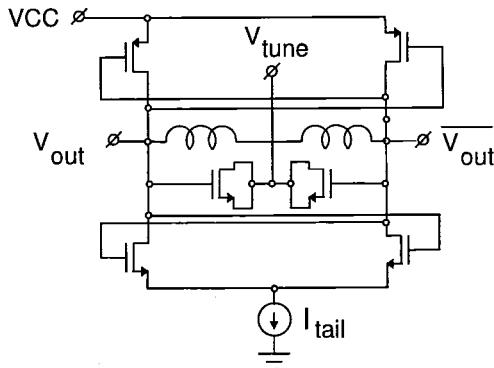
Figure 6.28. LC oscillator with complementary cross-coupled differential pair.

taking into account the nonlinearity. This approach can also be generalized for multiple noise sources.

The last approach mentioned studies the differential cross-coupled LC oscillator, and takes into account resonator noise, tail current noise, noise folding of the differential pair noise, and $1/f$-noise up-conversion [24]. Relatively simple closed-form expressions are obtained. A detailed analysis of noise folding is also presented in [30].

6.3.2 Topologies

The topology which is most often used in integrated circuits, the cross-coupled differential pair, has already been introduced. Single-ended oscillator structures are rarely used since the oscillator signal is then present in the ground and supply lines, and introduces spurious signals in other circuits on the chip. Many different resonator circuits are used in combination with the cross-coupled pair, using one or more of the tuning methods already described. Capacitive tapping can be used to give a high voltage swing over the tank, and only a certain ratio of this voltage over the cross-coupled pair. This can be very useful for low voltage circuits, but can also improve $\mathcal{L}(f_m)$ by providing a noise match between the resonator and cross-coupled pair [31].

A few more topologies will be described in this section but it should be noted that this is just a selection of some of the interesting topologies. A multitude of other LC oscillator topologies exists and it is the application which determines which one is best.

The complementary differential topology in Figure 6.28 has a maximum amplitude $(4/\pi \cdot I_{tail}R_p$ with R_p the effective tank impedance) which is twice the amplitude of an NMOS differential pair [32, 33]. Due to the push-pull
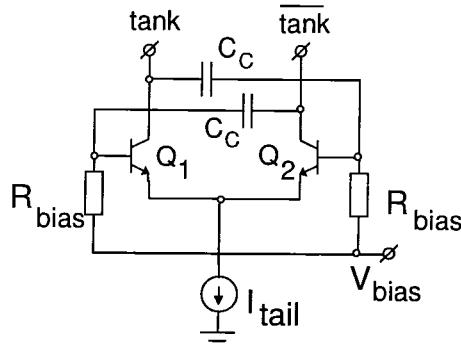
*Figure 6.29.* AC coupling to withstand high voltage swings.
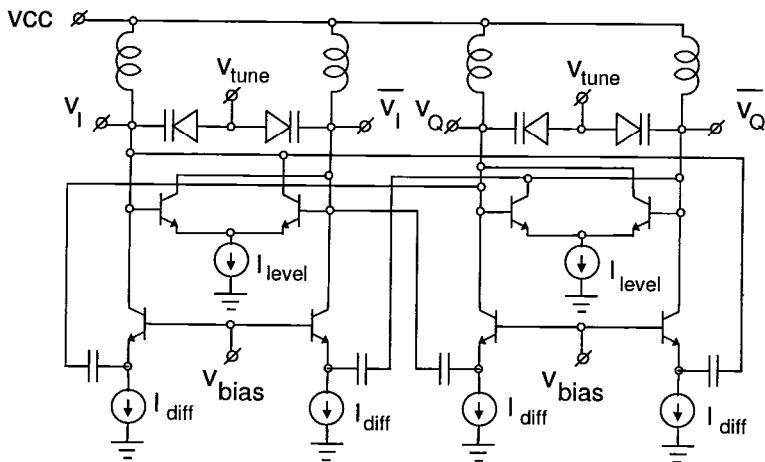
Figure 6.30. Quadrature LC oscillator with phase shifters.

structure, the oscillator waveform can be made more symmetrical by tuning the W/L ratios of the PMOS and NMOS transistors. As mentioned earlier, this symmetry can lower $\omega_{1/f}$ of the phase noise spectrum. Compared to the NMOS-only implementation, the PMOS adds additional parasitic capacitance which will decrease the tuning range.

If there is sufficient voltage room, the AC-coupled cross-coupled bipolar pair in Figure 6.29 can be used, which allows a large voltage swing. The collector-base junction limits the maximum voltage swing without AC-coupling. Once this junction comes into forward bias, the oscillator enters the voltage limited region and the phase noise performance becomes poor. By biasing the bases of the transistors lower than the collectors, the maximum allowable voltage swing can be increased.

An implementation of a quadrature LC oscillator with phase shifters is shown in Figure 6.30 [23]. The capacitors and the bottom transistors implement a differentiator which provides the 90° phase shift for optimal phase noise. At high frequencies the coupling differentiators will provide less than 90° phase shift, and emitter followers (not shown in the figure) can be added to compensate for this. The followers will add some parasitic phase shift. The carrier level can be controlled with tail current I_{level}. As mentioned earlier, parasitic phase shift can replace the need for intentionally added phase shift in optimally coupled N-stage multi-phase LC oscillators, provided that the parasitic phase shift is $\pm 180°/N$ per stage.

6.4 RC OSCILLATORS

RC oscillators, such as ring and relaxation oscillators are especially known for their ease of integration and large tuning ranges. The feedback model discussed in the previous section, will be used to shown why RC oscillators are normally much noisier than LC oscillators, given a certain power budget. Relaxation oscillators will not be discussed as they have in principle no particular advantages at high frequencies compared to ring oscillators. Analysis of relaxation oscillators and some implementations can be found in [34, 35, 36, 37, 38, 39, 40]. Analogous to the discussion of LC oscillators, frequency, tuning and phase noise to carrier ratio of RC oscillators will be analyzed first. Finally, two high-performance RC oscillator topologies are highlighted.

6.4.1   Frequency, Tuning and Phase Noise

During the discussion of phase noise of RC oscillator, it will become clear that a minimum number of stages is optimum for low phase noise given a certain power budget. Hence the two-integrator oscillator is interesting since it has only two stages and also provides quadrature signals.

The behavioral model of the two-integrator oscillator was shown in Figure 6.15 and is repeated in Figure 6.31 for convenience. In this model, losses in each stage are modeled by resistor R. If these losses are exactly compensated ($g_{ml} = 1/R$), then each stage is a perfect integrator and the phase condition for oscillation is fulfilled for all frequencies. The model in Figure 6.31 will be used to analyze the properties of the two-integrator oscillator. All noise sources in each stage are combined in noise current source i_n. Note that if a differential implementation is assumed, the inversion in the behavioral model is simply a cross-coupling of wires.

Figure 6.32 shows a general model for a multi-stage ring oscillator. This model will be used to analyze the properties of ring oscillators of three stages
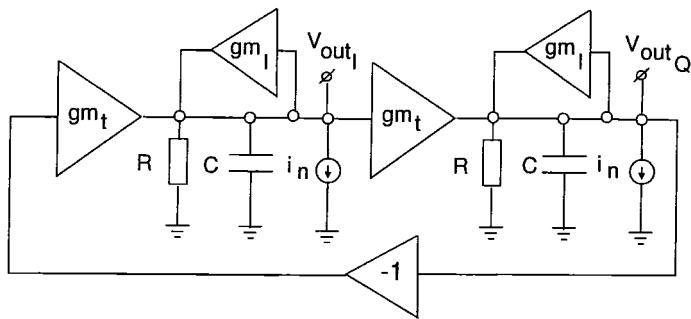
*Figure 6.31.*   Behavioral model of a two-integrator oscillator.
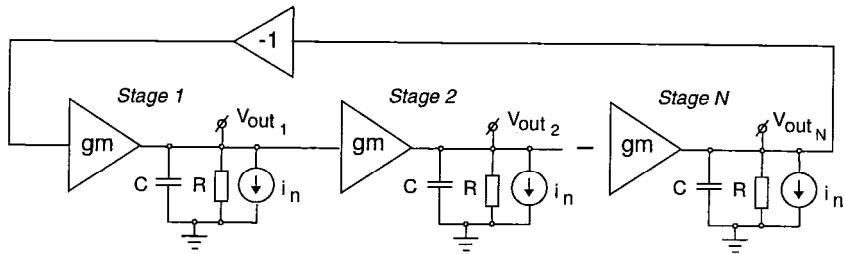
Figure 6.32.   Multi-phase ring oscillator model.

and more. In order to comply with the conditions for oscillation for a finite oscillation frequency, the minimum number of stages N is three. In Figure 6.32, all noise sources in a stage are combined in noise current source i_n.

Frequency

The RC oscillators discussed in this chapter, the two-integrator and ring oscillator, consist of two or more stages. The oscillation frequency is determined by the large signal delay τ_{delay} in each stage. Each half period, the oscillator signal propagates through the stages and forces a signal inversion. If the propagation delay τ_{delay} is equal for all parts of the propagating signal, the oscillation frequency is equal to:

$$f_{osc}|_{large-signal} = \frac{1}{2N\tau_{delay}} \qquad (6.41)$$

Alternatively, the oscillation frequency can be determined by looking at the open loop transfer of the RC oscillator, calculating the gain and phase characteristics, and seeing at which frequency the oscillation conditions are met. This linear analysis can accurately predict the actual oscillation frequency if the

oscillator works in the linear region. For oscillators with active devices work-
ing in the moderate or strongly nonlinear region, linear analysis will predict a
higher oscillation frequency than the actual (large signal) oscillation frequency.
Nevertheless, linear analysis is a convenient starting point and provides insight
into the basic parameters determining the frequency.

The frequency of the two-integrator oscillator is determined by the amplitude
condition of oscillation. Provided that the losses modeled by R in Figure 6.31
are completely compensated by g_{ml}, the phase condition for oscillation is met
for any frequency. Calculating the amplitude condition for oscillation gives:

$$\omega_{osc} = \frac{g_{mt}}{C} \tag{6.42}$$

Unlike the two-integrator oscillator, the oscillation frequency of the N-stage
ring oscillator in Figure 6.32 is determined by the phase condition of oscilla-
tion. The transconductances g_m should be large enough to meet the amplitude
conditions for oscillation. For $N \geq 3$ it holds that:

$$\omega_{osc} = \frac{\tan(\pi/N)}{R \, C} \tag{6.43}$$

The maximum frequency of operation of RC oscillators is determined by
technology parameters such as f_T or whatever time constant is dominant (for
example input or output bandwidth) in the active devices. The topology also has
a significant influence. In general, more components mean more inter-connect
and device capacitance, and therefore a lower maximum oscillation frequency.
Since the degree of nonlinearity determines how much the large signal oscil-
lation frequency drops compared to linear analysis, the carrier amplitude (and
therefore the $\mathcal{L}(f_m)$ specification) also has a significant influence.

Tuning

Tuning of the two-integrator oscillator can be realized by varying the transcon-
ductance g_{mt} or the capacitance C as indicated by (6.42). The g_m is normally
varied. If the transconductances are implemented by bipolar differential pairs,
the frequency will vary linearly with the tail current. If MOS differential pairs
are used in saturation, the frequency will be proportional to the square root of
the tail current. Figure 6.33 shows the tuning behavior of a bipolar implemen-
tation (see Figure 6.36) and the effect of transistor parasitics. A large part of
the nonlinearity is determined by the diffusion capacitance part of C_π, which is
current dependent. Since g_{mt} can be varied over a large range, the tuning range
can easily be made more than an octave, and can even be made larger than a
decade.
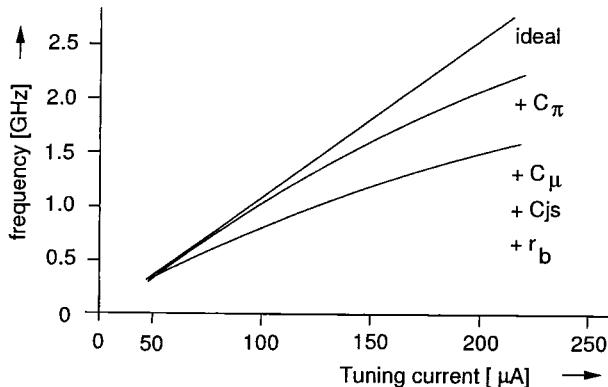
Figure 6.33.   Tuning characteristic of the two-integrator oscillator.

Ring oscillators can be tuned in four different ways. The first two are indicated by (6.43). By changing resistance R or capacitance C, the oscillation frequency changes. The third method is to change the delay in each stage by changing the transconductance g_m in Figure 6.32. The fourth possibility is to use two signal paths per stage with different delays, and summing the outputs of the separate paths with a controllable weighting factor. This last technique is called delay interpolation.

Capacitive tuning can be realized using the techniques described on page 208. The tuning range will then be rather limited and nonlinear compared to the other available techniques of tuning RC oscillators. Capacitive tuning will also limit the maximum oscillation frequency, since a varactor with a certain minimum capacitance must be added. For these reasons, capacitive tuning is rarely used in combination with ring oscillators. If the transconductances in Figure 6.32 are implemented by differential pairs, the parasitic delay (controlled by the tail current) can be used for tuning. The tuning range will be very small, especially when the dominant time constant is the output bandwidth formed by R and C.

Large linear tuning ranges are possible when employing resistive tuning. Two examples are given in Figure 6.34. The capacitors should be omitted in these circuit diagrams, for a maximum oscillation frequency. In Figure 6.34(a), the cross-coupled pair implements a negative resistance which is in parallel with the collector resistors and input resistance of the next stage. Therefore, by varying I_{tune} the total resistance and thus the frequency is varied. A second possibility is shown in Figure 6.34(b). The PMOS transistors operate as variable resistances controlled by V_{tune}, in the triode region. Since it is not only the frequency which varies with both methods but also the gain in each stage, the amplitude level
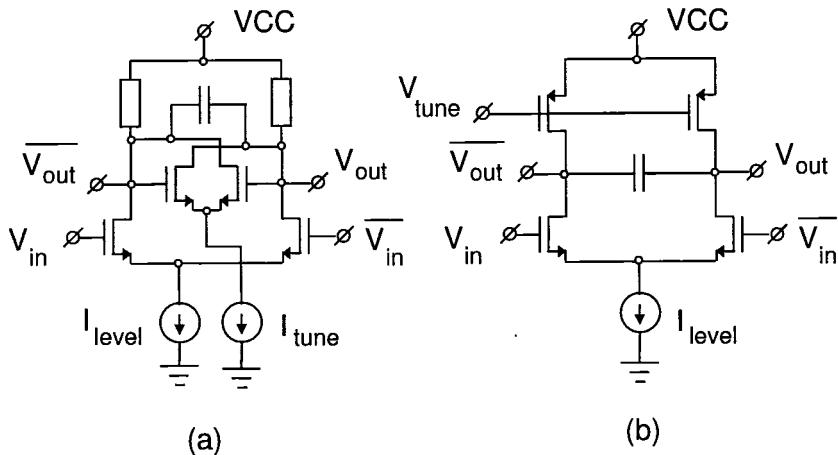
Figure 6.34. Examples of resistive tuning in a stage: using a negative controllable resistance (a) or MOS transistors in the triode region (b).
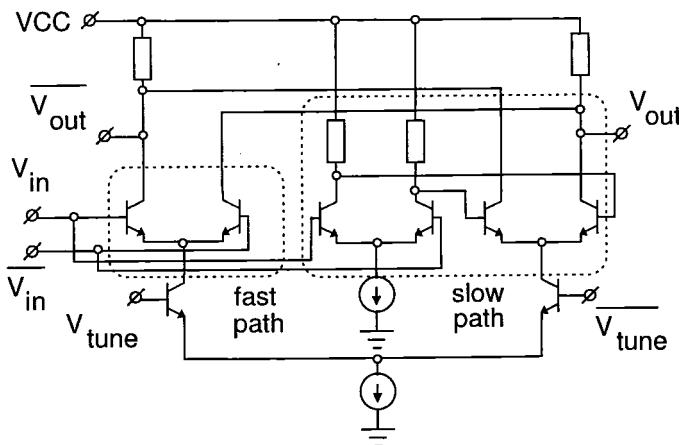

Figure 6.35. Bipolar implementation of a delay interpolation stage.

varies significantly over the tuning range. This means that $\mathcal{L}(f_m)$ will also vary. By increasing I_{level} in Figure 6.34(b) when V_{tune} decreases, the swing variation across the tuning range can be decreased [41].

Figure 6.35 shows an example of ring oscillator tuning applying delay interpolation [42]. The differential pair controlled by V_{tune} determines the ratio of the output current of the fast and the slow path. The slow path has one additional delay compared to the fast path implemented by a differential pair.

Phase Noise to Carrier Ratio

Linear Modeling

For the analysis of phase noise in RC oscillators the same route will be followed as for LC oscillators. First, the phase noise to carrier ratio will be analyzed using the transfer function for an oscillating feedback system shown in (6.24) and its accompanying quality factor formula in (6.25). In the second part of this section on phase noise in RC oscillators, nonlinear effects will be discussed.

Using (6.25) and the oscillation conditions, the quality factor for the ring oscillator model in Figure 6.32 can be derived:

$$Q_{RC} = \frac{1}{2} N \sin\left(\frac{\pi}{N}\right) \tag{6.44}$$

Equation (6.44) is also valid for the two-integrator oscillator shown in Figure 6.15. For $N = 2$, $Q_{RC} = 1$. When N goes to infinity Q_{rc} reaches its maximum which is $\pi/2$.

Now that the quality factor Q_{RC} has been derived, derivation of $\mathcal{L}(f_m)$ for ring oscillators is straightforward. The quality factor Q_{LC} in (6.29) has to be replaced by Q_{RC}, and i_n^2 has to be multiplied by N to account for the noise in every stage:

$$\mathcal{L}(f_m) = \frac{1}{2} \cdot \frac{1}{4\,Q_{RC}^2} \cdot \left(\frac{f_{osc}}{f_m}\right)^2 \cdot \frac{N\overline{i_n^2}}{i_{carrier-rms}^2} \tag{6.45}$$

Equation (6.45) is valid for $N \geq 2$. That is, it is also valid for the two-integrator oscillator. By choosing an RC oscillator topology, i_n^2 in (6.45) can be expanded in a similar way as was done in Section 6.3.1.

A number of conclusions and remarks can be made about the derived equations for RC oscillators:

- Equation (6.45) models the 6 dB per octave slope (proportional to $1/f^2$) of an RC oscillator. This slope continues for large offset frequencies until it hits the white noise floor. In practice, this white noise floor is determined by buffers cascaded after the oscillator.

- In a similar way as for the LC oscillator, the phase noise becomes 3 dB lower if the power is doubled. If two identical RC oscillators are combined, the phase noise lowers 3 dB, since the carrier adds correlated and the noise adds uncorrelated.

- The quality factor can be slightly improved by increasing the number of stages N. However, when N is increased, also the number of noise sources and the power dissipation are increased.

- Since $Q_{RC} = 1$ for $N = 2$ and is maximally $\pi/2$, it is clear that LC oscillators have superior phase noise for a given power budget. In practice, Q_{LC} can be made significantly larger than 10, and depends on the resonator technology, whereas Q_{RC} is technology-independent. In other words, given a certain $\mathcal{L}(f_m)$ specification, the power dissipation in an RC oscillator will be much higher than in an LC oscillator.

Nonlinear Modeling

In general the same nonlinear phase noise mechanisms which are present in LC oscillators contributing to the total $\mathcal{L}(f_m)$ in RC oscillators. The general phase noise theory outlined in on page 220, on nonlinear phase noise modeling in LC oscillators can also be applied to RC oscillators. Application of this theory on differential ring oscillators, shows that a minimum number of stages is optimum for a minimum $\mathcal{L}(f_m)$, given a certain power budget [29].

Instead of analyzing the phase deviations of an oscillator in the frequency domain, [43, 44] describe and model the phase deviation in the time domain (jitter). This results in a figure of merit for ring oscillators, which is independent of the number of stages N and links circuit noise to the total jitter performance.

In [19], phase noise is analyzed using linear theory and several contributions originating from nonlinear phase contributions, such as noise folding, are added to the linear results. The assumption here is that the oscillator works in the weakly nonlinear region. Modulation of the noise sources is not taken into account, for example.

A final remark on the phase noise of RC oscillators involves (6.40). Since the tuning range of an RC oscillator is usually large, and since maximum supply voltages are becoming lower and lower, the tuning constant K_{vco} is usually very large. Equation (6.40) shows that this implies that the RC oscillator is very sensitive to modulation by external noise sources. One measure which can be taken is to divide the tuning range into a coarse tuning range which uses low noise switchable current sources, and a fine tuning range which interpolates between the bands. The K_{vco} of the fine tune input is reduced in this way.

6.4.2   Topologies

In this section two RC oscillator circuit implementations will be highlighted. A number of topologies have already been introduced in Section 6.4.1. A number of references to interesting topologies is also given in the following section, which describes an RC oscillator design example.

An implementation of the two-integrator oscillator is shown in Figure 6.36. The cross-coupled pair in each section implements the negative resistance which is needed to compensate all losses (g_{ml} in Figure 6.31). The integration capaci-
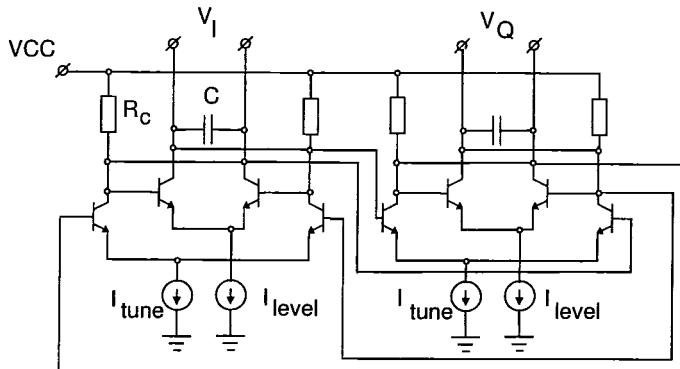
Figure 6.36.   Bipolar implementation of the two-integrator oscillator.
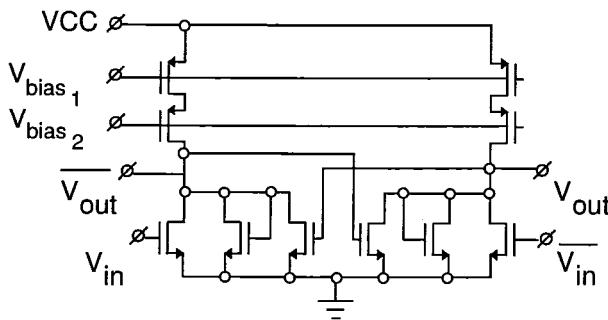

Figure 6.37.   Low voltage ring oscillator stage with high supply rejection ratio.

tance consists of device parasitics and interconnect capacitance. The differential pair at the bottom implements g_{mt} in the behavioral model from Figure 6.31. At high frequencies, the cross-coupled pair controlled by I_{level} does not need to compensate all losses, since the differential pairs have some parasitic phase shift. The lowest frequency of operation is determined by the ratio between I_{tune} and I_{level}. If I_{tune} approaches I_{level}, the circuit will start to act as a flip-flop, and latch-up will occur. The ratio at which this occurs depends on the mismatches in the circuit. As with all symmetric oscillators, symmetry in the layout is very important in order to obtain accurate quadrature signals from this oscillator.

Figure 6.37 shows a ring oscillator stage which has a good supply rejection ratio due to the cascoded current sources [45]. Only one stage is shown. The bias voltage for the current sources are most easily generated by a cascode current mirror. The minimum supply voltage of this architecture is one gate-source voltage plus two drain-source saturation voltages. Replacing the cascode current source by one device will save on saturation voltage, but at the expense
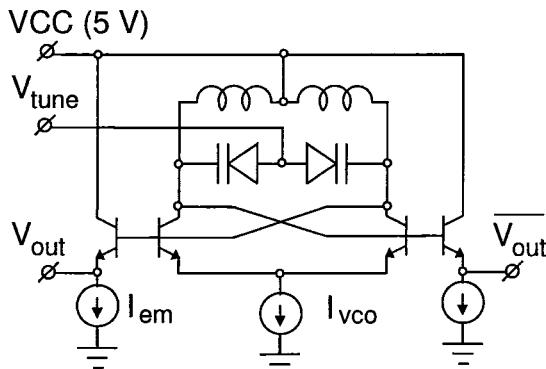
Figure 6.38.   Simplified schematic of the LC oscillator.

of decreasing the supply rejection ratio. Two diode-connected transistors limit the voltage swing and keep the current sources in the saturation region, which improves the speed.

6.5   DESIGN EXAMPLES

In the last section of this chapter, an LC oscillator and an RC oscillator design are highlighted.

6.5.1   An 830 MHz Monolithic LC Oscillator

In Chapter 1 the technology Silicon-on-Anything was discussed. This design example of a fully integrated LC oscillator exploits this technology and its high quality passives in order to realize a low phase noise to carrier ratio at a dissipation level below 1 mW.

Circuit Design

A balanced oscillator topology was chosen for the oscillator design, to maximize rejection of common mode disturbances. The design is shown in Figure 6.38. Common mode rejection is particularly important if the VCO is integrated with other sub-systems, like in a transceiver system. The disadvantage of a balanced LC oscillator design with integrated coils is a substantial increase of the chip area, since the inductor area often dominates.

The value of the on-chip inductors was 29.8 nH with a Q_{max} of ~ 17.3 at 1 GHz. The total parasitic capacitance of this coil was 150 fF. Seven varactor sections in parallel, each of 240 fF (zero bias) have been used.The capacitance ratio C_{max}/C_{min} for a bias voltage $V_{reverse}$ ranging from 0 to 5 volt, was ap-
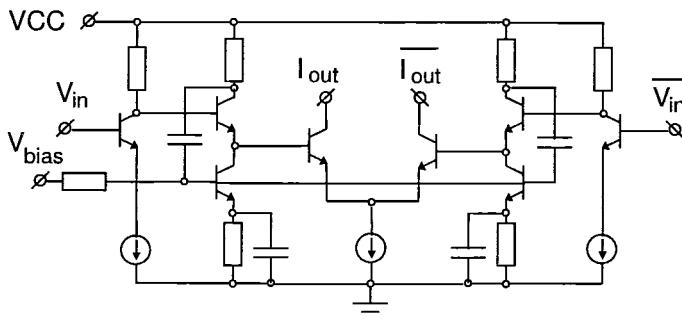
*Figure 6.39.*   Circuit diagram of the output buffer.
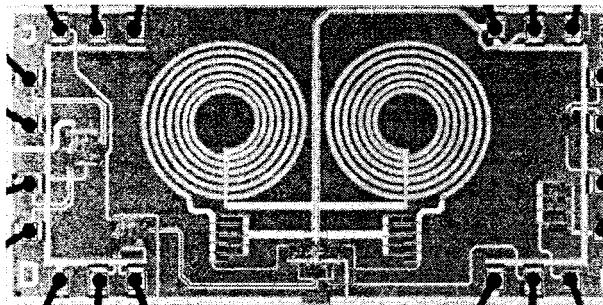
Figure 6.40.   Micrograph of the LC oscillator.

proximately 1.7. The quality factor of the PN-junction varactor is around 26 at 1 GHz. The total current consumption of the VCO consists of twice I_{em} (40 μA) plus I_{vco} (80 μA). Nominal supply voltage was 5 volt to have a maximum tuning range. Dissipation of the VCO can be lowered by lowering VCC at the expense of a slightly reduced tuning range.

An open collector output buffer is implemented, in cascade with the VCO and emitter followers, delivering -30 dBm in 50 Ω. The schematic of the differential buffer is shown in Figure 6.39. Without special measures in the output buffer, a multi-oscillation would occur as discussed in Section 6.1.4 of this chapter. In this VCO design, one of the two simultaneous oscillations was introduced by loading of the VCO core due to the buffer. This resulted in a severely distorted VCO output signal. The feedback in the second stage of the buffer substantially reduced the loading on the VCO circuitry, and eliminated the multi-oscillation phenomena. The total dissipation of the output buffer was 7.2 mW.
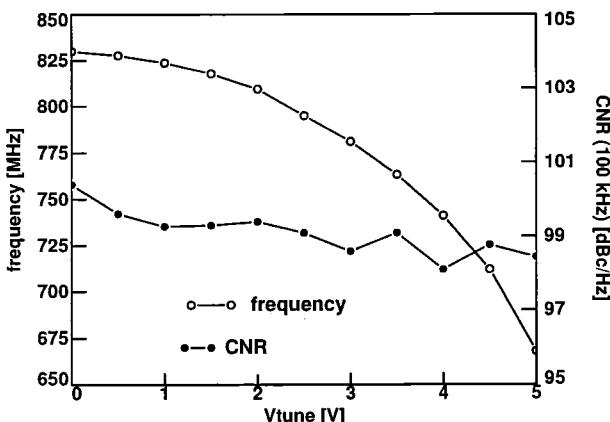
Figure 6.41. Frequency and CNR(100 kHz) versus V_{tune}.

Measurements

A micrograph of the monolithic LC oscillator is shown in Figure 6.40. The two inductors occupy a dominant part of the 1940 μm $\times$ 1280 μm active chip area. On the left two output buffers can be seen. One was used to measure the buffer characteristics.

Frequency and CNR measurements were performed using an HP8562E spectrum analyzer with pre-amplifier. Frequency versus tuning voltage V_{tune} measurements are presented in Figure 6.41. The measured tuning range is 668 MHz to 830 MHz when V_{tune} is varied from 5 to 0 volt, which is more than 21%. Phase noise measurements are also plotted in Figure 6.41. For the tuning voltage range from 5 to 0 volt, CNR(100 kHz) varies between 98.1 and 100.4 dBc/Hz.

The power spectrum of the oscillator with V_{tune} equal to 0 volt is shown in Figure 6.42.

6.5.2 A 10 GHz I/Q RC Oscillator with Active Inductors

The two-integrator oscillator discussed in this chapter can have a very large tuning range. Implementations of the two-integrator oscillator were realized with a CNR of 106 dBc/Hz at a 2 MHz offset (using 100 mW), with a wide tuning range from 0.9 GHz to 2.2 GHz [7]. However, when an RC oscillator implementation has to work at very high frequencies, for example at 1/3 of the f_T of an IC technology, other topologies are more suitable. The topology in this design example outperforms the two-integrator oscillator with respect to the phase-noise to carrier ratio and maximum oscillation frequency at the expense of tuning range reduction [46].
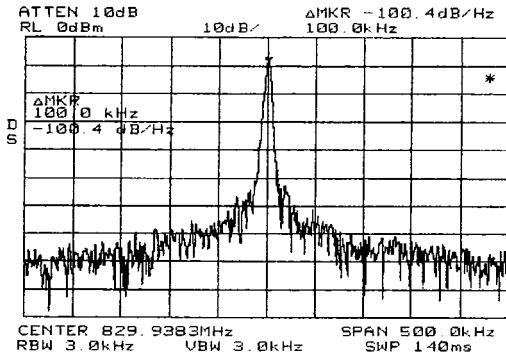
*Figure 6.42.*   Power spectrum of the SOA oscillator.
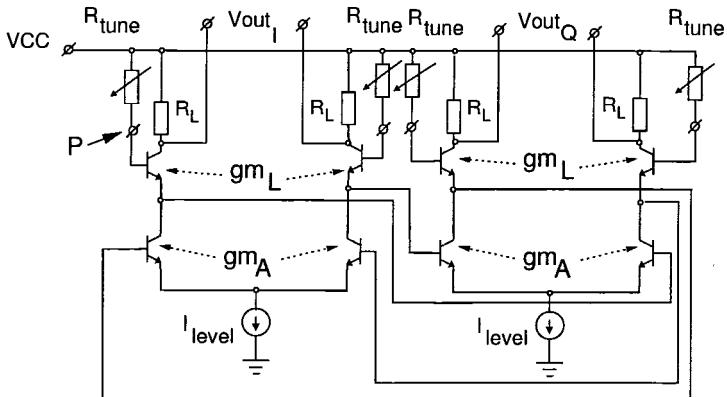
Figure 6.43.   Circuit implementation of the I/Q RC oscillator.

Circuit Design

The circuit diagram of the I/Q RC oscillator is shown in Figure 6.43. It is one possible circuit implementation of the multi-phase LC behavioral model shown in Figure 6.25. However, the inductors are implemented with active devices which gives this circuit the phase noise properties of an RC oscillator. Conceptually, this can be regarded as using an inductor with a quality factor close to unity.

The I/Q RC oscillator was realized without lumped capacitance: all capacitance consisted of device and interconnect parasitics. Differential pairs g_{mA} with tail current I_{level} provided the gain to comply with the amplitude condition for oscillation. The inductors in Figure 6.25 were implemented using transistors g_{mL}. When the base and collector of a bipolar transistor are shorted, this
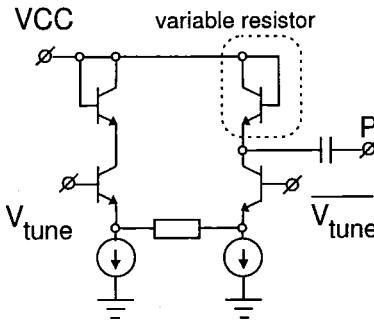
Figure 6.44. Simplified schematic of the implementation of R_{tune}.

two-terminal device becomes an active inductor for high frequencies [47]. The inductance can be varied by controlling the base resistance or by varying the current through the transistor. By adding load resistance R_L in the collectors of transconductance g_{mL}, this transistor not only realizes the active inductor, but also provides cascode buffering. This means that circuits with low impedance levels can be connected to the collectors of transconductance g_{mL}, without loading the oscillator core. Tuning was realized using variable resistor R_{tune} which is added in the base of g_{mL}.

The implementation of R_{tune} is shown in Figure 6.44. R_{tune} was realized using a diode-connected transistor (dimensioned to have little inductance), which was AC-coupled to the base of transistor g_{mL} (point P in Figure 6.43 and 6.44). The current of this transistor was controlled by a linearized differential pair.

Measurements

The RC oscillator from Figure 6.43 was realized in a 30 GHz f_T BiCMOS process. The IC micrograph of the realized I/Q RC oscillator is shown in Figure 6.45. The internal oscillator signal swing was set to 250 mV$_p$ by setting I_{level} to 12 mA, which was found to be the optimum for a maximum CNR. Making I_{level} smaller would result in a higher oscillation frequency but a smaller CNR. The active chip area of the RC oscillator with V/I-converter was less than 0.13 mm^2. The total chip area including bond pads was 1.5x1.5 mm^2. The power dissipation of the total IC was 230 mW of which 75 mW was dissipated by the VCO core. The power supply voltage was set to 2.7 V.

Measurements were performed on packaged samples (16 pin HTSSOP package). On-chip 50 Ω I/Q buffers provided the quadrature output signals with -20 dBm output power. The measured frequency and CNR versus differential tuning voltage V_{tune} are shown in Figure 6.46.

The tuning range is 9.8 GHz to 11.5 GHz, which is 16 %. The CNR was measured using a spectrum analyzer, and results were verified using HP3048
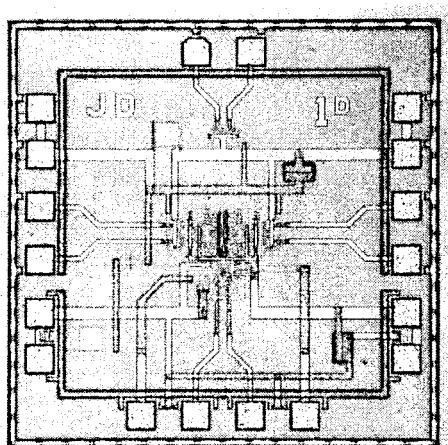
Figure 6.45.   Micrograph of the IC.
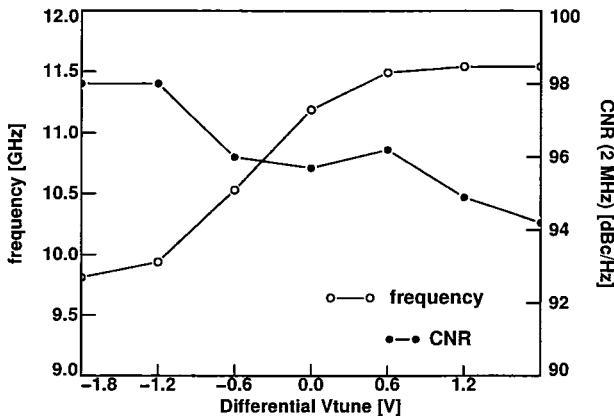
Figure 6.46.   Frequency and CNR (2 MHz) versus differential tuning voltage.

phase noise measurement equipment, which has an accuracy of ±2 dB. The measured CNR at 2 MHz offset was better than 94 dBc/Hz over the complete tuning range. Best case CNR was measured at 9.8 GHz and was 98 dBc/Hz.

The power spectrum of the ring oscillator at maximum frequency is shown in Figure 6.47.

In the literature, reported ring oscillators are realized in various IC technologies, ranging from CMOS, BiCMOS and SiGe to InP and GaAs implementations. In Table 6.2, these oscillators are compared with the presented quadrature oscillator design. At high frequencies, relatively close to the f_T, both the CNR and oscillation frequency should be benchmarked as they are exchangeable: lowering the oscillation frequency would increase the internal signal swing (because the gain per stage increases), and would thus improve the CNR. This effect is in addition to the first order dependence of the CNR on
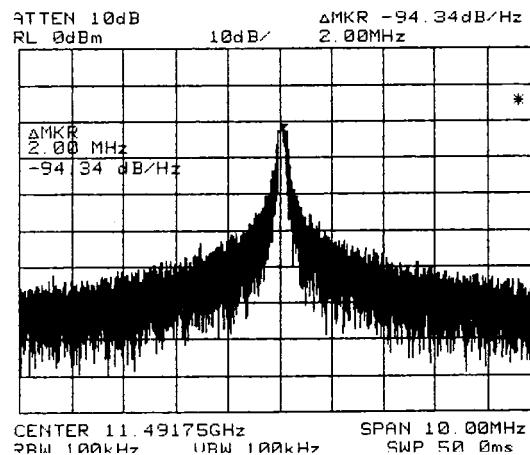
Figure 6.47.   Oscillator power spectrum at 11.5 GHz.

Table 6.2.   Ring oscillator benchmarking.

Ref.	f_{osc}	P_{DC}	I/Q	Process/ f_T	f_{osc}/f_T	CNR_{norm}
	GHz	mW		– / GHz		[dBc/Hz]
This Design	11.5	75	Yes	BiCMOS/ 30	0.38	150.4
[48]	10	250	No	InP HBT/ 53	0.19	153
[49]	6.5	–	Yes	AlGaAs, GaAS/ 22	0.3	–
[50]	9	850	Yes	Si Bipo./ 25	0.36	130
[51]	2.2	1.3	No	BiCMOS/ 18	0.13	140.7
[52]	10	–	Yes	SiGe/ 45	0.23	–
[19]	2.2	11.8	No	0.5 μm CMOS/ ~15	~ 0.15	150.2
[53]	5.43	40	Yes	0.25 μm CMOS/ ~40	~ 0.14	157.2
[54]	1.8	22.5	Yes	Bipolar/ 18	0.1	153.5
[7]	2.2	100	Yes	BiCMOS/ 11	0.2	146.8

the oscillation frequency (CNR $\sim 1/f_{osc}^2$).

The CNR is benchmarked using the figure of merit (FOM) in (6.46). The theoretical argumentation for the applied normalizations originates from (6.45).

$$CNR_{norm} = CNR_{measured} - 10\log\left[\left(\frac{f_m}{f_{osc}}\right)^2 \frac{P_{DC}}{1\text{mW}}\right] \qquad (6.46)$$

The FOM CNR_{norm} neglects second order effects (e.g. increased influence of parasitics at high frequencies), thus making it difficult to realize a high CNR at high f_{osc}/f_T ratios, in practice. It does normalize for first order dependency of CNR on power dissipation P_{DC} (normalized to 1 mW), CNR offset frequency f_m and f_{osc}. The comparison in Table 6.2 shows that a high f_{osc}/f_T ratio can be combined with a high CNR_{norm} using the I/Q RC oscillator architecture with active inductors.

REFERENCES

[1] J. van der Tang and S. Hahn, "A Monolithic 0.4 mW SOA LC Voltage-Controlled Oscillator," in *European Solid-State Circuits Conf. (ESS-CIRC)*, 1999, pp. 150–153.

[2] J. 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, no. 8, pp. 1238–1243, Aug. 2000.

[3] H. Wu and A. Hajimiri, "A 10 GHz CMOS Distributed Voltage Controlled Oscillator," in *IEEE Custom Integrated Circuits Conf. (CICC)*, 2000, pp. 581–584.

[4] E. H. Nordholt *et al.*, "Single-pin integrated crystal oscillators," *IEEE Transactions on Circuits and Systems*, vol. 37, no. 2, pp. 175–182, Feb. 1990.

[5] E. Vittoz *et al.*, "High-performance crystal oscillator circuits: Theory and Application," *IEEE Journal of Solid-State Circuits*, vol. 23, no. 3, pp. 774–783, Mar. 1988.

[6] U. Rohde, "Designing SAW Resonators and DRO Oscillators Using Nonlinear CAD tools," in *IEEE International Frequency Control Symposium*, 1995, pp. 379–395.

[7] J. van der Tang and D. Kasperkovitz, "A 0.9-2.2 GHz Monolithic Quadrature Mixer Oscillator for Direct-Conversion Satellite Receivers," in *IEEE International Solid-State Circuits Conf. (ISSCC)*, 1997, pp. 88–89.

[8] H. von Barkhausen, *Lehrbuch der Elektronen-Roehren, Band 3, Rueckkopplung*, Verlag S. Hirzel, 1935.

[9] H. Bode, *Network Analysis and Feedback Amplifier Design*, D. van Nostrand Company, Inc., 1945.

[10] W. Chen, *Active Network and Feedback Amplifier Theory*, Hemisphere Publishing Corporation, 1980.

[11] N. Nguyen, *Monolithic microwave oscillators and amplifiers*, Ph.D. thesis, University of California, 1991.

[12] A. N. Riddle, *Oscillator noise: Theory and characterization*, Ph.D. thesis, North Carolina State University, 1986.

[13] B. van der Pol, "The nonlinear theory of electric oscillations," *Proceedings of the IRE*, vol. 22, no. 9, pp. 1051–1086, Sept. 1934.

[14] F. L. Walls *et al.*, "Extending the Range and Accuracy of Phase Noise Measurements," in *IEEE International Frequency Control Symposium*, 1988, pp. 432–441.

[15] W.P. Robins, *Phase Noise in Signal Sources*, 9. IEE Telecomm., London, 2nd edition, 1996.

[16] B. Razavi, "Design Considerations for Direct-Conversion Receivers," *IEEE Transactions on Circuits and Systems II*, vol. 44, no. 6, pp. 428–435, June 1997.

[17] A. Kral *et al.*, "RF-CMOS Oscillators with Switched Tuning," in *IEEE Custom Integrated Circuits Conf. (CICC)*, 1998, pp. 555–558.

[18] D. B. Leeson, "A simple model of feedback oscillator noise spectrum," *Proceedings of the IEEE*, pp. 329–330, Feb. 1966.

[19] B. Razavi, "A study of phase noise in CMOS oscillators," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 3, pp. 331–343, Mar. 1996.

[20] D. Triantis *et al.*, "Thermal Noise Modeling for Short-Channel MOSFET's," *IEEE Transactions on Electron Devices*, vol. 43, no. 11, pp. 1950–1955, Nov. 1996.

[21] Y. Tsividis, *Operation and Modeling of the MOS Transistor*, McGraw-Hill, second edition, 1999.

[22] Q. Huang, "Phase Noise to Carrier Ratio in LC oscillators," *IEEE Transactions on Circuits and Systems I*, vol. 47, no. 7, pp. 965–980, July 2000.

[23] P. van de Ven *et al.*, "An Optimally Coupled 5 GHz Quadrature LC Oscillator," in *Symposium on VLSI Circuits*, 2001, pp. 115–118.

[24] J. J. Rael and A. A. Abidi, "Physical Processes of Phase Noise in Differential LC Oscillators," in *IEEE Custom Integrated Circuits Conf. (CICC)*, 2000, pp. 569–572.

[25] E. Hegazi *et al.*, "A Filtering Technique to Lower Oscillator Phase Noise," in *IEEE International Solid-State Circuits Conf. (ISSCC)*, 2001, pp. 364–365.

[26] P. Andreani and H. Sjöland, "A 2.2 GHz CMOS VCO with Inductive Degeneration Noise Suppression," in *IEEE Custom Integrated Circuits Conf. (CICC)*, 2001, pp. 197–200.

[27] A. Hajimiri *et al.*, "A General Theory of Phase Noise in Electrical oscillators," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 2, pp. 179–194, Feb. 1998.

[28] A. Hajimiri *et al.*, "Corrections to "A General Theory of Phase Noise in Electrical Oscillators"," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 6, June 1998.

[29] A. Hajimiri and T.H. Lee, *Low Noise Oscillators*, Kluwer Academic Publishers, 1999.

[30] C. Samori *et al.*, "Spectrum folding and phase noise in LC tuned oscillators," *IEEE Transactions on Circuits and Systems II*, vol. 45, no. 7, pp. 781–789, July 1998.

[31] A. van Staveren *et al.*, *Structure Electronic Design, High-Performance Harmonic Oscillators and Bandgap References*, Kluwer Academic Publishers, Boston, 2000.

[32] J. Craninckx *et al.*, "A Fully Integrated Spiral-LC CMOS VCO Set with Prescaler for GSM and DCS-1800 Systems," in *IEEE Custom Integrated Circuits Conf. (CICC)*, 1997, pp. 403–406.

[33] F. Svelto *et al.*, "A 1mA, -120.5 dBc/Hz at 600 kHz from 1.9 GHz fully tuneable LC CMOS VCO," in *IEEE Custom Integrated Circuits Conf. (CICC)*, 2000, pp. 577–580.

[34] A. A. Abidi, *Effects of random and periodic excitations on relaxation oscillators*, Ph.D. thesis, University of California, Berkeley, 1981.

[35] A. Abidi *et al.*, "Noise in relaxation oscillators," *IEEE Journal of Solid-State Circuits*, vol. 18, no. 6, pp. 794–802, June 1983.

[36] C. A. M. Boon *et al.*, "Modeling the phase noise of RC multivibrators," in *Midwest symposium on circuits and systems*, 1984, pp. 421–424.

[37] C. J. M. Verhoeven, "A New model for regenerative circuits," in *Midwest symposium on circuits and systems*, 1988, pp. 631–634.

[38] B. M. Fleischer, *Jitter in relaxation oscillators*, Ph.D. thesis, Stanford University, 1989.

[39] C. J. M. Verhoeven, *First-order Oscillators*, Ph.D. thesis, Delft University of Technology, 1990.

[40] C. J. M. Verhoeven, "A high-frequency electronically tunable quadrature oscillator," *IEEE Journal of Solid-State Circuits*, vol. 27, no. 7, pp. 1097–1100, July 1992.

[41] I. Young *et al.*, "A PLL clock generator with 5 to 110 MHz lock range for microprocessors," in *IEEE International Solid-State Circuits Conf. (ISSCC)*, 1992, pp. 50–51.

[42] B. Razavi, *Monolithic Phase-locked Loops and Clock Recovery Circuits*, IEEE Press, New York, 1996.

[43] J. A. McNeill, *Jitter in Ring Oscillators*, Ph.D. thesis, Boston University, 1994.

[44] J. McNeill, "Jitter in Ring Oscillators," *IEEE Journal of Solid-State Circuits*, vol. 32, no. 6, pp. 870–879, June 1997.

[45] L. Sun *et al.*, "A 1.0V GHz Range 0.13 um CMOS Frequency Synthesizer," in *IEEE Custom Integrated Circuits Conf. (CICC)*, 2001, pp. 327–330.

[46] J. van der Tang *et al.*, "A 9.8-11.5 GHz Quadrature Ring Oscillator for Optical Receivers," in *IEEE Custom Integrated Circuits Conf. (CICC)*, 2001, pp. 323–326.

[47] J. van der Tang *et al.*, "A 2.7 V, 8 GHz Monolithic I/Q RC Oscillator with Active Inductive Loads," in *European Solid-State Circuits Conf. (ESSCIRC)*, 2000, pp. 304–307.

[48] R. K. Montgomery *et al.*, "10 and 26 GHz differential VCOs using Inp HBTs," in *IEEE International Microwave Symposium*, 1996, pp. 1507–1510.

[49] A. W. Buchwald *et al.*, "A 6 GHz Integrated Phase-Locked Loop Using ALGaAS/GaAs Heterojunction Bipolar Transistors," in *IEEE International Solid-State Circuits Conf. (ISSCC)*, 1992, pp. 98–99.

[50] A. Pottbacker and U. Langmann, "An 8 GHz Silicon Bipolar Clock-Recovery and Data-Regenerator IC," in *IEEE International Solid-State Circuits Conf. (ISSCC)*, 1994, pp. 116–117.

[51] B. Razavi, "A 2-GHz 1.6 mW Phase-Locked Loop," *IEEE Journal of Solid-State Circuits*, vol. 32, no. 5, pp. 730–735, May 1997.

[52] M. Meghelli *et al.*, "SiGe BiCMOS 3.3V Clock and Data Recovery Circuits for 10 Gb/s Serial Transmission Systems," in *IEEE International Solid-State Circuits Conf. (ISSCC)*, 2000, pp. 56–57.

[53] A. Hajimiri *et al.*, "Jitter and Phase Noise in Ring Oscillators," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 6, pp. 790–804, June 1999.

[54] S. Finocchiaro *et al.*, "Design of Bipolar RF Ring Oscillator," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, 1999, vol. 1, pp. 5–8.

Chapter 7

Frequency Synthesizers

Frequency synthesizers are used to generate the local oscillator (LO) signal in transceiver systems. The frequency of the LO signal determines which RF channel will be received and to which RF channel the base-band information will be transferred, before being transmitted by the PA-antenna combination. In other words, the frequency synthesizer operates as the transceiver's tuning system; in practice, the frequency synthesizer is based on a phase-locked loop (PLL) control system. Important design aspects which need to be taken into account are the spectral purity of the PLL output signal and the power dissipation of the PLL building blocks. This chapter focuses on the spectral purity performance of PLL frequency synthesizers, and on the circuit implementations of low-power programmable frequency dividers and high operation frequency phase-frequency detector/charge-pump combinations.

7.1   INTRODUCTION

Modern frequency synthesizers for high-frequency applications invariably consist of a voltage controlled oscillator (VCO) which is incorporated in a feedback control loop. If the controlled variable is the phase of the oscillator then the combination of the VCO with the control system is referred to as a phase-locked loop (PLL) frequency synthesizer. The block-diagram of a single-loop PLL frequency synthesizer is presented in Figure 7.1. For the design of the VCO the reader is referred to Chapter 6. The main points of this chapter are the design of programmable frequency dividers and phase-frequency detector/charge-pump combinations, and the spectral purity of the PLL output signal. We start with a review of the single-loop PLL architecture, which is followed by a description of the main specification points of tuning systems employed in modern transceiver products. We then proceed with a discussion of the system-level aspects of the
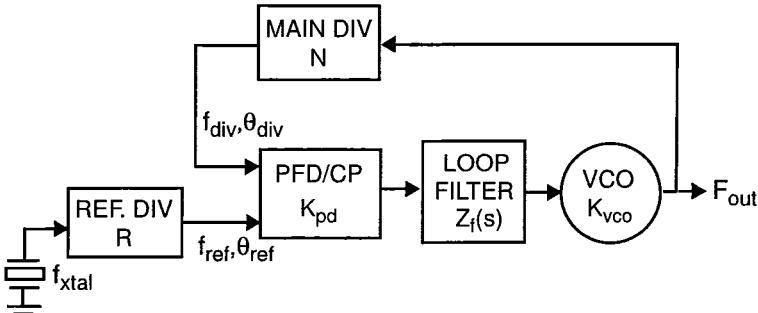
Figure 7.1. Block diagram of a single-loop PLL frequency synthesizer using a tri-state phase-frequency detector/charge-pump combination (PFD/CP).

different PLL building blocks. After that, the concepts of open-loop bandwidth and phase margin are introduced, and design equations for a type-2 third-order PLL are derived.[1] We continue with an analysis of the spectral purity performance of a single-loop PLL, namely spurious reference breakthrough and phase noise performance; the section is concluded with a design procedure for the loop filter components which leads to compliance to spectral purity requirements. The remaining of the chapter presents architectures and circuit implementations of building blocks for PLL frequency synthesizers: a truly-modular architecture for low-power fully programmable frequency dividers and an architecture for high operation frequency phase-frequency detector/charge-pump combinations.

7.2   INTEGER-N PLL ARCHITECTURE

Single loop PLLs are by far the most used synthesizer architectures in the industry. This is partly because of their simplicity in terms of external components and ease of application, but also because they can be produced with high reliability and occupy a small chip area.

A single-loop Integer-N PLL, as depicted in Figure 7.1, consists of a voltage controlled oscillator (VCO), a programmable frequency divider with a divider ratio N, a phase-frequency detector/charge-pump combination (PFD/CP) and a loop filter [1, 2]. In addition, the architecture also comprizes a reference crystal oscillator and a reference frequency divider of ratio R. When the loop is locked the phase θ_{div} of the divided output signal f_{div} accurately tracks the phase θ_{ref} of the reference signal f_{ref}. The phase-lock process therefore forces

---

[1] The "type" of a PLL indicates the number of perfect integrators in the loop, and the order of a PLL indicates the total number of poles in the open- and closed-loop transfer functions.

the frequencies of f_{div} and f_{ref} to be equal. Relating F_{out} to f_{div} and f_{ref} one readily obtains

$$F_{out} = N \cdot f_{ref} = N \cdot \frac{f_{xtal}}{R}. \qquad (7.1)$$

If the division ratio N is programmable in steps of 1, then the frequency of the LO signal F_{out} can be stepped with a minimum step size equal to f_{ref}. If the transceiver needs to receive and/or transmit in another RF channel then a different division ratio N is programmed in the frequency divider, and the feedback loop adjusts the frequency of the VCO to the new value.

The basic limitation of the Integer-N PLL is that the reference frequency f_{ref} is equal to the minimum step size of the frequency synthesizer (see Section 7.3.2 below). This property can result in a very small value for f_{ref} in some applications. Such a situation is undesirable, as small f_{ref} values can lead to high values of spurious reference breakthrough and to a high phase noise contribution from the "PLL blocks", as will be presented in Section 7.6. The Fractional-N PLL architecture decouples the reference frequency from the minimum realizable step size [2]. Fractional-N techniques have received extensive attention in the recent literature [3, 4, 5, 6, 7, 8, 9], yet they will not be treated in this chapter. Instead, we will concentrate on system aspects and building block architectures which are equally valid and applicable to Integer-N and to Fractional-N PLL architectures.

7.3 TUNING SYSTEM SPECIFICATIONS

This section describes the main specification points of tuning systems which are employed within modern receiver and transceiver products. The specification points determine important parameters of a PLL frequency synthesizer implementation. Furthermore, the concept of base-band phase noise power spectral density is introduced in Section 7.3.5.

7.3.1 Tuning Range

The tuning range denotes the range of frequencies to be generated by the synthesizer, see Figure 7.2 and Section 6.2.1. The tuning range is a function of the RF input frequency range in receiver applications, and of the input and output frequency ranges in transceiver systems. It is usual to classify a tuning system as being either a narrow-range or a wide-range (or large-range) tuning system. Cordless and cellular telephones are examples of narrow-range systems, whereas terrestrial TV and AM broadcasting are systems which operate with wide tuning ranges.

Minimum tuning step

(frequency resolution)

Minimum tuning step:

System	kHz
AM	1
PAGER	12.5
TV	62.5
FM	100
GSM	200
SAT. TV	1000
Bluetooth	1000
DECT	1728

f_1 f_2 f_3 f_n
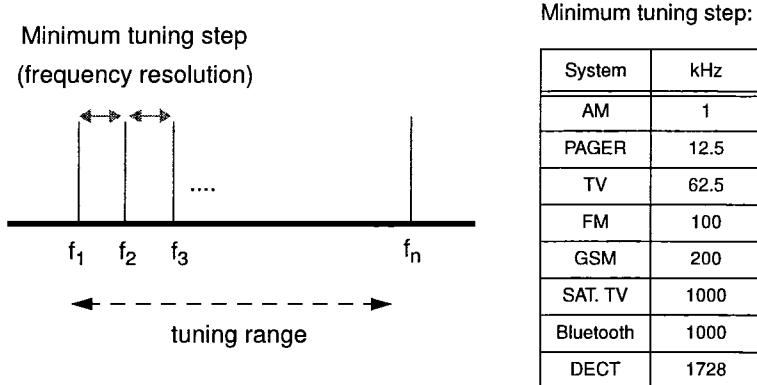
tuning range

Figure 7.2.   Tuning range and minimum frequency step.

7.3.2   Minimum Step Size

The minimum step size f_{min}, also known as frequency resolution, is the minimum frequency difference between two successive output frequencies that can be provided by the synthesizer. Figure 7.2 depicts the concept.

The required step size is dependent on the application, and it is very often equal to the frequency difference between two successive RF channels. With this situation, stepping the frequency of the synthesizer changes the received channel. Other applications require a minimum step size smaller than the channel width, for performing e.g. automatic frequency control (AFC) functions. For example, an undesired frequency offset can be detected between the frequency of the receiver input signal and the frequency generated by the synthesizer. The offset can then be decreased by generation of an output frequency which has a smaller offset with respect to the input frequency.

The required minimum step size value varies widely with the application. A few examples are given on the table inserted in Figure 7.2.

7.3.3   Settling Time

The settling time is the time necessary for the tuning system to settle within a frequency window delimited by $\pm f_{error}$ from the desired frequency f_{lock}, after a "change frequency" command has been received.[2] The settling time must be smaller than a certain locking time t_{lock}, which is a specification point defined by the intended application. The locking time t_{lock} is often defined for the largest frequency step, which is equal to the tuning range. An analysis of the settling

---

[2]That is, the division ratio N of the main divider in Figure 7.1 is reprogrammed.

behavior of a type-2 third-order PLL as a function of its open-loop bandwidth and phase-margin can be found in [10, 11].

The required locking time is a function of the application. For example, performing inaudible signal quality checks in a FM receiver equipped with Radio Data System (RDS) requires a synthesizer with a locking time smaller than 1 ms, defined as a residual settling error of 6 kHz for a 20 MHz frequency step [12]. Telecommunication systems which employ a combination of Time Division Duplex (TDD) and Frequency Division Duplex (FDD) techniques have the down-link frequencies (base station to hand-sets) placed in different bands as compared to up-link frequencies. In order to save cost and decrease the size of the hand-set, it is desirable to use the same frequency synthesizer to generate up-link and down-link frequencies. The settling requirements are that the tuning system has to switch between bands and settle to another frequency within a predetermined time ($\sim$ 1 ms for GSM and DCS-1800 systems [13]).

7.3.4 Spurious Signals

Spurious signals are undesired spectral components which appear at the output of the tuning system, in addition to the carrier signal with frequency f_{LO}. Figure 7.3, a typical output spectrum from a PLL frequency synthesizer, clearly shows the presence of spurious signals. In general, these signals originate either from unwanted coupling of signals present elsewhere in the PLL to the output node, or by modulation of the local oscillator by deterministic baseband signals.

Next, we will investigate the relationship of RF spurious signals to the phase modulation (PM) properties of an oscillator signal. The output signal S_{vco} of an oscillator can be generically expressed as follows

$$S_{vco} = A(t) \cos \left(2\pi f_{LO} t + \theta(t) \right), \tag{7.2}$$

where $A(t)$ is the amplitude of the VCO signal and $\theta(t)$ is the "excess phase", or the phase deviation with respect to the "ideal" phase of the signal, namely $2\pi f_{LO} t$. In most applications, the amplitude $A(t)$ of the VCO signal can be considered constant, i.e. free of undesired amplitude modulation. We consider next a signal with constant amplitude A_{LO} which is phase-modulated by a sine wave of frequency f_m

$$S_{vco} = A_{LO} \cos \left(2\pi f_{LO} t + \theta_p \sin 2\pi f_m t \right), \tag{7.3}$$

where θ_p is the peak phase deviation, also known as the modulation index β [14]. When $\theta_p \ll 1$, thus fulfilling the narrow band FM condition, S_{vco} can be
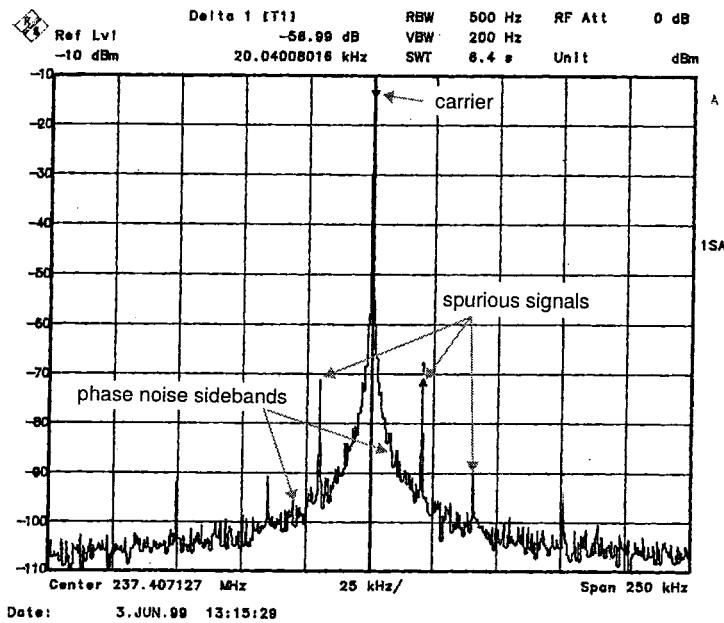
Figure 7.3.   Frequency source imperfections: spurious signals and phase noise sidebands.

approximated by [2, 14, 15]

$$S_{vco} = A_{LO}\left(\cos 2\pi f_{LO} t - \frac{\theta_p}{2} \cos 2\pi (f_{LO} + f_m)t + \frac{\theta_p}{2} \cos 2\pi (f_{LO} - f_m)t\right)$$

$$(7.4)$$

We see that the modulation process by a sine-wave of baseband frequency f_m generates a *pair* of frequency components — the spurious signals, in the present consideration — at a distance $\pm f_m$ from the carrier frequency f_{LO} (as seen, for instance, in Figure 7.3). Therefore, the offset frequency from the carrier equals the frequency of the modulating signal f_m. Equation (7.4) shows that the amplitude A_{sp} of the spurious signals is related to the amplitude of the carrier signal A_{LO} and to the peak phase deviation θ_p by

$$A_{sp} = A_{LO}\frac{\theta_p}{2}.$$

$$(7.5)$$

Conversely, we can conclude that the peak phase deviation θ_p associated with a pair of (PM) spurious signals with amplitude A_{sp} is given by

$$\theta_p = 2 \times \frac{A_{sp}}{A_{LO}}.$$

$$(7.6)$$

Equation (7.6) shows that the peak phase deviation θ_p does not depend on the absolute magnitude of the spurious signals A_{sp}. Instead, the peak deviation is determined by the ratio of the magnitudes of the spurs and the carrier signal.

The relative magnitude of undesired signal components in relation to the magnitude of the carrier is often expressed in dBc. Equation (7.6) can be modified to relate the peak phase deviation θ_p to the relative magnitude of a pair of spurious signals, each of magnitude a_{sp} (in dBc):

$$\theta_p = 2 \times 10^{\,a_{sp}/20}. \tag{7.7}$$

Spurious components which arise from undesired coupling of signals to the output node do not necessarily appear in pairs around the carrier, as they are not generated by a baseband modulation process. Nonetheless, it can be demonstrated that a "single" spurious signal produces AM *and* PM modulation of the carrier signal [15].

7.3.5 Phase Noise Sidebands

Phase noise sidebands is the common denomination given to the energy present in the power spectrum of (locked) oscillators, in addition to the carrier and to deterministic spurious signals, see Figure 7.3. Phase noise sidebands represent unintended phase modulation of the carrier signal [15], similar to the spurious signals described in the previous section; the phase modulation due to the phase noise sidebands is however an stochastic process.

In order to quantify the effect of the phase noise sidebands, the spectral components can be represented as a multitude of spurious signals (or sinusoids), each having the same average power as the phase noise — measured in the unity bandwidth — at the corresponding offset frequency from the carrier [15, 16]. A thorough discussion of the statistical assumptions and implications behind this model can be found in [15].

We have seen in Section 6.2.4 that the ratio of the power of a single-sideband (SSB) phase noise component at offset frequency f_m, measured in a 1 Hz bandwidth, to the total signal power is denoted $\mathcal{L}(f_m)$. In practice, the power of the carrier is taken as an approximation for the total signal power, so that $\mathcal{L}(f_m)$ can be written as

$$\mathcal{L}(f_m) = 10 \log \frac{v_{n,rms}^2(f_m)}{V_{c,rms}^2}, \quad \text{with } \mathcal{L}(f_m) \text{ in dBc/Hz}, \tag{7.8}$$

where $v_{n,rms}(f_m)$ is the rms value of the sinusoid representing the phase noise sideband at offset frequency f_m and $V_{c,rms}$ is the rms value of the carrier signal.

Conversely, the ratio of $v_{n,rms}(f_m)$ and $V_{c,rms}$ can be expressed as

$$\frac{v_{n,rms}(f_m)}{V_{c,rms}} = 10^{\mathcal{L}(f_m)/20}. \tag{7.9}$$

The sinusoidal representation for the phase noise sidebands enables a relationship between the (baseband) phase modulation of the carrier and the RF phase noise sidebands to be obtained. Applying the same reasoning as used for (7.6) and (7.7) leads to an equation describing the peak phase deviation $\psi_p(f_m)$ caused by phase noise sidebands at offset frequencies of $\pm f_m$ from the carrier. Similar to (7.6),

$$\psi_p(f_m) = 2 \times \frac{v_{n,rms}(f_m)}{V_{c,rms}} = 2 \times 10^{\mathcal{L}(f_m)/20}. \tag{7.10}$$

Note that $\psi_p(f_m)$ is the average *peak* phase deviation due to the RF phase noise sidebands at $\pm f_m$, which are expressed by their *rms* value. We see that $\psi_p(f_m)$ is a continuous function of f_m, and that the RF offset frequencies $\pm f_m$ are converted to the baseband variable f_m in a phase domain representation.

Phase noise calculations often require the phase noise power density $\phi_o^2(f_m)$ and its rms value $\phi_o(f_m)$ to be known. Equation (7.10) provides instead the average peak phase deviation due to the RF phase noise sidebands. With the peak phase deviation $\psi_p(f_m)$ related to the rms value as $\psi_p(f_m) = \sqrt{2} \times \phi_o(f_m)$ [15], we arrive at the following expression for $\phi_o(f_m)$:

$$\phi_o(f_m) = \frac{1}{\sqrt{2}} \psi_p(f_m)$$

$$= \sqrt{2} \times 10^{\mathcal{L}(f_m)/20} \quad [\text{rad}/\sqrt{\text{Hz}}]. \tag{7.11}$$

The baseband phase noise power spectral density $\phi_o^2(f_m)$ is therefore related to $\mathcal{L}(f_m)$ (in dBc/Hz) as follows

$$\phi_o^2(f_m) = 2 \times 10^{\mathcal{L}(f_m)/10} \quad [\text{rad}^2/\text{Hz}], \tag{7.12}$$

and consequently

$$\mathcal{L}(f_m) = 10 \log \left(\frac{\phi_o^2(f_m)}{2} \right) \quad [\text{dBc/Hz}]. \tag{7.13}$$

Equations (7.12) and (7.13) are of fundamental importance for the treatment, calculation and simulation of phase noise in PLLs and tuning systems. These relationships provide a direct conversion between RF and baseband representations as depicted in Figure 7.4. Furthermore, (7.12) enables calculation of integrated residual phase deviation over a given (baseband) bandwidth.

LO signal power spectral density (1Hz BW)

LO signal in baseband phase noise domain

$\phi_o{}^2$ (f_m) [rad^2/Hz]

$\mathcal{L}(f_m)$ dBc/Hz

f_{lo} -> 0 Hz

f_m

f_{LO} f_{RF}

0   f_m
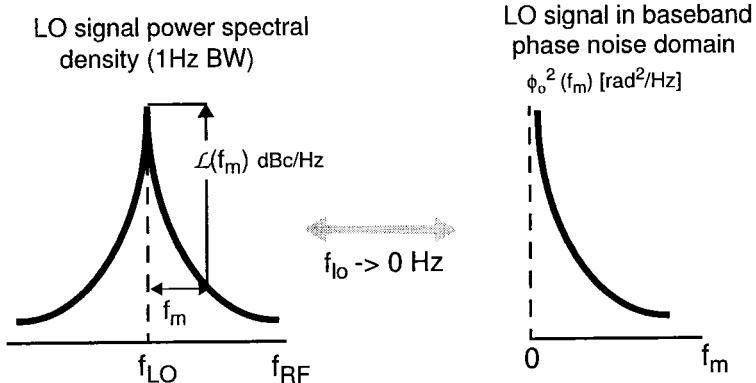
Figure 7.4. Conversion of given RF power spectrum to and from baseband phase noise power densities.

7.4 SYSTEM-LEVEL ASPECTS OF PLL BUILDING BLOCKS

7.4.1 Voltage Controlled Oscillators

The voltage controlled oscillator (VCO) generates the output signal of the PLL. As discussed in Chapter 6, the frequency of the VCO signal is dependent on the voltage V_{tune} at its tuning input. The relationship between the output frequency F_{out} and the tuning voltage V_{tune} can be written as $F_{out} = f_{center} + K_{vco}(V_{tune}) \cdot V_{tune}$, where $K_{vco}(V_{tune})$ is the VCO gain factor in [Hz/V] and f_{center} is the output frequency when the tuning voltage is 0 Volt.

The control action of the loop is based on a phase error signal, namely the difference of the θ_{div} and θ_{ref} signals in Figure 7.1. The relationship of the phase θ_o of the VCO signal with the tuning voltage V_{tune} can be derived as follows

$$\theta_o(t) = \int 2\pi F_{out}(t) \, dt$$

$$= \int 2\pi \left(f_{center} + K_{vco}(V_{tune}) \cdot V_{tune}(t) \right) dt. \qquad (7.14)$$

Dropping the first term of the integral, which is not dependent on V_{tune}, results in

$$\theta_o(t) = \int 2\pi K_{vco}(V_{tune}) \cdot V_{tune}(t) \, dt. \qquad (7.15)$$

Equation (7.15) shows that the phase of the VCO signal represents a perfect integration of the control signal V_{tune}.

After phase—and frequency—lock is achieved, the DC value of V_{tune} is (nearly) constant, so that the dependency of K_{vco} on V_{tune} can be neglected.

Taking the Laplace transform of (7.15) yields

$$\theta_o(s) = \frac{2\pi K_{vco} \cdot V_{tune}(s)}{s}. \tag{7.16}$$

7.4.2 Frequency Dividers

Digital frequency dividers are responsible for frequency scaling in the loop (see *main div* and *ref. div* in Figure 7.1). For example, the frequency f_{div} of the output signal of the main divider equals the frequency f_{in} of its input signal divided by an integer number.[3] The effect of the frequency division on the phase relationship between input and output signals is derived next. We will calculate the effect of frequency division on a signal which is phase modulated by a sine wave of frequency f_m. From this information we may derive a model for the frequency divider in the phase domain.

The phase θ_{in} of the input signal is given by

$$\theta_{in}(t) = 2\pi f_{in}t + \theta_p \sin 2\pi f_m t, \tag{7.17}$$

and the instantaneous frequency of the input signal is

$$f_{inst}(t) = \frac{1}{2\pi} \frac{d\theta_{in}(t)}{dt} = f_{in} + \theta_p f_m \cos 2\pi f_m t.$$

The frequency of the output signal, after division by an integer N, can be expressed as

$$f_{div} = \frac{f_{inst}}{N} = \frac{f_{in}}{N} + \frac{\theta_p f_m \cos 2\pi f_m t}{N},$$

and the phase of the output signal can now be found as

$$\theta_{div}(t) = \int 2\pi f_{div}(t)\, dt$$

$$= 2\pi \frac{f_{in}}{N}t + \frac{\theta_p}{N} \sin 2\pi f_m t, \tag{7.18}$$

which can also be expressed as

$$\theta_{div}(t) = \frac{\theta_{in}(t)}{N}, \tag{7.19}$$

with $\theta_{in}(t)$ as defined in (7.17).

---

[3] Frequency dividers that can divide by half-integer numbers have recently been reported [17]. The draw-back of the proposed implementation is high power dissipation, due to the increased complexity of the circuitry in the high frequency part of the divider.

Equation (7.18) has two terms. The first term represents the frequency scaling of the "average" input frequency f_{in} to f_{in}/N. The second term shows that the peak phase deviation θ_p is reduced ("divided") in proportion to the division ratio N. The modulation frequency f_m, on the other hand, is *not affected* by the division process. An important conclusion from (7.18) is that the frequency divider's phase transfer function $\theta_{div}(t)/\theta_{in}(t)$ is simply a gain factor with value $1/N$.

7.4.3 Phase-Frequency Detector/Charge-Pump Combination

The phase detector compares the phase of the f_{ref} and f_{div} signals and generates an error signal which is proportional to their phase difference, see Figure 7.1. Nowadays the most commonly used topology, within the context of PLL frequency synthesizers, is the sequential phase-frequency detector (PFD). The sequential phase-frequency detector (PFD) ensures frequency and phase-lock by itself, irrespective of the initial frequency error of the VCO. Besides, the output of the PFD is only active during a small fraction of the reference period (this behavior is elucidated in more detail in the next sub-section.) The small duty-cycle effectively attenuates the PFD noise contribution to the loop. A further advantage is that the spurious breakthrough generation of the phase-frequency detector is minimal, as it only delivers the amount of energy necessary to compensate for leakage currents in the loop filter or at the input of the op-amp, when an active loop filter configuration is used (see, for example, Figure 7.10).

The block diagram of a common implementation of the phase-frequency detector is presented in Figure 7.5. It consists of two D-type flip-flops (D-FF) which have their D inputs connected to the active level.[4]

The upper D-FF, which is clocked by f_{ref}, generates the *up* signal. The lower D-FF is clocked by f_{div}, and generates the *down* signal. The AND gate monitors the *up* and *down* signals, and generates the *reset* signal for the D-FFs at the moment both outputs become active. The *up* and *down* signals are used to switch the current sources in the charge-pump CP. When *up* is active, a current with magnitude of I_{cp} is sourced by the charge-pump; conversely, when *down* is active, current is sinked into the charge-pump. When both *up* and *down* are inactive, no current flows into or out the output node of the charge-pump. The output is a high impedance node, under all circumstances.

---

[4]There are many circuit implementations of a PFD which provide the same functionality [1, 18, 19, 2]. The main advantage of the D-FF based implementation is its compactness.
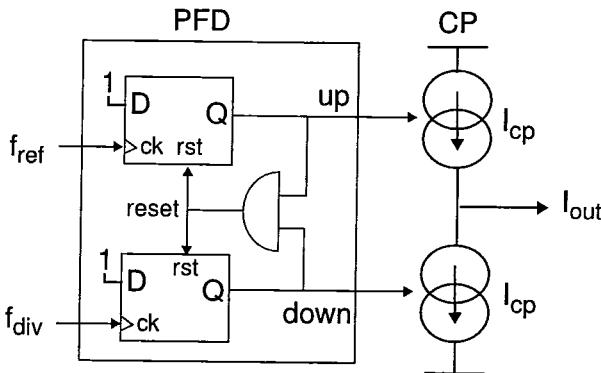
Figure 7.5.   Sequential phase-frequency detector combined with single-ended charge-pump.

Polarity of the Feedback Signal

The polarity of the output pulses of the charge-pump must be such that negative feedback in the loop is insured for each *specific* combination of VCOs, PFD/CPs and loop filter topologies. For example, the configuration depicted in Figure 7.5 suits a loop with a loop filter which does not invert the polarity of the input signal, and where the VCO has a positive K_{vco} gain (i.e., the oscillator frequency increases with an increased tuning voltage). The use of the configuration as depicted in Figure 7.5 with a negative gain oscillator *or* with an inverting loop filter (e.g. an active integrator) results in a *latch-up* situation, as the polarity of the resulting control-signal will in fact pull the loop out-of-lock. Alternately, the combination of a negative gain oscillator *and* an inverting loop filter again requires the configuration of Figure 7.5 to be stable. There are situations where the polarity of the oscillator gain or the type of loop filter are not a priori known. In that case, two (programmable) switches in between the PFD and the frequency dividers provide an elegant solution to the problem. The switches can be used to interchange the connections from the dividers to the PFD as a function of the polarity of the VCO gain in each specific situation.

Time-domain Operation

The time-domain operation of the PFD/CP combination will be elucidated with help of Figure 7.6. On the left side of Figure 7.6 we see the situation where the active (rising) edges of f_{div} and f_{ref} arrive simultaneously at the PFD. The input signals cause the *up* and *down* signals to become active at the same moment. The AND gate reacts to the simultaneous presence of *up* and *down* by generating the *reset* signal for the D-FFs, whereupon *up* and *down* become inactive again. We notice that there is no output current I_{out} from the charge-
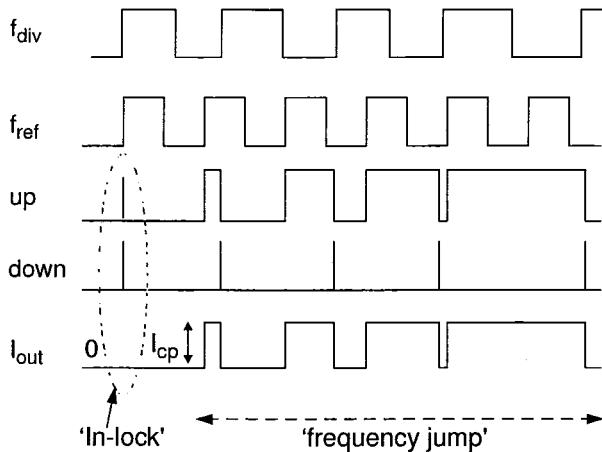
Figure 7.6.   Operation of sequential phase-frequency detector.

pump in this situation, as (ideally) the current sourced in response to the *up* signal is perfectly compensated by the current sinked in response to the *down* signal. This is a desirable behavior for an "in-lock" situation, as the absence of an output signal from the charge-pump avoids spectral purity degradation of the VCO.

It is important to remark that the *up*, the *down* and the *reset* signals do have a minimum width when the loop is phase-locked. In fact, the frequency-detection characteristics of the PFD for high-frequency input signals are *determined* by the width of the *up*, the *down* and the *reset* signals in phase-lock [18]. This aspect will be treated in more detail later in this section.

A large portion of Figure 7.6 is marked as "frequency jump". This part of the diagram represents the situation where the frequency of f_{div} becomes smaller than the frequency of f_{ref}. This might happen due to e.g. an increase of the division ratio N in the feedback loop, whereupon the PLL must react in such a way as to increase the frequency of the VCO, so that eventually frequency and phase lock between f_{ref} and f_{div} are restored. We see that the rising edge of f_{ref}, which sets *up* active, arrives earlier than the rising edge of f_{div}. The *up* signal stays active until the next rising edge of f_{div}. Now there is a net output current from the charge-pump. This current can be used to build up a VCO control voltage which at last will bring the frequency and phase of f_{div} to match those of f_{ref} again.

We observe, in the example of Figure 7.6, that the duty-cycle of the *up* and I_{out} signals grow in proportion to the phase difference $\Delta\theta = \theta_{ref} - \theta_{div}$ of the input signals. The relationship of the duty cycles δ_{up} and δ_{down} to the phase
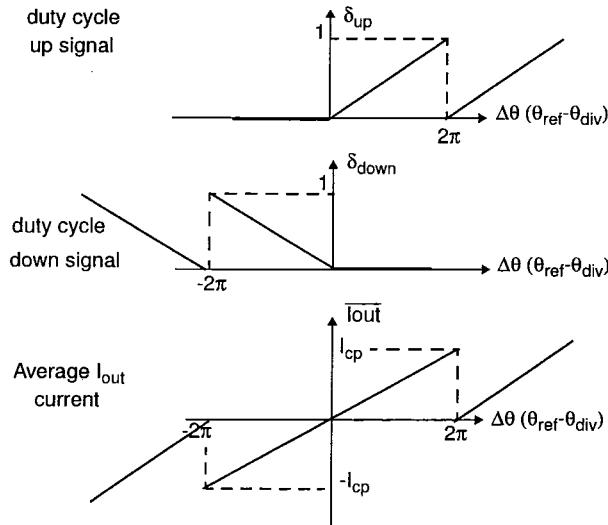
Figure 7.7. Duty cycle of the up and down signals as a function of the phase difference at the input of the PFD. The lower curve is the phase-to-current transfer of the PFD/CP combination.

difference $\theta_{ref} - \theta_{div}$ is shown in Figure 7.7. The duty cycle grows linearly with $\Delta\theta$, reaching a value of 1 for an absolute phase error of 2π. The average current $\overline{I_{out}}$ is also depicted in Figure 7.7. $\overline{I_{out}}$ reaches the nominal value of the charge-pump currents $\pm I_{cp}$ when $\Delta\theta$ equals $\pm 2\pi$ rad. The relationship between $\overline{I_{out}}$ and $\Delta\theta$ can therefore be written as follows

$$\overline{I_{out}} = I_{cp} \frac{\Delta\theta}{2\pi}$$

The gain K_{pd} of the PFD/CP combination, defined as the average charge pump output current for a given phase difference at the input of the PFD, can then be simply expressed as

$$K_{pd} = \frac{\overline{I_{out}}}{\Delta\theta} = \frac{I_{cp}}{2\pi} \quad [\text{A/rad}] \tag{7.20}$$

Equation (7.20) suffices to describe the behavior of the PFD/CP in a linearized time-continuous model, for the time being. In Section 7.8 we will find out that these blocks add parasitic poles to the loop transfer function. The effect of the parasitic poles is the largest in PLLs with large loop bandwidth.

High-Frequency Limitations of the Phase-Frequency Detector

The effect of the finite reset time for the logical elements composing the PFD leads to an upper-limit for its frequency discrimination capability. It has been
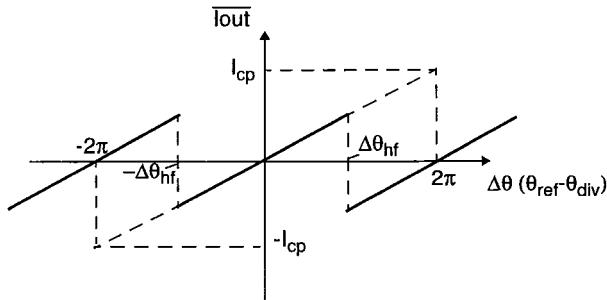
Figure 7.8.   High-frequency phase-to-current transfer of a sequential phase-frequency detector.

demonstrated in [18] that the maximum frequency $f_{ref,max}$ at which the frequency difference of the input signals can be discriminated is related to the reset time ΔR of the D-FFs when the loop is in phase-lock, as

$$f_{ref,max} = \frac{1}{2\Delta R}. \tag{7.21}$$

Equation (7.21) is somewhat un-practical because the reset time ΔR includes the propagation time inside the flip-flops and the delay of the logical elements. A simpler way to evaluate $f_{ref,max}$ is with help of the (high-frequency) phase-to-current transfer of a given PFD/CP circuit implementation. Figure 7.8 depicts the concept schematically. The effect of the finite reset time is to set an upper limit to the maximum phase-difference $\Delta\theta_{hf}$ which can be detected before the PFD/CP output erroneously switches polarity. The maximum phase-difference $\Delta\theta_{hf}(f_{in})$ can be expressed as a function of the period of the input signals $T_{in} = 1/f_{in}$ as follows:

$$\Delta\theta_{hf}(f_{in}) = \pm 2\pi \left(1 - \frac{\Delta R}{T_{in}}\right). \tag{7.22}$$

Combining (7.21) and (7.22) with $T_{in} = 1/f_{ref,max}$ shows that

$$\Delta\theta_{hf}(f_{ref,max}) = \pm\pi, \tag{7.23}$$

which means that the phase-frequency detector must have a linear phase detection range that goes beyond $\pm 180°$ at the highest operation frequency of interest. Practice shows that failure to comply to this requirement often translates itself in situations of permanent frequency-lock at *wrong* frequencies.[5]

---

[5]This phenomenon is sometimes referred to as a "false-lock" condition.

Spectral Components of the Charge-Pump Output Signal

Let us now calculate the spectral components of the output signal I_{out}, as a function of the phase error $\Delta\theta$ between f_{ref} and f_{div}. In the following analysis it is assumed that the output of the charge-pump consists of current pulses of amplitude I_{cp}, and that there is no mismatch between the current sources of Figure 7.5. Mismatch in the current sources is discussed in Section 7.6.1.

The duty cycle δ_{cp} of the output pulse equals $\Delta\theta/2\pi$, as depicted in Figure 7.7. Conversely, δ_{cp} can be written as τ/T_{ref}, where τ is the active time of the charge-pump output, and T_{ref} is the period of the reference signal. The Fourier series expression for a periodic train of pulses of amplitude I_{cp} and duration τ is [14]:

$$I_{out}(t) = \frac{I_{cp}\tau}{T_{ref}} + \frac{2I_{cp}\tau}{T_{ref}} \sum_{n=1}^{\infty} \frac{\sin(n\pi\tau/T_{ref})}{n\pi\tau/T_{ref}} \cos\frac{2\pi nt}{T_{ref}}. \qquad (7.24)$$

The equation above can be expressed as a function of the phase error $\Delta\theta$

$$I_{out}(t) = I_{cp}\frac{\Delta\theta}{2\pi} + 2I_{cp}\frac{\Delta\theta}{2\pi} \sum_{n=1}^{\infty} \frac{\sin(n\pi\frac{\Delta\theta}{2\pi})}{n\pi\frac{\Delta\theta}{2\pi}} \cos\frac{2\pi nt}{T_{ref}}, \qquad (7.25)$$

and also as a function of the duty cycle δ_{cp}

$$I_{out}(t) = I_{cp}\delta_{cp} + 2I_{cp}\delta_{cp} \sum_{n=1}^{\infty} \frac{\sin(n\pi\delta_{cp})}{n\pi\delta_{cp}} \cos\frac{2\pi nt}{T_{ref}}. \qquad (7.26)$$

For small values of duty cycle δ_{cp} the sinc function $\sin(n\pi\delta_{cp})/(n\pi\delta_{cp})$ can be approximated as unity. This results in the following, simplified expression for I_{out}:

$$I_{out}(t) = I_{cp}\delta_{cp} + 2I_{cp}\delta_{cp} \sum_{n=1}^{\infty} \cos\left(2\pi nf_{ref}t\right), \qquad (7.27)$$

which shows that the amplitude of the spectral components of I_{out}, namely the reference frequency f_{ref} and its harmonics, are twice as large as its DC value $I_{cp}\delta_{cp}$. Therefore, if $\delta_{cp} = \Delta\theta/2\pi$ equals zero the charge-pump output theoretically contains no DC or AC signal components whatsoever.

7.4.4 Loop Filter

The loop filter provides the current-to-voltage conversion from the charge-pump signal to the tuning voltage input of the VCO. The purity of the tuning voltage determines to a great extent the spectral components of the output signal F_{out}.

Equation (7.25) shows that the output of the charge-pump vanishes when the phase difference between the input signals of the PFD is zero. This is the ideal locking position for the loop: there is no current injection into the loop filter, and therefore no degradation of spectral purity performance of the VCO. Phase lock with zero phase error, for all possible output frequencies, requires a loop filter with infinite DC gain [1]. In other words, the loop filter must perform an integration operation on the charge-pump output signal. As the pump output is a node with an ideally infinite output impedance, a simple capacitor suffices to realize the integration function at the loop filter.

On the other hand, the loop "already" contains a perfect integrator—the VCO. So, the addition of another perfect integrator in the loop's transfer function leads to instability and oscillatory behavior, unless further measures are taken. In order to increase the loop's phase margin, very often a resistance is placed in series with the integrator capacitor. This adds a zero to the trans-impedance function of the loop filter $Z_f(s)$. Hence, the RC combination causes a phase advance in the PLL open-loop response, potentially solving the stability problem (provided that the open-loop bandwidth f_c is located in the region with the phase advance, as depicted in Figure 7.12). The RC combination is the simplest loop filter topology which yields a stable PLL output signal.

Very often the PLL has to cope with DC leakage currents in the tuning line of the VCO. The loop reacts to the leakage current by increasing the duty cycle δ_{cp} of the charge-pump output signal I_{out}. Equation (7.27) shows that the magnitude of the AC signal components of I_{out} grow in direct proportion to the duty cycle. These undesired signal components are converted to the voltage domain by the loop filter, and the resulting voltage ripple on the tuning line generates spurious signals at the output of the VCO. As a consequence, the minimum loop filter configuration found in practice includes an additional capacitor in parallel to the RC section (or in parallel to the resistance R). The purpose of this extra capacitor is to decrease the loop filter trans-impedance for higher frequencies, and therefore to decrease the magnitude of the ripple voltage for a given value of DC leakage current.

Based on the exposition of the previous paragraphs, we can write the transimpedance transfer function of the loop filter as[6]

$$Z_f(s) = \frac{k}{s} \frac{1 + s\tau_2}{1 + s\tau_3}$$
$$= \frac{k}{s} \frac{1 + s\tau_2}{1 + s\tau_2/b}, \tag{7.28}$$

---

[6]The notation for the time constants is consistent with [20].

LF1   LF2

Param.	LF1	LF2
τ_2	$R_1(C_1+C_2)$	R_1C_1
τ_3	R_1C_2	$R_1 \dfrac{C_1C_2}{C_1+C_2}$
k	$\dfrac{1}{C_1}$	$\dfrac{b\text{-}1}{b}\dfrac{1}{C_1}$
$b\ (\tau_2/\tau_3)$	$1 + \dfrac{C1}{C2}$	$1 + \dfrac{C1}{C2}$
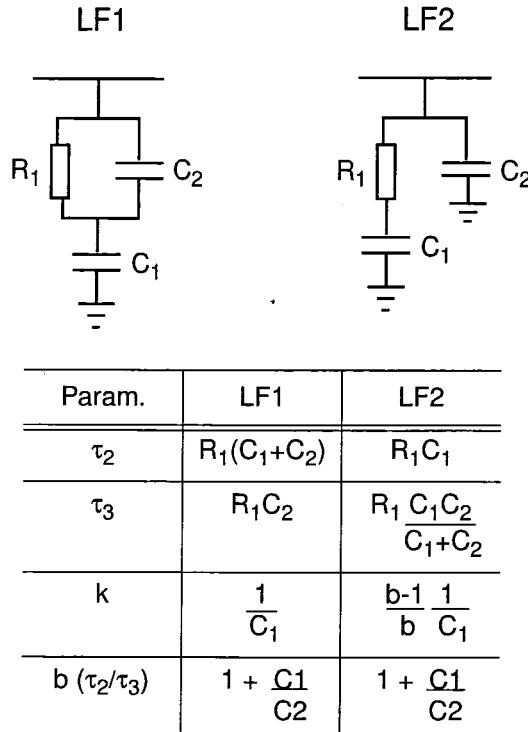
Figure 7.9. Two common passive loop filter topologies, and corresponding relationships between design parameters and component values.

where k is a gain factor which depends on the specific configuration of the loop filter, τ_2 is the time constant of the "stabilizing" zero, τ_3 is the time constant of the pole which is used to attenuate the reference frequency and its harmonics, and b is the ratio of the time constants τ_2/τ_3.

Passive Loop Filters

Two passive loop filter configurations which comply to (7.28) are shown in Figure 7.9. The trans-impedance functions $Z_{f1}(s)$ and $Z_{f2}(s)$, for loop filters LF1 and LF2 respectively, are given below:

$$Z_{f1}(s) = \frac{1 + s(R_1(C_1 + C_2))}{sC_1(1 + sR_1C_2)}$$

$$Z_{f2}(s) = \frac{1 + s(R_1C_1)}{s(C_1 + C_2)(1 + sR_1\frac{C_1C_2}{C_1+C_2})} \tag{7.29}$$
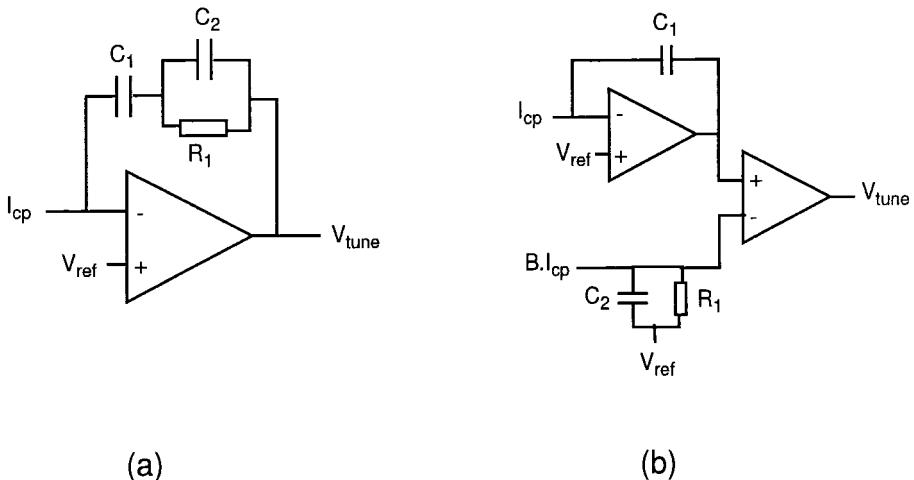
Figure 7.10.   Active loop filter implementations.

Configuration LF2 is preferable for fully integrated loop filter applications, as the bottom plates of C_1 and C_2 are both grounded. This property eliminates the possibility of substrate noise coupling into the filter's output node through the parasitic capacitance of C_2 to the substrate. The tuning voltage for the VCO is therefore cleaner, and the risk of phase noise degradation due to substrate noise is minimized.

Active Loop Filters

A counterpart for the passive loop filters are active filter implementations, which are displayed in Figure 7.10 (a) and (b). The use of active loop filters results in increased complexity and power dissipation, and introduces additional noise sources in the loop. On the other hand, there are situations where active loop filters are either unavoidable, or where they lead to a smaller chip-area for fully integrated loop filters.

The active loop filter configuration displayed in Figure 7.10(a) is often used when the charge-pump output can not directly provide the required voltage range for tuning of the VCO. For example, wide tuning range applications, such as terrestrial TV and satellite reception, often require tuning voltages of up to 33 V [21]. Such voltages are incompatible with charge-pumps built in standard IC technologies, so that a (partly external) active loop filter is used to isolate the charge-pump output from the VCO tuning input and to generate the high tuning voltages [22].

loop open here
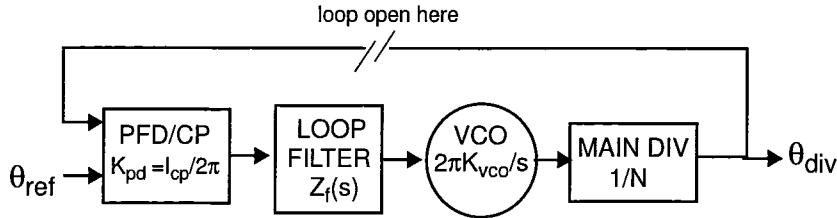
Figure 7.11.   Linear model of a PLL.

The loop filter depicted in Figure 7.10(b) is employed for a different purpose, namely to decrease the size of the capacitance C_1 in fully integrated loop filters

uration of Figure 7.10 (b), the trans-impedance function has a zero $(1/\tau_2)$ at a frequency of $\approx 1/(B R_1 C_1)$, where B is the ratio of the output currents of two charge-pumps, the one connected to the active integrator built around C_1 and the other to the parallel combination of R_1 and C_2. Therefore, the physical size of C_1 can be traded-off against the value of B. The added design freedom can be advantageous in applications where the loop filter is fully integrated.

The main draw-back of an active loop filter is the additional noise added to the loop. Its noise contribution can be designed to be non-dominant, yet at the cost of increased total power dissipation. This consideration is especially important for low-power applications, because the loop filter's circuitry can become the dominant source of power dissipation [24].

7.5 DIMENSIONING OF THE PLL PARAMETERS — BASIC CONCEPTS

7.5.1 Open- and Closed-loop Transfer Functions G(s) and H(s)

The loop considered here consists of a VCO of gain K_{vco} [Hz/V], a programmable frequency divider with a divider ratio N, a phase-frequency detector/single-ended charge-pump combination (PFD/CP) with a combined gain K_{pd} of $I_{cp}/2\pi$ (where I_{cp} is the nominal charge-pump current), and a loop filter with a trans-impedance transfer function $Z_f(s)$. When the loop is locked, the phase of the divided output signal θ_{div} accurately tracks the phase of the reference signal θ_{ref}. A linear, phase domain model for the loop is shown in Figure 7.11. The model can be obtained by combination of (7.16), (7.19) and (7.20), with $Z_f(s)$ the (trans)impedance of the loop filter.
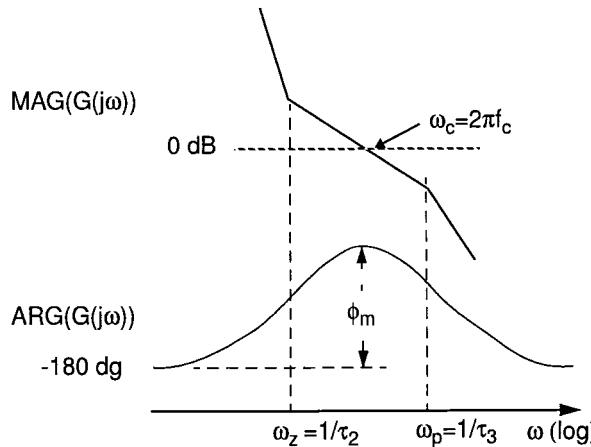
Figure 7.12.   The open-loop bandwidth f_c and the phase margin ϕ_m.

The open-loop transfer function $G(s) = \theta_{div}(s)/\theta_{ref}(s)$ is expressed as

$$G(s) = K_{pd} Z_f(s) \frac{2\pi K_{vco}}{s} \frac{1}{N}$$

$$= \frac{I_{cp}}{2\pi} Z_f(s) \frac{2\pi K_{vco}}{s} \frac{1}{N}, \qquad (7.30)$$

and the closed-loop transfer function $H(s) = \theta_{div}(s)/\theta_{ref}(s)$ as

$$H(s) = \frac{G(s)}{1 + G(s)}$$

$$= \frac{2\pi K_{pd} Z_f(s) K_{vco}/N}{s + 2\pi K_{pd} Z_f(s) K_{vco}/N}, \qquad (7.31)$$

which shows that $H(j2\pi f_m)$ has a low-pass transfer character.

7.5.2   Open-loop Bandwidth f_c and Phase Margin ϕ_m

For the design of the loop parameters we shall adopt the concepts of *open-loop bandwidth* f_c and *phase margin* ϕ_m. The open-loop bandwidth is defined by the condition

$$|G(j2\pi f_c)| = 1.$$

Note that f_c is also known as the 0 dB cross-over frequency. The phase margin is defined as

$$\phi_m = \arg\left(G(j2\pi f_c)\right) + \pi.$$

These concepts are depicted graphically in Figure 7.12.

The open-loop gain equation $G(j\omega)$ can be written as

$$
G(j\omega) = -\frac{I_{cp}K_{vco}k}{N\omega^2}\frac{1+j\omega\tau_2}{1+j\omega\tau_3}
$$
$$
= -\frac{I_{cp}K_{vco}k}{N\omega^2}\frac{1+j\omega\tau_2}{1+j\omega\tau_2/b},\tag{7.32}
$$

where k is a function of the of the loop filter used, see Figure 7.9.

The condition for having an open-loop bandwidth of ω_c is that $|G(j\omega_c)| = 1$. The open-loop transfer function from (7.32) yields

$$
|G(j\omega_c)| = \frac{I_{cp}K_{vco}}{kN\omega_c^2}\frac{|1+j\omega_c\tau_2|}{|1+j\omega_c\tau_3|} = 1.\tag{7.33}
$$

Solving (7.33) for I_{cp}, and incorporation of the constant k and of the time constants τ_2 and τ_3 from Figure 7.9 provide expressions for the charge-pump current as a function of the open-loop bandwidth $\omega_c = 2\pi f_c$ in [rad/s], with K_{vco} in [Hz/V]:

- for loop filter LF1:

$$
I_{cp} = \frac{C_1 N\omega_c^2}{K_{vco}}\frac{\sqrt{1+(\omega_c R_1 C_2)^2}}{\sqrt{1+(\omega_c R_1(C_1+C_2))^2}}.\tag{7.34}
$$

- for loop filter LF2:

$$
I_{cp} = \frac{(C_1+C_2)N\omega_c^2}{K_{vco}}\frac{\sqrt{1+\left(\omega_c R_1 \frac{C_1 \cdot C_2}{C_1+C_2}\right)^2}}{\sqrt{1+(\omega_c R_1 C_1)^2}}.\tag{7.35}
$$

The phase of the transfer function $G(j\omega)$ from (7.32) will be denoted as $\Psi(j\omega)$

$$
\Psi(j\omega) = -\pi + \arg(1+j\omega\tau_2) - \arg(1+j\omega\tau_3)
$$
$$
= -\pi + \arctan\omega\tau_2 - \arctan\omega\tau_3,\tag{7.36}
$$

and the point of zero derivative of the phase response will be called ω_{max}. This frequency corresponds to the maximum value of the $\Psi(j\omega)$ function, and therefore to the maximum (potential) value of phase margin for given values of τ_2 and τ_3.[7] The frequency ω_{max} can be found by differentiation of (7.36), and

---

[7] In reality, the finite output impedance of the charge-pump causes the phase of the transfer function at near DC to be $-90°$, and not $-180°$ as obtained with the assumption of an infinite output impedance for the CP. In this sense, the frequency ω_{max} corresponds to a local maximum in the phase of the transfer function $G(j\omega)$.
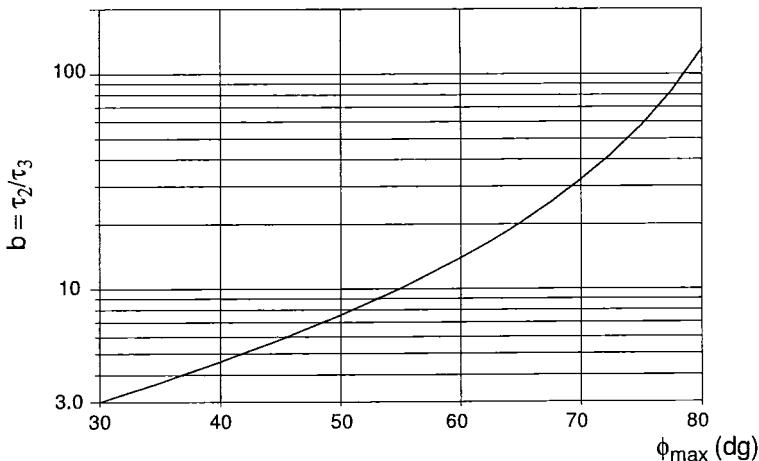
Figure 7.13. Value of the ratio of the time constants $b = \tau_2/\tau_3$ as a function of the maximum phase advance of $G(j\omega)$. The curve is asymptotical to $90°$.

by equating the resulting expression to zero. These operations provide

$$\omega_{max} = \sqrt{\frac{1}{\tau_2 \tau_3}}. \tag{7.37}$$

So, the frequency of maximum phase advance lies at the geometrical average of the inverse of time constants τ_2 and τ_3. The value of the maximum phase advance $\phi_{max} = \Psi(j\omega_{max}) + \pi$ at $\omega = \omega_{max}$ can be calculated as a function of τ_2, τ_3 and $b = \tau_2/\tau_3$ by inserting (7.37) in (7.36)

$$\phi_{max} = \arctan \frac{\tau_2 - \tau_3}{2\sqrt{\tau_2 \tau_3}}$$

$$= \arctan \frac{b - 1}{2\sqrt{b}}. \tag{7.38}$$

Solving (7.38) for b as a function of ϕ_{max} yields

$$b = \frac{1}{(-\tan \phi_{max} + 1/\cos \phi_{max})^2} \tag{7.39}$$

The numerical values of b as a function of ϕ_{max} are plotted in Figure 7.13.

If ω_c is dimensioned to be equal to ω_{max}, then the phase margin ϕ_m equals the value of ϕ_{max} given by (7.38) above. In the following derivations it is assumed that $\omega_c = \omega_{max}$ and therefore $\phi_m = \phi_{max}$. Now we can find τ_2 and τ_3 as a

function of b and ω_c. Equation (7.37) with $\omega_{max} = \omega_c$ directly results in

$$\tau_2 = \frac{\sqrt{b}}{\omega_c}$$

$$\tau_3 = \frac{1}{\sqrt{b}\omega_c} \tag{7.40}$$

The PLL open-loop bandwidth frequency $\omega_c = 2\pi f_c$ under the condition that $\omega_c = \omega_{max}$ is expressed below. Replacement of the time constants as given by (7.40) into (7.33) yields

- for loop filter LF1:

$$\omega_c = \frac{I_{cp}K_{vco}}{N} R_1 \frac{b}{b-1}$$
$$= \frac{I_{cp}K_{vco}}{N} R_1 \frac{C_1 + C_2}{C_1} \tag{7.41}$$

- for loop filter LF2:

$$\omega_c = \frac{I_{cp}K_{vco}}{N} R_1 \frac{b-1}{b}$$
$$= \frac{I_{cp}K_{vco}}{N} R_1 \frac{C_1}{C_1 + C_2} \tag{7.42}$$

The next step is to calculate the value of the loop filter components as depicted in Figure 7.9. The results of the calculations are given in (7.43) and (7.44) below:

- for loop filter LP1:

$$R_1 = \frac{2\pi N \omega_c}{I_{cp}2\pi K_{vco}} \frac{b-1}{b}$$
$$C_1 = \frac{\tau_2 - \tau_3}{R_1}$$
$$C_2 = \frac{\tau_3}{R_1} \tag{7.43}$$

ϕ_m	b	sqrt(b)
20	2.04	1.42
30	3.00	1.73
40	4.59	2.14
45	5.82	2.41
50	7.55	2.74
55	10.06	3.17
60	13.93	3.73
70	32.16	5.67
80	130.64	11.43
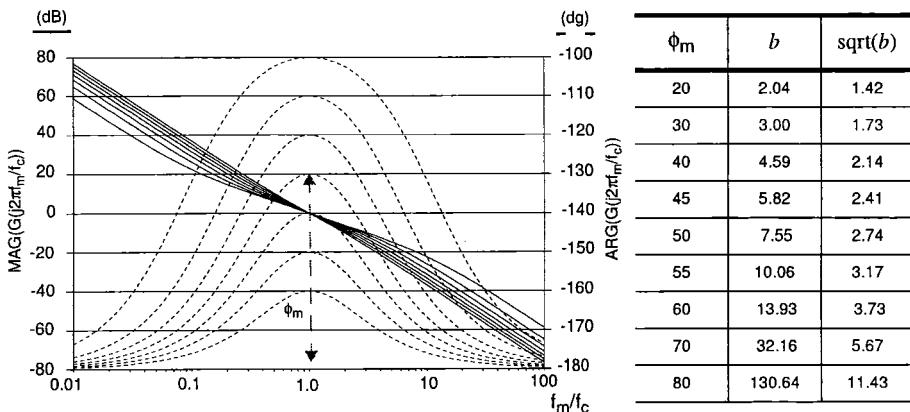
Figure 7.14. Open-loop frequency transfer for different values of phase margin ϕ_m. The table shows the relationship of ϕ_m to $b = \tau_2/\tau_3$ and to $\sqrt{b}$.

ϕ_m	peaking (dB)	f_{0dB}/f_c	f_{-3dB}/f_c
30	6.05	1.45	1.69
40	3.95	1.38	1.70
50	2.59	1.23	1.66
60	1.70	1.00	1.56
70	1.07	0.72	1.39
80	0.55	0.46	1.18
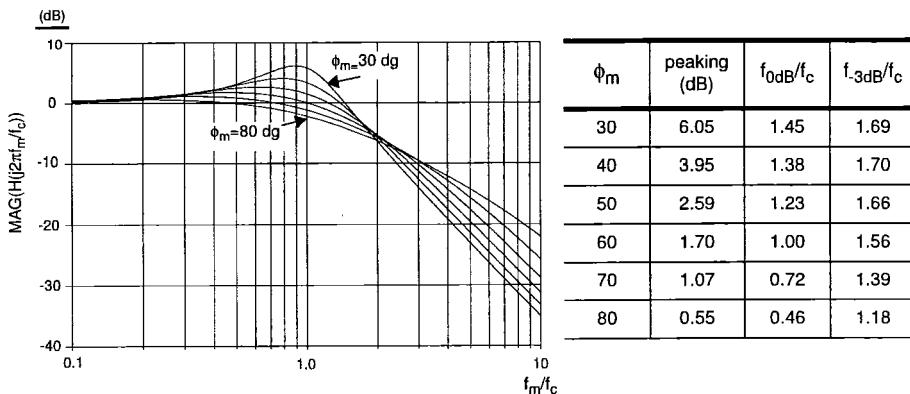
Figure 7.15. Closed-loop frequency transfer for different values of phase margin ϕ_m. The table shows the peaking (in dB), and the ratio of the 0 dB and -3 dB frequencies to f_c.

- for loop filter LP2:

$$R_1 = \frac{2\pi N\omega_c}{I_{cp}2\pi K_{vco}}\frac{b}{b-1}$$

$$C_1 = \frac{\tau_2}{R_1}$$

$$C_2 = \frac{1}{R_1}\frac{\tau_2\tau_3}{\tau_2-\tau_3} \tag{7.44}$$

The calculated magnitude and phase transfers of the open-loop transfer function $G(j\omega)$ are plotted in Figure 7.14, for different values of phase margin ϕ_m

and normalized to the value of the open-loop bandwidth f_c. The inserted table shows the relationship of ϕ_m to b and to $\sqrt{b}$.

Figure 7.15 shows the magnitude of the closed-loop transfer function $H(j\omega)$, for different values of ϕ_m ranging from 30° to 80°. The table shows the value of the closed-loop peaking, and the values of the closed-loop 0 dB and -3 dB frequencies, normalized to the value of the open-loop bandwidth f_c.

7.6   SPECTRAL PURITY PERFORMANCE

7.6.1   Spurious Reference Breakthrough

The nature of the sequential phase-frequency detector determines that correction pulses occur at discrete moments in time. This means that the correction signal contains the fundamental and the harmonics of the reference frequency f_{ref}. This was already demonstrated by (7.27), which is repeated below. The equation shows that the amplitude of the spectral components of I_{out} are twice as large as its DC value $I_{cp}\delta_{cp}$

$$I_{out}(t) = I_{cp}\delta_{cp} + 2I_{cp}\delta_{cp}\sum_{n=1}^{\infty}\cos\left(2\pi n f_{ref} t\right). \qquad (7.45)$$

The effect of these components on the output spectrum of a VCO can be seen in Figure 7.3, which clearly shows spurious reference breakthrough up to values of $n = 2$.

During the lock-in process, the loop builds up a tuning voltage for the VCO which results in frequency and phase lock of the signals at the input of the PFD. This voltage is "stored" in capacitor C_1 of the loop filters displayed in Figures 7.9 and 7.10.

In an ideal situation the phase of the VCO would stay perfectly locked to the phase of the reference signal and the duty cycle δ_{cp} of the charge-pump output signal would be zero. In that case there would be no signal components at the reference frequency nor its harmonics coming into the loop filter, and therefore there would be no spectral degradation of the oscillator's output signal. In practice, however, there are two main effects which can generate reference spurious breakthrough, namely:

- leakage currents in the loop filter,

- mismatch in the charge-pump *up* and *down* current sources [16, 25].

We start by considering the effect of leakage currents in the loop filter, with the assumption that the current sources are perfectly matched. The treatment leads to equations which are applicable to the effect of mismatch in the current sources as well.
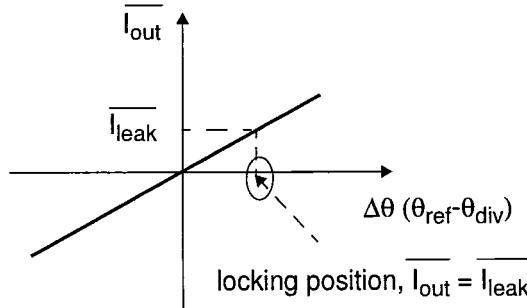
Figure 7.16. The locking position is such that the average charge-pump current is equal to the average leakage current.

Effect of Leakage Currents

Leakage currents in the tuning line alter the voltage stored in capacitor C_1. There are several sources of leakage currents which may shift the voltage in C_1: the capacitor C_1 itself, the input of the VCO, the charge-pump output and the input biasing current of the op-amp, when an active loop filter configuration is used (in that case, the leakage at the input of the VCO does not lead to a change of the C_1 voltage).

Phase-lock requires the average voltage on the tuning line to be constant over many periods of the reference signal. This is only accomplished if the average charge-pump current I_{out} equals the (average) value of I_{leak}. In other words, the loop reacts to the DC leakage current I_{leak} by restoring the charge lost during a reference period to the loop filter, at the next correction moment.

The "in-lock" situation is depicted in Figure 7.16, which shows that the loop locks with a phase difference $\Delta\theta$ at the input of the PFD which satisfies the condition $\overline{I_{out}} = \overline{I_{leak}}$. The duty cycle δ_{cp} of the charge-pump output signal becomes a function of the nominal charge-pump current I_{cp} and of I_{leak}. It follows that $I_{out} = I_{cp}\delta_{cp} = I_{leak}$, and therefore

$$\delta_{cp} = I_{leak}/I_{cp} \qquad (7.46)$$

Inserting (7.46) into (7.45) gives

$$I_{out}(t) = I_{leak} + 2I_{leak} \sum_{n=1}^{\infty} \cos\left(2\pi n f_{ref} t\right), \qquad (7.47)$$

from which we may take two important conclusions. First, the amplitude of the spectral components of I_{out} are twice the value of the DC leakage current I_{leak}. Second, the amplitudes are *not* dependent on the nominal charge-pump

current I_{cp} [26].[8] The next step is to link the leakage current to the magnitude of the spurious components at the output of the VCO. We know from standard modulation theory [14] that the relationship of the peak phase deviation $\theta_p(f_m)$ to the peak frequency deviation $\Delta f(f_m)$ and the modulation frequency f_m is given by

$$\theta_p(f_m) = \frac{\Delta f(f_m)}{f_m}. \tag{7.48}$$

The peak frequency deviation is the product of the *magnitude* of the spectral components $V_{ripple}(n \cdot f_{ref})$ of the ripple voltage at the tuning line with the VCO gain K_{vco} in [Hz/V]. The spectral components of the ripple voltage at the reference frequency f_{ref} and its harmonics can be expressed as follows, with help of (7.47):

$$V_{ripple}(n \cdot f_{ref}) = 2I_{leak}|Z_f(j2\pi n f_{ref})|, \tag{7.49}$$

with n ranging from 1 to ∞, and $|Z_f(j2\pi n f_{ref})|$ the magnitude of the transimpedance function of the loop filter at the corresponding frequency.

The peak phase deviation $\theta_p(n \cdot f_{ref})$ due to each of the frequency components $n \cdot f_{ref}$ of the ripple voltage can therefore be written as

$$\begin{aligned}
\theta_p(n \cdot f_{ref}) &= \frac{\Delta f(n \cdot f_{ref})}{n \cdot f_{ref}} \\
&= \frac{V_{ripple}(n \cdot f_{ref})K_{vco}}{n \cdot f_{ref}} \\
&= \frac{2I_{leak}|Z_f(j2\pi n f_{ref})|K_{vco}}{n \cdot f_{ref}}.
\end{aligned} \tag{7.50}$$

Each of the baseband modulation frequencies $n \cdot f_{ref}$ generates two RF spurious signals, which are located at offset frequencies $\pm n \cdot f_{ref}$ from the carrier frequency f_{LO}. The amplitude of each spurious signal A_{sp} is related to the magnitude of the carrier A_{LO} and to the peak phase deviation θ_p by

$$\begin{aligned}
A_{sp}(f_{LO} \pm n \cdot f_{ref}) &= A_{LO}\frac{\theta_p(n \cdot f_{ref})}{2} \\
&= A_{LO}\frac{I_{leak}|Z_f(j2\pi n f_{ref})|K_{vco}}{n \cdot f_{ref}},
\end{aligned} \tag{7.51}$$

---

[8]Except when the leakage current itself is proportional to the charge-pump current, for example when the charge-pump is the dominant source of leakage current and its output impedance is a function of the nominal output current.
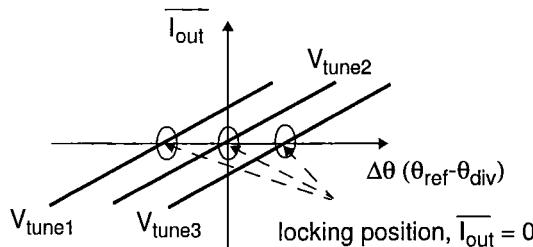
Figure 7.17. The locking position is such that the average charge-pump current is zero.

so that

$$\frac{A_{sp}(f_{LO} \pm n \cdot f_{ref})}{A_{LO}} = \frac{I_{leak}|Z_f(j2\pi n f_{ref})|K_{vco}}{n \cdot f_{ref}}. \tag{7.52}$$

It is common to express the magnitude of (undesired) signal components in decibel with respect to the magnitude of the carrier. In that case, (7.52) becomes

$$\left[\frac{A_{sp}}{A_{LO}}\right]_{dBc} = 20\log\frac{\theta_p(n \cdot f_{ref})}{2}$$

$$= 20\log\frac{I_{leak}|Z_f(j2\pi n f_{ref})|K_{vco}}{n \cdot f_{ref}} \quad [\text{dBc}]. \tag{7.53}$$

An important conclusion to be drawn from (7.53) is that the relative amplitude of the spurious signals is *not* dependent on the absolute value of loop bandwidth or on the nominal charge-pump current I_{cp}. Instead, they are determined by the trans-impedance of the loop filter, by the magnitude of the DC leakage current, by the VCO gain and by the value of the reference frequency. Theoretically, if $I_{leak} = 0$ there are *no* spurious reference breakthrough signals in the spectrum of the oscillator signal.

Effect of Mismatch in the Charge-pump Current Sources

The next step is to look at the effect of mismatch in the current sources. Mismatch, in the present context, originates in the different type of devices used to implement the N-type current source, which sinks current from the output node to ground, and the P-type source which sources current from the positive supply to the output node. Besides, the nominal current supplied by the N-type and P-type sources is likely to be a function of the voltage at the output node of the charge-pump. With a passive loop filter, this voltage is the tuning voltage V_{tune} to the oscillator, and therefore it is a function of the output frequency of the loop.

The assumption is made that there are no leakage currents in the loop filter. With this assumption, phase-lock occurs with a given phase difference at the input of the PFD which results in an *average* output current from the charge-pump of zero. The concept is depicted graphically in Figure 7.17, which represents three possible locking situations for three different values of the tuning voltage V_{tune}. Nevertheless, phase-lock always occurs at a phase difference which results in an average output current of zero. With this knowledge, the circuit designer can search for the locking position by means of transient simulations (perhaps for different values of the tuning voltage), and then perform a Fourier analysis on the output signal of the charge-pump in the "locked condition." The spectral analysis should provide the *magnitude* of the spectral components $I_{out}(n \cdot f_{ref})$ at the fundamental and harmonics of the reference frequency f_{ref}. With this information, the magnitude of the spectral components of the ripple voltage due to current-source mismatch can be found

$$V_{mismatch}(n \cdot f_{ref}) = I_{out}(n \cdot f_{ref}) \cdot |Z_f(j2\pi nf_{ref})|, \qquad (7.54)$$

with n ranging from 1 to ∞, and $|Z_f(j2\pi nf_{ref})|$ the magnitude of the transimpedance function of the loop filter at the corresponding frequency.

Using a similar approach as applied to the effect of leakage currents provides the following expression to the magnitude of the reference spurious breakthrough, expressed in decibel with respect to the carrier:

$$\left[\frac{A_{sp}(n \cdot f_{ref})}{A_{LO}}\right]_{dBc} = 20 \log \frac{I_{out}(n \cdot f_{ref})|Z_f(j2\pi nf_{ref})|K_{vco}}{2 \cdot n \cdot f_{ref}} \quad [dBc].$$

$$(7.55)$$

7.6.2   Phase Noise Performance

The phase noise generated by the PLL building blocks can be modelled with the help of additive (phase) noise sources [27]. Figure 7.18 shows the noise sources of a single loop PLL with a passive loop filter. The dependency of the (phase) noise sources on the modulation/offset frequency f_m is not explicitly shown in the figure, for convenience of notation. In the present treatment the oscillator is assumed to be free of amplitude (AM) noise.

The rms phase noise power density of the main and reference dividers are represented by $\phi_d(f_m)$ and $\phi_{ref}(f_m)$, respectively.[9] The phase noise of the phase detector is represented by $\phi_{pd}(f_m)$, and the phase noise of the reference crystal oscillator by $\phi_x(f_m)$. The dimension of the phase noise sources is

---

[9]More information on modeling and practical aspects of frequency divider phase noise can be found in [28, 29, 30].
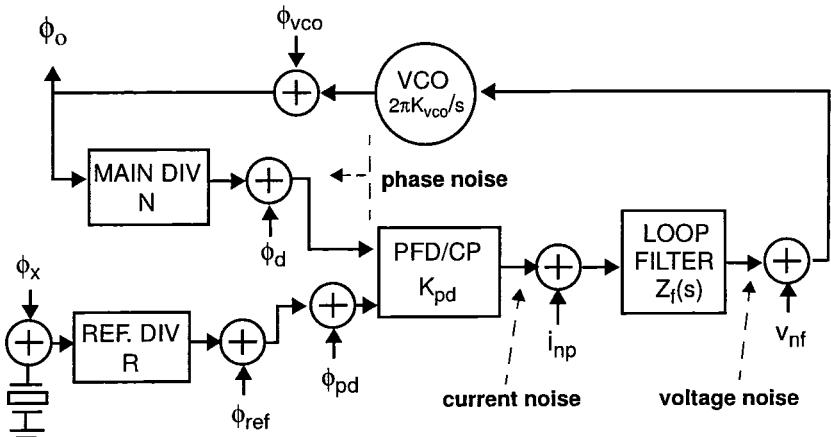
Figure 7.18.   Noise sources of the PLL.

[rad/$\sqrt{\text{Hz}}$]. The charge-pump noise is taken into account with the noise current source $i_{np}(f_m)$ [A/$\sqrt{\text{Hz}}$], and the noise of the loop filter (resistive) components is represented by the noise voltage source $v_{nf}(f_m)$ [V/$\sqrt{\text{Hz}}$]. Finally, the phase noise of the free-running VCO is modelled by $\phi_{vco}(f_m)$ with dimension of [rad/$\sqrt{\text{Hz}}$].

The rms phase noise power density of the loop's output signal is denoted $\phi_o(f_m)$ [rad/$\sqrt{\text{Hz}}$]. It consists of the noise contribution of all noise sources, modified by the action of the feedback loop upon them. The output phase noise power density $\phi_o^2(f_m)$ will be expressed as a function of two components

$$\phi_o^2(f_m) = \phi_{olp}^2(f_m) + \phi_{ohp}^2(f_m), \qquad (7.56)$$

where $\phi_{olp}^2(f_m)$ stands for the phase noise power density generated by (phase) noise sources which are subjected to a low-pass transfer function when transfered to the output node; $\phi_{ohp}^2(f_m)$ represents the effect of the (phase) noise sources which are subjected to a high-pass transfer function.

Phase noise Originated on Dividers, PFD/CP and Crystal Reference Source

The transfer function of the noise sources $\phi_d(f_m)$, $\phi_{ref}(f_m)$ and $\phi_{pd}(f_m)$ to the output node output node is the same, namely

$$\text{transfer} = N\frac{G(s)}{1 + G(s)} \qquad (7.57)$$

$$= N \cdot H(s), \qquad (7.58)$$

with $G(s)$ and $H(s)$ as given in (7.30) and (7.31), respectively.

The scaled contributions from the charge-pump $i_{np}(f_m)/K_{pd}$ and from the crystal oscillator $\phi_x(f_m)/R$ are also subjected to transfer function (7.58). Therefore, the "low-pass" phase noise power component $\phi^2_{olp}(f_m)$ of $\phi^2_o(f_m)$ can be expressed as follows, with s replaced by $j2\pi f_m$ in (7.58):

$$\phi^2_{olp}(f_m) = N^2|H(j2\pi f_m)|^2\big(\phi^2_d(f_m) + \phi^2_{ref}(f_m) + \phi^2_{pd}(f_m)$$
$$+ i^2_{np}(f_m)/K^2_{pd} + \phi^2_x(f_m)/R^2\big). \qquad (7.59)$$

We see that the loop transfers the power spectral density of the noise sources within brackets to the output with a multiplication by N^2 and by the squared magnitude of $H(j2\pi f_m)$, which is presented in Figure 7.19(a). It follows from (7.59) that the contribution from individual building blocks can not be readily evaluated from measurements on a closed-loop configuration.

To proceed, we define a function called *the equivalent synthesizer phase noise floor at the input of the phase detector*, which is denoted here as $\phi^2_{eq}(f_m)$:

$$\phi^2_{eq}(f_m) \equiv \phi^2_d(f_m) + \phi^2_{ref}(f_m) + \phi^2_{pd}(f_m) + i^2_{np}(f_m)/K^2_{pd} + \phi^2_x(f_m)/R^2.$$
$$(7.60)$$

Substitution of (7.60) in (7.59) gives

$$\phi^2_{olp}(f_m) = N^2|H(j2\pi f_m)|^2\phi^2_{eq}(f_m), \qquad (7.61)$$

which conveniently expresses the influence of several noise sources on the "low-pass" component $\phi^2_{olp}(f_m)$ of $\phi^2_o(f_m)$.

The Dependency of the Equivalent Phase Noise Floor on the Reference Frequency

The magnitude of the noise sources ϕ_d, ϕ_r, ϕ_{pd} and i_{ncp} are dependent on the offset frequency f_m *and* on the loop's reference frequency f_{ref} [31]. It is argued in [31] that the phase noise power densities of the above mentioned sources may present a 0 dB/octave, a 3 dB/octave or a 6 dB/octave dependency on the loop's reference frequency f_{ref}. The dependency is determined mainly by the spectral distribution of the dominant sources of thermal noise, and by the type of waveform which drives the circuit elements. In practice, the equivalent synthesizer phase noise floor of low-noise synthesizers is often dominated by a 3 dB/octave dependency on the reference frequency f_{ref}. Let us investigate this behavior in more detail. We start with the charge-pump contribution.

The magnitude of the charge-pump current noise power density $i^2_{np}(f_m)$ is proportional to the duty-cycle of the charge-pump output pulses when the loop is locked; in simplified form this dependency can be expressed as $i^2_{np}(f_m) \propto$

(a)

(b)
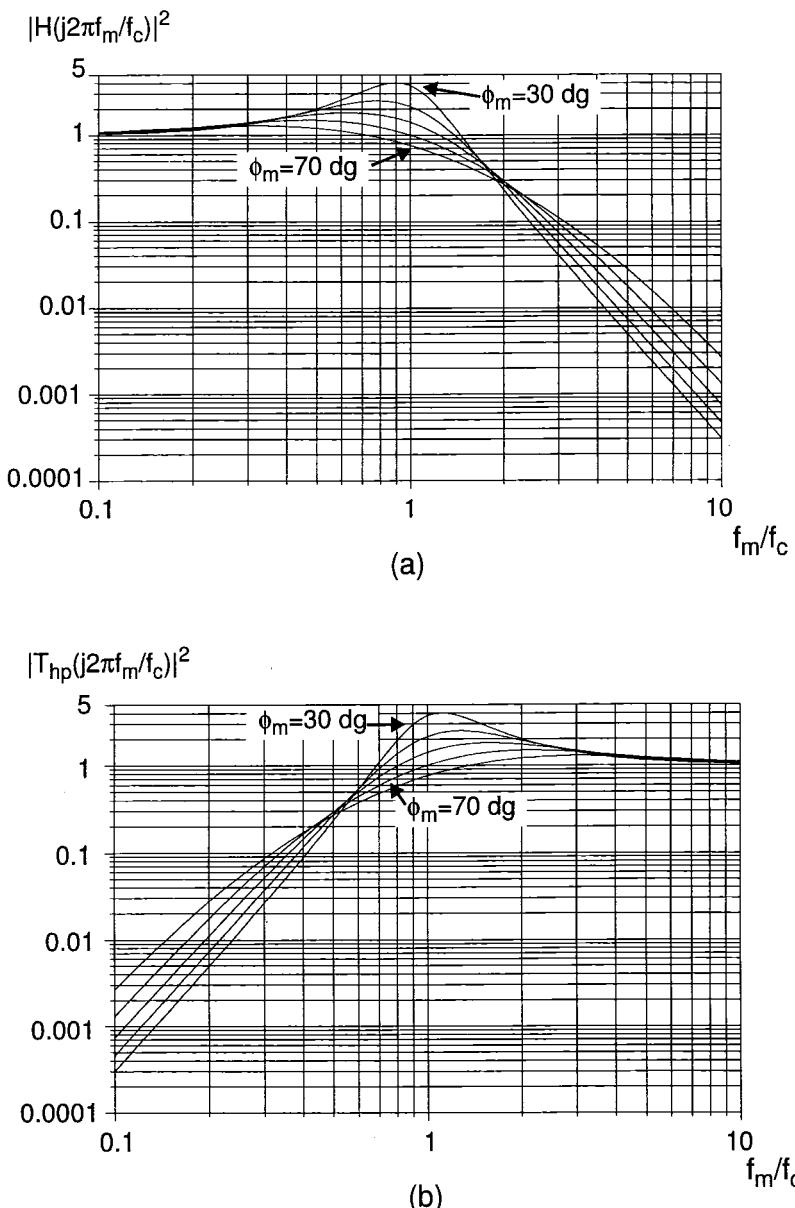
Figure 7.19. Normalized phase noise power transfer functions $|H(j2\pi f_m/f_c)|^2$ and $|T_{hp}(j2\pi f_m/f_c)|^2$, for several values of phase margin ranging from 30° to 70°.

t_{cp}/T_{ref}, with t_{cp} the active time of the charge-pump current sources at each reference period and $T_{ref} = 1/f_{ref}$; t_{cp} is linked to ΔR, the reset time of the phase-frequency detector, and it is not a function of the reference frequency f_{ref}. Replacing $f_{ref} = 1/T_{ref}$ into the above expression gives

$$i_{np}^2(f_m) \propto t_{cp} f_{ref},$$

which expresses a 3 dB/octave dependency of the charge-pump contribution to the equivalent synthesizer noise floor on the reference frequency.

Now let us consider the effect of wide-band noise sources in the frequency dividers and in the phase-frequency detector. During transitions of the logic gates and flip-flops a sampling action is performed on the noise from the wide-band sources. This results in aliasing of wide-band noise into the Nyquist bandwidth. Let us focus on the phase-frequency detector contribution $\phi_{pd}^2(f_m)$ to the equivalent synthesizer noise floor. The aliased voltage noise power will be denoted $v_{n,al}^2(f_m)$. It can be demonstrated that $v_{n,al}^2(f_m) \propto 1/f_{ref}$ [31]. The next step is to translate the voltage noise power spectral density $v_{n,al}^2(f_m)$ into the phase noise power spectral density $\phi_{pd}^2(f_m)$. The voltage noise power will cause time jitter in the logic transitions, which when converted to the phase domain provides the following dependency: $\phi_{pd}^2(f_m) \propto v_{n,al}^2(f_m) f_{ref}^2$. With the knowledge that $v_{n,al}^2(f_m) \propto 1/f_{ref}$ the "overall" dependency of the phase noise power spectral density $\phi_{pd}^2(f_m)$ on the reference frequency is of the form

$$\phi_{pd}^2(f_m) \propto f_{ref},$$

which expresses a 3 dB/octave relationship to the reference frequency. The same reasoning can be applied to the phase noise contribution from the frequency dividers.

Phase Noise due to Loop Filter and Free-running VCO Phase Noise

Oscillator. The free-running spectral purity of an oscillator has been discussed in Chapter 6. In this chapter the free-running VCO phase noise power density is modelled with help of phase noise source $\phi_{vco}(f_m)$ in Figure 7.18. If the phase noise power density of the free-running VCO is assumed to have a pure $1/f_m^2$ dependency on the modulation frequency f_m (i.e., a -6 dB/octave dependency on f_m) then we may write $\phi_{vco}^2(f_m)$ as a function of the phase noise power density at an offset frequency f_r as

$$\phi_{vco}^2(f_m) = \phi_{vco}^2(f_r)\frac{f_r^2}{f_m^2}. \tag{7.62}$$

Note that f_r is a *pre-defined offset frequency* at which the free-running VCO phase noise power density $\phi_{vco}^2(f_r)$ is specified. Equation (7.62) can be eas-

ily modified to include the region with the $1/f_m^3$ dependency — i.e., at small modulation frequencies f_m — and the flat phase noise floor at very large offset frequencies:

$$\phi_{vco}^2(f_m) = \phi_{vco}^2(f_r)\frac{f_r^2}{f_m^2}\left(1 + \frac{f_k}{f_m}\right) + \phi_{vco,nf}^2, \qquad (7.63)$$

where f_k is the corner frequency of the $1/f_m^3$ dependency and $\phi_{vco,nf}^2$ is the VCO phase noise floor power density.

Loop filter. Thermal noise voltage originated in the loop filter resistor causes unintended phase modulation of the VCO. The noise from the loop filter is modeled as v_{nf} in Figure 7.18. The open-loop phase noise power density $\phi_{lf}^2(f_m)$ due to $v_{nf}(f_m)$ can be expressed as

$$\phi_{lf}^2(f_m) = v_{nf}^2(f_m)\frac{K_{vco}^2}{f_m^2} \quad [\text{rad}^2/\text{Hz}], \qquad (7.64)$$

with K_{vco} the gain of the VCO in [Hz/V].

Influence of the loop. The influence of the feedback loop on the free-running VCO phase noise power density $\phi_{vco}^2(f_m)$ and on the open-loop phase noise power density generated by the loop filter elements $\phi_{lf}^2(f_m)$ will be expressed with the transfer function $T_{hp}(s)$,

$$T_{hp}(s) = \frac{1}{1 + G(s)}, \qquad (7.65)$$

where the subscript hp expresses the high-pass transfer character of $T_{hp}(j2\pi f_m)$.

The "high-pass" phase noise component $\phi_{ohp}^2(f_m)$ of $\phi_o^2(f_m)$ can therefore be written as

$$\phi_{ohp}^2(f_m) = |T_{hp}(j2\pi f_m)|^2 \left(\phi_{vco}^2(f_m) + \phi_{lf}^2(f_m)\right). \qquad (7.66)$$

The squared magnitude of $T_{hp}(j2\pi f_m)$ is depicted in Figure 7.19(b), normalized to the open-loop bandwidth f_c.

Total Phase Noise at Output of the PLL

Substitution of (7.66) and (7.61) in (7.56) provides the following expression for the total phase noise power spectral density $\phi_o^2(f_m)$:

$$\phi_o^2(f_m) = N^2\phi_{eq}^2(f_m)|H(j2\pi f_m)|^2$$
$$+ \left(\phi_{vco}^2(f_m) + \phi_{lf}^2(f_m)\right)|T_{hp}(j2\pi f_m)|^2 \quad [\text{rad}^2/\text{Hz}]. \qquad (7.67)$$
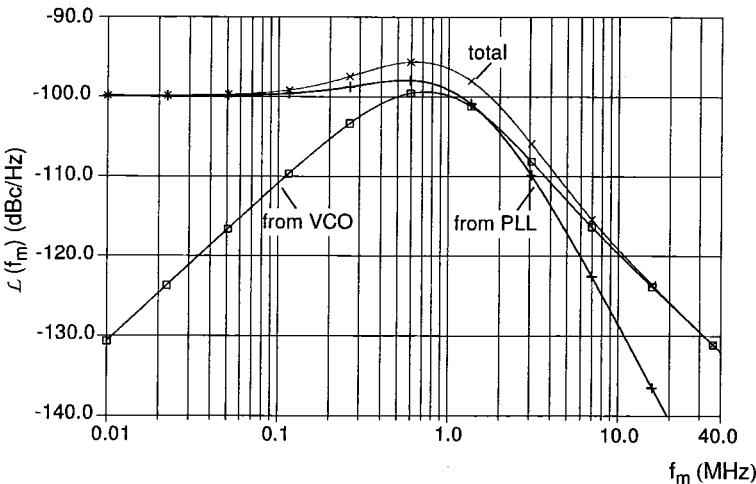
Figure 7.20. Simulated SSB phase noise power density depicting the contribution from the different phase noise sources.

Figure 7.20 shows the simulated closed-loop SSB phase-noise spectral density of a PLL frequency synthesizer which has an open-loop bandwidth $f_c \simeq 1$ MHz. In the simulation model the free-running VCO phase noise power density $\phi_{vco}^2(f_m)$ has a -6 dB/octave dependency on the offset frequency f_m, and the equivalent synthesizer noise floor $\phi_{eq}^2(f_m)$ has a flat spectral distribution. The phase noise contribution from the dividers and PFD/CP ("PLL blocks") dominates at modulation frequencies smaller than the open-loop loop bandwidth f_c. At these frequencies, the free-running VCO phase noise and the contribution from the loop filter are attenuated by the loop's feedback action. At offset frequencies larger than the loop bandwidth the noise from the "PLL blocks" is attenuated by the closed-loop transfer function, and the phase noise is virtually equal to the free-running VCO phase noise.

Estimation of the equivalent synthesizer noise floor from closed-loop measurements. The magnitude of the equivalent synthesizer phase noise floor ϕ_{eq}^2 can be assessed from measurements on a closed-loop PLL. We know that, close to the carrier, the phase noise power density can be attributed to ϕ_{eq}^2 and to the multiplication factor N^2 as expressed by (7.59).

For example, in Figure 7.20 the multiplied noise floor leads to a phase noise level of ~ -100 dBc/Hz at small offset frequencies from the carrier. With a division ratio $N = 200$ the equivalent phase noise floor (in dBc/Hz) can then be readily estimated as $\mathcal{L}_{eq} = -100 - 20 \log 200 = -146$ dBc/Hz. Sometimes, the closed-loop phase noise level close to the carrier is specified in the data-

sheets [32]; this value must then be added to $-20 \log N$ if the equivalent noise floor in dBc/Hz needs to be calculated. In other cases, the equivalent phase noise floor at a given reference frequency is directly specified [22].

7.6.3   Spectral Purity and the Dimensioning of the PLL Loop Filter

Next, the dimensioning of the loop filter components will be investigated. We have seen the PLL dimensioning concepts of open-loop bandwidth f_c and phase margin ϕ_m in Section 7.5. Basically, the choice of a certain open-loop bandwidth f_c fixes the product of the charge-pump current I_{cp} and the value of the loop filter resistance R_1, as expressed by (7.41) and (7.42). This means that the magnitude of the charge-pump current can be "traded-off" against the value of the loop filter resistor R_1 for a given value of f_c. The phase margin ϕ_m fixes the ratio of the loop filter time constants $b = \tau_2/\tau_3$. With $\tau_2 \approx R_1 \cdot C_1$ and $\tau_3 \approx R_1 \cdot C_2$ then $b \approx C_1/C_2$. These results show that there are, in principle, an infinity of values for I_{cp}, R_1, C_1 and C_2 which satisfy a given choice of f_c and ϕ_m. To establish the value of the loop filter elements one needs further knowledge of the loop spectral purity specifications and requirements.

In a practical situation the design targets are twofold. First, one wants to decrease I_{cp} to its lowest acceptable level,[10] in order to decrease the power dissipation and to simplify the design of the charge-pump circuitry. Second, the impedance level of the loop filter should be maximized, for the chip-area of fully integrated loop filters to be minimized. High impedance level means small capacitors C_1 and C_2, a relatively large R_1 and a small value for the charge-pump current I_{cp}. However, a high impedance level leads to a higher noise contribution from the loop filter.

The analysis will be focused on loop filter LF2 of Figure 7.9. Configuration LF2 is preferable for fully integrated loop filter applications as the bottom plates of C_1 and C_2 are both grounded. This property eliminates the possibility of substrate noise coupling into the filter output node through the parasitic capacitance of C_2 to the substrate. However, there are no qualitative differences between LF1 and LF2 from a "system" point-of-view, so that the considerations presented in this section are equally valid for LF1 and LF2, although the exact mathematic formulation can be slightly different for each configuration.

---

[10]The noise contribution from the charge-pump must remain within acceptable/specified levels, see Equation (7.59).

Spurious Reference Breakthrough

In a given application, there are specifications for the maximum value of the spurious signals *maxspurious* (in dBc) and for the maximum value of the leakage current I_{leak}. The design problem is to dimension the loop filter's transimpedance level $|Z_f(j2\pi n f_{ref})|$ to meet the spurious reference breakthrough specifications, as discussed in Section 7.6.1. Manipulation of (7.53) results in the following expression for $|Z_f(j2\pi f_{ref})|$, with $n = 1$ as the worst case situation:

$$|Z_f(j2\pi f_{ref})| < \frac{f_{ref}}{I_{leak} \cdot K_{vco}} 10^{\frac{maxspurious}{20}} \qquad (7.68)$$

With loop filter LF2 of Figure 7.9, the impedance of the capacitor C_2 can be taken as an approximation for $|Z_f(j2\pi f_{ref})|$, if $f_{ref} \gg 1/(2\pi\tau_3)$. This condition is satisfied when $f_c \ll f_{ref}$, which is the case in many practical applications. Substitution of $|Z_f(j2\pi f_{ref})| \approx 1/(2\pi f_{ref} C_2)$ into (7.68) leads to the smallest value for C_2 that satisfies the reference breakthrough specification,

$$C_{2,min} = \frac{I_{leak} \cdot K_{vco}}{2\pi f_{ref}^2} 10^{-\frac{maxspurious}{20}}. \qquad (7.69)$$

We observe that $C_{2,min}$ is directly proportional to I_{leak} in [A] and K_{vco} in [Hz/V], and inversely proportional to the spurious level specification in [dBc] and to the *second power* of the reference frequency f_{ref}.

We have seen in Section 7.6.1 that the effect of leakage currents in the loop filter is equivalent to the effect of mismatch in the charge-pump. If the leakage current is much smaller than the expected amplitude of the spectral components $I_{out}(n \cdot f_{ref})$ due to mismatches, then we should use a slightly modified variant of (7.69), where $C_{2,min}$ is given as a function of the magnitude of the spectral component $I_{out}(f_{ref})$ at the reference frequency:

$$C_{2,min} = \frac{I_{out}(f_{ref}) \cdot K_{vco}}{4\pi f_{ref}^2} 10^{-\frac{maxspurious}{20}}. \qquad (7.70)$$

Phase Noise Contribution from the Loop Filter Resistor

We have seen in Section 7.6.2 that (thermal) noise voltages originated in the loop filter elements cause phase modulation of the VCO. These noise sources are modelled with help of $v_{nf}(f_m)$ in Figure 7.18. In the loop filters of Figure 7.9, the source of thermal noise is the resistor R_1. For loop filter LF2, the relationship of $v_{nf}^2(f_m)$ and R_1 is

$$v_{nf}^2(f_m) = \left(\frac{b-1}{b}\right)^2 \cdot \frac{4kTR_1}{|1 + j2\pi f_m\tau_3|^2}, \qquad (7.71)$$

In (7.71) b and τ_3 are as defined on the table inserted in Figure 7.9, $k = 1.37 \times 10^{-23}$ J/K is the Boltzmann constant, and T is the absolute temperature in K. The expression shows that the thermal noise power from R_1 is "shunted" at modulation frequencies f_m larger than $\approx 1/(2\pi\tau_3)$ by capacitor C_2.

Use of (7.64) provides the following expression for the *open-loop* phase noise power spectral density $\phi_{lf}^2(f_m)$, due to the loop filter resistor

$$\phi_{lf}^2(f_m) = \left(\frac{b-1}{b}\right)^2 \cdot \frac{4kTR_1}{|1 + j2\pi f_m\tau_3|^2} \frac{K_{vco}^2}{f_m^2}. \tag{7.72}$$

To proceed, the relative degradation in phase noise power spectral density at the output of the VCO will be calculated as a function of the relative magnitude of $\phi_{lf}^2(f_m)$ with respect to the free-running VCO phase noise spectral density $\phi_{vco}^2(f_m)$. A function $\alpha_{lf}(f_m)$ will be defined as

$$\alpha_{lf}(f_m) \equiv \frac{\phi_{lf}^2(f_m)}{\phi_{vco}^2(f_m)}, \tag{7.73}$$

so that $\alpha_{lf}(f_m)$ can be used to relate $\phi_{lf}^2(f_m)$ to $\phi_{vco}^2(f_m)$

$$\phi_{lf}^2(f_m) = \alpha_{lf}(f_m) \cdot \phi_{vco}^2(f_m). \tag{7.74}$$

Accordingly, the sum of $\phi_{lf}^2(f_m)$ and $\phi_{vco}^2(f_m)$ can be written as

$$\phi_{vco,lf}^2(f_m) = \phi_{vco}^2(f_m)\left(1 + \alpha_{lf}(f_m)\right), \tag{7.75}$$

where $\phi_{vco,lf}^2(f_m)$ represents the "combined" *open-loop* phase noise power density at the output of the oscillator. The degradation of $\phi_{vco,lf}^2(f_m)$ with respect to $\phi_{vco}^2(f_m)$ is thus expressed by the factor $\left(1 + \alpha_{lf}(f_m)\right)$. An equivalent expression to (7.75) is

$$\mathcal{L}_{vco,lf}(f_m) = \mathcal{L}_{vco}(f_m) + \Delta\mathcal{L}_{vco}(f_m), \tag{7.76}$$

with the "combined" SSB phase noise power density $\mathcal{L}_{vco,lf}(f_m)$ expressed in dBc/Hz, and the degradation $\Delta\mathcal{L}_{vco}(f_m) = 10\log\left(1 + \alpha_{lf}(f_m)\right)$ in dB. Figure 7.21 presents the numerical values of $\Delta\mathcal{L}_{vco}(f_m)$ as a function of $\alpha_{lf}(f_m)$. As can be expected a small value of $\alpha_{lf}(f_m)$ corresponds to a small degradation with respect to the oscillator's free-running phase noise power density. It is a part of the designer's job to define a function $\alpha_{lf}(f_m)$ which satisfies specific requirements of the intended application.

Let us investigate next the particulars of the function $\alpha_{lf}(f_m)$ within the context of the passive loop filters of Figure 7.9. Substitution of (7.62) and
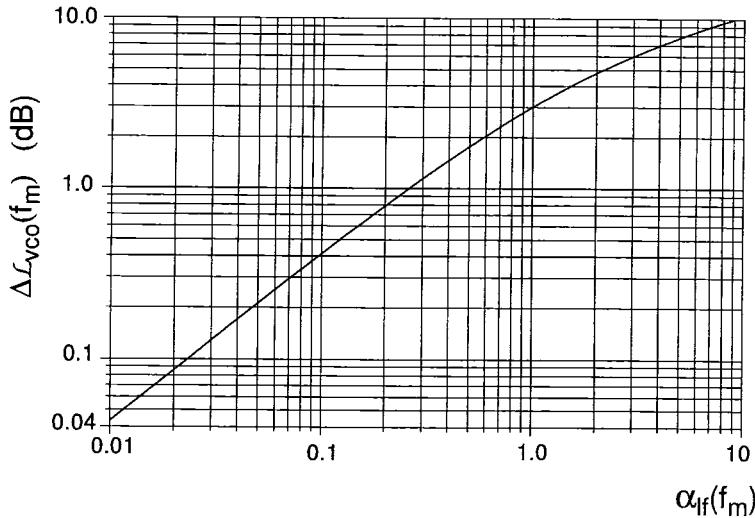
Figure 7.21. Degradation in the phase noise power density at the output of the VCO in dB, as a function of $\alpha_{lf}(f_m) = \phi_{lf}^2(f_m)/\phi_{vco}^2(f_m)$.

(7.72) in (7.73), and simplification leads to

$$\alpha_{lf}(f_m) = \left(\frac{b-1}{b}\right)^2 \cdot \frac{4kTR_1K_{vco}^2}{\phi_{vco}^2(f_r)f_r^2} \cdot \frac{1}{|1 + j2\pi f_m \tau_3|^2}. \tag{7.77}$$

This expression shows that $\alpha_{lf}(f_m)$ has a "low-pass" character with a -3 dB corner frequency of $1/(2\pi\tau_3)$. The "low-frequency" value of $\alpha_{lf}(f_m)$ will be called $\alpha_{lf,dc}$, so that (7.77) can be written as

$$\alpha_{lf}(f_m) = \alpha_{lf,dc} \cdot \frac{1}{|1 + j2\pi f_m \tau_3|^2}, \tag{7.78}$$

with

$$\alpha_{lf,dc} = \left(\frac{b-1}{b}\right)^2 \cdot \frac{4kTR_1K_{vco}^2}{\phi_{vco}^2(f_r)f_r^2}. \tag{7.79}$$

In practice, $\alpha_{lf,dc}$ determines an upper limit for the degradation in the phase noise of the free-running oscillator. Once a given value for $\alpha_{lf,dc}$ has been chosen by the designer, for example with the help of Figure 7.21, then the range of values of R_1 that results in an acceptable performance can be calculated. Solving (7.79) for R_1 gives the maximum value $R_{1,max}$ that can be used in the

loop filter:

$$R_{1,max} = \alpha_{lf,dc} \left(\frac{b}{b-1} \right)^2 \cdot \frac{\phi_{vco}^2(f_r) f_r^2}{4kT K_{vco}^2}$$

$$= \alpha_{lf,dc} \left(\frac{b}{b-1} \right)^2 \cdot \frac{2 \cdot 10^{\frac{\mathcal{L}_{vco}(f_r)}{10}} \cdot f_r^2}{4kT K_{vco}^2}. \tag{7.80}$$

A smaller value than $R_{1,max}$ for the loop filter resistor obviously leads to a smaller $\alpha_{lf,dc}$, which is always an acceptable situation from the point of view of phase noise performance. We observe that $R_{1,max}$ is inversely proportional to the *second power* of the oscillator gain K_{vco}.

Dimensioning of Time Constant τ_2 and Capacitance C_1

With the minimum value of C_2 determined with (7.69), the maximum value of R_1 with (7.80) and with the target open-loop bandwidth f_c and phase margin ϕ_m known in advance, the parameters which still need to be defined are the values of the capacitance C_1 and of the charge-pump current I_{cp}. The value of C_1 will be derived from knowledge of the time constant τ_2, which in combination with time constant τ_3 leads to the desired value of the phase margin ϕ_m at the open-loop bandwidth $\omega_c = .2\pi f_c$. The phase margin ϕ_m is defined as

$$\begin{aligned} \phi_m &= \Psi(j\omega_c) + \pi \\ &= -\pi + \arg(1 + j\tau_2\omega_c) - \arg(1 + j\tau_3\omega_c) + \pi \\ &= \arctan \tau_2\omega_c - \arctan \tau_3\omega_c, \end{aligned} \tag{7.81}$$

where $\Psi(j\omega_c)$ is the phase of the open-loop transfer function $G(j\omega_c)$ as given in (7.36). Solving for $\tau_2\omega_c$ gives

$$\tau_2\omega_c = \tan(\phi_m + \arctan \tau_3\omega_c). \tag{7.82}$$

The relationship expressed by (7.82) is plotted in Figure 7.22 as a function of the product $\tau_3\omega_c$ for different values of the phase margin ϕ_m. For the product $\tau_2\omega_c$ to be positive and finite, which is a requirement for the physical implementation of the passive loop filter, then the argument of the tangent function in (7.82) must be smaller than $\pi/2$ and therefore

$$\tau_3\omega_c < \arctan\left(\frac{\pi}{2} - \phi_m\right). \tag{7.83}$$

The limiting values of $\tau_3\omega_c$ are clearly seen in Figure 7.22 for the different values of the (target) phase margin.
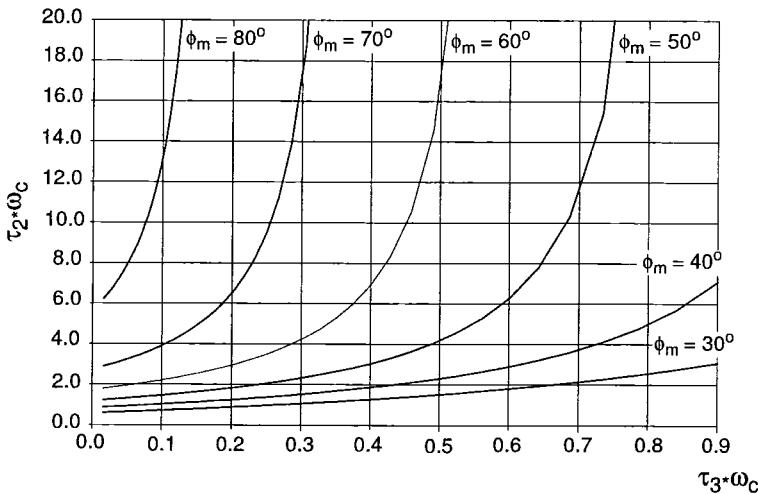
Figure 7.22. Relationship of the product $\tau_2\omega_c$ to the product $\tau_3\omega_c$.

For loop filter LF2 $\tau_2 = R_1 C_1$ (see Figure 7.9), so that the value of the capacitance C_1 can be calculated from substitution of $\tau_2 = R_1 C_1$ in (7.82)

$$C_1 = \frac{1}{\omega_c R_1} \tan\left(\phi_m + \arctan \tau_3 \omega_c\right). \tag{7.84}$$

We see that the dependency of C_1 on the product $\tau_3\omega_c$ and on the phase margin ϕ_m is of the same nature as that of the product $\tau_2\omega_c$ expressed in (7.82) and plotted in Figure 7.22. Therefore, the size of capacitor C_1 has a strong dependency on the product $\tau_3\omega_c$ and on the choice of the phase margin ϕ_m. It is convenient to choose a relatively high value of phase margin ϕ_m, as this results in a smaller peaking in the noise transfer functions as can be seen in Figure 7.19. However, higher values of phase margin demand larger values of C_1, what in turn can generate a chip-area "penalty" if the loop filter is fully integrated. For example, Figure 7.22 shows that an increase in the phase margin from 50° to 60° leads to a larger value for τ_2, and therefore for C_1, by a factor of 2 for $\tau_3\omega_c = 0.3$ and by a factor of 4 for $\tau_3\omega_c = 0.5$.

Let us investigate next the time constant τ_3. In loop filter LF2 τ_3 consists of the product of R_1 and the series combination of capacitors C_1 and C_2. Normally $C_1 \gg C_2$, so that in the subsequent discussion τ_3 will be approximated as the product $R_1 C_2$.

With the minimum value of C_2 determined with (7.69) and the maximum value of R_1 with (7.80), a time constant $\tau_{3,sp}$ can be defined as

$$\tau_{3,sp} = R_{1,max} C_{2,min}, \tag{7.85}$$

with the subscript indicating that time constant $\tau_{3,sp}$ is determined from spectral purity considerations. Incorporating (7.85) into (7.84) gives for the value of C_1

$$C_1 = \frac{1}{\omega_c R_{1,max}} \tan\left(\phi_m + \arctan \omega_c R_{1,max} C_{2,min}\right). \qquad (7.86)$$

We have seen that (7.83) expresses an upper limit for the product $\tau_3 \omega_c$. Therefore, $\tau_{3,sp} = R_{1,max} C_{2,min}$ must comply to the following inequality

$$R_{1,max} C_{2,min} < \frac{1}{\omega_c} \arctan\left(\frac{\pi}{2} - \phi_m\right). \qquad (7.87)$$

If the condition imposed by (7.87) is not satisfied then the designer needs to decrease time constant $\tau_3 = R_1 C_2$. With the minimum value of C_2 fixed by (7.69), the only possible way to reduce τ_3 is with a reduction of R_1 with respect to $R_{1,max}$. The consequence of reducing R_1 is that the impedance level of the loop filter decreases, which must be compensated by an increase in the value of the charge-pump current for a given open-loop bandwidth ω_c. An alternative approach to attain enough spurious suppression without an excessive high charge-pump current is to add an extra pole to the loop filter, whose transfer function then becomes third-order.

7.7 DESIGN OF PROGRAMMABLE FREQUENCY DIVIDERS
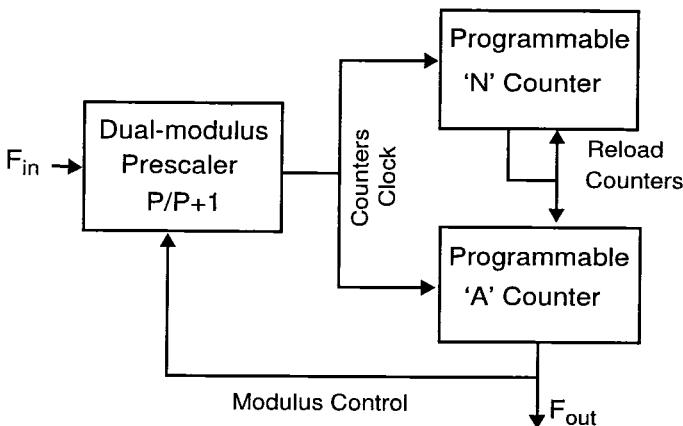
7.7.1 Programmable Divider Architectures

This section describes architectures and circuit implementations of programmable frequency dividers. We start with a discussion of the programmable divider architecture based on a dual-modulus prescaler. Next, the functionality of the "basic" programmable prescaler is analyzed. The truly-modular "extended" programmable prescaler architecture, which overcomes the limitations of the basic programmable prescaler, is then presented in the last sub-section.

The programmable prescaler architectures presented in this section lead to implementations of fully-programmable frequency dividers, so that the denomination "programmable prescaler" is equivalent to "programmable frequency divider." Use of each designation in different parts of the text refers to the same circuits and architectures.

Architecture Based on a Dual-modulus Prescaler

Figure 7.23 depicts the divider architecture based on a dual-modulus prescaler [2, 33]. The architecture consists of a dual-modulus prescaler of division ratios P and $P + 1$, and of two programmable counters, the "N" counter and the "A" counter. These counters are "down-counters", i.e. they are "loaded" with a

Figure 7.23.   Fully programmable divider based on a dual-modulus prescaler.

given number at a certain moment and then count-down by one at each cycle of their input signal. The counters have an output line which toggles state when the counter reaches a final count of, let us say, 0. The output of the "*A*" counter controls the *Modulus Control* input of the dual-modulus prescaler. When the modulus-control signal is high, the prescaler divides its input signal by $P + 1$, otherwise it divides by P.

The operation principle of the fully-programmable divider depicted in Figure 7.23 is as follows. The output signal of the prescaler drives the inputs of the "*N*" and "*A*" counters. When the "*N*" counter reaches 0, it generates a *reload counters* signal for itself and for the "*A*" counter. At that moment, the counters are re-loaded with digits N and A respectively, which are stored on local registers. After being re-loaded, the output line of the counters goes high, the *reload counters* signal vanishes, and the counters continue on counting-down from the programmed values N and A. The dual-modulus prescaler divides by $P + 1$ until the "*A*" counter reaches 0, and then switches its ratio to P. Following this event, there are $N - A$ cycles of its output signal before the "*N*" counter reaches 0 as well, whereupon the whole cycle takes place again. So, the period T_{out} of the output signal can be expressed as a function of the period T_{in} of the input signal as follows:

$$T_{out} = \left(A \cdot (P + 1) + (N - A) \cdot P \right) \cdot T_{in}$$
$$= \left(N \cdot P + A \right) \cdot T_{in}, \tag{7.88}$$

where the term within brackets is the realized division ratio of the input signal frequency F_{in}. So when the counter "A" is programmable is steps of 1, the dual-modulus prescaler based frequency divider realizes integer division ratios with an unity step size.
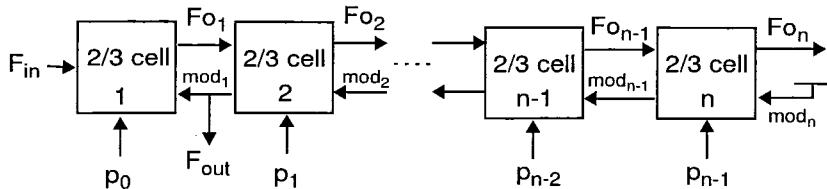
The design of the dual-modulus prescaler itself has been extensively treated in the literature [33, 34, 35, 36]. On the other hand, less attention has been given to the implications of using the programmable divider architecture of Figure 7.23.

One readily notices the lack of modularity of the concept: besides the dual-modulus prescaler, the architecture requires two additional counters for the generation of a given division ratio. The programmable counters — which are, in fact, fully programmable dividers, albeit not operating at the full RF frequency — represent a significant load at the output of the dual-modulus prescaler, and lead to increased power dissipation. Besides, the time-to-market of new products is increased, due to the additional design and layout effort required for the programmable counters. The lack of modularity, the high fan-out at the output of the dual-modulus prescaler and the extra design effort make the dual-modulus based architecture a less interesting option, within the context of high reusability and short time-to-market of new products.

Basic programmable Prescaler

The "basic" programmable prescaler architecture is depicted in Figure 7.24. The modular structure consists of a chain of 2/3 divider cells connected like a ripple counter [26]. The structure of Figure 7.24 is characterized by the absence of long delay loops, as feedback lines are only present between adjacent cells. This "local feedback" enables simple optimization of power dissipation. Another advantage is that the topology of the different cells in the prescaler is the same, therefore facilitating layout work. The architecture of Figure 7.24 resembles the one presented in [37], which is also based on 2/3 divider cells. Yet there are two fundamental differences. First, in [37] all cells operate at the same (high) current level. Second, the architecture of [37] relies on a common strobe signal shared by all cells. This leads to high power dissipation, because of high requirements on the slope of the strobe signal, in combination with the high load presented by all cells in parallel.

The programmable prescaler operates as follows. Once in a division period, the last cell on the chain generates the signal mod_{n-1}. This signal then propagates "up" the chain, being reclocked by each cell along the way. An active *mod* signal enables a cell to divide by 3, once in a division cycle, provided that its programming input p is set to 1. If the programming input is set to 0 then the cell keeps on dividing by 2. Despite the state of the p input, the *mod* signal

Minimum div. ratio = 2^n

Maximum div. ratio = 2^{n+1} -1

Figure 7.24.   Modular programmable prescaler architecture.

is reclocked and output towards the higher frequency cells. Division by 3 adds one extra period of each cell's input signal to the period of the output signal. Hence, a chain of n 2/3 cells provides an output signal with a period of
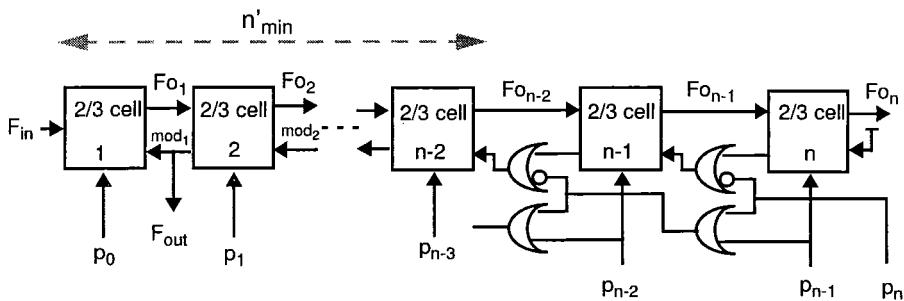
$$
\begin{aligned}
T_{out} &= 2^n \cdot T_{in} + 2^{n-1} \cdot T_{in} \cdot p_{n-1} + 2^{n-2} \cdot T_{in} \cdot p_{n-2} + \ldots \\
&\quad + 2 \cdot T_{in} \cdot p_1 + T_{in} \cdot p_0 \\
&= (2^n + 2^{n-1} \cdot p_{n-1} + 2^{n-2} \cdot p_{n-2} + \ldots + 2 \cdot p_1 + p_0) \cdot T_{in} \quad (7.89)
\end{aligned}
$$

In (7.89), T_{in} is the period of the input signal F_{in}, and p_0, ..., p_{n-1} are the binary programming values of the cells 1 to n, respectively. The equation shows that all integer division ratios ranging from 2^n (if all $p_i = 0$) to $2^{n+1} - 1$ (if all $p_i = 1$) can be realized. The division range is thus rather limited, amounting to roughly a factor two between maximum and minimum division ratios.[11] This limitation renders the "basic" programmable prescaler useless for wide-band and multi-band applications. The division range can be extended by combining the prescaler with a presettable programmable counter; this approach was taken during the design of the adaptive loop synthesizer described in [10]. In that case, however, the resulting architecture is no longer modular. The lack of modularity leads to longer design time and to decreased reusability of an optimized design.

Prescaler with Extended Programmability

The divider implementation presented in Figure 7.25 extends the division range of the basic prescaler, whilst maintaining the modularity of the basic architecture [38, 39]. The operation of the new architecture is based on

---

[11] In principle, it is also possible to divide by 3^n, but the gap between this value and the continuous division range makes it useless in standard synthesizer applications.

Minimum div. ratio $= 2^{n'min}$

Maximum div. ratio $= 2^{n+1} - 1$

Figure 7.25.   Modular programmable prescaler with extended division range.

Table 7.1.   Relationship between divider ratio, binary programming word and the required effective length of the divider chain.

Division ratio N_{div}	Binary prog. word $\cdots\ P_4\,P_3\,P_2\,P_1\,P_0$	Effective length n'
3	0 0 0 1 1	1
4	0 0 1 0 0	2
8	0 1 0 0 0	3
15	0 1 1 1 1	3
31	1 1 1 1 1	4

a—very convenient—property of the basic prescaler structure: the direct relation of the performed division ratio to the bus programmed division word $p_n, p_{n-1}, \cdots, p_1, p_0$.

Let us introduce the concept of *effective length n'* of the chain. It is the number of divider cells that are effectively influencing the division cycle. Deliberately setting the *mod* input of a certain 2/3 cell to the active level overrules the influence of all cells to the right of that cell. The divider chain behaves as if it has been shortened. The effective length of the chain can therefore be easily "adjusted".

The required *effective length n'* corresponds to the index of the most significative (and active) bit of the programmed division word. This property is clarified with help of Table 7.1. Only a few extra OR gates are required to adapt n' to the programmed division word, as depicted on the right side of Figure 7.25.

Divider	Minimum division ratio	Maximum division ratio
Reference	8	8191
UHF	256	262143
L - band	512	524287
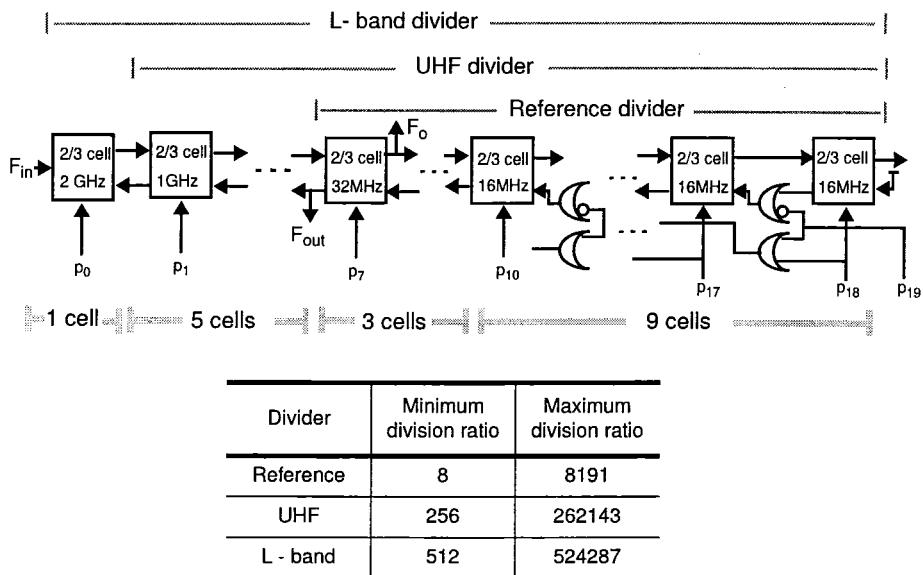
Figure 7.26.   Family of truly-modular programmable dividers.

With the additional logic the division range becomes:

- Minimum division ratio: $2^{n'_{min}}$

- Maximum division ratio: $2^{n+1} - 1$

We see that the minimum and maximum division ratios can be set independently, by choice of n'_{min} and n respectively. Subsequent changes in an optimized design can be realized with low risk. Decreasing n'_{min}, for instance, can be performed by adding additional OR gates to an existing design.

7.7.2   Low Power Truly-modular programmable Dividers in a Standard CMOS Technology

The modular structure presented in Figure 7.25 allows an existing design to be easily adapted to different input frequency requirements, simply by adding or removing divider cells in the high frequency part of the chain.

The concept of Figure 7.25 was applied in the realization of a family of low-power fully programmable frequency dividers in a standard 0.35μm CMOS Technology. Three circuits were implemented, an 18-bit "L-band divider", a 17-bit "UHF divider", and a low input frequency 12-bit "reference divider." The architecture and the division range of the dividers is presented in Figure 7.26. The L-band divider was used as the basis for the UHF and for the reference
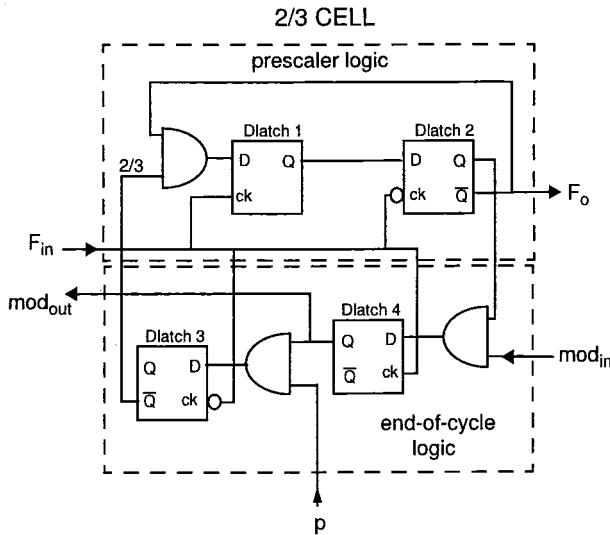
2/3 CELL

Figure 7.27. Functional blocks and logical implementation of a 2/3 divider cell.

divider. The UHF divider consists of the same circuitry as the L-band divider, except for the first 2/3 cell, which was removed. The reference divider is simply the L-band divider stripped off its 6 high frequency cells. The divider output signal is indicated by F_{out}.

The UHF divider and the reference divider were designed to be used in a low-power, PLL-synthesized CMOS pager receiver operating in the UHF frequency range. Therefore, low-power dissipation and low-interference generation were the main goals of this work. The L-band divider, on its turn, could be applied in low-power tuning systems for dual-band applications. For example, in the receiver concept described in [40], where the tuning frequency lies in the middle of the GSM-DCS1800 bands (i.e., at 1.35 GHz).

Logic Implementation of the Divider Cells

A 2/3 divider cell contains two functional blocks, as depicted in Figure 7.27. The *prescaler logic* block divides, upon control by the *end-of-cycle logic*, the frequency of the F_{in} input signal either by 2 or by 3, and outputs the divided clock signal to the next cell in the chain. The end-of-cycle logic controls the momentaneous division ratio of the cell. The division ratio depends on the state of the *mod*$_{in}$ and p signals. The *mod*$_{in}$ signal becomes active once in a division cycle. At that moment, the state of the p input is checked, and if $p = 1$, the end-of-cycle logic forces the prescaler to swallow one extra period of the input signal. In other words, the cell divides by 3. If $p = 0$, the cell stays
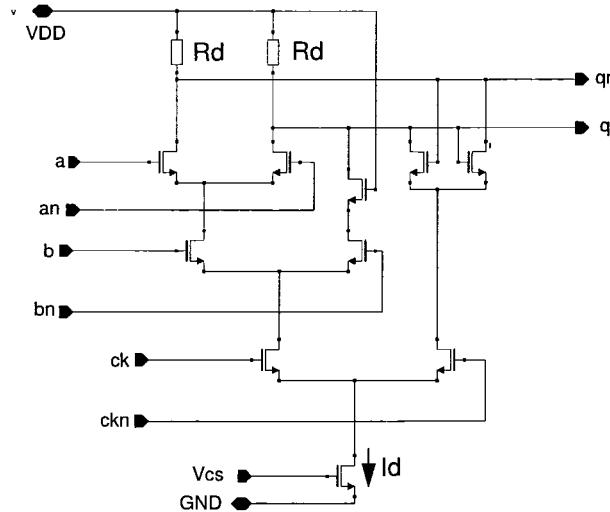
Figure 7.28.   Source coupled logic implementation of an AND gate combined with a latch function.

in division by 2 mode. Regardless of the state of the p input, the end-of-cycle logic reclocks the mod_{in} signal, and outputs it to the preceding cell in the chain (mod_{out} signal).

Circuit Implementation of the Divider Cells

The use of standard rail-to-rail CMOS logic techniques makes the integration of digital functions with sensitive RF signal processing blocks difficult, due to the generation of large supply and substrate disturbances during logic transitions. Source Coupled Logic (SCL), often referred to as MOS Current Mode Logic (MCML), has better EMC properties, because of the constant supply current and differential voltage switching operation [36, 41]. Besides, Source Coupled Logic has lower power dissipation than rail-to-rail logic, for (very) high input frequencies [42].

   The logic functions of the 2/3 cells are implemented with the SCL structure presented in Figure 7.28. The logic tree combines an AND gate with a latch function. Three "AND_latch" circuits are used to implement Dlatch1, Dlatch3, Dlatch4 and the AND gates of the 2/3 cells (see Figure 7.27). Therefore, 6 logic functions are achieved at the expense of 3 tail currents only. Dlatch2 is implemented as a "normal" D latch (without the differential pair connected to the b-bn inputs).

Divider with 3 2/3 cells - N_{div} = 15 - Fin = 2GHz
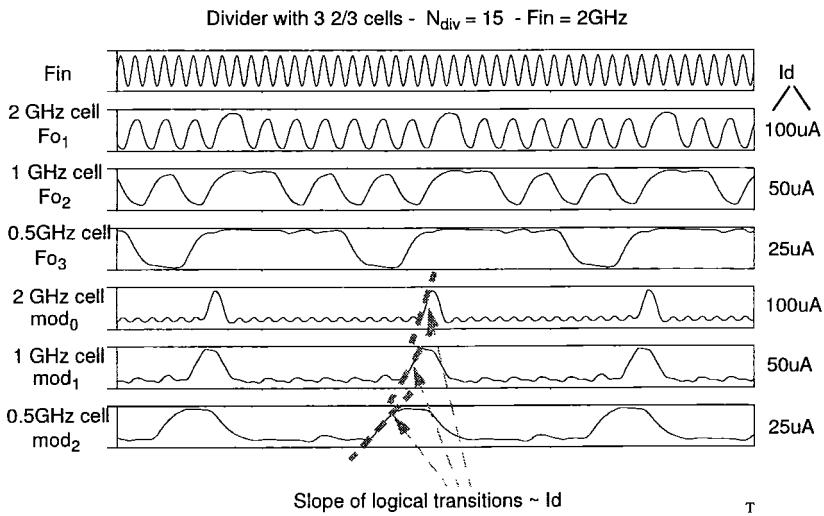
Figure 7.29. Transient simulation of optimized L-band divider.

The nominal voltage swing is set to 500 mV in the high frequency (and high current) cells, and to 300 mV in the low current cells ($Id \leq 2\ \mu A$). The voltage is generated by the tail current, set by the current source Id, and by the load resistances Rd. The W/L of the transistors is 3/0.35.

Power Dissipation Optimization

The modular structure of the programmable prescaler architecture (see Figure 7.24) enables fast and reliable optimization of power dissipation. There are no complicated delay loops, as feedback lines are only present between adjacent cells. This leads to a relatively simple and fast optimization procedure, because simulation runs can be done for clusters of two cells each time.

The critical point in the operation of the programmable prescaler are the divide by 3 actions [26]. There is a maximum delay between the *mod* and the clock signals in a given cell that still allows properly timed division by 3. The maximum delay is $\tau_{max} = 1.5 \cdot T_{in}$, where T_{in} is the period of the cell's input signal. The input frequency for each cell is scaled down by the previous one. As a consequence, the maximum allowed delay increases as one moves "down" the chain. Because the delay in a cell is inverse proportional to the cell's current consumption (which is a property of current mode logic circuits), the currents in the cells may be scaled down as well.

The results of a transient simulation with the optimized high frequency cells of the L-band divider are presented in Figure 7.29. The influence of current

Table 7.2.   Scaling of currents in the 2/3 divider cells.

Cell	Nominal current Id (μA)	Nominal load res. Rd (kΩ)
2 GHz	100	5
1 GHz	50	10
500 MHz	25	20
250 MHz	12.5	40
125 MHz	6.25	80
62 MHz	3	150
32 MHz	2	150
16 MHz	1	300

consumption on the slope of the digital signals (and hence on the time delay) is clearly observed. Layout optimization took about three iteration cycles. Transient simulations, including extracted parasitics, showed that layout parasitics caused a decrease of about 30% in the highest operation frequency, when compared to the original simulations.

Table 7.2 presents the tail current and the resistance values of the optimized divider cells. It has been demonstrated in the literature [42] that, for (very) high input frequencies, CMOS Source Coupled Logic (SCL) has lower power dissipation than standard CMOS rail-to-rail logic. On the other hand, rail-to-rail logic is much more compact than SCL logic operating at low-current levels, because of the absence of the large load resistances required in a low-frequency SCL gate. Furthermore, rail-to-rail logic has no DC current flow, and does not require an analogue biasing signal (e.g., node Vcs in Figure 7.28).

The main draw-back of a rail-to-rail gate at lower operating frequencies remains the "spiky" nature of its supply current, which leads to signal components in the supply line with much higher frequencies than the switching frequency of the gate. Besides, charge injection in the substrate also needs to be considered in some applications. These effects may lead to interference into other circuits, such as the highly sensitive input of a receiver's low-noise amplifier (LNA), and therefore to a decrease in the receiver's sensitivity. Based on these considerations, the dividers described here were fully implemented with Source Coupled Logic techniques.
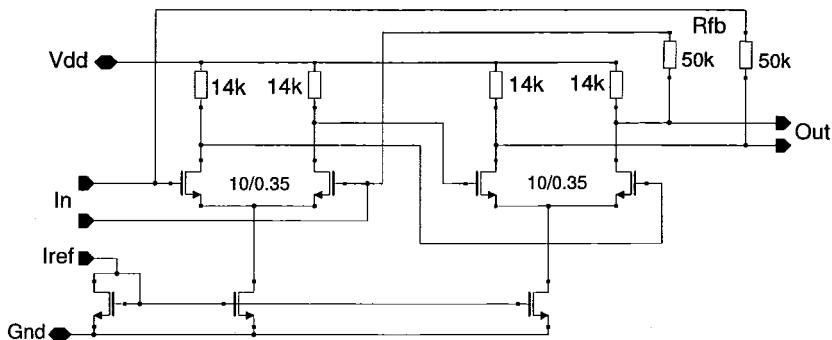
Figure 7.30.   Circuit diagram of the UHF input amplifier.

Input Amplifier

A low power dissipation frequency synthesizer requires a frequency divider with high input sensitivity. The input sensitivity determines the minimum VCO signal level that must be applied to the divider, so that correct division is performed. High input sensitivity enables the divider to be directly coupled to a wide range of VCO's, without the need for external (discrete) buffers. The total power dissipation of the frequency synthesizer is therefore (much) lower. For this reason, the L-band and UHF dividers contain integrated input amplifiers.

The input amplifiers provide the required amplification of the VCO signal to "digital" levels, determined by the sensitivity specifications and by the divider circuitry. Moreover, the input amplifiers perform other functions, which are listed below:

- to provide single-ended to differential conversion of the (very often) single-ended VCO signal

- to enable the VCO to be AC coupled to the divider function, and to provide a signal to the first divider cell with the proper DC level

- to provide reverse isolation, to prevent the divider activity from "kicking-back" and disturbing the VCO

Figure 7.30 presents the circuit implementation of the UHF amplifier. The required amplification, set by sensitivity requirements (-20 dBm) has been split into two differential stages. Each differential pair operates with 50 μA nominal current, so that the total circuit draws 125 μA from the power supply, including the input reference current.
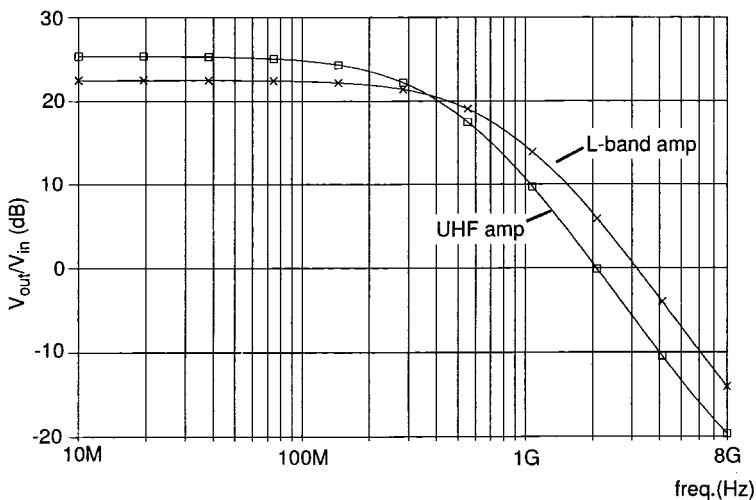
Figure 7.31.   AC small signal gain of the UHF and L-band input amplifiers.

The amplifier of Figure 7.30 is designed to be driven by a relatively low-impedance (voltage) source. The negative feedback, implemented with the 50 kΩ resistances, provides DC biasing to the first stage, and allows AC coupling of the VCO signal to the first differential pair. Furthermore, the negative feedback improves the input sensitivity at high frequencies. Without feedback, mismatch in the differential pairs can lead to a large offset in the output voltage of the amplifier, due to the high DC gain. The offset is particularly harmful for the divider sensitivity at high frequencies, because the amplifier is subject to the usual gain roll-off (see Figure 7.31). As the input frequency increases, "more" input signal is required to overcome the DC offset, which means that the sensitivity at high frequencies is decreased. In applications where the VCO is AC coupled to the amplifier—which is very often the case—, the feedback decreases the DC gain of the amplifier, hence suppressing the amplification of the differential pair offset voltage. For high frequencies, the feedback does not decrease the amplifier gain, if the output resistance R_s of the source driving the amplifier is much smaller than the feedback resistors R_{fb}. A small ratio R_s/R_{fb} also ensures stability of the amplifier, as the effect of the feedback at higher frequencies becomes negligible.

The L-band input amplifier is a scaled version of the UHF input amplifier. The tail currents were doubled, and the drain resistances were halved. The AC
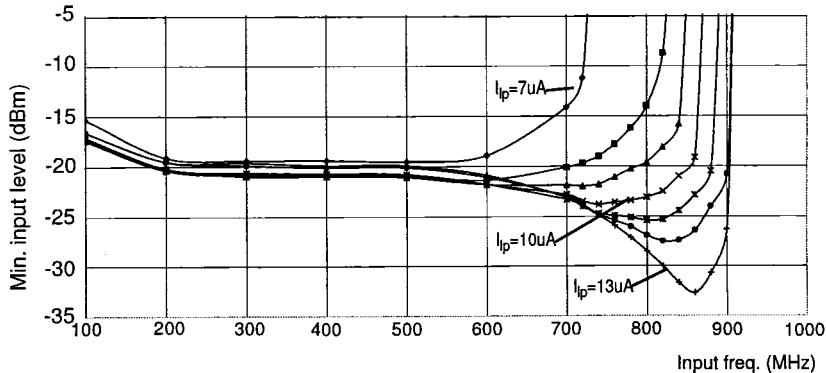
Figure 7.32. Sensitivity of the UHF divider, for different divider current settings. Division ratio = 511, nominal current is $I_{lp} = 10 \ \mu$A.

gain of the amplifiers is presented in Figure 7.31. The simulations include the first cell of the dividers, so that nominal load conditions are taken into account.

Input Sensitivity Measurements and Maximum Operation Frequencies

The supply current flowing into the amplifier and dividers can be set externally by means of control currents I_{bf} and I_{lp}. The internal biasing current I_{ref} of the input amplifier (see Figure 7.30) is controlled by current I_{bf}, and the current in the divider cells are controlled by current I_{lp}. The curves presented in this section were obtained with the nominal supply voltage of 2.2 V, except where otherwise noted.

Figure 7.32 presents sensitivity curves of UHF divider test chip for different current settings of the divider control current I_{lp}. The curve shows the minimum amplitude of a *sine-wave* input signal which is properly divided by the frequency divider, as a function of the input signal frequency.[12] The nominal value of I_{lp} is 10 μA. The "flat" portion of the curves show that the circuit is highly sensitive over a large frequency range. Hence, it is well suited for multi-band applications in the VHF/UHF range. Figure 7.33 displays the effect of the input amplifier current on the input sensitivity. One sees that the amplifier current influences mostly the sensitivity on the "flat" part of the sensitivity curve. The nominal value of I_{bf} is 12.5 μA, corresponding to a total current in the input buffer of 125 μA.

---

[12]The power of the input signal is expressed in dBm, i.e. dB with respect to 1 mW, dissipated into a 50 Ω load; 0 dBm = 223.6 mVrms.
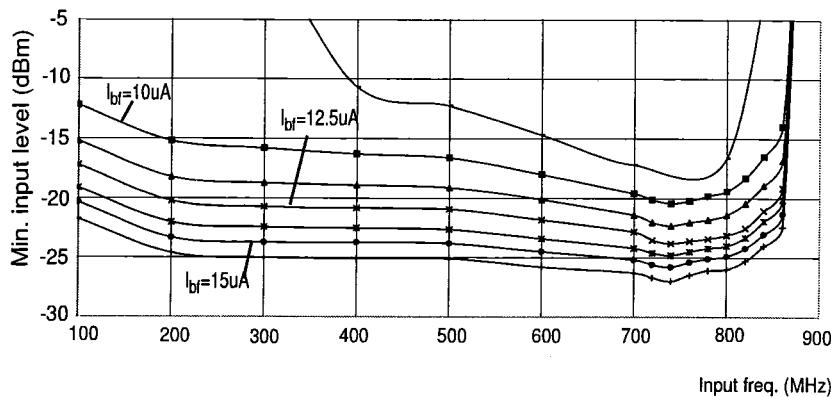
Figure 7.33. Sensitivity of the UHF divider, for different input buffer current settings. Division ratio = 511, nominal current is $I_{bf} = 12.5 \ \mu A$.
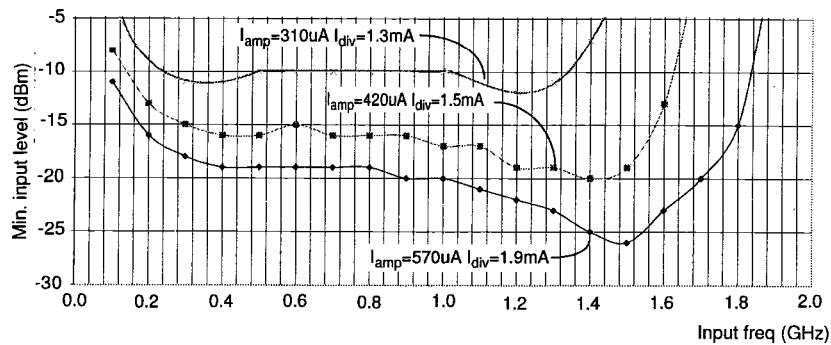
Figure 7.34. Sensitivity curves of the L-band divider, for a few divider and amplifier current settings. Division ratio = 1023.

Figure 7.34 presents measured sensitivity curves of the L-band divider, for different current settings in the divider and input amplifier. Such as the UHF divider, the L-band divider is highly sensitive over a large frequency range (> 1 GHz). The dividers were driven with a differential input signal, carried over PCB microstrip lines with a characteristic impedance of 50 Ω. The microstrip lines were loaded with discrete resistances of 50 Ω to ground, which were set close to the input leads of the input amplifiers.

The maximum operation frequencies of the UHF and L-band dividers, as a function of the current consumption, are plotted in Figure 7.35. We see that the 17-bit divider operates correctly at frequencies in excess of 800 MHz, using not more than 0.5 mA. Setting the nominal current to 0.7 mA enables operation up to 900 MHz. It is interesting to note that the current consumption of the L-band
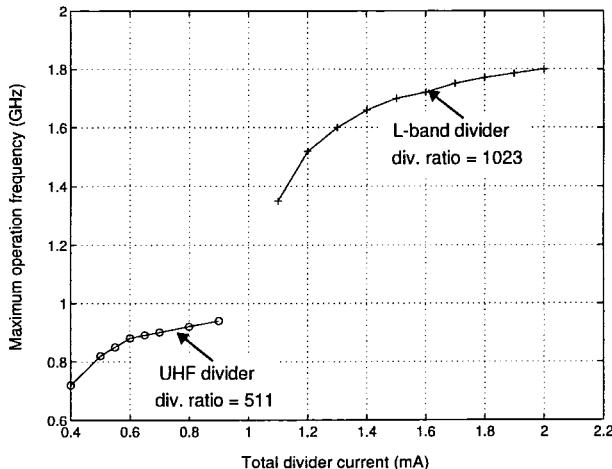
Figure 7.35. Maximum operation frequency of the UHF and L-band dividers, as a function of the current consumption (excluding input amplifiers).
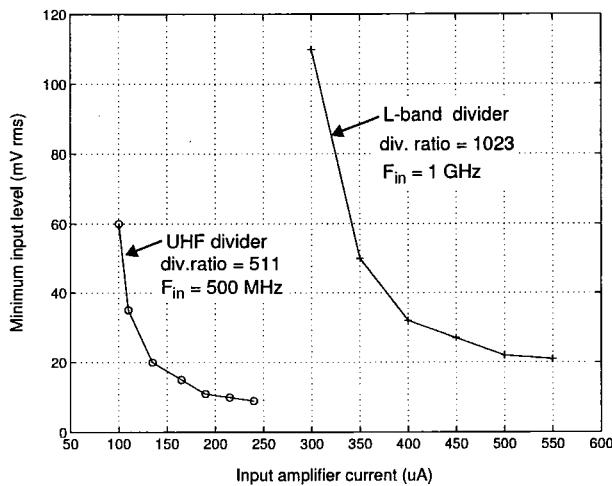
Figure 7.36. Sensitivity of the UHF and L-band dividers on the "flat" portion of the sensitivity curve, expressed in mVrms, as a function of the input buffer current consumption.

divider is about a factor 2 higher than the UHF divider's consumption, which means that the first 2/3 cell of the L-band divider consumes as much power as the whole 17-bit UHF divider. This property follows from the current scaling strategy described on page 293.

The effect of the input amplifiers current consumption on the input sensitivities is displayed in Figure 7.36. We see that the sensitivity on the "flat" portion

of the curve can be readily traded-off against the input amplifier power dissipation. Setting the UHF amplifier current to 230 μA yields a sensitivity value in excess of 10 mVrms (-27 dBm). This is, to the best of the author's knowledge, the highest sensitivity value for CMOS frequency dividers published in the literature (outside the typical self-oscillation region of divider circuits, of course).

The influence of the supply voltage on the maximum operating frequency was found to be small ($\approx$ 5% for Vdd decreased from 2.2 V down to 1.8 V). It is interesting to mention that MOS Current Mode Logic circuits have been demonstrated to operate with supply voltages as low as 1.2 V without significant loss of speed [42].

7.8   DESIGN OF HIGH FREQUENCY PHASE-FREQUENCY DETECTOR/CHARGE-PUMP COMBINATIONS

The phase-frequency detector/charge-pump (PFD/CP) combination is the most widely used phase detector configuration within the context of PLL frequency synthesizers. Nevertheless, the PFD/CP circuitry poses some challenges to the circuit designer:

- in practice, many charge-pump circuits create a dead-zone in the combined transfer function of the PFD/CP, which results in a decreased spectral purity performance when the loop is (nearly) phase-locked.

- usual PFD implementations require a more complex circuitry than e.g., the double-balanced mixer, which in turn limits the maximum operation frequency. Some techniques used to solve the dead-zone problem limit the maximum operation frequency as well.

These two aspects are dealt with in this section. The architecture and circuit design of a PFD/CP combination that operates well into VHF frequencies will be described.

7.8.1   The Dead-zone Phenomenon

The dead-zone is the region of the transfer curve of the PFD/CP combination where there is no output from the charge-pump in response to the phase error at the input of the PFD. The dead-zone phenomenon is schematically depicted in Figure 7.37.

Within the dead-zone the loop is essentially open and the VCO is effectively "free-running;" as a consequence, no reduction of the oscillator phase noise can be performed and leakage currents in the loop filter will lead to a chaotic behavior of the PLL output frequency. In an ideal situation, without leakage
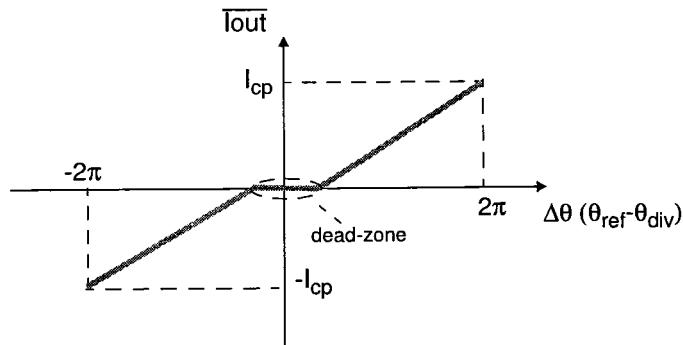
Figure 7.37.   The dead-zone region of the PFD/CP combination is marked by a non-linear relationship of the output current to input phase error.

currents, the oscillator frequency would be stable and the phase error $\Delta\theta$ might stay indefinitely within the dead-zone. In practice, however, leakage currents are always present. Leakage results in a drift of the tuning voltage of the VCO, and eventually the phase error will reach a boundary of the dead-zone. At that moment, a correction pulse from the PFD/CP will "push" the phase error back into the dead-zone. The phase error then may or may not reach the other end of the dead-zone, where it will be "pushed back" towards the center of the dead-zone, and so on. Irrespective of the exact phase error behavior, the irregular pattern at the output signal of the PFD/CP will seriously disturb the spectral purity of the VCO.

The usual cause of the dead-zone is the impossibility of the (slow) charge-pump current switches to react to the narrow *up* and *down* signals coming from the PFD when the loop is (nearly) in-lock, see Section 7.4.3 and Figure 7.7. The usual way of eliminating the dead-zone is by increasing the minimum width of the *up* and *down* signals from the PFD when the phase error is close to zero. This can be done by adding a fixed delay element is series with the AND gate which generates the *reset* signal to the D-FFs [24]. A second possibility is to monitor the output of the current switches and to delay generation of the reset signal until the current switches deliver current to the output node. Use of these techniques implies that PFD/CPs implemented in standard IC technologies are limited in their high frequency operation by the slow switching speed of the pnp or P-MOST transistors used in the charge-pump. For example, in the circuit that will be described in this section the f_T of the pnp transistor was about 200 MHz.
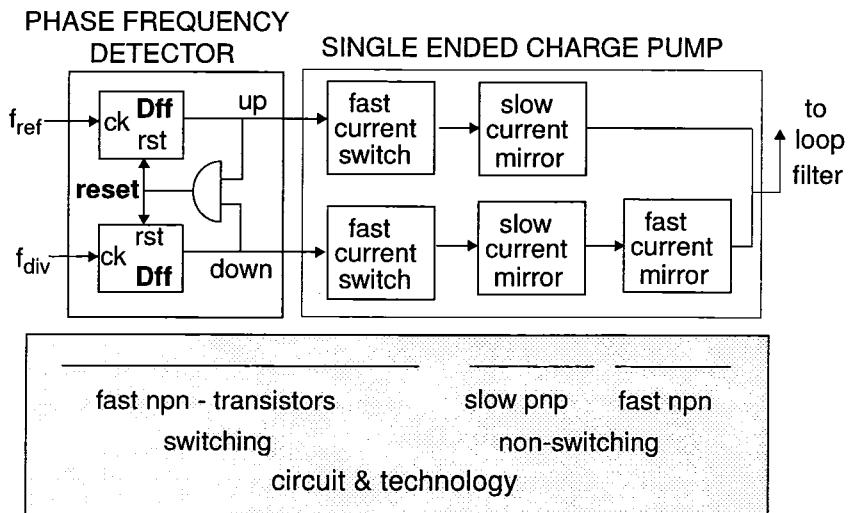
Figure 7.38.　Structure of a 300 MHz phase-frequency detector/charge-pump.

7.8.2　VHF PFD/CP: Architecture

The problem of combining high speed operation with the absence of a dead-zone can be solved with the architecture presented in Figure 7.38. The phase-frequency detector consists of two D-FFs and an AND gate, as described in Section 7.4.3. To avoid the dead-zone the switching part of the charge-pump is implemented with fast npn-transistors. These fast switches can follow the narrow *up* and *down* signals from the PFD when the loop is in-lock. (The speed of the transistors which compose the PFD is the same as the speed of the switching part of the CP.) The current switches are succeeded by two high-performance matched current mirrors using, in the present case, slow pnp transistors with $f_T \sim 200$ MHz. The operation of the current mirrors may distort the shape of the pulses at their output, yet *it keeps the average charge intact.* In order to realize a single ended charge-pump function an additional npn current mirror is added in the down-branch. In principle this function introduces an asymmetry, but due to the large difference in the cut-off frequencies of the transistors its influence is negligible. The charges provided by the current mirrors are subtracted in the output node of the charge-pump before reaching the loop filter. As the phase error information is, in fact, present on the average charge difference the combination presents no dead-zone. The equivalent low-pass filtering operation performed by the slow pnp current mirrors on the *up* and *down* current signals can be incorporated in the loop transfer function as an additional pole in the loop filter.

out

| fast current switch | slow current mirror | fast current mirror | slow current mirror | fast current switch |

up → | | | ↑ | | | ← down
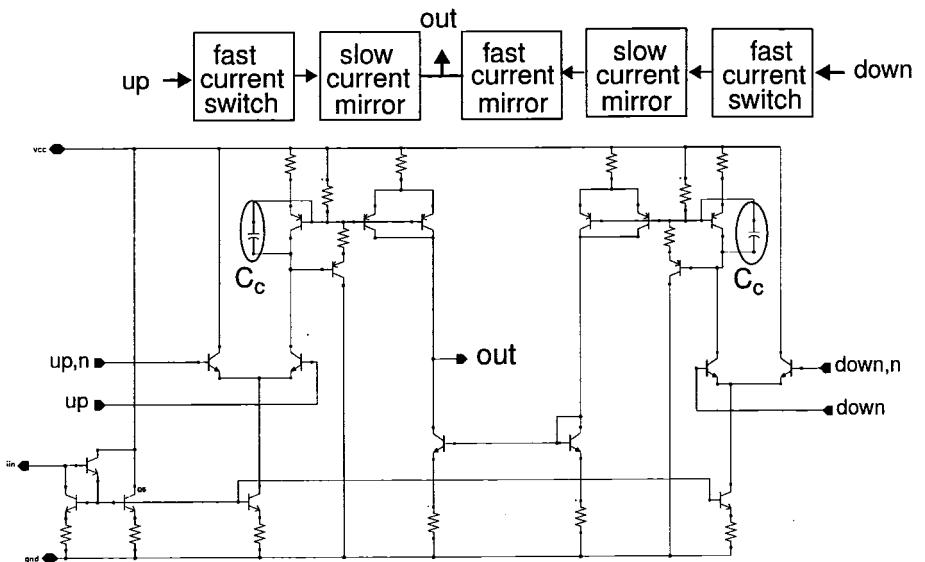
Figure 7.39.   Circuit implementation of wide band charge-pump.

7.8.3   VHF PFD/CP: Circuit Implementation

The PFD circuit design has been optimized for high speed and low power dissipation by using dedicated emitter-coupled logic (ECL) circuit design. The PFD optimization criteria was to achieve a linear static transfer for phase differences up to $\pm 180°$ at the highest operation frequency of 307 MHz; this is a requirement for frequency discrimination at the highest operating frequency, or—in other words—to guarantee that the loop locks at the correct frequency under all circumstances, see Section 7.4.3 and Ref. [18].

The circuit implementation of the charge-pump is shown in Figure 7.39. The two high performance pnp current mirrors dominate the dynamic transfer function of the PFD/CP. The pnp current mirrors are designed to combine maximum DC accuracy with maximum bandwidth. Each pnp mirror has base current compensation and a compensation capacitor C_c for stabilizing the feedback loop of the base current compensation.

The resulting (simulated) magnitude and phase transfer of the precision mirror is shown in Figure 7.40, with and without the compensation capacitor of 2 pF. In the simulations, the back-plate parasitic capacitance of C_c was taken into account. The bandwidth of the mirror with compensation capacitor is about 40 MHz, which is large enough to ensure stable closed-loop operation in the intended application (i.e., with a closed-loop bandwidth in the order of 2 MHz).
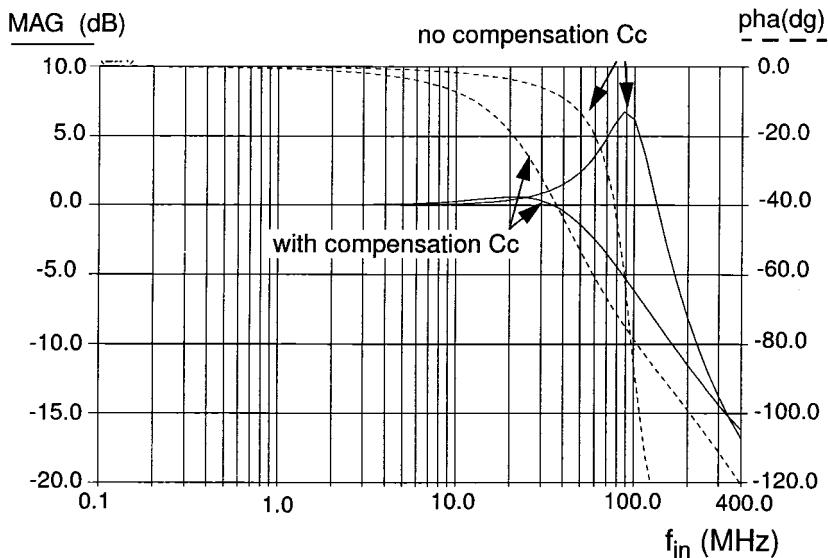
Figure 7.40. Simulated magnitude and phase transfer of the precision PNP current mirrors; with and without compensation capacitor C_c.

A micrograph of the realized PFD/CP combination test-chip is shown in Figure 7.41. The IC included input buffers, a stabilizer, the high-frequency phase-frequency detector and three charge-pumps, two with a nominal output current level of 200 μA and a third charge-pump with an output current level of 10 μA.

7.8.4   VHF PFD/CP: Measurement Results

Figure 7.42 shows the measured static transfer function of the PFD/CP vs. phase error for two different reference frequencies. This measurement was performed with a pulse generator which provided two output signals with a programmable time delay between them. From Figure 7.42 it can be seen that the PFD/CP works correctly within $\pm 160°$ at 200 MHz and within $\pm 60°$ at 300 MHz. Furthermore, it shows that the PFD/CP has no dead zone, although none of the usual dead zone compensation techniques, which would prevent operation in the VHF frequency range, were used. It was expected that the linear region of operation would extend to $\pm 180°$ at the maximum operation frequency of 307 MHz. The disagreement between measurements and simulations was traced to an additional 1 ns delay in the reset line of the DFFs, which originated in a parasitic capacitive loading of the reset line. The problem can be solved by decreasing the physical length of the reset line and by a slight increase in the
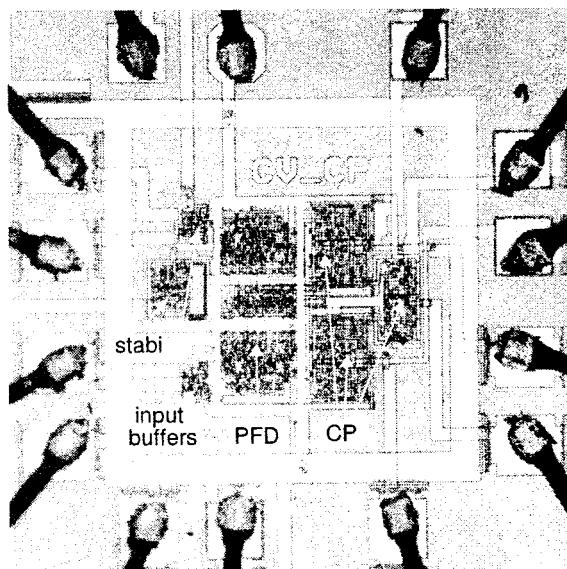
Figure 7.41.   Micrograph of the 300 MHz phase-frequency detector / charge-pump test-chip.
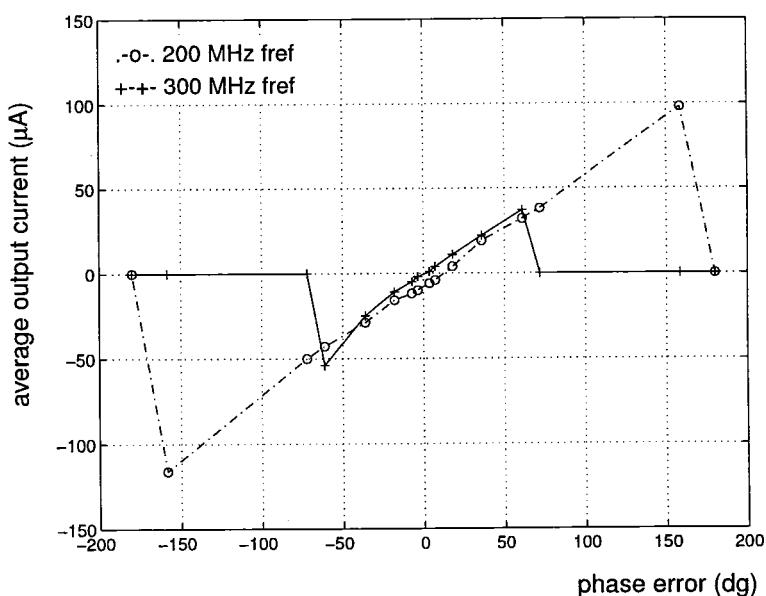
Figure 7.42.   Measured static transfer of the VHF PFD/CP.

Range: -20 dBm   Oversweep: Off   29-Aug-1996 12:23
Res BW: 17 000 Hz   VBW: Off   Swp Time: 811.2 mSec
                    ΔMkr   40 057 500 Hz   -3.82 dB

-59.3
dBm

LogMag

1
dB
/div

-69.3
dBm
        Start: 1 000 000 Hz   Stop: 50 000 000 Hz
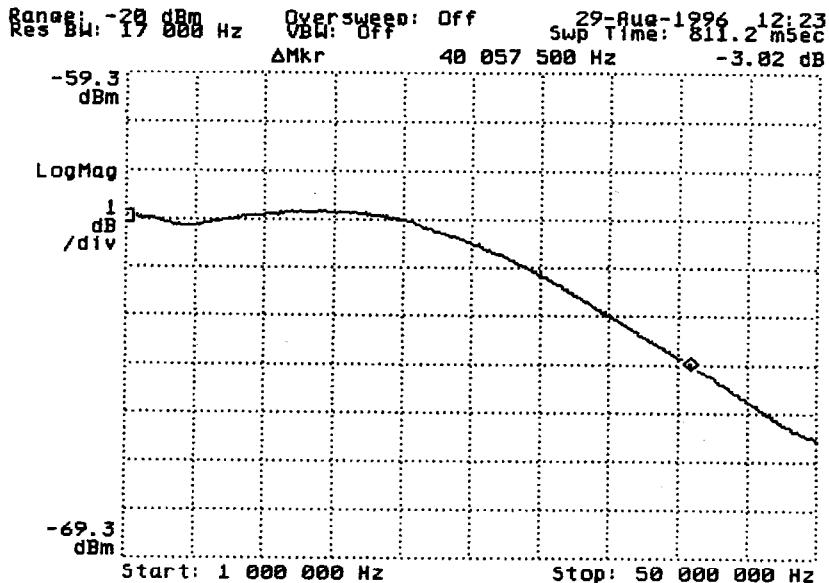
Figure 7.43.   Measured dynamic transfer (gain) of the PFD/CP at a reference frequency of 300 MHz, as a function of the modulation frequency f_m.

AND gate current level. The PFD operated correctly as phase detector up to 380 MHz, with a nominal power dissipation of 10 mW (5 V, 2 mA).

Figure 7.43 shows the dynamic transfer function of the PFD/CP. The curve shows the (scaled) magnitude of the output current of the charge-pump as a function of the modulation frequency of the phase error at the input of the PFD. The measurement was performed as follows: the 300 MHz output signal from a sine-wave signal generator was split into two paths. One of the paths was applied directly to the f_{ref} input of the PFD/CP test-chip. The signal on the second path was added, with a power combiner, to the sine-wave output of the tracking generator of a base-band spectrum analyzer. The output of the power combiner was then applied to the f_{div} input of the PFD/CP. The output level of the tracking generator was kept much smaller than the level of the signal generator. As the input buffers of the PFD perform a limiting operation on their input signals, the AM component of the combined signal at the f_{div} input was suppressed. The resulting signal is a 300 MHz phase modulated signal with a constant phase deviation θ_p and with a modulation frequency f_m equal to the frequency of the tracking generator. The output signal of the charge-pump was applied to the input of the baseband spectrum analyzer, whose output is presented in Figure 7.43.
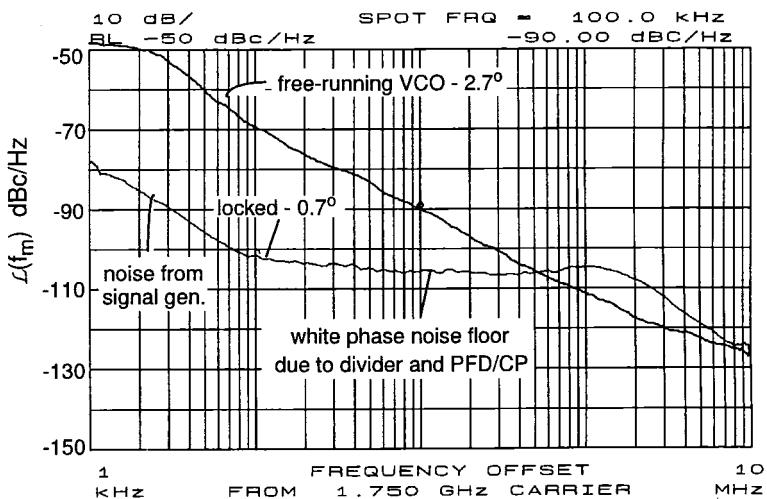
Figure 7.44. Measured phase noise performance of a wide-band loop implemented with the test chips described in this chapter, and of a free-running VCO.

The dynamic transfer function is dominated by the slow pnp current mirrors and by the capacitive load at their input resulting from the Miller capacitance of the npn current switch. The bandwidth of 41 MHz is independent of the modulation index, and large enough for the implementation of, for example, wide-loop bandwidths for suppression of the phase noise of (integrated) VCOs.

The closed-loop performance of the integrated PFD/CP and of a two-bit bipolar programmable divider is presented in Figure 7.44. To assess the phase noise floor of the building blocks a good LC VCO and a signal generator were used. The LC VCO operated at 1.75 GHz. The reference frequency, coming from the signal generator, was set to 350 MHz with a corresponding division ratio in the divider of 5. The loop bandwidth, implemented with a discrete loop filter, was set to 2 MHz, which is much smaller than the internal bandwidth of the PFD/CP. With this configuration the residual phase deviation of the VCOlor the square root of the integral of the phase noise power density between 10 kHz and 100 MHzlwas reduced from 2.7° rms to 0.7° rms. Figure 7.44 shows that the equivalent phase noise floor $\mathcal{L}_{eq}(f_{ref})$ of the divider and PFD/CP combination is $\mathcal{L}_{eq}(f_{ref}) \sim -120$ dBc/Hz (i.e., $-106 - 20\log 5$) at a reference frequency $f_{ref} = 350$ MHz.

REFERENCES

[1] F.M. Gardner, *Phase-lock Techniques*, Wiley, New York, 2nd. edition, 1979.

[2] U.L. Rohde, *RF and Microwave Digital Frequency Synthesizers*, Wiley, New York, 1997.

[3] B. Miller and B. Conley, "A Multi-modulator Fractional Divider," in *Annual Frequency Control Symposium*, 1990, vol. 44, pp. 559–567.

[4] B. Miller and R.J. Conley, "A Multiple Modulator Fractional Divider," *IEEE Transactions on Instrumentation and Measurement*, vol. 40, no. 3, pp. 578–583, June 1991.

[5] T.A.D. Riley, M.A. Copeland and T.A. Kwasniewski, "Delta-Sigma Modulation in Fractional-N Frequency Synthesis," *IEEE Journal of Solid-State Circuits*, vol. 28, no. 5, pp. 553–559, May 1993.

[6] H. Adachi *et al.*, "High-Speed Frequency-Switching Synthesizer Using Fractional N Phase-Locked Loop," *Electronics and Communication in Japan (IEICE Transactions on Electronics),Part 2*, vol. 77, no. 4, pp. 20–28, 1994.

[7] M.H. Perrot, T.L. Tewksbury III and C.G. Sodini, "A 27-mW CMOS Fractional-N Synthesizer Using Digital Compensation for 2.5-Mb/s GFSK Modulation," *IEEE Journal of Solid-State Circuits*, vol. 32, no. 12, pp. 2048–2060, Dec. 1997.

[8] N.M. Filiol *et al.*, "An Agile ISM Band Frequency Synthesizer with Built-in GMSK Data Modulation," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 7, pp. 998–1008, July 1998.

[9] L. Sun *et al.*, "Reduced Complexity, High Performance Digital Delta-Sigma Modulator for Fractional-N Frequency Synthesis," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, 1999, vol. 2, pp. 152–155.

[10] C.S. Vaucher, "An Adaptive PLL Tuning System Architecture Combining High Spectral Purity and Fast Settling Time," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 4, pp. 490–502, Apr. 2000.

[11] C.S. Vaucher, *Architectures for RF Frequency Synthesizers*, Ph.D. Thesis, University of Twente, The Netherlands, 2001.

[12] C.S. Vaucher and K. Kianush, "A Global Car-radio IC with Inaudible Frequency Jumps," in *IEEE International Conf. on Consumer Electronics (ICCE)*, 1998, vol. 17, pp. 218–219.

[13] B. Razavi, "A 900MHz/1.8GHz CMOS Transmitter for Dual-Band Applications," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 5, pp. 573 – 579, May 1999.

[14] H. Taub and D.L. Schilling, *Principles of Communication Systems*, McGraw-Hill, New York, 2nd. edition, 1986.

[15] W.P. Robins, *Phase Noise in Signal Sources*, 9. IEE Telecomm., London, 2nd edition, 1996.

[16] B. Razavi, *RF Microelectronics*, Prentice Hall, New York, 1998.

[17] Y. Sumi *et al.*, "PLL Synthesizer with Multi-Programmable Divider and Multi-Phase Detector," *IEEE Transactions on Consumer Electronics*, vol. 45, no. 3, pp. 950–955, Aug. 1999.

[18] M. Soyuer and R.G. Meyer, "Frequency Limitations of a Conventional Phase-Frequency Detector," *IEEE Journal of Solid-State Circuits*, vol. 25, no. 4, pp. 1019–1022, Aug. 1990.

[19] B. Razavi, *Monolithic Phase-locked Loops and Clock Recovery Circuits*, IEEE Press, New York, 1996.

[20] F.M. Gardner, "Charge-Pump Phase-lock Loops," *IEEE Transactions on Communications*, vol. 28, no. 11, pp. 1849–1858, Nov. 1980.

[21] A.K. Hadjizada *et al.*, "TV and TVSAT Mixer-oscillator PLL IC," *IEEE Transactions on Consumer Electronics*, vol. 41, no. 3, pp. 942–945, Aug. 1995.

[22] Philips Semiconductors, *TSA5059 Datasheet - 2.7 GHz I²C-bus controlled low phase noise frequency synthesizer*, 2000.

[23] D. Mijuskovic *et al.*, "Cell-Based Fully Integrated CMOS Frequency Synthesizers," *IEEE Journal of Solid-State Circuits*, vol. 29, no. 3, pp. 271–279, Mar. 1994.

[24] J. Craninckx and M. Steyaert, "A Fully Integrated CMOS DCS-1800 Frequency Synthesizer," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 12, pp. 2054–2065, Dec. 1998.

[25] W. Rhee, "Design of High-Performance CMOS Charge Pumps in Phase-Locked Loops," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, 1999, vol. 2, pp. 545–548.

[26] C.S. Vaucher and D. Kasperkovitz, "A Wide-Band Tuning System for Fully Integrated Satellite receivers," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 7, pp. 987–997, July 1998.

[27] V. F. Kroupa, "Noise Properties of PLL systems," *IEEE Transactions on Communications*, vol. 30, no. 10, pp. 2244–2552, Oct. 1982.

[28] D.E. Phillips, "Ramdom Noise in Digital Gates and Dividers," in *Annual Frequency Control Symposium*, 1987, vol. 41, pp. 507–511.

[29] W.F. Egan, "Modeling Phase Noise in Frequency Dividers," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 37, no. 4, pp. 307–315, July 1990.

[30] M.R. McClure, "Residual Phase Noise of Digital Frequency Dividers," *Microwave Journal*, vol. 35, no. 3, pp. 124–128, Mar. 1992.

[31] M.Q. Tavares, *PLL frequency synthesizers: phase noise issues and wide band loops*, Ph.D. Thesis, Institut National des Sciences Appliquees de Lyon, France, 1999.

[32] Philips Semiconductors, *UMA1022M Datasheet - Low cost dual frequency synthesizer for radio telephones*, 1998.

[33] Y. Kado *et al.*, "An Ultralow Power CMOS/SIMOX Programmable Counter LSI," *IEEE Journal of Solid-State Circuits*, vol. 32, no. 10, pp. 1582–1587, Oct. 1997.

[34] T. Senef *et al.*, "A Sub-1mA 1.5GHz Silicon Bipolar Dual Modulus Prescaler," *IEEE Journal of Solid-State Circuits*, vol. 29, no. 10, pp. 1206–1211, Oct. 1994.

[35] J. Craninckx and M. Steyaert, "A 1.75GHz/3V Dual-modulus Divide-by-128/129 Prescaler in 0.7 μm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 7, pp. 890–897, July 1996.

[36] F. Piazza and Q. Huang, "A Low Power CMOS Dual Modulus Prescaler for Frequency Synthesizers," *IEICE Transactions on Electronics*, vol. E80-C, no. 2, pp. 314–319, Feb. 1997.

[37] N.-H. Sheng *et al.*, "A High-Speed Multimodulus HBT Prescaler for Frequency Synthesizer Applications," *IEEE Journal of Solid-State Circuits*, vol. 26, no. 10, pp. 1362–1367, Oct. 1991.

[38] C.S. Vaucher and Z. Wang, "A Low-power Truly-modular 1.8GHz Programmable Divider in Standard CMOS Technology," in *Proc. of the 25th European Solid-State Circuits Conference (ESSCIRC)*, 1999, vol. 25, pp. 406–409.

[39] C.S. Vaucher *et al.*, "A Family of low power truly-modular Programmable Dividers in Standard 0.35μm CMOS Technology," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 7, pp. 1039–1045, July 2000.

[40] S. Wu and B. Razavi, "A 900MHz/1.8GHz CMOS Receiver for Dual-Band Applications," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 12, pp. 2178 – 2185, Dec. 1998.

[41] S.G. Kwaaitaal, *Preliminary Study of a Source-Coupled Logic Frequency Divider for RF Applications*, M.Sc. Thesis, University of Twente, The Netherlands, 1999.

[42] M. Mizuno *et al.*, "A GHz MOS Adaptive Pipeline Technique Using MOS Current-Mode Logic," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 6, pp. 784–791, June 1996.

Appendix A
Behavioral Models

All the behavioral models are written in Matlab language.

MODEL FOR A LOW NOISE AMPLIFIER

The code for the LNA is given below.

```
function Output_LNA=lna(PowerGain,NoiseFigure,OIP3dbm,Qs,Input_Signal,Fs);

% This program represents a RF LNA model:
% Output_LNA=lna(PowerGain,NoiseFigure,OIP3dbm,Qs,BW);
% PowerGain      Gain of LNA in dB
% NoiseFigure    NoiseFigure of LNA in dB
% OIP3dbm        Output IP3 in dBm
% Qs             Quality factor input LNA
% Input_Signal   (Noisy) Signal at input LNA
% Fs=30e9;       sampling frequency of Input_Signal
%
% program written by Leenaerts, de Vreede and Sylla
% Philips Research Laboratories, 1999
%%%%%%%%%%%%%%%%%%%%%%%UserDefined%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

kBoltzmann=1.38e-23;
T=290;          %%Temperature in Kelvin
R=50; %%Matching resistance
Fn=Fs/2; %Nyquist frequency
BW=Fn; %BT bandwidth


%%%%%%%%%%%%%%%%%%% Low Noise Amplifier%%%%%%%%%%%%%%%%%%%%%%

% NONLINEARITY

    Gain=10^(PowerGain/20);     %voltage
    IIP3dbm=OIP3dbm-PowerGain;
    IIP3=10^((IIP3dbm-30)/10);  %power (Watts)
    IIP3=IIP3/(Qs^2);
    IIP3v=(IIP3*R)^0.5;         %voltage
    A2=0;
    A3=4*Gain/(3*IIP3v*IIP3v);
    A1=Gain;
    nonlin=[A3 A2 A1 0];
    NonLinearLNA=polyval(nonlin,Input_Signal);

% NOISE
% NF=(total output noise power)/(output noise due to input source)
% The mean square noise voltage due to input source is 4kTR*BW with BW the sampled band.

    NoiseFactor=10^(NoiseFigure/10);   %noise ratio in numbers
    InputNoisems=4*kBoltzmann*T*BW*R;  %impedance is 50 Ohms
    InputNoiserms=InputNoisems^0.5;    %rms noise voltage at input
```

313

```
    OutputNoiseDueToInputrms=polyval(nonlin,InputNoiserms); %rms output noise voltage
%due to input noise
    OutputNoisems=NoiseFactor*((OutputNoiseDueToInputrms)^2);
    OutputNoiserms=OutputNoisems^0.5; %Total output noise root mean square voltage

% OutputNoiserms represents the root mean square noise voltage at the output.
% We have to generate normal distributed noise with a root mean square noise voltage
% equal to OutputNoiserms and add this time signal to NonLinearLNA

    SignalSize=size(Input_Signal);
    Noise=randn(1,SignalSize(2)); %normal distribution
    NoiseVoltage=OutputNoiserms*Noise; %normal distributed noise at output according
%to NoiseFactor
    Output_LNA=NonLinearLNA+NoiseVoltage;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%% End of Low Noise Amplifier%%%%%%%%%%%%
```

MODEL FOR A MIXER

The code for the mixer is given below.

```
function Output_Mixer=mixer(RFsignal,LOsignal,PowerConversionGain,NoiseFigure,OIP3dbm,Qs,Fs);

% This program represents a RF Mixer model:
% Output_Mixer=mixer(RFsignal,LOsignal,PowerConversionGain,NoiseFigure,OIP3dbm,Fs);
% RFsignal                (Noisy) RF signal at the RF input of the mixer
% LOsignal                (Noisy) LO signal at the LO input of the mixer
% PowerConversionGain     PowerConversionGain of Mixer in dB
% NoiseFigure             NoiseFigure of Mixer in dB
% OIP3dbm                 SSB (Single Side Band) Output IP3 in dBm
% Qs                      Quallity factor matching network at the input
% Fs              sampling frequency of RFsignal
%
% program written by Leenaerts, de Vreede and Sylla
% Philips Research Laboratories, 1999
%%%%%%%%%%%%%%%%%%%%UserDefined%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

kBoltzmann=1.38e-23;
T=290;          %%Temperature in Kelvin
R=50; %%Matching resistance
Fn=Fs/2; %Nyquist frequency
BW=Fn; %BT bandwidth

IdealMixSignal=RFsignal.*LOsignal;




%%%%%%%%%%%%%%%%% Mixer %%%%%%%%%%%%%%%%%%%

% NONLINEARITY

    Gain=10^(PowerConversionGain/20);      %voltage
    IIP3dbm=OIP3dbm-PowerConversionGain;
    IIP3=10^((IIP3dbm-30)/10);   %power (Watts)
    IIP3=IIP3/(Qs^2);
    IIP3v=(IIP3*R)^0.5;          %voltage
    A2=0;
    A3=4*Gain/(3*IIP3v*IIP3v);
    A1=Gain;
```

```
    nonlin=[A3 A2 A1 0];
    NonLinearMixer=polyval(nonlin,IdealMixSignal);

% NOISE
% NF=(total output noise power)/(output noise due to input source)
% The mean square noise voltage due to input source is 4kTR*BW with BW the sampled band.

    NoiseFactor=10^(NoiseFigure/10);   %noise ratio in numbers
    InputNoisems=4*kBoltzmann*T*BW*R;  %impedance is 50 Ohms
    InputNoiserms=InputNoisems^0.5;    %rms noise voltage at input
    OutputNoiseDueToInputrms=polyval(nonlin,InputNoiserms); %rms output noise voltage
     %due to input noise
    OutputNoisems=NoiseFactor*((OutputNoiseDueToInputrms)^2);
    OutputNoiserms=OutputNoisems^0.5; %Total output noise root mean square voltage

% OutputNoiserms represents the root mean square noise voltage at the output.
% We have to generate normal distributed noise with a root mean square noise voltage
% equal to OutputNoiserms and add this time signal to NonLinearLNA

    SignalSize=size(IdealMixSignal);
    Noise=randn(1,SignalSize(2)); %normal distribution
    NoiseVoltage=OutputNoiserms*Noise; %normal distributed noise at output
   %according to NoiseFactor
    Output_Mixer=NonLinearMixer+NoiseVoltage;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%% End of Mixer %%%%%%%%%%%%%%%%%
```

MODEL FOR A POWER AMPLIFIER

The code for the PA is given below.

```
function Output_PA=pa(PowerGain,NoiseFigure,OIP3dbm,Qs,Input_Signal,Powsupp,PAE,Fs);

% This program represents a RF Linear PA model:
% Output_PA=pa(PowerGain,NoiseFigure,OIP3dbm,Qs,BW,Powsupp,PAE);
% PAE           Power Added Efficiency of the PA
% Powsupp       Supply Voltage
% NoiseFigure   NoiseFigure of PA in dB
% OIP3dbm       Output IP3 in dBm
% Qs            Quality factor input PA
% Input_Signal  (Noisy) Signal at input PA
% Fs=30e9;   sampling frequency of Input_Signal

% program written by Leenaerts, de Vreede and Sylla
% Philips Research Laboratories, 1999
%%%%%%%%%%%%%%%%%%%%%UserDefined%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

kBoltzmann=1.38e-23;
T=290;          %Temperature in Kelvin
R=50; %Matching resistance
Fn=Fs/2; %Nyquist frequency
BW=Fn; %BT bandwidth

%%%%%%%%%%%%%%%%%%% Power Amplifier%%%%%%%%%%%%%%%%%%%%%%%

%we first have to compute the averaged input power. This can be find by integration
%over the input signal over the full time period;

    LocalPower=(Input_signal.^2)/R; %power in Watts per time unit
    [n,m]=size(LocalPower);
```

```
    powervector=reshape(LocalPower,1,n*m); %generate vector
    Pin_PA=cumsum(powervector); %cumulative power
    Pin_PA=Pin_PA(n*m)/(n*m);    %averaged power in Watts

% NONLINEARITY

    Pin_PAdbm= 10*log10(Pin_PA/1e-3);           % Input power in dBm
    Pout_PA= ((Powsupp*PAE))+Pin_PA;            % output power of the PA in watt
    Pout_PAdbm=10*log10(Pout_PA/1e-3);          % Output power in dbm
    Gain=((Powsupp*PAE)/Pin_PA)+1;
    PowerGain=10*log10(Gain);                   %  Gain in db
    IIP3dbm=OIP3dbm-PowerGain;
    IIP3=10^((IIP3dbm-30)/10);                  %power (Watts)
    IIP3=IIP3/(Qs^2);
    IIP3v=(IIP3*R)^0.5;                         %voltage
    A2=0;%0.01;
    A3=0;%4*Gain/(3*IIP3v*IIP3v);
    A1=Gain;
    nonlin=[A3 A2 A1 0];
    NonLinearPA=polyval(nonlin,Input_Signal);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%NOISE%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% NF=(total output noise power)/(output noise due to input source)
% The mean square noise voltage due to input source is 4kTR*BW with BW the sampled band.

    NoiseFactor=10^(NoiseFigure/10);    %noise ratio in numbers
    InputNoisems=4*kBoltzmann*T*BW*R;   %impedance is 50 Ohms
    InputNoiserms=InputNoisems^0.5;     %rms noise voltage at input
    OutputNoiseDueToInputrms=polyval(nonlin,InputNoiserms); %rms output noise voltage
   %due to input noise
    OutputNoisems=NoiseFactor*((OutputNoiseDueToInputrms)^2);
    OutputNoiserms=OutputNoisems^0.5; %Total output noise root mean square voltage

% OutputNoiserms represents the root mean square noise voltage at the output.
% We have to generate normal distributed noise with a root mean square noise voltage
% equal to OutputNoiserms and add this time signal to NonLinearPA

    SignalSize=size(Input_Signal);
    Noise=randn(1,SignalSize(2)); %normal distribution
    NoiseVoltage=OutputNoiserms*Noise; %normal distributed noise at output according
   %to NoiseFactor
    Output_PA=NonLinearPA;%+NoiseVoltage;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%% End of Power Amplifier%%%%%%%%%%%%%
```

About the Authors

Dr. Domine Leenaerts studied electrical engineering at Eindhoven University of Technology. He gained his degree in 1987 and his Ph.D. in 1992. Between 1992 and 1999 he worked at this university as assistant and associate professor in the micro-electronic circuit design group. In 1995, he was a Visiting Scholar at the Department of Electrical Engineering and Computer Science at the University of California, Berkeley, and at the Electronic Research Laboratory of the same department. In 1997, he was a visiting professor at the Technical University of Lausanne (EPFL). He has been a senior research scientist at Philips Research Laboratories Eindhoven since 1999. Currently, he is a principal research scientist, responsible for the RF Telecom circuit activities in the Integrated Transceivers department.

Ir. Johan van der Tang received his Ing. degree in electrical engineering from the Technical College in Leeuwarden in 1992. He received his Ir. degree in the same field from the University of Twente in Enschede in 1995. From 1995 until 2000 he was a research scientist in the field of integrated transceivers at Philips Research Laboratories Eindhoven. In this function he worked on analog integrated HF key building blocks for satellite, radio and optical front-ends. Since 2000, he is an assistant professor in the Mixed-signal Microelectronics group at Eindhoven University of Technology and responsible for the RF transceiver activities. His research interests are design methodology for High-Frequency integrated oscillators and in general, design of integrated transceivers.

Dr. Cicero Vaucher graduated in electrical engineering from the Federal University of Rio Grande do Sul, Porto Alegre, Brazil, in 1989. He received the Ph.D. degree from the University of Twente, Enschede, The Netherlands, in 2001. Since 1990 he is with Philips Research Laboratories Eindhoven, where he is a senior research scientist in the Integrated Transceivers department. His research activities have been mainly focused on PLL frequency synthesizers, with emphasis on low-power high-speed PLL building blocks and on low phase noise, fast settling time PLL architectures. At the moment, he is responsible for activities on analogue IC design for microwave applications.

Index